

Gebhard Böckle · Wolfram Decker  
Gunter Malle *Editors*

# Algorithmic and Experimental Methods in Algebra, Geometry, and Number Theory

 Springer

# Algorithmic and Experimental Methods in Algebra, Geometry, and Number Theory

Gebhard Böckle • Wolfram Decker • Gunter Malle  
Editors

# Algorithmic and Experimental Methods in Algebra, Geometry, and Number Theory

 Springer

*Editors*

Gebhard Böckle  
IWR  
Heidelberg University  
Heidelberg, Germany

Wolfram Decker  
Department of Mathematics  
Technische Universität Kaiserslautern  
Kaiserslautern, Germany

Gunter Malle  
Department of Mathematics  
Technische Universität Kaiserslautern  
Kaiserslautern, Germany

ISBN 978-3-319-70565-1      ISBN 978-3-319-70566-8 (eBook)  
<https://doi.org/10.1007/978-3-319-70566-8>

Library of Congress Control Number: 2018932356

Mathematics Subject Classification (2010): 06Bxx, 11Fxx, 11Gxx, 11G15, 13Pxx, 13P10, 14G05, 14H40, 14Qxx, 14Txx, 16Exx, 16Gxx, 20Cxx, 20C40, 20Exx, 51F15, 52Bxx

© Springer International Publishing AG, part of Springer Nature 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

Experiments based on calculating examples have always played a key role in mathematical research. Today, modern computers paired with sophisticated mathematical software tools allow for far-reaching experiments that were previously unimaginable. They enable mathematicians to test working hypotheses or conjectures in a large number of instances, to find counterexamples or sufficient mathematical evidence to refine a conjecture, to arrive at new conjectures in the first place, and to verify theorems whose proofs have been reduced to handling a finite number of special cases.

In the realm of algebra and its applications, where exact calculations are essential, the desired software tools are implemented in computer algebra systems that are large, complex pieces of software and contain and rely on a vast amount of mathematical reasoning. Driven by intended applications, they are created by collaborative efforts involving specialists in many different fields. Importantly, these systems also allow non-experts to access and apply a virtual treasure trove of mathematical knowledge.

Over the last few decades, computer algebra has evolved as a mathematical discipline in its own right. Its algorithms have opened up new ways of accessing some of the key disciplines of pure mathematics and are fundamental to the practical applications of these disciplines. A decisive feature of current developments is that more and more of the abstract concepts of pure mathematics are being made constructive, with interdisciplinary methods playing a significant role.

In this context, the German Research Foundation (DFG) established the Priority Programme SPP 1489 on Algorithmic and Experimental Methods in Algebra, Geometry, and Number Theory, which was running from July 2010 until June 2016. The overall goal of the programme was to considerably advance the algorithmic and experimental methods in these disciplines, to combine the different methods where needed, and to apply them to central questions in theory and practice. In particular, the programme was meant to support the further development of leading open-source computer algebra systems based (to a large extent) within its boundaries, and to interconnect these systems, supplemented by a number of smaller libraries

and packages, in order to create cutting-edge software tools for interdisciplinary research.

This proceedings volume reports on selected aspects of the work done during the Priority Programme. It contains original research articles as well as survey papers that reflect all levels of computer algebra—providing computational access to mathematical concepts, designing algorithms, implementing the algorithms, and applying them to profound mathematical questions. The mathematical themes are taken from group and representation theory, algebraic geometry, polyhedral and tropical geometry, and number theory. Specific topics include finite group theory, reflection arrangements, associative algebras, algebraic curves, moduli spaces, lattices, modular forms, Jacobians and Abelian varieties, rational points, and real and complex multiplication.

The editors would like to take this opportunity to thank the DFG for its generous support, which made considerable progress in the field possible and helped launch the careers of several young researchers, as witnessed by this proceedings volume.

Heidelberg, Germany  
Kaiserslautern, Germany  
Kaiserslautern, Germany  
September 2017

Gebhard Böckle  
Wolfram Decker  
Gunter Malle

# Contents

<b>Algorithmic Aspects of Units in Group Rings</b> .....	1
Andreas Bächle, Wolfgang Kimmerle, and Leo Margolis	
<b>A Constructive Approach to the Module of Twisted Global Sections on Relative Projective Spaces</b> .....	23
Mohamed Barakat and Markus Lange-Hegermann	
<b>Local to Global Algorithms for the Gorenstein Adjoint Ideal of a Curve</b> .....	51
Janko Böhm, Wolfram Decker, Santiago Laplagne, and Gerhard Pfister	
<b>Picard Curves with Small Conductor</b> .....	97
Michel Börner, Irene I. Bouw, and Stefan Wewers	
<b>Normaliz 2013–2016</b> .....	123
Winfried Bruns, Richard Sieg, and Christof Söger	
<b>Integral Frobenius for Abelian Varieties with Real Multiplication</b> .....	147
Tommaso Giorgio Centeleghe and Christian Theisen	
<b>Monodromy of the Multiplicative and the Additive Convolution</b> .....	177
Michael Dettweiler and Mirjam Jöllenbeck	
<b>Constructing Groups of ‘Small’ Order: Recent Results and Open Problems</b> .....	199
Bettina Eick, Max Horn, and Alexander Hulpke	
<b>Classifying Nilpotent Associative Algebras: Small Coclass and Finite Fields</b> .....	213
Bettina Eick and Tobias Moede	
<b>Desingularization of Arithmetic Surfaces: Algorithmic Aspects</b> .....	231
Anne Frühbis-Krüger and Stefan Wewers	
<b>Moduli Spaces of Curves in Tropical Varieties</b> .....	253
Andreas Gathmann and Dennis Ochse	

<b>Tropical Moduli Spaces of Stable Maps to a Curve</b> .....	287
Andreas Gathmann, Hannah Markwig, and Dennis Ochse	
<b>Invariant Bilinear Forms on <math>W</math>-Graph Representations and Linear Algebra Over Integral Domains</b> .....	311
Meinolf Geck and Jürgen Müller	
<b>Tropical Computations in <code>polymake</code></b> .....	361
Simon Hampe and Michael Joswig	
<b>Focal Schemes to Families of Secant Spaces to Canonical Curves</b> .....	387
Michael Hoff	
<b>Inductive and Recursive Freeness of Localizations of Multiarrangements</b> .....	403
Torsten Hoge, Gerhard Röhrle, and Anne Schauenburg	
<b>Toric Ext and Tor in <code>polymake</code> and Singular: The Two-Dimensional Case and Beyond</b> .....	423
Lars Kastner	
<b>The Differential Dimension Polynomial for Characterizable Differential Ideals</b> .....	443
Markus Lange-Hegermann	
<b>Factorization of <math>\mathbb{Z}</math>-Homogeneous Polynomials in the First <math>q</math>-Weyl Algebra</b> .....	455
Albert Heinle and Viktor Levandovskyy	
<b>Complexity of Membership Problems of Different Types of Polynomial Ideals</b> .....	481
Ernst W. Mayr and Stefan Toman	
<b>Localizations of Inductively Factored Arrangements</b> .....	495
Tilman Möller and Gerhard Röhrle	
<b>One Class Genera of Lattice Chains Over Number Fields</b> .....	503
Markus Kirschmer and Gabriele Nebe	
<b><code>polyDB</code>: A Database for Polytopes and Related Objects</b> .....	533
Andreas Paffenholz	
<b>Construction of Neron Desingularization for Two Dimensional Rings</b> .....	549
Gerhard Pfister and Dorin Popescu	
<b>A Framework for Computing Zeta Functions of Groups, Algebras, and Modules</b> .....	561
Tobias Rossmann	
<b>On Decomposition Numbers of Diagram Algebras</b> .....	587
Armin Shalile	



<b>Koblitz’s Conjecture for Abelian Varieties</b> .....	611
Ute Spreckels and Andreas Stein	
<b>Chabauty Without the Mordell-Weil Group</b> .....	623
Michael Stoll	
<b>An Explicit Theory of Heights for Hyperelliptic Jacobians of Genus Three</b> .....	665
Michael Stoll	
<b>Some Recent Developments in Spectrahedral Computation</b> .....	717
Thorsten Theobald	
<b>Topics on Modular Galois Representations Modulo Prime Powers</b> .....	741
Panagiotis Tsaknias and Gabor Wiese	

# Algorithmic Aspects of Units in Group Rings



Andreas Bächle, Wolfgang Kimmerle, and Leo Margolis

**Abstract** We describe the main questions connected to torsion subgroups in the unit group of integral group rings of finite groups and algorithmic methods to attack these questions. We then prove the Zassenhaus Conjecture for Amitsur groups and prove that any normalized torsion subgroup in the unit group of an integral group of a Frobenius complement is isomorphic to a subgroup of the group base. Moreover we study the orders of torsion units in integral group rings of finite almost quasisimple groups and the existence of torsion-free normal subgroups of finite index in the unit group.

**Keywords** Units • Integral group rings • Zassenhaus conjectures • Computational character methods

**Subject Classifications** 16S34, 16U60, 20C05, 20C40

---

The first author is a postdoctoral researcher of the FWO (Research Foundation Flanders). The third is supported by a Marie Skłodowska-Curie Individual Fellowship from EU project 705112-ZC.

A. Bächle  
Vakgroep Wiskunde, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium  
e-mail: [abachle@vub.ac.be](mailto:abachle@vub.ac.be)

W. Kimmerle (✉)  
Fachbereich Mathematik, IGT, Universität Stuttgart, Pfaffenwaldring 57, 70550 Stuttgart, Germany  
e-mail: [kimmerle@mathematik.uni-stuttgart.de](mailto:kimmerle@mathematik.uni-stuttgart.de)

L. Margolis  
Departamento de matemáticas, Facultad de matemáticas, Universidad de Murcia, 30100 Murcia, Spain  
e-mail: [leo.margolis@um.es](mailto:leo.margolis@um.es)

# 1 Introduction

The study of the units of an integral group ring  $\mathbb{Z}G$  for a finite group  $G$  has begun in Higman's thesis [29]. Higman classified the finite groups  $G$  whose integral group ring has only trivial units. The aim of this article is to present recent work on the structure of torsion subgroups on the unit group of  $\mathbb{Z}G$  which has been achieved especially with the aid of algorithmic tools.

Denote by

$$\varepsilon : \mathbb{Z}G \rightarrow \mathbb{Z}, \quad \sum_{g \in G} z_g g \mapsto \sum_{g \in G} z_g$$

the augmentation map. Being a ring homomorphism,  $\varepsilon$  maps units of  $\mathbb{Z}G$  to units of  $\mathbb{Z}$ , so up to multiplication with  $-1$  any unit in  $\mathbb{Z}G$  has augmentation 1 and it suffices to study the units of augmentation 1 in  $\mathbb{Z}G$ . The set of these so called normalized units will be denoted by  $V(\mathbb{Z}G)$ .

Although extensively studied, very few general theorems on the behaviour of finite subgroups in  $V(\mathbb{Z}G)$  are available. It is known that the order of a finite subgroup of  $V(\mathbb{Z}G)$  divides the order of  $G$  [62] and that the order of a torsion unit in  $V(\mathbb{Z}G)$  divides the exponent of  $G$  [15]. But it is not even known, whether the orders of torsion units in  $V(\mathbb{Z}G)$  coincide with the orders of elements in  $G$ . For a long time the Isomorphism Problem, which asks whether a ring isomorphism  $\mathbb{Z}G \cong \mathbb{Z}H$  implies a group isomorphism  $G \cong H$ , was the focus of attention in the area, see e.g. [52]. A negative answer to this problem was finally given by Hertweck [21].

The main questions in the area of torsion units of integral group rings are given and inspired by three conjectures of Zassenhaus. A subgroup  $U \leq V(\mathbb{Z}G)$  is called rationally conjugate to a subgroup of  $G$ , if there exist a subgroup  $U' \leq G$  and a unit  $x \in \mathbb{Q}G$  such that  $x^{-1}Ux = U'$ . Let  $G$  be finite group. The conjectures put forward by Zassenhaus [61] read as follows.

- (ZC1)** Units of finite order of  $V(\mathbb{Z}G)$  are rationally conjugate to elements of  $G$ .
- (ZC2)** Group bases, i.e. subgroups of  $V(\mathbb{Z}G)$  of the same order as  $G$ , are rationally conjugate.
- (ZC3)** A finite subgroup  $H$  of  $V(\mathbb{Z}G)$  is rationally conjugate to a subgroup of  $G$ .

Note that (ZC1) is still open. For (ZC2) and (ZC3) counterexamples are known (not only by the counterexample to the Isomorphism Problem). However they hold for important classes of finite groups, cf. Sect. 4.

In the last 10 years computational tools have been developed to attack these questions. The fundamental tool is the HeLP method which is now available as a GAP package [4]. This package makes use of two external solvers for integral linear inequalities, namely 4ti2 [1] and normaliz [14] which substantially improved the efficiency of HeLP. In Sect. 2 we describe the HeLP method as well as other algorithmic methods which have been developed in the last years to handle cases left open by HeLP. The remaining part of the article is organized as follows. In

Sect. 3 we prove (ZC1) for Amitsur groups. In the next section we survey recent results circulating around the Zassenhaus conjectures and isomorphism questions. We exhibit specifically those results which have been established with the help of computational tools. We present the major research problems on torsion units of integral group rings as well. In Sect. 5 we show that finite subgroups of  $V(\mathbb{Z}G)$  are isomorphic to a subgroup of  $G$  if  $G$  is a Frobenius complement. In the last two sections further classes of finite groups (in particular almost simple groups) are investigated especially with the help of character theory. For the question whether the projection of  $V(\mathbb{Z}G)$  onto a faithful Wedderburn component (of  $\mathbb{Q}G$  or  $\mathbb{C}G$ ) has a torsion free kernel, it becomes transparent that the use of generic characters and generic character tables is extremely useful. This underlines the connection to other topics of computational representation theory of finite groups which are in the focus of recent research. It also shows that this is related with the construction of large torsion free normal subgroups of  $V(\mathbb{Z}G)$ . This connects the investigation of torsion units of integral group rings with the other main topic in the area, the description of the whole unit group of  $\mathbb{Z}G$  in terms of generators and relations. This involves questions on the generation of units of infinite order, free non-abelian subgroups of  $V(\mathbb{Z}G)$ , generators of subgroups of finite index in  $V(\mathbb{Z}G)$  and others. For a recent detailed monograph on the latter topics see [34, 35].

## 2 Tools and Available Algorithms

Important tools to study the questions mentioned above are so-called partial augmentations. For an element  $u = \sum_{g \in G} u_g g \in \mathbb{Z}G$  and a conjugacy class  $x^G$  in  $G$  the integer

$$\varepsilon_x(u) = \sum_{g \in x^G} u_g$$

is called the partial augmentation of  $u$  at the conjugacy class of  $x$ . Being class functions of  $G$ , partial augmentations are a natural object to study using representation theory. The connection between the questions mentioned in the introduction and partial augmentations is established by the following result.

**Proposition 2.1** (Marciniak et al. [46, Theorem 2.5]) *A torsion unit  $u \in V(\mathbb{Z}G)$  is rationally conjugate to a group element if and only if  $\varepsilon_x(u^d) \geq 0$  for all divisors  $d$  of  $o(u)$  and all  $x \in G$ .*

Note that for  $u \in V(\mathbb{Z}G)$  the condition  $\varepsilon_x(u) \geq 0$  for all  $x \in G$  is equivalent to the fact that one partial augmentation of  $u$  is 1 while all other partial augmentations are 0—a situation which clearly applies for an element  $g \in G$ .

Thus it is of major interest to find restrictions on the partial augmentations of torsion units. For the orders of elements providing possibly non-vanishing partial augmentations the following is known.

**Proposition 2.2** *Let  $u \in V(\mathbb{Z}G)$  be a torsion unit of order  $n$ .*

- i)  $\varepsilon_1(u) = 0$ , unless  $u = 1$ . (Berman-Higman Theorem [34, Proposition 1.5.1]).
- ii) If  $\varepsilon_x(u) \neq 0$  for some  $x \in G$  then the order of  $x$  divides  $n$  [22, Theorem 2.3].

From the properties of the  $p$ -power map in group algebras of characteristic  $p$  one can obtain more restrictions on the partial augmentations given in terms of congruences modulo  $p$ .

**Lemma 2.3 (cf. [4, Proposition 3.1] for a Proof)** *Let  $s$  be some element in  $G$  and  $u \in V(\mathbb{Z}G)$  a unit of order  $p^j \cdot m$  with  $p$  a prime and  $m \neq 1$ . Then*

$$\sum_{x^G, x^{p^j} \sim s} \varepsilon_x(u) \equiv \varepsilon_s(u^{p^j}) \pmod{p}.$$

In some special situations there are more theoretical restrictions on the partial augmentations of torsion units. We will only mention one of them which has not been used frequently yet, but turns out to be quite useful for our results.

**Proposition 2.4 ([23, Proposition 2])** *Suppose that  $G$  has a normal  $p$ -subgroup  $N$ , and that  $u$  is a torsion unit in  $V(\mathbb{Z}G)$  whose image under the natural homomorphism  $\mathbb{Z}G \rightarrow \mathbb{Z}G/N$  has strictly smaller order than  $u$ . Then  $\varepsilon_g(u) = 0$  for every element  $g$  of  $G$  whose  $p$ -part has strictly smaller order than the  $p$ -part of  $u$ .*

*Remark 2.5* An easy but often useful observation when working with quotient groups is the following. Let  $N$  be a normal subgroup of  $G$ , and denote by  $\varphi : \mathbb{Z}G \rightarrow \mathbb{Z}G/N$  the linear extension of the natural projection from  $G$  to  $G/N$ . Then for an element  $g \in G$  and a unit  $u \in V(\mathbb{Z}G)$  we have

$$\varepsilon_{\varphi(g)}(\varphi(u)) = \sum_{\substack{x^G \\ \varphi(x) \sim \varphi(g)}} \varepsilon_x(u)$$

where the sum runs over the conjugacy classes of  $G$ .

## 2.1 HeLP

An idea to obtain more restrictions on the partial augmentations of torsion units in  $V(\mathbb{Z}G)$  using the values of ordinary characters of  $G$  was introduced by Luthar and Passi [44]. If  $\chi$  denotes an ordinary character of  $G$  and  $D$  a representation of  $G$  realizing  $\chi$  then  $D$  can be linearly extended to the group ring  $\mathbb{Z}G$ . This provides a ring homomorphism from  $\mathbb{Z}G$  to a matrix ring and thus units of  $\mathbb{Z}G$  are mapped to

invertible matrices. Hence  $D$  extends to a representation of  $V(\mathbb{Z}G)$  and  $\chi$  extends to a character of  $V(\mathbb{Z}G)$ . Let  $x_1, \dots, x_h$  be representatives of the conjugacy classes of elements in  $G$ . Since  $\chi$  is a  $\mathbb{Z}$ -linear function we obtain

$$\chi(u) = \sum_{i=1}^h \varepsilon_{x_i}(u) \chi(x_i) \text{ for } u \in V(\mathbb{Z}G).$$

Denote by  $\chi_1, \dots, \chi_h$  the irreducible complex characters of  $G$  and by  $X(G)$  the character table of  $G$ . So from the arguments above we see

$$\begin{pmatrix} \chi_1(u) \\ \chi_2(u) \\ \vdots \\ \chi_h(u) \end{pmatrix} = \begin{pmatrix} \chi_1(x_1) & \chi_1(x_2) & \dots & \chi_1(x_h) \\ \chi_2(x_1) & \chi_2(x_2) & \dots & \chi_2(x_h) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_h(x_1) & \chi_h(x_2) & \dots & \chi_h(x_h) \end{pmatrix} \begin{pmatrix} \varepsilon_{x_1}(u) \\ \varepsilon_{x_2}(u) \\ \vdots \\ \varepsilon_{x_h}(u) \end{pmatrix} = X(G) \begin{pmatrix} \varepsilon_{x_1}(u) \\ \varepsilon_{x_2}(u) \\ \vdots \\ \varepsilon_{x_h}(u) \end{pmatrix}. \quad (1)$$

Since the character table of a group is an invertible matrix, Eq. (1) provides restrictions on the partial augmentations of  $u$  once we obtain restrictions on the character values  $\chi_1(u), \dots, \chi_h(u)$ .

For a unit  $u$  of finite order these restrictions follow from the fact that  $D(u)$  is a matrix of order dividing the order of  $u$ . Thus  $D(u)$  is diagonalizable and its eigenvalues are  $o(u)$ -th roots of unity. So there are only finitely many possibilities for the values of  $\chi(u)$ . Hence going through these possibilities for all the irreducible complex characters of  $G$  and applying Eq. (1), one obtains finitely many possibilities for the partial augmentations of  $u$ . If we assume moreover that the partial augmentations of proper powers of  $u$  are known, say by induction, then the restrictions on the eigenvalues of  $D(u)$  can be significantly strengthened since they may be obtained as the pairwise product of the eigenvalues of  $D(u^d)$  and  $D(u^e)$  where  $d, e$  denote integers not coprime with  $o(u)$  such that  $d + e \equiv 1 \pmod{o(u)}$ .

Clearly if one assumes that  $K$  is an algebraically closed field of characteristic  $p$  not dividing  $o(u)$  then the arguments of the last paragraph still apply. Fixing a correspondence between the complex roots of unity of order not divisible by  $p$  and the roots of unity in  $K$ , as it is custom in modular representation theory, one can view the character  $\chi$  as a  $p$ -Brauer character having complex values. It was shown by Hertweck [22, Section 3] that if one takes  $x_1, \dots, x_h$  to be only the representatives of  $p$ -regular conjugacy classes in  $G$  and  $\chi_1, \dots, \chi_h$  to be the irreducible  $p$ -Brauer characters of  $G$  then Eq. (1) also applies. This modular extension of the idea of Luthar and Passi is in particular useful for simple non-abelian groups. It does however not provide new restrictions for solvable groups by the Fong-Swan-Rukolaine Theorem [16, Theorem 22.1].

The method described above is nowadays referred to as HeLP method (the name is an acronym of the names of the originators of the method: *HertweckLutharPassi*). The HeLP method can be implemented into a computer program as it has been done in the GAP package HeLP [4]. The HeLP method has been applied for single

groups, e.g. in the study of the Zassenhaus Conjecture for small groups as in [11, 31] or [9] or to study non-solvable groups as e.g. in [12] or [42]. It might also be used to study infinite series of groups possessing generic character tables as it was done in [7, 22] or [49]. In this paper we will apply the HeLP method in Sects. 6 and 7.

Note that when one knows the partial augmentations of a torsion unit  $u \in V(\mathbb{Z}G)$  and all its powers, as e.g. after the application of the HeLP method, one may compute the eigenvalues, with multiplicities, of  $D(u)$  for any ordinary representation  $D$  of  $G$ . This observation is often useful when combining the HeLP method with other ideas described below.

## 2.2 Other Algorithmic Methods: Quotients, Partially Central Units and the Lattice Method

An inductive approach to questions about torsion units in  $V(\mathbb{Z}G)$  may be taken when one possesses information on the torsion units in  $V(\mathbb{Z}G/N)$  where  $N$  is some normal subgroup in  $G$ , since a homomorphism from  $G$  to  $G/N$  naturally extends to a homomorphism from  $V(\mathbb{Z}G)$  to  $V(\mathbb{Z}G/N)$ . If one can control the fusion of conjugacy classes in the projection from  $G$  to  $G/N$  then one can also obtain restrictions on the partial augmentations of elements in  $V(\mathbb{Z}G)$  assuming some knowledge about the partial augmentations of the units in  $V(\mathbb{Z}G/N)$ , cf. Remark 2.5. This approach was taken by many authors in particular when studying classes of groups closed under quotients as e.g. in [24]. In this paper also quotients play a significant role in all our results.

Assume that some torsion unit  $u \in V(\mathbb{Z}G)$  is central in some Wedderburn component  $B$  of the complex group algebra  $\mathbb{C}G$ , but its spectrum in this component does not coincide with the spectrum of any element in  $G$ . An idea to disprove the existence of such units was first used manually by Höfert [30] and recently developed as a GAP program by Herman and Singh [20]. This is sometimes called the Partially Central Method and uses an explicit representation of  $G$  to show that no element in  $\mathbb{C}G$  having only integral coefficients can realize the given central unit in  $B$ . Since  $u$  has no other conjugates inside  $B$ , this also proves that  $u$  can not be globally conjugate to an element in  $\mathbb{Z}G$ . This method turns out to be useful for the study of small groups as demonstrated in [9], but is also not always successful.

A further algorithmic method, particularly useful for the study of the Prime Graph Question (cf. Problem 4.4 of Sect. 4), was introduced in [8] and is known as the Lattice Method. Let  $p$  be a prime,  $(K, R, k)$  be a  $p$ -modular system for  $G$  and  $u \in V(\mathbb{Z}G)$  a torsion unit of order divisible by  $p$ . The idea of the Lattice method is that when  $B$  is a block of the modular group algebra  $kG$ ,  $D$  an ordinary irreducible representation of  $G$  belonging to  $B$  with corresponding  $RG$ -lattice  $L$  and  $S$  a simple  $kG$ -composition factor of  $\bar{L}$ , where  $\bar{\phantom{x}}$  denotes the projection from  $R$  onto  $k$ , then the spectrum of  $D(u)$  provides restrictions on the isomorphism type of  $L$  as  $R\langle u \rangle$ -lattice and thus on the isomorphism type of  $S$  as  $k\langle \bar{u} \rangle$ -module. The use of other

$RG$ -lattices whose reduction to  $kG$  involve  $S$  as a composition factor may finally lead to a contradiction to the existence of  $u$ , since in some situations the restrictions obtained on the isomorphism type of  $S$  as  $k(\bar{u})$ -module may contradict each other. The Lattice Method has successfully been applied in the study of the Prime Graph Question and the Zassenhaus Conjecture for non-solvable groups in [6, 8]. We will apply the Lattice Method in Sect. 6.

### 3 Amitsur Groups

Herstein pointed out that all finite subgroups of division rings in positive characteristic  $p$  are cyclic  $p'$ -groups. In [2] Amitsur described the finite subgroups of division algebras in characteristic 0; these groups are nowadays often called *Amitsur groups*. Recall that a group is a  $Z$ -group if all its Sylow subgroups are cyclic. We will use a weaker version of Amitsur's classification suitable for us.

**Theorem 3.1 (Amitsur [55, Theorem 2.1.4])** *If a finite group  $G$  is a subgroup of a division algebra of characteristic 0 then*

- (Z)  $G$  is a  $Z$ -group  
 (NZ) or  $G$  is isomorphic to one of the following groups
- $\mathcal{O}^* = \langle s, t \mid (st)^2 = s^3 = t^4 \rangle$  (binary octahedral group),
  - $\mathrm{SL}(2, 5)$ ,
  - $\mathrm{SL}(2, 3) \times M$ , with  $M$  a group in (Z) of order coprime to 6 and 2 has odd order modulo  $|M|$ ,
  - $C_m \rtimes Q$ , where  $m$  is odd,  $Q$  a quaternion group of order  $2^l$  such that an element of order  $2^{l-1}$  centralizes  $C_m$  and an element of order 4 inverts  $C_m$ ,
  - $Q_8 \times M$  with  $M$  a group in (Z) of odd order and 2 has odd order modulo  $|M|$ .

In [18, Theorem 3.5] it was proved that for an Amitsur group  $G$ , the order of a normalized torsion unit in  $\mathbb{Z}G$  coincides with the order of an element in  $G$ . We now verify that even the first Zassenhaus Conjecture holds for these groups.

**Theorem 3.2** *Let  $G$  be a finite subgroup of a division ring, then (ZC1) holds for  $G$ .*

*Proof* If  $G$  is a  $Z$ -group then it is either cyclic or metacyclic and hence (ZC1) holds for  $G$  [24, Theorem 1.1]. The binary octahedral group was handled by Dokuchaev and Juriaans [17, Proposition 4.2] and  $\mathrm{SL}(2, 5)$  by Juriaans and Polcino-Milies [36, Proposition 4.2] (for those two groups even (ZC3) was verified). The groups in (NZ)(d) have cyclic normal subgroups of order  $2^{l-1}m$  with an abelian quotient of order 2, so we can again apply [24, Theorem 1.1]. If  $G = Q_8 \times M$  for some  $Z$ -group  $M$  of odd order, then it is a direct product of a nilpotent group with a group for which the Zassenhaus Conjecture is known with coprime orders and the claim follows from [24, Proposition 8.1]. So we are left with the groups in (NZ)(c).

Assume from now on that  $G = \mathrm{SL}(2, 3) \times M$ , with  $M$  a group in (Z) of order coprime to 6, and let  $u \in V(\mathbb{Z}G)$  be a torsion unit. If the order of  $u$  is a divisor of



$|M|$ , then  $(o(u), |\mathrm{SL}(2, 3)|) = 1$ . Then we can consider the ring homomorphism  $\mathbb{Z}G \rightarrow \mathbb{Z}G/\mathrm{SL}(2, 3) \simeq \mathbb{Z}M$  induced by the projection  $G \rightarrow G/\mathrm{SL}(2, 3)$  to a group ring of a group for which the Zassenhaus Conjecture is known and apply [17, Theorem 2.2] to conclude that  $u$  is rationally conjugate to an element of  $G$ . If  $u$  has an order a divisor of  $|\mathrm{SL}(2, 3)|$  then a similar argument applies.

We now consider units having an order which has a common divisor with both, 6 and the order of  $M$ . Note that  $G$  has a normal Sylow 2-subgroup  $P$ . Denote by  $\varphi$  the natural ring homomorphism  $\mathbb{Z}G \rightarrow \mathbb{Z}G/P$ , which will also be denoted by  $\bar{\phantom{x}}$ , i.e.  $\bar{x} = \varphi(x)$  for  $x \in \mathbb{Z}G$ .

Assume first that  $u \in \mathrm{V}(\mathbb{Z}G)$  is of order  $2m$  with  $1 \neq m$  a divisor of  $|M|$ . If  $\varepsilon_x(u) \neq 0$ , then  $o(x) \mid 2m$  by Proposition 2.2. The image  $\bar{u}$  has order a divisor of  $m$ , in particular strictly smaller order than  $u$ . So by Proposition 2.4 the only partial augmentations of  $u$  that are potentially non-zero are those at classes of group elements of order  $2m_0$  for  $m_0$  a divisor of  $m$ . Let  $w \in G$  be an element whose natural projection onto  $\mathrm{SL}(2, 3)$  is trivial. Then the conjugacy classes that are mapped under  $\varphi$  onto the conjugacy class of  $\bar{w}$  are exactly the classes of  $w, zw$  and  $tw$ , where  $z$  and  $t$  denote elements of  $G$  of order 2 and 4, respectively. Hence  $\varepsilon_{\bar{w}}(\bar{u}) = \varepsilon_w(u) + \varepsilon_{zw}(u) + \varepsilon_{tw}(u) = \varepsilon_{zw}(u)$ , since the order of  $w$  and  $tw$  is not of the form  $2m_0$ . As the Zassenhaus Conjecture holds for the metacyclic group  $G/P$  we conclude that  $u$  has exactly one non-vanishing partial augmentation and is rationally conjugate to a group element by Proposition 2.1.

If  $u$  is of order  $4m$  where  $1 \neq m$  is a divisor of  $|M|$  then analogous arguments as in the case  $2m$  show that  $u$  is rationally conjugate to an element of  $G$ . Now assume that  $u \in \mathrm{V}(\mathbb{Z}G)$  is of order  $3m$  with  $1 \neq m \mid |M|$ . Then  $\bar{u}$  is conjugate within  $\mathbb{Q}\bar{G}$  to an element of  $\bar{G}$ . As  $(o(u), |P|) = 1$  we can use [17, Theorem 2.2] to conclude that  $u$  is rationally conjugate to an element of  $G$ .

Finally assume that  $o(u) = 6m$  with  $1 \neq m$  a divisor of  $|M|$ . If  $\varepsilon_x(u) \neq 0$  for some  $x \in G$ , then  $o(x) \mid 6m$  by Proposition 2.2. The image  $\bar{u}$  has strictly smaller order than  $u$ , so again by Proposition 2.4 the only partial augmentations of  $u$  that are potentially non-zero are those at classes of group elements of order  $2m_0$  for  $m_0$  a divisor of  $3m$ . First consider an element  $w \in G$  that projects to 1 when mapped to  $\mathrm{SL}(2, 3)$ . Then there are three conjugacy classes of  $G$  that are mapped on the conjugacy class of  $\bar{w}$  in  $\bar{G}$ , namely those of  $w, zw$  and  $tw$ , where  $z$  and  $t$  are elements of  $G$  of order 2 and 4, respectively. Then  $\varepsilon_{\bar{w}}(\bar{u}) = \varepsilon_w(u) + \varepsilon_{zw}(u) + \varepsilon_{tw}(u) = \varepsilon_{zw}(u)$ . Now assume that  $w \in G$  maps to an element of order 3 in  $\mathrm{SL}(2, 3)$ . Observe that for each element  $s \in P$ ,  $ws$  is either conjugate to  $w$  or to  $zw$ . Hence exactly those two conjugacy classes map to the conjugacy class of  $\bar{w}$ . Thus  $\varepsilon_{\bar{w}}(\bar{u}) = \varepsilon_w(u) + \varepsilon_{zw}(u) = \varepsilon_{zw}(u)$ . As the Zassenhaus Conjecture holds for  $G/P$  we can conclude that  $u$  has exactly one non-trivial partial augmentation. By Proposition 2.1,  $u$  is rationally conjugate to an element of  $G$ . The theorem is proved.

## 4 From (IP) to (SIP) and (PQ)

From the point of view of the unit group  $V(\mathbb{Z}G)$  the counterexample to the isomorphism problem (IP) simply shows that different group bases may be non-isomorphic. Nevertheless a lot of positive results have been established and it is justified to say that (IP) has almost a positive answer. Indeed for each finite group  $G$  there is an abelian extension  $E := A \rtimes G$  such that (ZC2), and so also (IP), has a positive answer, i.e. different group bases of  $\mathbb{Z}E$  are conjugate within  $\mathbb{Q}G$ . This follows from the  $F^*$ -theorem which has been discovered by Roggenkamp and Scott [53], [51, Theorem 19] and has now finally a published account [26, Theorem A and p. 350], see also [27, p. 180]. With respect to semilocal coefficient rings the  $F^*$ -theorem (in its automorphism version) may be stated as follows.

**$F^*$ -Theorem** Let  $G$  be a finite group. Denote by  $\pi(G)$  the set of primes dividing the order of  $G$ . Let  $S$  be the semilocal ring  $\mathbb{Z}_{\pi(G)}$ . Suppose that the generalized Fitting subgroup  $F^*(G)$  is a  $p$ -group and let  $\alpha$  be an  $S$ -algebra automorphism of  $SG$  preserving augmentation. Then  $\alpha$  is given as the composition of an automorphism induced from a group automorphism of  $G$  followed by a central automorphism (i.e. given by conjugation with a unit of  $\mathbb{Q}G$ ).

The assumption on  $G$  in the preceding theorem holds for all group bases of  $\mathbb{Z}G$  and for all group bases of  $\mathbb{Z}(G \times G)$ . Thus it follows for groups whose generalized Fitting subgroup  $F^*(G)$  is a  $p$ -group that group bases of  $\mathbb{Z}G$  are rationally conjugate, cf. [37, 5.3]. For a given group  $G$  let  $A$  be the additive group of  $\mathbb{F}_pG$ . Consider the semidirect product  $E = A \rtimes G$ , where the action of  $G$  is just given by the multiplication of  $G$  on  $A$ . Clearly  $C_E(A) = A$  and thus the  $F^*$ -theorem establishes (ZC2) and therefore a strong answer to (IP) for  $\mathbb{Z}E$ .

The preceding paragraph shows that (IP) has a positive solution for many important classes of finite groups. Thus the following subgroup variation came in the focus of research within the last years.

**Problem 4.1 (Subgroup Isomorphism Problem (SIP))** Classify all finite groups  $H$  such that whenever  $H$  occurs for a group  $G$  as subgroup of  $V(\mathbb{Z}G)$  then  $H$  is isomorphic to a subgroup of  $G$ .

If  $H$  has this property we say that (SIP) holds for  $H$ .

(SIP) holds for the following groups:

- 4.1. cyclic groups of prime power order [15].
- 4.2.  $C_p \times C_p$ ,  $p$  a prime [25, 39].
- 4.3.  $C_4 \times C_2$  [48].

This shows that with respect to general finite groups very limited general facts are known about torsion units of the integral group ring. Much more is known on the following related question. Let  $G$  be a specific finite group. Are all torsion subgroups of  $V(\mathbb{Z}G)$  isomorphic to a subgroup of  $G$ ? In Sect. 5 we settle this question for all groups occurring as Frobenius complements.

Whether (SIP) holds for finite  $p$ -groups is certainly one of the major open questions. Clearly this leads to Sylow like theorems for  $\mathbb{Z}G$ . Even for conjugacy of finite  $p$ -groups within  $\mathbb{Q}G$  no counterexample is known.

**Problem 4.2 ([17, p-ZC3, p. 1170])** Is a Sylow like theorem (SLT) valid in  $V(\mathbb{Z}G)$ , i.e. is each  $p$ -subgroup of  $V(\mathbb{Z}G)$  rationally conjugate to a subgroup of  $G$ ?

We say that (SLT) holds for a given group  $G$  if in  $V(\mathbb{Z}G)$  Problem 4.2 has an affirmative answer for all primes  $p$  and that  $(\text{SLT})_p$  holds if this is the case for a specific prime  $p$ . If each subgroup of  $V(\mathbb{Z}G)$  of prime power order is isomorphic to a subgroup of  $G$  we speak of a weak Sylow like theorem (WSLT) and use the notion  $(\text{WSLT})_p$  if this holds for a specific prime  $p$ .

Denote by  $G_p$  a Sylow  $p$ -subgroup of  $G$ . Summary of known results on (SLT).

- 4.4.  $(\text{SLT})_p$  holds when  $G_p$  is normal [54, (41.12)].
- 4.5. (SLT) holds when  $G$  is nilpotent-by-nilpotent [17].
- 4.6.  $(\text{SLT})_p$  holds when  $G_p$  is abelian and  $G$  is  $p$ -constrained [3, Proposition 3.2].
- 4.7. (SLT) holds for  $\text{PSL}(2, p^f)$  where  $p$  denotes a prime [47] if  $f = 1$  or  $p = 2$ . It also holds for  $\text{PSL}(2, p^2)$  if  $p \leq 5$  [43]. Moreover (WSLT) holds if  $f \leq 3$  [5, 28].
- 4.8.  $(\text{SLT})_2$  is valid if  $|G_2| \leq 8$ , unless  $G \cong A_7$  [48].  $(\text{WSLT})_p$  is valid if  $G_p$  is cyclic [25, 39] and  $(\text{WSLT})_2$  is proved if  $G_2$  is generalized quaternion [40, Theorem 4.1] or a dihedral group [48].

For Frobenius groups we refer to the next section. The following two further special cases of (SIP) have been studied extensively in the last decade.

**Problem 4.3 ((SIP-C), Problem 8 in [54])** Let  $G$  be a finite group. Is each cyclic subgroup of  $V(\mathbb{Z}G)$  isomorphic to a subgroup of  $G$ ?

**Problem 4.4 (Prime Graph Question (PQ))** Let  $G$  be a finite group. Do  $G$  and  $V(\mathbb{Z}G)$  have the same prime graphs? Equivalently, is (SIP) valid for cyclic groups of order  $p \cdot q$ , where  $p$  and  $q$  are different primes?

We say that (SIP-C) or (PQ) holds for a group  $G$  if Problem 4.3 or Problem 4.4 respectively has a positive answer for  $G$ . Note that (ZC1) implies (SIP-C) and this in turn implies (PQ). So both problems may be also considered as test problems for the first Zassenhaus conjecture.

Summary of known results on (SIP-C) and (PQ).

- 4.9. (SIP-C) holds for soluble groups [23]. Moreover (SIP-C) is valid for any soluble extension of a group  $Q$  for which each torsion unit of order  $n$  has non-vanishing partial augmentations on a class of elements of  $Q$  of order  $n$ , cf. Lemma 6.3. This is the case when (ZC1) holds for  $\mathbb{Z}Q$ .  
(PQ) is valid for any soluble extension of a group  $Q$  for which (PQ) holds. [38, Proposition 4.3].
- 4.10. (SIP-C) holds for Frobenius groups [42, Corollary 2.5].
- 4.11. (PQ) holds for all simple groups  $\text{PSL}(2, p)$ , for  $p$  a prime [22].  
(PQ) also holds for any almost simple group with socle isomorphic to  $\text{PSL}(2, p)$  or  $\text{PSL}(2, p^2)$  [7].

- 4.12. If each almost simple image of the group  $G$  has an order divisible by three primes then (PQ) has an affirmative answer [42, Theorem 3.1], [8].
- 4.13. (PQ) holds for many almost simple groups whose socle has an order divisible by at most 4 different primes [6].

We like to point out that the computational tools explained in Sect. 2 play a prominent role in proving these results. A typical example for this is 4.12. By theoretical arguments the proof is reduced to almost simple groups whose orders are divisible by exactly three primes, cf. [41, §4]. CFSG shows that there are only eight such simple groups with this property. Now a computer examination of the almost simple groups arising from those simple groups yields successfully the result. However, the HeLP method does not suffice to deal with all cases. In the case of automorphism groups of  $A_6$  the final piece is obtained by the Lattice method.

For further results on almost simple groups of small order see Sect. 6.

## 5 Frobenius Groups and Complements

The torsion units of the integral group rings of Frobenius groups were considered in [10, 18, 36, 38, 43]. In particular (SLT) holds (cf. [43]) and (PQ) is known [38] for Frobenius groups. However, none of the Zassenhaus conjectures has been established completely. For many specific Frobenius groups (ZC1) and (ZC3) are known. The following Theorem 5.1 should be seen as a first important step towards (ZC3) for Frobenius groups. By the notation

$$G = \frac{Q}{N}$$

we indicate that  $G$  has a normal subgroup  $N$  such that  $G/N$  is isomorphic to  $Q$ .

**Theorem 5.1** *Let  $G$  be a Frobenius complement. Then each torsion subgroup of  $V(\mathbb{Z}G)$  is isomorphic to a subgroup of  $G$ .*

*Proof* By Passman [50, §18] the structure of Frobenius complements  $G$  is as follows.

Denote by  $W$  the Fitting subgroup of  $G$ .

- (1) If  $G_2$  is cyclic then  $G$  is a  $\mathbb{Z}$ -group.
- (2a) Suppose that  $W_2$  is cyclic. Then  $G$  is metabelian.
- (2b) Suppose that  $W_2 \cong Q_8$ . Then

$$(i) \ G = \frac{C_2}{\text{SL}(2, 3) \times M} \quad \text{or} \quad (ii) \ G = \text{SL}(2, 3) \times M$$

$$\text{or} \ (iii) \ G = Q_8 \times M,$$

where  $M$  is a metacyclic  $Z$ -group of odd order coprime to the order of  $\mathrm{SL}(2, 3)$  and  $Q_8$  respectively.

(2c) Suppose that  $W_2 \cong Q_{2^n}$  with  $n \geq 4$ . Then

$$G = \frac{C_2}{C_{2^{n-1}} \times M}$$

where  $M$  is a metacyclic  $Z$ -group of odd order and  $G_2 \cong Q_{2^n}$ . So  $G \cong W_2 \times M$ .

(3) If  $G$  is insoluble then

$$(i) \ G = \frac{C_2}{\mathrm{SL}(2, 5) \times M} \quad \text{or} \quad (ii) \ G = \frac{\mathrm{SL}(2, 5) \times M}{C_2}$$

where  $M$  is a metacyclic  $Z$ -group of odd order coprime to the order of  $\mathrm{SL}(2, 5)$ . The following results settle several of these cases immediately.

- (5.1) If  $G$  is a direct product of two groups  $H_1$  and  $H_2$  of coprime order then each torsion subgroup of  $V(\mathbb{Z}G)$  is isomorphic to a subgroup of  $G$  if and only if the same holds in  $V(\mathbb{Z}H_1)$  and  $V(\mathbb{Z}H_2)$ .
- (5.2) For  $Z$ -groups (ZC3) is valid [59].
- (5.3) If  $G$  has an abelian normal subgroup  $A$  such that  $G/A$  is abelian then each torsion subgroup of  $V(\mathbb{Z}G)$  is isomorphic to a subgroup of  $G$ . This follows from the small group ring sequence

$$0 \longrightarrow A \cong \mathbb{Z}G \cdot I(A)/I(G) \cdot I(A) \longrightarrow \mathbb{Z}G/I(G) \cdot I(A) \longrightarrow \mathbb{Z}G/A \longrightarrow 0$$

together with the well known facts that torsion subgroups of  $V(\mathbb{Z}G)$  are trivial provided  $G$  is an abelian torsion group and that by Marciniak and Sehgal [45]

$$V(\mathbb{Z}G) \cap (1 + I(A)I(G))$$

is torsion-free.

So Case 1 follows from 5.2, Cases 2a and 2b(iii) from 5.3. Moreover (ZC3) is valid for  $\mathrm{SL}(2, 3)$ ,  $\mathrm{SL}(2, 5)$  [18, Theorem 4.3] and for  $p$ -groups [58, 60]. Thus applying 5.1 and 5.2 we see that the theorem is valid in the Cases 2b(ii), 2c, 3(ii) resp.

*Case 3(i)* We first consider subgroups of  $V(\mathbb{Z}G)$  whose order divide 240. Then factoring out the normal metacyclic group  $M$  these subgroups are subgroups of  $V(\mathbb{Z}G/M)$ . Then  $G_1 = G/M$  is the double cover of the  $S_5$  occurring as Frobenius complement. (ZC1) holds for  $V(\mathbb{Z}G_1)$  [10], (SIP-C) holds for  $G$  by Kimmerle and Konovalov [42, Corollary 2.5] and a Sylow like theorem by Kimmerle and Margolis [43]. Thus because the order of a torsion subgroup  $H$  of  $V(\mathbb{Z}G_1)$  has to divide  $|G_1| = 240$  the remaining orders of subgroups of  $V(\mathbb{Z}G_1)$  are:

$$240, 120, 80, 60, 48, 40, 30, 24, 20, 15, 12, 10 \text{ and } 6.$$

If  $|H| = 240$  then  $H$  is a group basis. So we have to show that (IP) holds for  $G_1$ . This may be easily seen looking at the ordinary character table of  $G_1$ . The normal subgroup correspondence shows that  $\mathbb{Z}G_1 \cong \mathbb{Z}H$  implies that  $H$  has to be as well a double cover of  $S_5$ . But the character tables of the two double covers of  $S_5$  are different. Thus  $H \cong G_1$ . Each subgroup of even order of  $V(\mathbb{Z}G_1)$  contains the centre  $Z$  of  $G_1$  which is isomorphic to  $C_2$ . Thus subgroups of order 10 and 6 have to be cyclic. Factoring out  $Z$  we see that a subgroup  $H$  of order 120 has to map onto a subgroup  $\bar{H}$  of  $V(\mathbb{Z}S_5)$  of order 60. Because (ZC3) holds for  $S_5$  by Dokuchaev and Juriaans [17] we see that  $\bar{H} \cong A_5$ . There are only two insoluble groups of order 120 which map onto  $A_5$ . The group  $C_2 \times A_5$  has more than one involution. Thus it follows that  $H \cong \text{SL}(2, 5)$ .

Also there are no subgroups of order 80, 60, 30, 15 resp. because  $G_1/Z = S_5$  has no subgroups of order 40, 30, 15 resp.

Assume now that  $|H| = 40$ . Because a Sylow like theorem is valid in  $\mathbb{Z}G_1$  we know that  $H$  has a Sylow 2-subgroup  $H_2$  isomorphic to  $Q_8$  or to  $C_8$ . Suppose that  $H_2 \cong Q_8$ . Then  $H_2/Z \cong C_2 \times C_2$ . But  $S_5$  has no subgroup of type  $C_5 \times C_2 \times C_2$ . Thus  $H_2 \cong C_8$ . Assume that  $H$  is isomorphic to a dihedral group of order 40. Then  $H/Z \cong D_{10}$ . But  $H/Z$  has to be a Frobenius group of order 20 we conclude that  $H \cong C_5:C_8$  which is indeed isomorphic a subgroup of  $G_1$ . Similarly one sees that subgroups of order 20 are isomorphic to  $C_5:C_4$ , a subgroup of index 2 in  $C_5:C_8$ .

Let  $|H| = 48$ . Then we know that  $H/Z \cong S_4$ . Moreover  $Q_{16} \cong H_2$ . Suppose that  $H \cong Q_{16} \times C_3$ . Then  $H/Z \cong D_4 \times C_3 \not\cong S_4$ . Clearly  $H$  must have a normal subgroup of order 8 and contains a non-split central extension of  $A_4$  of order 24. This must be the binary tetrahedral group. Hence an examination of the groups of order 48 (e.g. with GAP [56]) shows that  $H$  is a binary octahedral group of order 48.

Similarly one sees that subgroups of order 24 are binary tetrahedral groups or  $C_3:Q_8$  which maps onto a subgroup of  $S_5$  isomorphic to  $S_3 \times C_2$ . Both occur as subgroups of  $G_1$ .

If  $H$  has order 12,  $H_2 \cong C_4$ . Thus  $H \cong C_{12}$  which must occur in  $G_1$  because (SIP-C) holds.

Now let  $H$  be a subgroup of  $V(\mathbb{Z}G)$  whose order is not divisible by 2, 3 or 5. Then by reduction modulo  $\text{SL}(2, 5)$  we get that  $H$  is isomorphic to a subgroup of  $\mathbb{Z}(G/\text{SL}(2, 5))$ . But  $\bar{G} = G/\text{SL}(2, 5)$  is a  $Z$ -group. Thus (ZC3) holds for  $\bar{G}$  and  $H$  is isomorphic to a subgroup of  $M$ .

Finally, if  $H$  is a subgroup such that  $H$  maps onto a subgroup  $\bar{H}$  of  $V(\mathbb{Z}\bar{G})$  (with  $\bar{G} = G/\text{SL}(2, 5)$ ) of even order  $m > 2$ . Then as before  $\bar{H}$  is isomorphic to a subgroup of  $\bar{G}$ . Let  $K$  be the image of  $H$  under the map onto  $\mathbb{Z}G_1$  and  $M_H$  the kernel of  $H$  under this map. Clearly  $M_H$  is isomorphic to the subgroup of index 2 of  $\bar{H}$ .

$H$  is a semidirect product of the form  $M_H \rtimes K$ . The action of  $K$  on  $M_H$  is determined modulo  $K/C_K(M_H)$  and therefore given by  $\bar{H}$ . Thus its isomorphism type is given by  $\bar{H}$  and  $M_H$ .

*Case 2b(i)* (ZC3) holds for the binary octahedral group by Dokuchaev and Juriaans [17, Theorem 4.7]. Now we can argue as in the case before and the proof is complete.

## 6 (SIP-C) for Almost Quasisimple Groups

In this section we will consider non-solvable groups and analyse how much the known methods can provide for our problems. We will focus on automorphic and central non-split extensions of non-abelian simple groups. Recall that a group  $G$  is called *almost simple* if there is a simple non-abelian group  $S$  such that  $G$  is isomorphic to a subgroup of the automorphism group of  $S$  containing the inner automorphisms of  $S$ , i.e.  $S \cong \text{Inn}(S) \leq G \leq \text{Aut}(S)$ . Moreover a group is called *quasisimple* if it is a central non-split extension of a non-abelian simple group  $S$ . We will call a group *almost quasisimple* if it is a central non-split extension of an almost simple group.

*Example 6.1* One of the smallest almost quasisimple groups for which the Zassenhaus Conjecture is open is the symmetric group of degree 6. The other group of the same size for which the Zassenhaus Conjecture is also open is the Mathieu group of degree 10. In this example we will concentrate on the example of a possible involution in  $V(\mathbb{Z}S_6)$  which is of particular interest since its existence would also provide a counterexample to the Torsionfree Kernels Question, cf. Problem 7.1 in Sect. 7, and even more so since the involution would lie in the kernel of the most natural representation of the group—the permutation representation on six points.

Let  $G$  be the symmetric group of degree 6. Denote by  $2a$  the conjugacy class of involutions in  $G$  which have no fixed points in the natural action, i.e. elements of cycle type  $(2, 2, 2)$ , by  $2b$  the class of involutions of cycle type  $(2, 2, 1, 1)$  and by  $2c$  the conjugacy class of involutions of cycle type  $(2, 1, 1, 1, 1)$ . The HeLP method is not sufficient to exclude the existence of an involution  $u \in V(\mathbb{Z}G)$  satisfying  $(\varepsilon_{2a}, \varepsilon_{2b}, \varepsilon_{2c}) = (-1, 1, 1)$ . Moreover the GAP function `HeLP_MultiplicitiesOfEigenvalues` provided by the HeLP package allows to construct an element of  $\mathbb{Q}G$  having the partial augmentations of  $u$ . Thus to show that  $u$  does not exist in  $\mathbb{Z}G$  it must be shown that the conjugacy class of this element in  $\mathbb{Q}G$  has trivial intersection with the  $\mathbb{Z}$ -order  $\mathbb{Z}G$  in  $\mathbb{Q}G$ . We give this element explicitly. For that let

$$\begin{aligned} \mathbb{Q}G \cong & \mathbb{Q} \times \mathbb{Q} \times \mathbb{Q}^{5 \times 5} \times \mathbb{Q}^{5 \times 5} \times \mathbb{Q}^{5 \times 5} \times \mathbb{Q}^{5 \times 5} \\ & \times \mathbb{Q}^{9 \times 9} \times \mathbb{Q}^{9 \times 9} \times \mathbb{Q}^{10 \times 10} \times \mathbb{Q}^{10 \times 10} \times \mathbb{Q}^{16 \times 16} \end{aligned}$$

be the Wedderburn decomposition of  $\mathbb{Q}G$ . Here the first factor of the form  $\mathbb{Q}^{5 \times 5}$  is understood to correspond to the representation of  $G$  obtained by cancelling out the trivial module from the 6-dimensional natural permutation module of  $G$ . Moreover the fourth factor of the form  $\mathbb{Q}^{5 \times 5}$  corresponds to the representation obtained from cancelling out the trivial module from the permutation module obtained by the other 6-transitive action of  $G$  (i.e. the one corresponding to the other conjugacy class of subgroups isomorphic to  $S_5$  in  $G$ ) and tensoring this module with the signum representation. Moreover the first factor of the form  $\mathbb{Q}^{10 \times 10}$  is understood to correspond to an irreducible representation of  $G$  which has character values  $-2$

on the class  $2a$ . In this setting a representative of the conjugacy class of  $u$  in  $\mathbb{Q}G$  is given by the following element:

$$\begin{aligned}
 & ((1), (1), \\
 & \text{diag}(1, 1, 1, 1, 1), \text{diag}(1, -1, -1, -1, -1), \text{diag}(1, -1, -1, -1, -1), \text{diag}(1, 1, 1, 1, 1), \\
 & \text{diag}(1, 1, 1, 1, 1, -1, -1, -1, -1), \text{diag}(1, 1, 1, 1, 1, -1, -1, -1, -1), \\
 & \text{diag}(1, 1, 1, 1, 1, 1, -1, -1, -1, -1), \text{diag}(1, 1, -1, -1, -1, -1, -1, -1, -1, -1), \\
 & \text{diag}(1, 1, 1, 1, 1, 1, 1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1)).
 \end{aligned}$$

Regarding (SIP-C) more can be achieved for almost quasisimple groups.

**Theorem 6.2** *Let  $G$  be an almost quasisimple group and let  $S$  be the only non-abelian composition factor of  $G$ . If  $S$  has smaller order than  $\text{PSL}(3, 3)$  then (SIP-C) holds for  $G$ .*

To prove Theorem 6.2 we will use the HeLP and the Lattice method. For some of the groups we study, we will use the following result.

**Lemma 6.3** *Let  $G$  and  $H$  be groups such that  $G$  contains a normal  $p$ -subgroup  $N$  and  $G/N \cong H$ . Assume moreover that when  $u \in \mathbb{V}(\mathbb{Z}H)$  is a torsion unit of order  $k$  then there exists an element  $h \in H$  of order  $k$  such that  $\varepsilon_h(u) \neq 0$ . Then (SIP-C) holds for  $G$ .*

*Proof* The proof follows the line of the proof of [23, Theorem]. Denote by  $\varphi : \mathbb{Z}G \rightarrow \mathbb{Z}G/N \cong \mathbb{Z}H$  the linear extension of the natural homomorphism from  $G$  to  $G/N$ ; as is common, it will also be denoted by  $\bar{\cdot}$ . We will apply the bar-convention to this ring homomorphism. So let  $u \in \mathbb{V}(\mathbb{Z}G)$  be a torsion unit. If  $\bar{u}$  has the same order as  $u$  then  $u$  has the same order as an element in  $G$  as by assumption  $\bar{u}$  has the same order as an element in  $H$ . Assume on the other hand that  $\bar{u}$  has strictly smaller order than the order of  $u$ . By assumption there is an element  $h \in H$  such that  $\varepsilon_h(\bar{u}) \neq 0$  and  $h$  has the same order as  $\bar{u}$ . So there is  $g \in G$  with  $\bar{g} = h$  and  $\varepsilon_g(u) \neq 0$ . Then the  $p'$ -part of the order of  $g$  and the  $p'$ -part of the order of  $u$  coincide. But by Proposition 2.4 the  $p$ -parts of the order of  $g$  and the order of  $u$  also coincide and so  $g$  has the same order as  $u$ .

*Proof (of Theorem 6.2)* There are ten non-abelian simple groups of order smaller than the order of  $\text{PSL}(3, 3)$  which give rise to 50 almost quasisimple groups. All of these groups are listed in the ATLAS and all their character and Brauer tables are available in the GAP character table library [13]. So in principle we can apply the HeLP package to all these groups. However it turns out that for central extension of the alternating and symmetric group of degree 7 these computations do not finish in a day, while finishing in a few minutes for all other groups. Among the groups not having  $A_7$  as a composition factor it turns out that HeLP is sufficient to prove (SIP-C) for all groups except groups containing  $A_6$  as a normal subgroup of index 2, the groups  $\text{PSL}(2, 16)$  and a group containing  $\text{PSL}(2, 16)$  as a normal subgroup of index 2. We handle these cases separately.



The two groups containing  $A_6$  as a normal subgroup of index 2 for which the HeLP method is not sufficient to prove (SIP-C) are  $\mathrm{PGL}(2, 9)$  and the Mathieu group of degree 10. For both these groups it remains to rule out the existence of units of order 6 in their normalized unit group of the integral group ring. This has been already done in [8] using the lattice method. For the group  $G = \mathrm{PSL}(2, 16)$  it remains to rule out the existence of units of order 6 in  $V(\mathbb{Z}G)$  and this has been also achieved using the lattice method in [6, Theorem C].

Next let  $G$  be a group of automorphisms of  $\mathrm{PSL}(2, 16)$  in which the group of inner automorphisms of  $\mathrm{PSL}(2, 16)$  has index 2. To prove (SIP-C) for  $G$  it remains to show that there are no torsion units of order 12 in  $V(\mathbb{Z}G)$ . HeLP provides us with two possibilities for the partial augmentations of a unit  $u \in V(\mathbb{Z}G)$  of order 12, both of which have the same partial augmentations on  $u^2$ . So ruling out the latter partial augmentations for units of order 6 will prove (SIP-C) for  $G$ . We will use the lattice method to do so. Denote by  $2a$  the conjugacy class of involutions in  $G$  which lies in  $\mathrm{PSL}(2, 16)$  and by  $3a$  the conjugacy class of elements of order 3 in  $G$ . The critical unit  $u$  of order 6 has partial augmentations equal to 0 on all conjugacy classes except  $2a$  and  $3a$  and furthermore  $(\varepsilon_{2a}(u), \varepsilon_{3a}(u)) = (4, -3)$  and the only class on which the partial augmentation of  $u^3$  does not vanish is  $2a$ . For an ordinary character  $\chi$  of  $G$  denote by  $\chi'$  its 3-modular reduction. Denote by  $\mathbf{1}$  the trivial character of  $G$ . There are irreducible complex characters  $\chi$  (a constituent of the lift of the Steinberg character of  $\mathrm{PSL}(2, 16)$ ) and  $\psi$  of degree 16 and 17 respectively such that  $\chi'$  is also irreducible as a 3-modular Brauer character and  $\psi' = \mathbf{1}' + \chi'$ . Both characters  $\chi$  and  $\psi$  only take integral values. Thus by a theorem of Fong [33, Corollary 10.13] there exists a 3-adically complete discrete valuation ring  $R$  unramified over the 3-adic integers such that there are  $R$ -representations  $D_\chi$  and  $D_\psi$  of  $G$  realizing  $\chi$  and  $\psi$  respectively. The partial augmentations of  $u$  and its powers allow us to compute the eigenvalues of  $u$  under these representations, e.g. using the GAP command `HeLP_MultiplicitiesOfEigenvalues` from the HeLP package. Denote by  $\zeta$  a primitive 3rd root of unity. Then

$$D_\chi(u) \sim \mathrm{diag}(1, 1, \zeta, \zeta, \zeta, \zeta^2, \zeta^2, \zeta^2, -1, -1, -1, -1, -\zeta, -\zeta, -\zeta^2, -\zeta^2),$$

$$D_\psi(u) \sim \mathrm{diag}(1, 1, 1, 1, 1, \zeta, \zeta, \zeta^2, \zeta^2, -\zeta, -\zeta, -\zeta, -\zeta, -\zeta^2, -\zeta^2, \zeta^2, -\zeta^2).$$

Denote by  $L_\chi$  and  $L_\psi$  full  $RG$ -lattices corresponding to  $D_\chi$  and  $D_\psi$  respectively. When an  $RG$ -lattice  $L$  is considered as an  $R\langle u \rangle$ -lattice it decomposes into a direct sum  $L \cong L^+ \oplus L^-$  such that all direct summands of  $L^+$  as  $R\langle u^3 \rangle$ -module are trivial while all direct summands of  $L^-$  as  $R\langle u^3 \rangle$ -module are non-trivial by Bächle and Margolis [8, Proposition 1.3]. Denote by  $\bar{\phantom{x}}$  the reduction modulo the maximal ideal of  $R$ , also with respect to modules, and let  $k$  be the field obtained by factoring out the maximal ideal from  $R$ . Then from the eigenvalues given above and [8, Proposition 1.4] we obtain that  $\bar{L}_\chi^-$  has exactly two indecomposable direct summands of  $k$ -dimension at least 2 while  $\bar{L}_\psi^-$  has four such summands. But from  $\psi' = \mathbf{1}' + \chi'$  we know that  $\bar{L}_\chi^-$  and  $\bar{L}_\psi^-$  must be isomorphic, since if  $S$  denotes a simple  $kG$ -module corresponding to  $\chi'$  then both these modules are isomorphic to

$S^-$ , i.e. the direct summand of  $S$  as  $k\langle\bar{u}\rangle$ -module consisting of the non-trivial direct summands of  $S$  as  $k\langle\bar{u}^3\rangle$ -module. This provides a final contradiction to the existence of  $u$ .

It remains to show (SIP-C) for non-split central extensions of the alternating and symmetric group of degree 7. Denote by  $2a$  the conjugacy class of double transpositions in  $A_7$  and  $S_7$ , i.e. elements of cycle type  $(2, 2, 1, 1, 1)$  and by  $3a$  and  $3b$  the conjugacy classes of elements of order 3 where the latter is of cycle type  $(3, 3, 1)$ . Note that all these three classes are the same in  $A_7$  and  $S_7$ . The Schur multiplier of both groups is cyclic of order 6, i.e. the maximal cyclic non-split extension is by a cyclic group of order 6. Thus by Lemma 6.3 it will be enough to show that when  $u \in V(\mathbb{Z}A_7)$  or  $u \in V(\mathbb{Z}S_7)$  is a unit of order  $k$  then there exists a  $g \in A_7$  or  $g \in S_7$  respectively such that  $\varepsilon_g(u) \neq 0$  and  $g$  is of order  $k$ . Applying HeLP to  $A_7$  and  $S_7$  one finds that if  $u$  is a unit not satisfying this condition then  $u$  is of order 6. Moreover  $u$  satisfies

$$(\varepsilon_{2a}(u), \varepsilon_{3a}(u), \varepsilon_{3b}(u)) \in \{(-2, 2, 1), (-2, 1, 2)\},$$

the partial augmentations of  $u$  at all other elements vanish and  $u^3$  is rationally conjugate to an element of  $2a$  while  $u^2$  is rationally conjugate to a 3-element in the conjugacy class  $C$  of  $G$  such that  $\varepsilon_C(u) = 1$ . In particular we obtain that showing the non-existence of such a unit in  $V(\mathbb{Z}S_7)$  implies the non-existence of such a unit in  $V(\mathbb{Z}A_7)$ .

So assume that  $G = S_7$  and assume that  $u$  is a unit of order 6 in  $V(\mathbb{Z}G)$  as described in the last paragraph. Again we will use the lattice method to show that  $u$  does not exist. The case when  $\varepsilon_{3a}(u) = 2$  will be called Case (i) and the case that  $\varepsilon_{3b}(u) = 2$  will be called (ii). Denote by  $\text{sig}$  the character of  $G$  corresponding to the signum representation. For an ordinary characters  $\chi$  denote once more by  $\chi'$  the corresponding 3-Brauer character.  $G$  possesses an irreducible 3-Brauer character  $\varphi$  of degree 13 and two irreducible characters  $\chi$  and  $\psi$  of degree 14 such that

$$\chi' = \mathbf{1}' + \varphi \quad \text{and} \quad \psi' = \text{sig}' + \varphi.$$

Let again  $D_\chi$  and  $D_\psi$  be  $R$ -representations of  $G$  corresponding to  $\chi$  and  $\psi$  respectively where  $R$  is a 3-adically complete discrete valuation ring unramified over the 3-adic integers. From the given partial augmentations of  $u$  and its powers we can compute the eigenvalues of  $u$  under these representations. Denote by  $\zeta$  a primitive 3rd root of unity.

In case (i) the eigenvalues of  $u$  under the representations of interest are:

$$D_\chi(u) \sim \text{diag}(1, 1, \zeta, \zeta, \zeta, \zeta^2, \zeta^2, \zeta^2, -1, -1, -\zeta, -\zeta, -\zeta^2, -\zeta^2),$$

$$D_\psi(u) \sim \text{diag}(1, 1, \zeta, \zeta, \zeta, \zeta^2, \zeta^2, \zeta^2, -1, -1, -1, -1, -\zeta, -\zeta^2).$$

While in case (ii) the eigenvalues of  $u$  are:

$$D_\chi(u) \sim \text{diag}(1, 1, \zeta, \zeta, \zeta^2, \zeta^2, \zeta^2, \zeta^2, -1, -1, -1, -1, -\zeta, -\zeta^2),$$

$$D_\psi(u) \sim \text{diag}(1, 1, \zeta, \zeta, \zeta^2, \zeta^2, \zeta^2, \zeta^2, -1, -1, -\zeta, -\zeta, -\zeta^2, -\zeta^2).$$

And moreover in both cases  $\text{sig}(u) = 1$ . Let  $L_\chi$  and  $L_\psi$  be full  $RG$ -lattices corresponding to  $\chi$  and  $\psi$  respectively and denote by  $\bar{\phantom{x}}$  the reduction modulo the maximal ideal of  $R$ . Let  $k$  be the quotient of  $R$  by its maximal ideal and let  $S$  be a simple  $kG$ -module corresponding to  $\varphi$ . When viewed as  $k\langle \bar{u} \rangle$ -module  $S$  decomposes into a direct sum  $S \cong S^+ \oplus S^-$  such that  $S^-$  contains all direct summands of  $S$  as  $k\langle \bar{u}^3 \rangle$ -module which are not trivial. An analogous decomposition applies for  $\bar{L}_\chi$  and  $\bar{L}_\psi$ . Then from [8, Propositions 1.3], the 3-modular decomposition behaviour and the eigenvalues of  $u$  under  $\mathbf{1}$  and  $\text{sig}$  we conclude that  $\bar{L}_\chi^- \cong S^- \cong \bar{L}_\psi^-$ . However from [8, Proposition 1.4] we know that in Case (i)  $\bar{L}_\chi^-$  has exactly two indecomposable summands of degree at least 2 while  $\bar{L}_\psi^-$  has only one such summand and in Case (ii)  $\bar{L}_\chi^-$  has exactly one indecomposable summands of degree at least 2 while  $\bar{L}_\psi^-$  has two such summands. This contradicts the existence of  $u$  and finishes the proof.

## 7 On Large Normal Subgroups

The following question has not yet been systematically studied, but it appears naturally in the questions mentioned above and might be of independent interest.

**Problem 7.1 (Torsionfree Kernel Question (TKQ))** Let  $K$  be a field of characteristic zero and  $G$  a finite group. Let  $B$  be a faithful block of  $KG$  and  $\pi$  be the projection from the units of  $KG$  onto  $B$ . Is  $\text{Ker } \pi \cap \mathbf{V}(\mathbb{Z}G)$  torsion free?

We remark that a big area in the study of units in integral group rings of finite groups is devoted to the study of “large subgroups” of  $\mathbf{V}(\mathbb{Z}G)$ . This involves questions on the generation of units of infinite order, free non-abelian subgroups of  $\mathbf{V}(\mathbb{Z}G)$ , generators of subgroups of finite index in  $\mathbf{V}(\mathbb{Z}G)$  and others. For more details we refer to recent monograph on these topics [34, 35].

In this section we present a little idea how torsion units may be used to find such large normal subgroups which are torsion free and of finite index. We also make transparent how HeLP and its companions may be used to answer (TKQ). A related question on the existence of a torsion free complement has been studied extensively in the 1980s.

Note that the hypothesis of the next lemma is valid if (ZC1) holds for  $\mathbb{Z}G$ .

**Lemma 7.2** *Let  $G$  be a finite group. Suppose that elements of prime order of  $\mathbf{V}(\mathbb{Z}G)$  are rationally conjugate to elements of  $G$ .*

Let  $R$  be a field of characteristic zero. Suppose that  $G$  is a subgroup of  $GL(n, R)$  and let  $\tau : \mathbb{Q}G \rightarrow M_n(R)$  be the ring homomorphism which is the unique extension of a given injective group homomorphism  $G \rightarrow GL(n, R)$ . Let  $K$  be the kernel of the group homomorphism  $\tau|_{U(\mathbb{Q}G)} : U(\mathbb{Q}G) \rightarrow GL(n, R)$ . Then  $K \cap V(\mathbb{Z}G)$  is torsion free.

*Proof* By assumption  $M_n(R)$  is a  $\mathbb{Q}$ -vector space. Thus  $\tau$  extends uniquely. If  $K$  is not torsion free then it has an element  $u$  of prime order  $p$ . By assumption there is a unit  $v \in \mathbb{Q}G$  such that  $v^{-1}uv = g \in G$ . But then  $\tau(u) = 1 \neq \tau(g)$ .

*Example 7.3* Let  $G = \text{PSL}(3, 3)$ . It can be checked using the GAP package HeLP that normalized units of  $V(\mathbb{Z}G)$  of order a prime  $r$  are rationally conjugate to elements of  $G$ , except possibly for  $r = 3$ . However in this situation the command HeLP\_MultiplicitiesOfEigenvalues can be used to see that no torsion unit of order 3 is contained in the kernel of any irreducible representation of  $G$  of degree larger than 1. Hence the previous lemma can be applied with any irreducible representation of  $G$  different from the principal one and (TKQ) has a positive answer for each block of  $\mathbb{C}G$ , while (ZC1) is unknown for this group.

**Proposition 7.4** *Let  $G$  be a finite group. Let  $B \cong M_n(\mathbb{Q})$  be a faithful block of the Wedderburn decomposition of  $\mathbb{Q}G$ . Assume that either*

- i)  $p$  is an odd prime or  $p = 4$  or*
- ii)  $p = 2$  and  $|G|$  is odd.*

*Let  $\pi$  be the projection of  $U(\mathbb{Q}G)$  onto  $B$ . Then*

$$(\text{Ker } \pi \cap V(\mathbb{Z}G)) \cdot ((1 + p\mathbb{Z}G) \cap V(\mathbb{Z}G))$$

*is a torsion free normal subgroup of  $V(\mathbb{Z}G)$  of finite index.*

*Proof* We may choose an integral representation of  $G$ , i.e.  $\pi$  maps  $\mathbb{Z}G$  into  $GL(n, \mathbb{Z})$ . Consider the reduction  $\kappa : GL(n, \mathbb{Z}) \rightarrow GL(n, \mathbb{Z}/p\mathbb{Z})$ . By a classical result of Minkowski [19, Lemma 9] the map  $\kappa$  is injective on torsion elements if  $p \neq 2$ . For  $p = 2$  the kernel is an elementary-abelian 2-group. Thus under the assumptions each torsion element of  $V(\mathbb{Z}G)$  injects into the finite group  $GL(n, \mathbb{Z}/p\mathbb{Z})$ . This finishes the proof.

The preceding construction may be applied especially in the situation of symmetric groups (with respect to almost each non-trivial block) because  $\mathbb{Q}$  is a splitting field for  $S_n$  and only few blocks are not faithful.

**Proposition 7.5** *Let  $G$  be a minimal simple group. Then  $V(\mathbb{Z}G)$  has a torsion-free normal subgroup of finite index constructed as in Proposition 7.4.*

*Proof* In [57], Thompson proved that a minimal simple group is isomorphic to  $\text{PSL}(2, q)$ ,  $\text{Sz}(q)$  or  $\text{PSL}(3, 3)$ . For the two series generic character tables are known, cf. e.g. [32, XI, §5]. Let  $G$  be one of these groups and let  $\chi$  be the Steinberg character. The character  $\chi$  takes the same value  $t$  on all elements of order a fixed

prime  $r$ . Let  $x_1, \dots, x_s$  be representatives of the conjugacy classes of  $G$  of elements of order  $r$  and let  $u \in V(\mathbb{Z}G)$  be a torsion unit of order  $r$ . Then

$$\chi(u) = \sum_{j=1}^s \varepsilon_{x_j}(u) \chi(x_j) = t \neq \chi(1).$$

Hence  $u$  is not in the kernel of a representation  $D$  affording  $\chi$ .  $D$  can be realized over the rationals, hence by Proposition 7.4 we obtain a torsion-free normal subgroup of  $V(\mathbb{Z}G)$  of finite index.

In case  $G = \text{PSL}(3, 3)$  Example 7.3 can be used to find a suitable block.

## References

1. 4ti2 team. 4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces. Available at [www.4ti2.de](http://www.4ti2.de) (2015). Version 1.6.7
2. S.A. Amitsur, Finite subgroups of division rings. *Trans. Am. Math. Soc.* **80**, 361–386 (1955)
3. A. Bächle, W. Kimmerle, On torsion subgroups in integral group rings of finite groups. *J. Algebra* **326**, 34–46 (2011)
4. A. Bächle, L. Margolis, HeLP – a GAP package for torsion units in integral group rings, 6 pp. (2015, submitted). [arXiv:1507.08174\[math.RT\]](https://arxiv.org/abs/1507.08174)
5. A. Bächle, L. Margolis, Torsion subgroups in the units of the integral group ring of  $\text{PSL}(2, p^3)$ . *Arch. Math. (Basel)* **105**(1), 1–11 (2015)
6. A. Bächle, L. Margolis, On the prime graph question for integral group rings of 4-primary groups II. 17 pp. (2016, submitted). [arXiv:1606.01506\[math.RT\]](https://arxiv.org/abs/1606.01506)
7. A. Bächle, L. Margolis, On the prime graph question for integral group rings of 4-primary groups I. *Int. J. Algebra Comput.* **27**(6), 731–767 (2017)
8. A. Bächle, L. Margolis, Rational conjugacy of torsion units in integral group rings of non-solvable groups. *Proc. Edinb. Math. Soc. (2)* **60**(4), 813–830 (2017)
9. A. Bächle, A. Herman, A. Konovalov, L. Margolis, G. Singh, The status of the Zassenhaus conjecture for small groups. *Exp. Math.* 6 pp. (2017). <https://dx.doi.org/10.1080/10586458.2017.1306814>
10. V.A. Bovdi, M. Hertweck, Zassenhaus conjecture for central extensions of  $S_5$ . *J. Group Theory* **11**(1), 63–74 (2008)
11. V. Bovdi, C. Höfert, W. Kimmerle, On the first Zassenhaus conjecture for integral group rings. *Publ. Math. Debr.* **65**(3–4), 291–303 (2004)
12. V.A. Bovdi, A.B. Konovalov, S. Linton, Torsion units in integral group ring of the Mathieu simple group  $M_{22}$ . *LMS J. Comput. Math.* **11**, 28–39 (2008)
13. T. Breuer, The GAP character table library, version 1.2.1 (2012). <http://www.math.rwth-aachen.de/~Thomas.Breuer/ctbllib>. GAP package
14. W. Bruns, B. Ichim, C. Söger, The power of pyramid decomposition in Normaliz. *J. Symb. Comput.* **74**, 513–536 (2016)
15. J.A. Cohn, D. Livingstone, On the structure of group algebras. I. *Can. J. Math.* **17**, 583–593 (1965)
16. C.W. Curtis, I. Reiner, *Methods of Representation Theory*, vol. I. Wiley Classics Library (Wiley, New York, 1990). With applications to finite groups and orders, Reprint of the 1981 original, A Wiley-Interscience Publication
17. M.A. Dokuchaev, S.O. Juriaans, Finite subgroups in integral group rings. *Can. J. Math.* **48**(6), 1170–1179 (1996)

18. M.A. Dokuchaev, S.O. Juriaans, C. Polcino Milies, Integral group rings of Frobenius groups and the conjectures of H. J. Zassenhaus. *Commun. Algebra* **25**(7), 2311–2325 (1997)
19. R.M. Guralnick, M. Lorenz, Orders of finite groups of matrices, in *Groups, Rings and Algebras*. Contemporary Mathematics, vol. 420 (American Mathematical Society, Providence, RI, 2006), pp. 141–161
20. A. Herman, G. Singh, Revisiting the Zassenhaus conjecture on torsion units for the integral group rings of small groups. *Proc. Indian Acad. Sci. Math. Sci.* **125**(2), 167–172 (2015)
21. M. Hertweck, A counterexample to the isomorphism problem for integral group rings. *Ann. Math. (2)* **154**(1), 115–138 (2001)
22. M. Hertweck, Partial augmentations and Brauer character values of torsion units in group rings. Manuscript, 16 pp. (2007). [arXiv:math.RA/0612429v2\[math.RA\]](https://arxiv.org/abs/math.RA/0612429v2)
23. M. Hertweck, The orders of torsion units in integral group rings of finite solvable groups. *Commun. Algebra* **36**(10), 3585–3588 (2008)
24. M. Hertweck, Torsion units in integral group rings of certain metabelian groups. *Proc. Edinb. Math. Soc. (2)* **51**(2), 363–385 (2008)
25. M. Hertweck, Unit groups of integral finite group rings with no noncyclic Abelian finite  $p$ -subgroups. *Commun. Algebra* **36**(9), 3224–3229 (2008)
26. M. Hertweck, Units of  $p$ -power order in principal  $p$ -blocks of  $p$ -constrained groups. *J. Algebra* **464**, 348–356 (2016)
27. M. Hertweck, W. Kimmerle, On principal blocks of  $p$ -constrained groups. *Proc. Lond. Math. Soc. (3)* **84**(1), 179–193 (2002)
28. M. Hertweck, C.R. Höfert, W. Kimmerle, Finite groups of units and their composition factors in the integral group rings of the group  $\mathrm{PSL}(2, q)$ . *J. Group Theory* **12**(6), 873–882 (2009)
29. G. Higman, Units in group rings. D. Phil. thesis, Oxford University (1940)
30. C. Höfert, Die erste Vermutung von Zassenhaus für Gruppen kleiner Ordnung. Diplomarbeit, Universität Stuttgart (2004)
31. C. Höfert, W. Kimmerle, On torsion units of integral group rings of groups of small order, in *Groups, Rings and Group Rings*. Lecture Notes in Pure and Applied Mathematics, vol. 248 (Chapman & Hall/CRC, Boca Raton, FL, 2006), pp. 243–252
32. B. Huppert, N. Blackburn, *Finite Groups. III*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 243 (Springer, Berlin-New York, 1982)
33. I.M. Isaacs, *Character Theory of Finite Groups*. Pure and Applied Mathematics, vol. 69 (Academic [Harcourt Brace Jovanovich, Publishers], New York-London, 1976)
34. E. Jespers, Á. del Río, *Group Ring Groups. Volume 1: Orders and Generic Constructions of Units* (De Gruyter, Berlin, 2016)
35. E. Jespers, Á. del Río, *Group Ring Groups. Volume 2: Structure Theorems of Unit Groups* (De Gruyter, Berlin, 2016)
36. S.O. Juriaans, C. Polcino Milies, Units of integral group rings of Frobenius groups. *J. Group Theory* **3**(3), 277–284 (2000)
37. W. Kimmerle, Beiträge zur ganzzahligen Darstellungstheorie endlicher Gruppen. Bayreuth. Math. Schr. **36**, 139 (1991)
38. W. Kimmerle, On the prime graph of the unit group of integral group rings of finite groups, in *Groups, Rings and Algebras*. Contemporary Mathematics, vol. 420 (American Mathematical Society, Providence, RI, 2006), pp. 215–228
39. W. Kimmerle, Torsion units in integral group rings of finite insoluble groups. Oberwolfach Rep. **4**(4), 3229–3230 (2007). Abstracts from the mini-workshop held November 25–December 1, 2007, organized by Eric Jespers, Zbigniew Marciniak, Gabriele Nebe, and Wolfgang Kimmerle
40. W. Kimmerle, Sylow like theorems for  $V(\mathbb{Z}G)$ . *Int. J. Group Theory* **4**(4), 49–59 (2015)
41. W. Kimmerle, A.B. Kononov, Recent advances on torsion subgroups of integral group rings, in *Groups St. Andrews 2013*. London Mathematical Society Lecture Note Series, vol. 422 (Cambridge University Press, Cambridge, 2015), pp. 331–347

42. W. Kimmerle, A.B. Kononov, On the Gruenberg - Kegel graph of integral group rings of finite groups. *Int. J. Algebra Comput.* **27**(6), 619–631 (2017)
43. W. Kimmerle, L. Margolis,  $p$ -Subgroups of units in  $\mathbb{Z}G$ , in *Groups, Rings, Group Rings. Contemporary Mathematics*, vol. 688 (American Mathematical Society, Providence, RI, 2017), pp. 169–179
44. I.S. Luthar, I.B.S. Passi, Zassenhaus conjecture for  $A_5$ . *Proc. Indian Acad. Sci. Math. Sci.* **99**(1), 1–5 (1989)
45. Z. Marciniak, S.K. Sehgal, The unit group of  $1 + \Delta(G)\Delta(A)$  is torsion free. *J. Group Theory* **6**(2), 223–228 (1987)
46. Z. Marciniak, J. Ritter, S.K. Sehgal, A. Weiss, Torsion units in integral group rings of some metabelian groups. II. *J. Number Theory* **25**(3), 340–352 (1987)
47. L. Margolis, A Sylow theorem for the integral group ring of  $\text{PSL}(2, q)$ . *J. Algebra* **445**, 295–306 (2016)
48. L. Margolis, Subgroup isomorphism problem for units of integral group rings. *J. Group Theory* **20**(2), 289–307 (2017)
49. L. Margolis, M. Serrano, Á. del Río, Zassenhaus conjecture on torsion units holds for  $\text{PSL}(2, p)$  with  $p$  a Fermat or Mersenne prime. Preprint, 32 pp. (2016). [arXiv:1608.05797\[math.RA\]](https://arxiv.org/abs/1608.05797)
50. D.S. Passman, *Permutation Groups* (W. A. Benjamin, New York-Amsterdam, 1968)
51. K.W. Roggenkamp, The isomorphism problem for integral group rings of finite groups, in *Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990)* (Mathematical Society of Japan, Tokyo, 1991), pp. 369–380
52. K.W. Roggenkamp, L.L. Scott, Isomorphisms of  $p$ -adic group rings. *Ann. Math. (2)* **126**(3), 593–647 (1987)
53. L.L. Scott, Recent progress on the isomorphism problem, in *The Arcata Conference on Representations of Finite Groups (Arcata, CA, 1986)*. Proceedings of Symposia in Pure Mathematics, vol. 47 (American Mathematical Society, Providence, RI, 1987), pp. 259–273
54. S.K. Sehgal, *Units in Integral Group Rings*. Pitman Monographs and Surveys in Pure and Applied Mathematics, vol. 69 (Longman Scientific & Technical, Harlow, 1993)
55. M. Shirvani, B.A.F. Wehrfritz, *Skew Linear Groups*. London Mathematical Society Lecture Note Series, vol. 118 (Cambridge University Press, Cambridge, 1986)
56. The GAP Group. GAP – Groups, Algorithms, and Programming, Version 4.8.5 (2016). <http://www.gap-system.org>
57. J.G. Thompson, Nonsolvable finite groups all of whose local subgroups are solvable. *Bull. Am. Math. Soc.* **74**, 383–437 (1968)
58. G. Thompson, Subgroup rigidity in finite - dimensional group algebras over  $p$ -groups. *Trans. Am. Math. Soc.* **341**, 423–447 (1994)
59. A. Valenti, Torsion units in integral group rings. *Proc. Am. Math. Soc.* **120**(1), 1–4 (1994)
60. A. Weiss, Rigidity of  $p$ -adic  $p$ -torsion. *Ann. Math. (2)* **127**(2), 317–332 (1988)
61. H.J. Zassenhaus, On the torsion units of finite group rings, in *Studies in Mathematics (In Honor of A. Almeida Costa)* (Instituto de Alta Cultura, Lisbon, 1974), pp. 119–126
62. È.M. Žmud, G.Č. Kurennoi, The finite groups of units of an integral group ring. *Vestnik Har'kov. Gos. Univ.* **1967**(26), 20–26 (1967)

# A Constructive Approach to the Module of Twisted Global Sections on Relative Projective Spaces



Mohamed Barakat and Markus Lange-Hegermann

**Abstract** The ideal transform of a graded module  $M$  is known to compute the module of twisted global sections of the sheafification of  $M$  over a relative projective space. We introduce a second description motivated by the relative BGG-correspondence. However, our approach avoids the full BGG-correspondence by replacing the Tate resolution with the computationally more efficient purely linear saturation and the Castelnuovo-Mumford regularity with the often enough much smaller linear regularity. This paper provides elementary, constructive, and unified proofs that these two descriptions compute the (truncated) modules of twisted global sections. The main argument relies on an established characterization of Gabriel monads.

**Keywords** Serre quotient category • Reflective localization of Abelian categories • Gabriel monad • (Truncated) Module of twisted global sections • Direct image functor • Linear regularity • Gröbner bases • Saturation

**Subject Classifications** 13D02, 13D07, 13D45, 13P10, 13P20, 18E10, 18E35, 18A40, 68W30, 14Q99

## 1 Introduction

We consider coherent sheaves over the projective space  $\mathbb{P}_B^n$  for a suitable ring  $B$ . Any such coherent sheaf  $\mathcal{F}$  can be described by a graded module over the polynomial ring  $S := B[x_0, \dots, x_n]$ . Even though this representation is not unique, among the

---

M. Barakat (✉)  
University of Siegen, 57068 Siegen, Germany  
e-mail: [mohamed.barakat@uni-siegen.de](mailto:mohamed.barakat@uni-siegen.de)

M. Lange-Hegermann  
Lehrstuhl B für Mathematik, RWTH Aachen University, Aachen, Germany  
e-mail: [markus.lange.hegermann@rwth-aachen.de](mailto:markus.lange.hegermann@rwth-aachen.de)



different graded  $S$ -modules representing  $\mathcal{F}$  there is the distinguished representative  $H_{\bullet}^0(\mathcal{F}) := \bigoplus_{p \in \mathbb{Z}} H^0(\mathbb{P}_B^n, \mathcal{F}(p))$ , the module of twisted global sections. In general the module of twisted global sections is not finitely generated, but any of its truncations  $H_{\geq d}^0(\mathcal{F})$  is.

In this paper we treat the functor  $H_{\geq d}^0$  constructively. It is well-known that the ideal transform functor computes  $H_{\geq d}^0$ , which we state as Theorem 4.6. Furthermore, in Theorem 6.7 we present a new recursive algorithm, inspired by [10, 11], to compute  $H_{\geq d}^0$ . The Tate resolution in loc. cit. incorporates all higher cohomologies,<sup>1</sup> whereas our new algorithm introduces a smaller complex, called the purely linear saturation, which is tailored to  $H_{\bullet}^0$  and computationally more efficient as it discards all higher cohomologies. This paper presents a categorical setup which yields unified elementary proofs of both theorems.

A central notion in this paper is that of the linear regularity. We use it in Corollary 4.7 for the convergence analysis of the inductive limit defining the ideal transform and in Corollary 6.8 to give the number of recursion steps in our new algorithm. Like the Tate resolution, the Castelnuovo-Mumford regularity incorporates all higher cohomologies. And again, the linear regularity is an adaption to  $H_{\bullet}^0$  which discards all higher cohomologies. It follows that the linear regularity is smaller or at most equal to the Castelnuovo-Mumford regularity. This provides another reason why computing the purely linear saturation is more efficient than computing the Tate resolution.

One application is the computation of global sections. More precisely, let  $\mathcal{F}$  be a coherent sheaf on  $\mathbb{P}_B^n$  and  $\pi : \mathbb{P}_B^n \rightarrow \text{Spec } B$  the natural projection.<sup>2</sup> The direct image sheaf  $H^0(\mathcal{F}) := \pi_* \mathcal{F}$  over  $\text{Spec } B$  is the degree zero part (cf. Algorithm 3.1) of  $H_{\geq 0}^0(\mathcal{F})$ . For example, if  $\mathcal{O}_X \subset \mathcal{O}_{\mathbb{P}_B^n}$  denotes the structure sheaf of a subscheme  $X \subset \mathbb{P}_B^n$ , then computing  $\pi_* \mathcal{O}_X$  is the geometric form of eliminating all  $n + 1$  homogeneous coordinates  $x_0, \dots, x_n$  from the defining equations of  $X \subset \mathbb{P}_B^n$ .

A computer model of the Abelian category  $\mathcal{Coh} \mathbb{P}_B^n$  of coherent sheaves on  $\mathbb{P}_B^n$  must incorporate the objects *and* the morphisms.<sup>3</sup> We represent an object  $\mathcal{F} \in \mathcal{Coh} \mathbb{P}_B^n$  by a finitely presented graded  $S$ -modules  $M$ , such that  $\mathcal{F} := \bar{M}$  is the sheafification of  $M$ .

The ideal transform is defined to be the inductive limit  $D_{\mathfrak{m}}(M)$  of the graded modules  $\text{Hom}_{\bullet}(\mathfrak{m}^{\ell}, M)$ , where  $\mathfrak{m} = \langle x_0, \dots, x_n \rangle \triangleleft S$  is the irrelevant ideal. The equivalence  $H_{\bullet}^0(\tilde{\cdot}) \simeq \varinjlim_{\ell} \text{Hom}_{\bullet}(\mathfrak{m}^{\ell}, \cdot)$ , reproved elementarily as Theorem 4.6, implies that  $H_{\bullet}^q(\tilde{\cdot}) \simeq \varinjlim_{\ell} \text{Ext}_{\bullet}^q(\mathfrak{m}^{\ell}, \cdot)$  for the higher derived cohomology functors [8, 20.4.4].

<sup>1</sup>See the introduction to Sect. 6 and Remark 7.4.

<sup>2</sup>The base  $\text{Spec } B$  might even serve as the ambient space of a geometric quotient, e.g., if  $B$  is the Cox ring of a toric variety.

<sup>3</sup>The more involved modeling of the morphisms is of no relevance for this paper (cf. [5]).

Another description of the cohomology functors  $H^q_\bullet$  arose from the BGG-correspondence [7]. It is a triangle equivalence between the bounded derived category of  $\mathcal{Coh} \mathbb{P}_B^n$  (originally over a base field  $B$ ) and the stable category of finitely generated graded  $E$ -modules over the exterior algebra  $E$ , the Koszul dual  $B$ -algebra of  $S$ . Since  $E$  is a Frobenius algebra, this stable category is easily seen to be triangle equivalent to the homotopy category of so-called Tate complexes. A constructive description of the composition of these two triangle equivalences was given in [9, 11]. The treatment of the relative BGG-correspondence in [10] does not only describe the coherent sheaf cohomologies  $H^q(\tilde{M}) = R^q\pi_*\tilde{M}$  as  $B$ -modules, but also provides a concrete realization of the direct image complex  $R\pi_*\tilde{M}$ . However, in this approach even the computation of global sections  $H^0(\tilde{M})$  in the relative case relies a priori on the entire Tate resolution.

The bottom complex  $E_1^{\geq d,0}(\mathbf{T}^{\geq d}(M))$  on the first page of the spectral sequence of the Tate resolution  $\mathbf{T}^{\geq d}(M)$  is a linear complex which corresponds to  $H_{\geq d}^0(\tilde{M})$ . We define a new so-called purely linear saturation functor  $\mathbf{S}^{\geq d}$ , which is computationally far more economic than the Tate functor  $\mathbf{T}^{\geq d}$ . In Theorem 6.7, Proposition 7.2, and Corollary 7.3 we prove that  $\mathbf{S}^{\geq d}$  computes  $H_{\geq d}^0(\tilde{M})$ , and hence  $E_1^{\geq d,0}(\mathbf{T}^{\geq d}(M))$ . The point is that we can compute  $\mathbf{S}^{\geq d}$  *without* computing  $\mathbf{T}^{\geq d}$ . This statement is not obvious in the relative case (cf. Remark 7.4). Furthermore, the linear regularity of  $M$  gives the precise number of recursion steps needed to achieve saturation. Since computing  $\mathbf{S}^{\geq d}$  relies on Gröbner bases over the exterior algebra  $E$  of *finite* rank over  $B$  the involved algorithms are, for many examples, faster than the ones for the ideal transform. The latter involve Gröbner bases over the polynomial ring  $S$  of *infinite* rank over  $B$ .

In order to develop a unified proof that both functors  $D_{m,\geq d}$  and  $\mathbf{S}^{\geq d}$  compute  $H_{\geq d}^0(\tilde{\phantom{M}})$  we need an appropriate categorical setup. Abstractly, the category  $\mathcal{Coh} \mathbb{P}_B^n$  of coherent sheaves on  $\mathbb{P}_B^n$  is equivalent to the Serre quotient category  $\mathcal{A}/\mathcal{C}$  of the Abelian category  $\mathcal{A}$  of finitely presented graded  $S$ -modules modulo a certain subcategory  $\mathcal{C}$ . The necessary categorical language is summarized in Sect. 2. In Sect. 7 we show that the categories  $\mathcal{A}$  and  $\mathcal{C}$  can be replaced by their respective full subcategories of modules which vanish in degrees  $< d$  for an arbitrary but fixed  $d \in \mathbb{Z}$  (cf. Proposition 7.1). The  $\mathcal{A}$ -endofunctor  $M \mapsto H_{\geq d}^0(\tilde{M})$  is a special case of what we call a Gabriel monad, which we characterized in [3] by a short set of properties. By verifying this short list of properties for the two functors  $D_{m,\geq d}$  and  $\mathbf{S}^{\geq d}$  we prove that they compute  $H_{\geq d}^0(\tilde{\phantom{M}})$ .

Two further applications rely on constructivity of the Gabriel monad, and now become algorithmically accessible for the category  $\mathcal{Coh} \mathbb{P}_B^n$ : First, the Serre quotient category  $\mathcal{A}/\mathcal{C}$  becomes constructively Abelian once  $\mathcal{A}$  is constructively Abelian [5, Appendix D].<sup>4</sup> Second, the computability of the bivariate Hom and Ext<sup>*i*</sup> functors in  $\mathcal{A}/\mathcal{C}$  now reduces to the computability of Hom and Ext<sup>*i*</sup> in  $\mathcal{A}$  (modulo a directed colimit process if  $i > 0$ ) [6].

---

<sup>4</sup>However, the approach using Gabriel morphisms in [5] seems computationally faster.

## 2 Preliminaries on Serre Quotient Categories

A non-empty full subcategory  $\mathcal{C}$  of an Abelian category  $\mathcal{A}$  is called **thick** if it is closed under passing to subobjects, factor objects, and extensions. In this case the **Serre quotient category**  $\mathcal{A}/\mathcal{C}$  is a category with the same objects as  $\mathcal{A}$  and Hom-groups defined by the directed colimit

$$\mathrm{Hom}_{\mathcal{A}/\mathcal{C}}(M, N) := \varinjlim_{\substack{M' \leq M, N' \leq N, \\ M/M', N'/\mathcal{C} \in \mathcal{C}}} \mathrm{Hom}_{\mathcal{A}}(M', N/N').$$

The **canonical functor**  $\mathcal{Q} : \mathcal{A} \rightarrow \mathcal{A}/\mathcal{C}$  is defined to be the identity on objects and maps a morphism  $\varphi \in \mathrm{Hom}_{\mathcal{A}}(M, N)$  to its class in the directed colimit  $\mathrm{Hom}_{\mathcal{A}/\mathcal{C}}(M, N)$ . The category  $\mathcal{A}/\mathcal{C}$  is Abelian and the canonical functor  $\mathcal{Q} : \mathcal{A} \rightarrow \mathcal{A}/\mathcal{C}$  is exact. An object  $M \in \mathcal{A}$  is called  **$\mathcal{C}$ -saturated** if  $\mathrm{Ext}_{\mathcal{A}}^0(C, M) \cong \mathrm{Ext}_{\mathcal{A}}^1(C, M) \cong 0$  for all  $C \in \mathcal{C}$ , i.e.,  $M$  has no nonzero subobjects in  $\mathcal{C}$  and every extension of an object  $C \in \mathcal{C}$  by  $M$  is trivial. Denote by  $\mathrm{Sat}_{\mathcal{C}}(\mathcal{A}) \subset \mathcal{A}$  the full subcategory of  $\mathcal{C}$ -saturated objects and by  $\iota : \mathrm{Sat}_{\mathcal{C}}(\mathcal{A}) \hookrightarrow \mathcal{A}$  its full embedding. A complex  $F$  in  $\mathrm{Sat}_{\mathcal{C}}(\mathcal{A})$  is exact if and only if  $\iota(F)$  has homology in  $\mathcal{C}$ .

A thick subcategory  $\mathcal{C} \subset \mathcal{A}$  is called **localizing** if the canonical functor  $\mathcal{Q} : \mathcal{A} \rightarrow \mathcal{A}/\mathcal{C}$  admits a right adjoint  $\mathcal{S} : \mathcal{A}/\mathcal{C} \rightarrow \mathcal{A}$ , called the **section functor** of  $\mathcal{Q}$ . In this case, the image of  $\mathcal{S}$  is contained in  $\mathrm{Sat}_{\mathcal{C}}(\mathcal{A})$  and  $\mathcal{A}/\mathcal{C} \xrightarrow{\mathcal{S}} \mathcal{S}(\mathcal{A}/\mathcal{C}) \hookrightarrow \mathrm{Sat}_{\mathcal{C}}(\mathcal{A})$  are equivalences of categories. The Hom-adjunction

$$\mathrm{Hom}_{\mathcal{A}/\mathcal{C}}(\mathcal{Q}(M), \mathcal{Q}(N)) \cong \mathrm{Hom}_{\mathcal{A}}(M, (\mathcal{S} \circ \mathcal{Q})(N))$$

allows to compute Hom-groups in  $\mathcal{A}/\mathcal{C}$  if they are computable in  $\mathcal{A}$  and the monad  $\mathcal{S} \circ \mathcal{Q}$  is computable. In particular, this avoids computing the directed colimit in the definition of  $\mathrm{Hom}_{\mathcal{A}/\mathcal{C}}$ . We call any monad equivalent to  $\mathcal{S} \circ \mathcal{Q}$  a **Gabriel monad** (of  $\mathcal{A}$  w.r.t.  $\mathcal{C}$ ). The following theorem characterizes Gabriel monads.

**Theorem 2.1** ([3, Thm. 3.6]<sup>5</sup>) *Let  $\mathcal{C} \subset \mathcal{A}$  be a localizing subcategory of the Abelian category  $\mathcal{A}$  and  $\iota : \mathrm{Sat}_{\mathcal{C}}(\mathcal{A}) \hookrightarrow \mathcal{A}$  the full embedding. An endofunctor  $\mathcal{W} : \mathcal{A} \rightarrow \mathcal{A}$  together with a natural transformation  $\eta : \mathrm{Id}_{\mathcal{A}} \rightarrow \mathcal{W}$  is a Gabriel monad (of  $\mathcal{A}$  w.r.t.  $\mathcal{C}$ ) if and only if the following five conditions hold:*

1.  $\mathcal{C} \subset \ker \mathcal{W}$ ,
2.  $\mathcal{W}(\mathcal{A}) \subset \mathrm{Sat}_{\mathcal{C}}(\mathcal{A})$ ,
3. the corestriction  $\mathrm{co}\text{-res}_{\mathrm{Sat}_{\mathcal{C}}(\mathcal{A})} \mathcal{W}$  of  $\mathcal{W}$  to  $\mathrm{Sat}_{\mathcal{C}}(\mathcal{A})$  is exact,
4.  $\eta \mathcal{W} = \mathcal{W} \eta : \mathcal{W} \rightarrow \mathcal{W}^2$ , and
5.  $\eta \iota : \iota \rightarrow \mathcal{W} \iota$  is a natural isomorphism.

---

<sup>5</sup>Thm. 4.6 in arXiv version.

In Sects. 4 and 6 we utilize this theorem to prove that certain functors are Gabriel monads of the category of coherent sheaves on the relative projective space  $\mathbb{P}_B^n$ , and thus compute the (truncated) module of twisted global sections. However, this theorem, abstract as it is, can be applied to categories of coherent sheaves of more general schemes.

### 3 Graded Modules over the Free Polynomial Ring

For the rest of the paper let  $B$  denote a Noetherian commutative ring with 1,  $V$  a free  $B$ -module of rank  $n + 1$ ,  $W := V^* = \text{Hom}_B(V, B)$  its  $B$ -dual, and  $x_0, \dots, x_n$  a free generating set of the  $B$ -module  $W$ . Set

$$S := \text{Sym}_B(W) = B[V] = B[x_0, \dots, x_n]$$

to be the free polynomial ring over  $B$  in the  $n + 1$  indeterminates  $x_0, \dots, x_n$ . Setting  $\deg(x_j) = 1$  turns  $S$  into a positively graded ring  $S = \bigoplus_{i \geq 0} S_i$  where  $S_i$  is the set of homogeneous polynomials of degree  $i$  in  $S$ . Define the irrelevant ideal

$$\mathfrak{m} := S_{>0} = \langle x_0, \dots, x_n \rangle \triangleleft S.$$

The isomorphism  $B = S_0 \cong S/\mathfrak{m}$  endows  $B$  with a natural graded  $S$ -module structure.

To make the statements of this paper constructive, the ring  $S$  needs to have a Gröbner bases algorithm. This is the case if  $B$  has effective coset representatives [1, §4.3], i.e., for every ideal  $I \subset B$  we can determine a set  $T$  of coset representatives of  $B/I$ , such that for every  $b \in B$  we can compute a unique  $t \in T$  with  $b + I = t + I$ .

We denote by  $S\text{-mod}$  the category of (non-graded) finitely presented  $S$ -modules and by  $S\text{-grmod}$  the category of finitely presented *graded*  $S$ -modules. Further we denote by

$$S\text{-grmod}_{\geq d} \subset S\text{-grmod}$$

the full subcategory of all modules  $M$  with  $M = M_{\geq d}$ . Define the shift autoequivalence on  $S\text{-grmod}$  by  $M(i)_j := M_{i+j}$  for all  $i \in \mathbb{Z}$ ; it induces an endofunctor on the subcategory  $S\text{-grmod}_{\geq d}$  for  $i \leq 0$ .

**Algorithm 3.1** *We briefly describe how to compute the  $i$ -th homogeneous part of an  $M \in S\text{-grmod}$ : Such a module is realized on the computer as the cokernel of a graded free  $S$ -presentation*

$$M \xleftarrow{\pi} \bigoplus_k S(g_k) \xleftarrow{\mathfrak{m}} \bigoplus_\ell S(r_\ell).$$

The image of the graded submodule  $\bigoplus_{k,i+g_k \geq 0} S_{\geq i+g_k}(g_k) \leq \bigoplus_k S(g_k)$  under  $\pi$  is the graded submodule  $\langle M_i \rangle \leq M$ , which we compute as the kernel of the cokernel of the restricted map  $M \leftarrow \bigoplus_{k,i+g_k \geq 0} S_{\geq i+g_k}(g_k)$ . Computing a free  $S$ -presentation  $\langle M_i \rangle \leftarrow \bigoplus_k S(i) \xleftarrow{m_i} \bigoplus_\ell S(r'_\ell)$  of  $\langle M_i \rangle$  thus involves two successive syzygy computations as explained in [2, (10) in the proof of Theorem 3.4]. To get a free  $B$ -presentation of  $M_i$  we just need to tensor the last exact sequence with  $B = S/\mathfrak{m}$  over  $S$ , which corresponds to extracting the degree 0 relations in the reduced Gröbner basis of the  $S$ -matrix of relations  $m_i$ .

### 3.1 Internal and External Hom Functors

Let  $M, N \in S\text{-grmod}$ . Then the Hom-group  $\text{Hom}_{S\text{-mod}}(M, N)$  of their underlying modules in  $S\text{-mod}$  is again naturally graded. This induces internal Hom functors

$$\text{Hom}_\bullet : S\text{-grmod}^{\text{op}} \times S\text{-grmod} \rightarrow S\text{-grmod}$$

in the category  $S\text{-grmod}$  and

$$\text{Hom}_{\geq d} : S\text{-grmod}_{\geq d}^{\text{op}} \times S\text{-grmod}_{\geq d} \rightarrow S\text{-grmod}_{\geq d}$$

in  $S\text{-grmod}_{\geq d}$ . These internal Hom functors are algorithmically computable if  $B$  has effective coset representatives (cf., e.g., [1, §4.3] and [2, §3.3]).

The (external) Hom-groups of the category  $S\text{-grmod}$  are finitely generated  $B$ -modules. They can be recovered as the graded part of degree 0 of the corresponding internal Hom's:

$$\text{Hom}(M, N) := \text{Hom}_{S\text{-grmod}}(M, N) \cong \text{Hom}_\bullet(M, N)_0,$$

$$\text{Hom}_{S\text{-grmod}_{\geq d}}(M, N) \cong \text{Hom}_{\geq d}(M, N)_0 \quad \text{for } d \leq 0.$$

In particular,  $\text{Hom}_{S\text{-grmod}}(S, M) \cong M_0$  and  $\text{Hom}_{S\text{-grmod}_{\geq d}}(S, M) \cong M_0$  for  $d \leq 0$ .

Dealing with  $d > 0$  would enforce further case distinctions. For example,  $B \cong S/\mathfrak{m}$  lies in  $S\text{-grmod}_{\geq d}$  only if  $d \leq 0$ .

Till the end of Sect. 4 we assume that  $d \leq 0$ .

*Remark 3.2* Applying  $\text{Hom}_\bullet(-, M)$  to the short exact sequence  $S/\mathfrak{m}^\ell \leftarrow S \leftrightarrow \mathfrak{m}^\ell$  yields

$$\text{Hom}_\bullet(S/\mathfrak{m}^\ell, M) \hookrightarrow M \xrightarrow{\eta_M^\ell} \text{Hom}_\bullet(\mathfrak{m}^\ell, M) \twoheadrightarrow \text{Ext}_\bullet^1(S/\mathfrak{m}^\ell, M) \quad (\eta_M^\ell)$$

as part of the long exact contravariant  $\text{Ext}_\bullet$ -sequence. We will repeatedly refer to this exact sequence as well as to the  $\ell = 1$  case

$$\text{Hom}_\bullet(B, M) \hookrightarrow M \xrightarrow{\eta_M^!} \text{Hom}_\bullet(\mathfrak{m}, M) \twoheadrightarrow \text{Ext}_\bullet^1(B, M). \quad (\eta_M^!)$$

### 3.2 Quasi-Zero Modules

Let  $S\text{-grmod}^0$  denote the thick subcategory of **quasi-zero** modules, i.e., those with  $M_{\geq \ell} = 0$  for  $\ell$  large enough. Analogously, we denote by  $S\text{-grmod}_{\geq d}^0$  the *localizing* (cf. Theorem 4.6) subcategory of quasi-zero modules in  $S\text{-grmod}_{\geq d}$ .

*Remark 3.3* For  $M \in S\text{-grmod}$ . Then for all  $\ell \geq 0$

1.  $\text{Tor}_i^S(S/\mathfrak{m}^\ell, M)_\bullet \in S\text{-grmod}^0$  for all  $i \geq 0$ .
2.  $\text{Ext}_\bullet^j(S/\mathfrak{m}^\ell, M) \in S\text{-grmod}^0$  for all  $j \geq 0$ .
3.  $\text{Ext}_\bullet^j(\mathfrak{m}^\ell, M) \in S\text{-grmod}^0$  for all  $j \geq 1$ .

*Proof* The existence of a finitely generated free resolution of the first argument  $S/\mathfrak{m}^\ell$  (and hence of  $\mathfrak{m}^\ell$ ) implies that all the above derived modules lie in  $S\text{-grmod}$ . By applying  $S/\mathfrak{m}^\ell \otimes_S -$  to a projective resolution of  $M$  and  $\text{Hom}_\bullet(S/\mathfrak{m}^\ell, -)$  to an injective resolution of  $M$  shows that the ideal  $\mathfrak{m}^\ell \triangleleft S$  annihilates  $\text{Tor}_i^S(S/\mathfrak{m}^\ell, M)_\bullet$  and  $\text{Ext}_\bullet^j(S/\mathfrak{m}^\ell, M)$ , which implies that they are also finitely generated  $S/\mathfrak{m}^\ell$ -modules, proving (1) and (2). The existence of the connecting isomorphisms  $\text{Ext}_\bullet^j(\mathfrak{m}^\ell, M) \cong \text{Ext}_\bullet^{j+1}(S/\mathfrak{m}^\ell, M)$  ( $j \geq 1$ ) finally implies (3).  $\square$

*Remark 3.4* The use of the nonconstructive injective resolution in the previous proof is an example of an admissible use of nonconstructive arguments in an otherwise constructive setup to prove statements which neither involve existential quantifiers nor disjunctions (so-called negative formulae):  $\text{Ext}_\bullet(S/\mathfrak{m}^\ell, M)$  has two descriptions. The nonconstructive one in the proof and the constructive one in which  $\text{Hom}(-, M)$  is applied to a finite free resolution of  $S/\mathfrak{m}^\ell$ . Although the isomorphism between the two descriptions is not constructive it is “good enough” for transferring the property we want to establish.

### 3.3 Regularity, Linear Regularity, and Relation to Tor and Ext

For convenience of the reader we recall the definition of the **Castelnuovo-Mumford regularity** in the relative case from [10, §2]. For any quasi-zero graded  $S$ -module  $N$  define

$$\text{reg } N := \max\{d \in \mathbb{Z} \mid N_d \neq 0\}.$$

The regularity of the zero module is set to  $-\infty$ . Then, for  $M \in S\text{-gmod}$  the  $S$ -module  $\text{Tor}_i^S(B, M)_\bullet$  is quasi-zero and

$$\text{reg } M := \max\{\text{reg } \text{Tor}_i^S(B, M)_\bullet - i \mid i = 0, \dots, n+1\}.$$

Equivalently, one can define

$$\text{reg } M := \max\{\text{reg } H_m^j(M) + j \mid j = 0, \dots, n+1\}$$

using the local cohomology modules  $H_m^j(M) = \varinjlim_\ell \text{Ext}_\bullet^j(S/m^\ell, M)$  (cf., e.g., [10, Prop. 2.1]).<sup>6</sup> In fact only  $\ell = 1$  in this sequential colimit is relevant for us. To see this we need the following result, which we also use in the proof of our key Lemma 5.4.

**Lemma 3.5** *There exists a natural isomorphism*

$$\text{Tor}_i^S(B, M)_\bullet \cong \text{Ext}_\bullet^{n+1-i}(\wedge^{n+1}V, M).$$

*Proof* The Tor-Ext spectral sequence

$$\text{Tor}_{-p}^S(\text{Ext}_\bullet^q(\wedge^{n+1}V, S), M)_\bullet \Rightarrow \text{Ext}_\bullet^{p+q}(\wedge^{n+1}V, M)$$

collapses since  $\text{Ext}_\bullet^q(\wedge^{n+1}V, S) = 0$  for  $q \neq n+1$  and  $\text{Ext}_\bullet^{n+1}(\wedge^{n+1}V, S) = B$ .  $\square$

When  $B = k$  is a field this Lemma becomes the intrinsic and rather generalizable form of the equality between the graded Betti numbers  $\beta_{ij} := \dim_k \text{Tor}_i^S(B, M)_j$  and the graded **Bass numbers**:

$$\mu_{n+1-ij-n-1} := \dim_k \text{Ext}_\bullet^{n+1-i}(\wedge^{n+1}V, M)_j.$$

*Remark 3.6* Lemma 3.5 and the noncanonical isomorphism  $\wedge^{n+1}V \cong B(n+1)$  yield

$$\text{reg } M = \max\{\text{reg } \text{Ext}_\bullet^j(B, M) + j \mid j = 0, \dots, n+1\}.$$

The value of the following definition will start to become obvious in Proposition 3.9 in the next subsection.

**Definition 3.7** Define the **linear regularity** of  $M \in S\text{-gmod}$  to be

$$\text{linreg } M = \max\{\text{reg } \text{Ext}_\bullet^j(B, M) \mid j = 0, 1\} \in \mathbb{Z} \cup \{-\infty\}.$$

---

<sup>6</sup>This definition clarifies the relation to two other regularity notions: The **geometric regularity** is defined by  $\text{g-reg } M := \max\{\text{reg } H_m^j(M) + j \mid j = 1, \dots, n+1\}$  and the **regularity of the sheafification**  $\text{reg } \widetilde{M} := \max\{\text{reg } H_m^j(M) + j \mid j = 2, \dots, n+1\}$ .

Analogously, the  $d$ -th **truncated linear regularity** of  $M \in S\text{-grmod}_{\geq d}$  is defined by

$$\text{linreg}_{\geq d} M = \max\{\text{reg Ext}_{\geq d}^j(B, M) \mid j = 0, 1\} \in \mathbb{Z}_{\geq d} \cup \{-\infty\},$$

for  $d \leq 0$  where  $\text{Ext}_{\geq d}^j := \text{Ext}_{S\text{-grmod}_{\geq d}}^j \simeq (\text{Ext}_{\bullet}^j)_{\geq d}$ .

Note that  $\text{linreg} = \text{reg}$  on  $S\text{-grmod}^0$  and  $\text{linreg} \leq \text{reg}$  on  $S\text{-grmod}$ .

*Example 3.8*  $\text{linreg } S/\mathfrak{m}^{\ell+1} = \text{reg } S/\mathfrak{m}^{\ell+1} = \ell = \text{linreg } \mathfrak{m}^{\ell+1} < \text{reg } \mathfrak{m}^{\ell+1} = \ell + 1$ .

The motivation behind introducing  $\text{linreg}$  is that it offers an upper bound in the saturation algorithms discussed below, where the use of the (often enough much larger) regularity would be a waste of computational resources.

### 3.4 Saturated Modules

The equivalent conditions (4) and (5) in the following proposition are computationally effective characterizations of saturated modules.

**Proposition 3.9** *For  $M \in S\text{-grmod}$  the following are equivalent:*

1.  $M$  is saturated w.r.t.  $S\text{-grmod}^0$ ;
2.  $\text{Ext}_{\bullet}^0(S/\mathfrak{m}^{\ell}, M) = 0$  and  $\text{Ext}_{\bullet}^1(S/\mathfrak{m}^{\ell}, M) = 0$  for all  $\ell \geq 0$ ;
3. The natural map  $\eta_M^{\ell} := \text{Hom}_{\bullet}(S \leftarrow \mathfrak{m}^{\ell}, M) : M \rightarrow \text{Hom}_{\bullet}(\mathfrak{m}^{\ell}, M)$  is an isomorphism for all  $\ell \geq 0$ ;
4.  $\text{Ext}_{\bullet}^0(B, M) = 0$  and  $\text{Ext}_{\bullet}^1(B, M) = 0$ <sup>7</sup>;
5. The natural map  $\eta_M^1 := \text{Hom}_{\bullet}(S \leftarrow \mathfrak{m}, M) : M \rightarrow \text{Hom}_{\bullet}(\mathfrak{m}, M)$  is an isomorphism;
6.  $\text{Tor}_{n+1}^S(B, M)_{\bullet} = 0$  and  $\text{Tor}_n^S(B, M)_{\bullet} = 0$ ;
7.  $\text{linreg } M = -\infty$ .

And if the base ring  $B$  is a field the above is also equivalent to:

8. The projective dimension  $\text{pd } M \leq n - 1$ .

In the proof of this proposition, we use the following simple remark.

*Remark 3.10* The kernel  $K$  of the epimorphism  $\mathfrak{m}^{\ell} \leftarrow \otimes^{\ell} \mathfrak{m}$  is concentrated in degree  $\ell$ . To see this note that any homogeneous element in  $\otimes^{\ell} \mathfrak{m}$  of degree  $m > \ell$

---

<sup>7</sup>Conditions (4) and (5) are in their use of Gröbner bases algorithmically equivalent. Computing them only involves the first two morphisms in the Koszul resolution of  $B$  (and then tensoring their duals with  $M$ ). One might be tempted to expect that (4) is always algorithmically superior to condition (6), which seem to involve an  $n + 1$ -term resolution of either  $B$  or of  $M$ . However, one can easily construct examples of  $M \in S\text{-grmod}$ , where condition (6) is algorithmically superior, e.g., if the resolution of  $M$  terminates after few steps, long before reaching step  $n$ .



which is the tensor product of monomials can be brought to the normal form  $x_{i_1} \otimes_B \cdots \otimes_B x_{i_{\ell-1}} \otimes_B x^\mu$  with  $i_1 \leq \cdots \leq i_{m-1} \leq \min\{i \mid \mu_i \neq 0\}$  and  $|\mu| = m - \ell + 1$ . This kernel  $K$  is free over  $B$  of rank  $(n+1)^\ell - \binom{n+\ell}{n}$  as the kernel of the  $B$ -epimorphism  $\text{Sym}^\ell W \leftarrow \otimes^\ell W$ .

*Proof (of Proposition 3.9)*

- (2)  $\Leftrightarrow$  (3): The claim is obvious from the  $(\eta_M^\ell)$ -sequence in Remark 3.2.  
(4)  $\Leftrightarrow$  (5): This is a special case of the equivalence (2)  $\Leftrightarrow$  (3) for  $\ell = 1$ .  
(4)  $\Leftrightarrow$  (6): This is the statement of Lemma 3.5 for  $i = n + 1$  and  $i = n$ .  
(4)  $\Leftrightarrow$  (7): By definition of  $\text{linreg}$ .  
(1)  $\Rightarrow$  (4): This follows directly from the definition of saturated objects (cf. Sect. 2), as  $B \in \mathcal{C} = S\text{-grmod}^0$ .  
(5)  $\Rightarrow$  (3): Applying the  $\ell$ -th power of  $\text{Hom}_\bullet(S \leftarrow \mathfrak{m}, -)$  to  $M$  and taking the diagonal in the  $\ell$ -dimensional cube yields the isomorphism

$$\varphi := M \xrightarrow{\sim} \text{Hom}_\bullet(\otimes^\ell \mathfrak{m}, M)$$

by the adjunction between  $\otimes$  and  $\text{Hom}_\bullet$ . This isomorphism can be written as the composition

$$\text{Hom}_\bullet(S \leftarrow \mathfrak{m}^\ell \leftarrow \otimes^\ell \mathfrak{m}, M) = \left( M \xrightarrow{\psi} \text{Hom}_\bullet(\mathfrak{m}^\ell, M) \xrightarrow{\chi} \text{Hom}_\bullet(\otimes^\ell \mathfrak{m}, M) \right).$$

The homomorphism  $\chi$  is a monomorphism since  $\text{Hom}_\bullet$  is left exact and an epimorphism since  $\chi \circ \psi = \varphi$  is an isomorphism. Hence,  $\chi$  is isomorphism and thus  $\psi$  is an isomorphism.

- (2)  $\Rightarrow$  (1): Clearly, any  $N \in S\text{-grmod}^0$  is an epimorphic image of  $\bigoplus_{i \in I} (S/\mathfrak{m}^{a_i})(b_i)$  for a finite set  $I$  and suitable  $a_i$  and  $b_i$ . Denote the kernel of  $N \leftarrow \bigoplus_i (S/\mathfrak{m}^{a_i})(b_i)$  by  $K$ . Applying  $\text{Hom}_\bullet(-, M)$  to  $N \leftarrow \bigoplus_i (S/\mathfrak{m}^{a_i})(b_i) \leftarrow K$  yields as parts of the long exact sequence

$$\text{Hom}_\bullet(N, M) \hookrightarrow \underbrace{\text{Hom}_\bullet\left(\bigoplus_i (S/\mathfrak{m}^{a_i})(b_i), M\right)}_{\cong 0},$$

and

$$\text{Hom}_\bullet(K, M) \rightarrow \text{Ext}_\bullet^1(N, M) \rightarrow \underbrace{\text{Ext}_\bullet^1\left(\bigoplus_i (S/\mathfrak{m}^{a_i})(b_i), M\right)}_{\cong 0}.$$

The first part implies  $\text{Hom}_\bullet(-, M) = 0$  on  $S\text{-grmod}^0$ . In particular,  $\text{Hom}_\bullet(K, M) = 0$  since  $K \in S\text{-grmod}^0$ . Combining this and the second part implies that  $\text{Ext}_\bullet^1(-, M)$  vanishes on  $S\text{-grmod}^0$ .

(6)  $\Leftrightarrow$  (8): If  $B$  is a field then there exists a finite free (and not merely relatively free) presentation  $M \leftarrow F_\bullet$  with  $\mathrm{Tor}_i^S(B, M)_\bullet$  isomorphic to the head of  $F_i$ .  $\square$

**Corollary 3.11** *For  $M \in S\text{-grmod}_{\geq d}$  the following are equivalent (recall,  $d \leq 0$ ):*

1.  $M$  is saturated w.r.t.  $S\text{-grmod}_{\geq d}^0$ ;
2.  $\mathrm{Ext}_{\geq d}^0(S/\mathfrak{m}^\ell, M) = 0$  and  $\mathrm{Ext}_{\geq d}^1(S/\mathfrak{m}^\ell, M) = 0$  for all  $\ell \geq 0$ ;
3. The natural map  $\eta_M^\ell := \mathrm{Hom}_{\geq d}(S \leftarrow \mathfrak{m}^\ell, M) : M \rightarrow \mathrm{Hom}_{\geq d}(\mathfrak{m}^\ell, M)$  is an isomorphism for all  $\ell \geq 0$ ;
4.  $\mathrm{Ext}_{\geq d}^0(B, M) = 0$  and  $\mathrm{Ext}_{\geq d}^1(B, M) = 0$ ;
5. The natural map  $\eta_M^1 := \mathrm{Hom}_{\geq d}(S \leftarrow \mathfrak{m}, M) : M \rightarrow \mathrm{Hom}_{\geq d}(\mathfrak{m}, M)$  is an isomorphism;
6.  $\mathrm{Tor}_{n+1}^S(B(n+1), M)_{\geq d} = 0$  and  $\mathrm{Tor}_n^S(B(n+1), M)_{\geq d} = 0$ ;
7.  $\mathrm{linreg}_{\geq d} M = -\infty$ .

## 4 Ideal Transforms

Recall, the  **$\mathfrak{m}$ -transform** of  $M \in S\text{-grmod}$  is the (not necessarily finitely generated) graded  $S$ -module defined by the sequential colimit

$$D_{\mathfrak{m}} := \varinjlim_{\ell} \mathrm{Hom}_\bullet(\mathfrak{m}^\ell, -) : S\text{-grmod} \rightarrow S\text{-grMod}.$$

On  $S\text{-grmod}_{\geq d}$  the  **$\mathfrak{d}$ -truncated  $\mathfrak{m}$ -transform** (recall,  $d \leq 0$ )

$$D_{\mathfrak{m}, \geq d} := \varinjlim_{\ell} \mathrm{Hom}_{\geq d}(\mathfrak{m}^\ell, -) : S\text{-grmod}_{\geq d} \rightarrow S\text{-grmod}_{\geq d}$$

is an endofunctor. This is a simple corollary of the Lemma 4.2 below.

**Definition 4.1** We define the **saturation interval** of  $M \in S\text{-grmod}_{\geq d}$  to be

$$I_{\geq d}(M) := [\delta_{M,d}^0, \delta_{M,d}^1] \cap \mathbb{Z} \subset \mathbb{Z}_{\geq 0},$$

where  $\delta_{M,d}^0 := \max\{\mathrm{reg} \mathrm{Hom}_{\geq d}(B, M) - d + 1, 0\}$  and  $\delta_{M,d}^1 := \max\{\mathrm{linreg}_{\geq d} - d + 1, 0\}$ .

The saturation interval plays a role in the following convergence analysis and the definition of its upper bound is a further motivation for the linear regularity.

**Lemma 4.2** *For each  $M \in S\text{-grmod}_{\geq d}$  the sequential colimit defining the  $\mathfrak{m}$ -transform is finite. More precisely, there exists a nonnegative integer  $\delta_{M,d} \in I_{\geq d}(M)$  such that the induced maps  $\mathrm{Hom}_{\geq d}(\mathfrak{m}^\ell, M) \rightarrow \mathrm{Hom}_{\geq d}(\mathfrak{m}^{\ell+1}, M)$  are*

isomorphisms for all  $\ell \geq t$  iff  $t \geq \delta_{M,d}$ . In particular, the natural map

$$\mathrm{Hom}_{\geq d}(\mathfrak{m}^\ell, M) \rightarrow D_{\mathfrak{m}, \geq d}(M)$$

is an isomorphism iff  $\ell \geq \delta_{M,d}$ .

*Proof* The short exact sequence  $B(-\ell)^{\oplus ?} \cong \mathfrak{m}^\ell / \mathfrak{m}^{\ell+1} \leftarrow \mathfrak{m}^\ell \leftrightarrow \mathfrak{m}^{\ell+1}$  induces for  $M \in S\text{-grmod}$  the exact contravariant  $\mathrm{Ext}_\bullet$ -sequence of which the first four terms are

$$\mathrm{Hom}_\bullet(B, M)^{\oplus ?}(\ell) \hookrightarrow \mathrm{Hom}_\bullet(\mathfrak{m}^\ell, M) \rightarrow \mathrm{Hom}_\bullet(\mathfrak{m}^{\ell+1}, M) \rightarrow \mathrm{Ext}_\bullet^1(B, M)^{\oplus ?}(\ell).$$

By Remark 3.3.(2) both  $\mathrm{Hom}_\bullet(B, M)$  and  $\mathrm{Ext}_\bullet(B, M)$  are quasi-zero. Hence, there exists a  $\delta_{M,d} \in I_{\geq d}(M)$  such that the truncated morphisms

$$\mathrm{Hom}_{\geq d}(\mathfrak{m}^\ell, M) \rightarrow \mathrm{Hom}_{\geq d}(\mathfrak{m}^{\ell+1}, M)$$

become isomorphisms in  $S\text{-grmod}_{\geq d}$  for  $\ell \geq t$  iff  $t \geq \delta_{M,d}$ .  $\square$

In particular, once  $B$  has effective coset representatives,  $D_{\mathfrak{m}, \geq d}$  is algorithmically computable on objects and morphisms, since the internal Hom functor  $\mathrm{Hom}_{\geq d}$  is.

**Definition 4.3** We call  $\delta_{M,d} \in I_{\geq d}(M)$  from Lemma 4.2 the **defect of saturation** of  $M$ .

*Example 4.4* Note that  $1 = \delta_{\mathfrak{m}(-t), 0} \in I_{\geq 0}(\mathfrak{m}(-t)) = [0, \mathrm{linreg}_{\geq 0} \mathfrak{m}(-t) + 1] = [0, t+1]$  for all  $t \in \mathbb{Z}_{\geq 0}$ . In other words, the maximum of  $I_{\geq d}(M)$  can be an arbitrarily bad upper bound for  $\delta_{M,d}$ .

*Example 4.5* For  $M = S \oplus B(-t)$  and  $t \geq 0$  we compute  $\mathrm{Hom}_\bullet(B, M) = B(-t)$  and  $\mathrm{Ext}_\bullet^1(B, M) = B(-t+1)^{n+1}$  (for  $n > 0$ ). Hence  $\delta_{M,0}^0 = t+1 = \delta_{M,0}^1 = \delta_{M,0}$  is the defect of saturation. Thus, for certain examples factoring out the  $S\text{-grmod}_{\geq d}^0$ -torsion part of  $M$  a priori could be beneficial.

The natural transformation

$$\eta_M := \varinjlim_{\ell} (\eta_M^\ell : M \rightarrow \mathrm{Hom}_{\geq d}(\mathfrak{m}^\ell, M)) : M \rightarrow D_{\mathfrak{m}, \geq d}(M)$$

is induced by applying  $\mathrm{Hom}_{\geq d}(-, M)$  to the embeddings  $(S \hookrightarrow \mathfrak{m}^\ell)_{\geq d}$ .

Now we reprove that the ideal transform computes the module of twisted global sections (cf., e.g., [14, §C.3]).

**Theorem 4.6** *The  $d$ -truncated  $\mathfrak{m}$ -transform  $D_{\mathfrak{m}, \geq d}$  together with the natural transformation  $\eta : \mathrm{Id}_{\mathcal{A}} \rightarrow D_{\mathfrak{m}, \geq d}$  is a Gabriel monad of  $\mathcal{A} = S\text{-grmod}_{\geq d}$  w.r.t.  $\mathcal{C} := S\text{-grmod}_{\geq d}^0$ .*

In fact, the theorem holds for all  $d \in \mathbb{Z}$ . The proof below assumes  $d \leq 0$  to avoid case distinctions.

**Corollary 4.7**  $\text{Hom}_{\geq d}(\mathfrak{m}^\ell, M)$  is  $S\text{-grmod}_{\geq d}^0$ -saturated iff  $\ell \geq \delta_{M,d}$ .

Before proving the theorem we state some simple facts about ideal transforms.

*Remark 4.8*

1. Any  $N \in S\text{-grmod}^0$  vanishes in degrees greater than  $\text{reg } N$ . Thus,

$$\text{Hom}_\bullet(L_{\geq \ell}, N)_{\geq \text{reg } N + 1 - \ell} = 0$$

for all  $\ell \in \mathbb{Z}$  and  $L \in S\text{-grmod}$ .

2. The embedding  $M_{\geq t} \hookrightarrow M \in S\text{-grmod}_{\geq d}$  induces (by simple degree considerations) an isomorphism

$$\text{Hom}_{\geq d}(L_{\geq \ell}, M_{\geq t}) \xrightarrow{\sim} \text{Hom}_{\geq d}(L_{\geq \ell}, M) \quad \text{for all } d \leq t \leq \ell + d.$$

In particular,  $D_{\mathfrak{m}, \geq d}(M) \cong D_{\mathfrak{m}, \geq d}(M_{\geq t})$  for any  $t \geq d$  and we are allowed to replace  $M$  by any of its truncations.

3. For  $M \in S\text{-grmod}_{\geq d}$  take  $t \geq d$  large enough such that the submodule  $M_{\geq t}$  has no  $S\text{-grmod}_{\geq d}^0$ -torsion. Then

$$\text{Hom}_{\geq d}(\mathfrak{m}^\ell, M) \cong \text{Hom}_{\geq d}(\mathfrak{m}^\ell, M_{\geq t}) \cong \text{Hom}_{\geq d}(\otimes^\ell \mathfrak{m}, M_{\geq t})$$

for all  $\ell \geq t - d$  by (2) and Remark 3.10. An admissible choice is  $t := \text{linreg}_{\geq d} M + 1$ , then  $\ell \geq t - d \geq \delta_{M,d}$  (cf. Lemma 4.2). In particular, after replacing  $M$  by a high enough truncation we can assume that  $\text{Hom}_{\geq d}(\mathfrak{m}^\ell, M) \cong \text{Hom}_{\geq d}(\otimes^\ell \mathfrak{m}, M)$ .

4. Since the shift functor  $(1) : S\text{-grmod}_{\geq d} \rightarrow S\text{-grmod}_{\geq d+1}$ ,  $M \mapsto M(1)$ ,  $\varphi \mapsto \varphi(1)$  is (quasi-)inverse to the shift functor  $(-1) : S\text{-grmod}_{\geq d+1} \rightarrow S\text{-grmod}_{\geq d}$  and  $D_{\mathfrak{m}, \geq d} \circ (-1) = (-1) \circ D_{\mathfrak{m}, \geq d+1}$  we can restrict the following proofs to  $D_{\mathfrak{m}, \geq 0}$ .

*Proof (of Theorem 4.6)* We use Theorem 2.1. Due to Remark 4.8.(4) we only need to consider the case  $d = 0$ .

- 2.1.(1)  $\mathcal{C} \subset \ker D_{\mathfrak{m}, \geq 0}$ :

Applying Remark 4.8.(1) with  $L = S$  (and  $L_{\geq L} = S_{\geq \ell} = \mathfrak{m}^\ell$ ) we conclude that  $D_{\mathfrak{m}}$  vanishes<sup>8</sup> on  $S\text{-grmod}^0$  and  $D_{\mathfrak{m}, \geq 0}$  on  $S\text{-grmod}_{\geq 0}^0$ .

- 2.1.(2)  $D_{\mathfrak{m}, \geq 0}(\mathcal{A}) \subset \text{Sat}_{\mathcal{C}}(\mathcal{A})$ :

For any  $M \in \mathcal{A}$ , the map

$$\begin{aligned} \text{Hom}_{\geq 0}(S \hookrightarrow \mathfrak{m}, D_{\mathfrak{m}, \geq 0}(M)) &= \text{Hom}_{\geq 0}(S \hookrightarrow \mathfrak{m}, \text{Hom}_{\geq 0}(\mathfrak{m}^{\delta_{M,0}}, M)) \\ &= \text{Hom}_{\geq 0}(S \hookrightarrow \mathfrak{m}, \text{Hom}_{\geq 0}(\otimes^{\delta_{M,0}} \mathfrak{m}, M)) \end{aligned}$$

<sup>8</sup>For  $N \in \mathcal{C}$  all modules in the sequential colimit defining  $D_{\mathfrak{m}, \geq 0}(N)$  vanish for  $\ell \geq \delta_{N,0} < \infty$ .

$$\begin{aligned}
&= \mathrm{Hom}_{\geq 0} \left( \otimes^{\delta_{M,0}} \mathfrak{m} \leftrightarrow \otimes^{\delta_{M,0}+1} \mathfrak{m}, M \right) \\
&= \mathrm{Hom}_{\geq 0} \left( \otimes^{\delta_{M,0}} \mathfrak{m}, M \right) \rightarrow \mathrm{Hom}_{\geq 0} \left( \otimes^{\delta_{M,0}+1} \mathfrak{m}, M \right) \\
&= \mathrm{Hom}_{\geq 0} \left( \mathfrak{m}^{\delta_{M,0}}, M \right) \rightarrow \mathrm{Hom}_{\geq 0} \left( \mathfrak{m}^{\delta_{M,0}+1}, M \right)
\end{aligned}$$

is an isomorphism by Lemma 4.2 proving statement (5) of Corollary 3.11. We have repeatedly used Remark 4.8.(3) and the adjunction between  $\otimes$  and  $\mathrm{Hom}_{\geq 0}$ .

2.1.(3)  $G := \mathrm{co}\text{-res}_{\mathrm{Sat}_{\mathcal{G}}(\mathcal{A})} D_{\mathfrak{m}, \geq 0}$  is exact:

Applying  $\mathrm{Hom}_{\geq 0}(\mathfrak{m}^{\ell}, -)$  to the short exact sequence  $L \hookrightarrow M \twoheadrightarrow N$  in  $S\text{-grmod}_{\geq 0}$  yields the exact sequence

$$\mathrm{Hom}_{\geq 0}(\mathfrak{m}^{\ell}, L) \hookrightarrow \mathrm{Hom}_{\geq 0}(\mathfrak{m}^{\ell}, M) \rightarrow \mathrm{Hom}_{\geq 0}(\mathfrak{m}^{\ell}, N) \rightarrow \mathrm{Ext}_{\geq 0}^1(\mathfrak{m}^{\ell}, L)$$

as part of the long exact covariant  $\mathrm{Ext}_{\geq 0}$ -sequence. Since  $\mathrm{Ext}_{\geq 0}^1(\mathfrak{m}^{\ell}, L)$  is quasi-zero by Remark 3.3.(3) the sequence is exact up to defects in  $S\text{-grmod}_{\geq 0}^0$ .

2.1.(4)  $\eta D_{\mathfrak{m}, \geq 0} = D_{\mathfrak{m}, \geq 0} \eta$ :

We repeatedly use the adjunction between  $\otimes$  and  $\mathrm{Hom}_{\geq 0}$  and Lemma 4.2 to interchange the involved sequential colimits over  $\ell'$  and  $\ell''$  by a common  $\ell \geq \ell', \ell''$ , high enough to stabilize both colimits:

$$\begin{aligned}
\eta D_{\mathfrak{m}, \geq 0}(M) &= \lim_{\ell'} \mathrm{Hom}_{\geq 0}(S \leftrightarrow \mathfrak{m}^{\ell'}, \lim_{\ell''} \mathrm{Hom}_{\geq 0}(\mathfrak{m}^{\ell''}, M)) \\
&= \mathrm{Hom}_{\geq 0}(S \leftrightarrow \mathfrak{m}^{\ell}, \mathrm{Hom}_{\geq 0}(\mathfrak{m}^{\ell}, M)) \\
&= \mathrm{Hom}_{\geq 0}((S \leftrightarrow \mathfrak{m}^{\ell}) \otimes_S \mathfrak{m}^{\ell}, M) \\
&= \mathrm{Hom}_{\geq 0}(\mathfrak{m}^{\ell}, \mathrm{Hom}_{\geq 0}(S \leftrightarrow \mathfrak{m}^{\ell}, M)) \\
&= \lim_{\ell''} \mathrm{Hom}_{\geq 0}(\mathfrak{m}^{\ell''}, \lim_{\ell'} \mathrm{Hom}_{\geq 0}(S \leftrightarrow \mathfrak{m}^{\ell'}, M)) \\
&= D_{\mathfrak{m}, \geq 0}(\eta M).
\end{aligned}$$

The proof implicitly uses commuting diagrams of morphisms in  $S\text{-grmod}_{\geq 0}$  to justify the equality signs.<sup>9</sup>

2.1.(5)  $\eta \iota$  is a natural isomorphism:

<sup>9</sup>We could have used the fact that  $D_{\mathfrak{m}, \geq 0} = \lim_{\ell} \mathrm{Hom}_{\geq 0}(\mathfrak{m}^{\ell}, -)$  commutes with directed colimits. However, the general form of the second statement is not quite trivial [8, Coro. 3.4.11] (the directed colimit is called direct limit in [8, Terminology 3.4.1]). Note that although the ideal transform commutes with *directed* colimits, it does not generally commute with arbitrary finite colimits, for otherwise it would be right exact and hence exact.

Let  $M \in S\text{-grmod}_{\geq 0}$  be saturated w.r.t.  $S\text{-grmod}_{\geq 0}^0$ . Applying  $\text{Hom}_{\geq 0}(-, M)$  to the short exact sequence  $S/\mathfrak{m}^\ell \leftarrow S \leftrightarrow \mathfrak{m}^\ell$  yields

$$\underbrace{\text{Hom}_{\geq 0}(S/\mathfrak{m}^\ell, M)}_{\cong 0} \hookrightarrow M \xrightarrow{\eta_M^\ell} \text{Hom}_{\geq 0}(\mathfrak{m}^\ell, M) \twoheadrightarrow \underbrace{\text{Ext}_{\geq 0}^1(S/\mathfrak{m}^\ell, M)}_{\cong 0}$$

since  $S/\mathfrak{m}^\ell \in S\text{-grmod}_{\geq 0}^0$ . In other words,  $\eta_M^\ell$  is an isomorphism for all  $\ell$ .  $\square$

*Remark 4.9* The saturation process of  $M \in S\text{-grmod}$  conducted by  $D_m$  brings  $\text{linreg}$  to  $-\infty$ , whereas  $\text{reg}$  is only brought down to the regularity of the sheafification.

Since the Frobenius powers  $\mathfrak{m}^{[\ell]} := \langle x_0^\ell, \dots, x_n^\ell \rangle$  satisfy  $\mathfrak{m}^\ell \geq \mathfrak{m}^{[\ell]} \geq \mathfrak{m}^{(n+1)\ell}$  we can use them instead of  $\mathfrak{m}^\ell$  them in the above sequential colimits. They are computationally superior since their number of generators does not increase with  $\ell$ . In other words, the module  $\text{Hom}_{\geq d}(\mathfrak{m}^{[\delta_{M,d}]}, M)$  is  $S\text{-grmod}_{\geq d}^0$ -saturated. Alternatively, one could iteratively ( $\delta_{M,d}$  times) apply  $\text{Hom}_{\geq d}(\mathfrak{m}, -)$  to (the  $S\text{-grmod}_{\geq d}^0$ -torsion-free factor of)  $M$ . It depends on the example which algorithm is faster.

## 5 Graded $S$ -Modules and Linear $E$ -Complexes

In this section we describe how to translate the module structure of  $M \in S\text{-grmod}$  into the structure of a linear complex  $\mathbf{R}(M)$  over the exterior algebra  $E := \wedge V$ , which is Koszul dual to  $S = \text{Sym } V^*$ . This translation turns out to be functorial, algorithmic, and an adjoint equivalence of categories. We denote the category of finitely generated graded  $E$ -modules by  $E\text{-grmod}$ .

Let  $e_0, \dots, e_n$  denote a  $B$ -basis  $V$  of which the indeterminates  $x_0, \dots, x_n$  of  $S$  form the dual  $B$ -basis of  $W = V^* = \text{Hom}_B(V, B)$ . We set  $\text{deg}(e_i) = -1$  for all  $i = 0, \dots, n$ .

### 5.1 The Functor $\mathbf{R}$

The  $B$ -linear maps

$$\mu^i(x_j) : M_i \rightarrow M_{i+1}, m \mapsto x_j m, \quad \text{for } j = 0, \dots, n, \text{ and } i \in \mathbb{Z}$$

induced by the indeterminates  $x_j$  encode the graded  $S$ -module structure of an  $M \in S\text{-grmod}$  (cf. Algorithm 3.1 for an algorithm to compute  $M_i$ ).

*Example 5.1* For  $S := B[x_0, x_1]$  consider the free  $S$ -module  $M := S = S(0)$  of rank 1. Each graded part  $M_i$  is a free  $B$ -module for which we fix a basis of monomials, e.g.,  $M_0 = \langle 1 \rangle_B$ ,  $M_1 = \langle x_0, x_1 \rangle_B$ ,  $M_2 = \langle x_0^2, x_0x_1, x_1^2 \rangle_B$ . Then the matrices

$$\begin{aligned} 0 : & \quad \mu^0(x_0) = (1 \ 0), \mu^0(x_1) = (0 \ 1), \\ 1 : & \quad \mu^1(x_0) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \mu^1(x_1) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \\ & \quad \vdots \end{aligned}$$

represent the maps  $\mu^i(x_j)$ .

Using the  $B$ -basis  $(e_0, \dots, e_n)$  of  $V$  define for each  $i \in \mathbb{Z}$  the map  $\mu^i$  as the composition

$$\mu^i : \begin{cases} M_i \rightarrow \text{End}_B(V) \otimes_B M_i \rightarrow V \otimes_B M_{i+1}, \\ m \mapsto \text{id}_V \otimes m \quad \mapsto \sum_{j=0}^n e_j \otimes x_j m \end{cases}.$$

By the natural isomorphism  $\text{Hom}_B(M_i, V \otimes_B M_{i+1}) \cong \text{Hom}_{E\text{-grmod}}(E \otimes_B M_i, E \otimes_B M_{i+1})$  each  $\mu^i$  can equally be understood as a map of *graded*  $E$ -modules

$$\mu^i : E \otimes_B M_i \rightarrow E \otimes_B M_{i+1},$$

where the  $B$ -module  $M_j$  is considered as a graded  $B$ -module concentrated in degree  $j$  and, therefore,  $E \otimes_B M_j$  is generated by (a generating set of)  $M_j$  in degree  $j$ .

For a better functorial behavior we replace  $E$  by its  $B$ -dual [12, §16C]

$$\omega_E := \text{Hom}_B(E, B) \cong \wedge W \cong \wedge^{n+1} W \otimes_B E$$

in the above maps.<sup>10</sup> In particular,  $\omega_E$  lives in the degree interval  $0, \dots, n+1$  and its socle  $(\omega_E)_0$ , which is naturally isomorphic to  $B$ , is concentrated in degree 0. We denote the distinguished generator of the socle corresponding to  $1_B$  by  $1_{\omega_E}$ .

This change of language is justified by reinterpreting  $\mu^i : M_i \rightarrow V \otimes_B M_{i+1}$  as a map  $\mu^i : W \otimes_B M_i \rightarrow M_{i+1}$  using the adjunction

$$\text{Hom}_B(W \otimes_B X, Y) \cong \text{Hom}_B(X, \text{Hom}_B(W, Y)) \cong \text{Hom}_B(X, W^* \otimes_B Y).$$

The graded  $E$ -module  $\omega_E \otimes_B M_j$  has (compared with  $E \otimes_B M_j$ ) the advantage of having the  $B$ -module  $M_j$  as its socle interpreted as a graded  $E$ -module concentrated in degree  $j$ .

<sup>10</sup>It is again a free graded  $E$ -module which is *non*naturally isomorphic to  $E(-n-1)$ .

The commutativity of  $S$  implies that the composed map  $\mu^{i+1} \circ \mu^i : \omega_E \otimes_B M_i \rightarrow \omega_E \otimes_B M_{i+2}$  is zero. Thus, the sequence of  $\mu_i$ 's yields the so-called **R-complex** (cf. [11, §2] and [10, §2])

$$\mathbf{R}(M) : \cdots \rightarrow \omega_E \otimes_B M_i \xrightarrow{\mu^i} \omega_E \otimes_B M_{i+1} \xrightarrow{\mu^{i+1}} \omega_E \otimes_B M_{i+2} \rightarrow \cdots$$

*Example 5.2 (Example 5.1 Continued)* For  $M = S(0)$  we obtain the **R-complex**

$$0 \rightarrow \omega_E(0) \xrightarrow{\begin{pmatrix} e_0 & e_1 \end{pmatrix}} \omega_E(-1) \xrightarrow{\begin{pmatrix} e_0 & e_1 \\ e_0 & e_1 \end{pmatrix}} \omega_E(-2) \xrightarrow{\begin{pmatrix} e_0 & e_1 & \cdot \\ \cdot & e_0 & e_1 \end{pmatrix}} \omega_E(-3) \xrightarrow{\cdot} \omega_E(-4) \rightarrow \cdots$$

**Lemma 5.3** ([11, Prop. 2.3]) *There exists a natural isomorphism*

$$H^a(\mathbf{R}(M))_{a+i} \cong \mathrm{Tor}_i^S(B, M)_{a+i}.$$

*Proof* The idea is to interpret the bigraded differential module  $\omega_E \otimes_B M$  either as  $\mathbf{R}(M)$  or as the Koszul resolution of  $B$  tensored with  $M$  over  $S$ .  $\square$

Lemmas 5.3 and 3.5 imply the following lemma, an important technical insight for the rest of this paper.

**Lemma 5.4 (Key Lemma)** *There exists a natural isomorphism*

$$H^a(\mathbf{R}(M))_{a+i} \cong \mathrm{Ext}_{\bullet}^{n+1-i}(\wedge^{n+1} V, M)_{a+i}.$$

*Hence, there is a noncanonical isomorphism  $H^a(\mathbf{R}(M))_{a+n+1-j} \cong \mathrm{Ext}_{\bullet}^j(B, M)_{a-j}$ .*

Let  $A$  be either  $S$  or  $E$ . An epimorphism in  $A$ -grmod is said to be  $B$ -split if it splits as a morphism over  $B$ . A graded module  $P \in A$ -grmod is said to be **relatively projective** (with respect to  $B$ ) if  $\mathrm{Hom}_S(P, -)$  sends  $B$ -split epis to surjections. Any module of the form  $A \otimes_B M$ , where  $M$  is a  $B$ -module, is called **relatively free** (with respect to  $B$ ). By Eisenbud and Schreyer [10, Proposition 1.1], an  $N \in A$ -grmod is relatively projective if and only if it is relatively free.

A complex  $C = C^\bullet$  of graded  $E$ -modules is called **linear** if each  $C^i$  is relatively free (with respect to  $B$ ) with socle concentrated in degree  $i$ .<sup>11</sup> The **regularity** of a linear complex  $C$  is defined as

$$\mathrm{reg} C := \sup\{a \in \mathbb{Z} \mid H^a(C) \neq 0\} \in \mathbb{Z} \cup \{-\infty, \infty\}.$$

Lemma 5.3 or 5.4 connects the regularity of a graded module with that of its **R-complex**.

**Corollary 5.5** *For  $M \in S$ -grmod the equality  $\mathrm{reg} M = \mathrm{reg} \mathbf{R}(M)$  holds.*

These definitions allow us to describe the image of **R**.

<sup>11</sup>And hence  $C^i$  is generated in degree  $i + n + 1$ .



**Definition 5.6** We denote by  $E\text{-grlin}$  the full subcategory of complexes  $C$  of graded  $E$ -modules satisfying

1.  $C$  is linear;
2. each  $C^i$  is finitely generated;
3.  $C$  is left bounded;
4.  $\text{reg } C < \infty$ .

By  $E\text{-grlin}^0$  we denote the thick subcategory of bounded complexes.

Finally, for any  $d \in \mathbb{Z}$ , denote by  $E\text{-grlin}^{\geq d}$  the full subcategory of complexes in  $E\text{-grlin}$  with  $C^{<d} = 0$  and by  $E\text{-grlin}^{\geq d,0} := E\text{-grlin}^{\geq d} \cap E\text{-grlin}^0$ .

*Remark 5.7* An object  $C \in E\text{-grlin}^{\geq d}$  can be represented on a computer by the finite complex  $C^d \rightarrow C^{d+1} \rightarrow \dots \rightarrow C^{j-1} \rightarrow C^j$  provided that  $j > \text{reg } C$ . The part  $C^{>j}$  of  $C$  can be recovered by an injective resolution of  $\text{coker}(C^{j-1} \rightarrow C^j)$ . This relatively injective resolution is isomorphic to  $\text{Hom}_\bullet(-, E)$  applied to a relatively projective resolution of  $\text{Hom}_\bullet(\text{coker}(C^{j-1} \rightarrow C^j), E)$ . A morphisms in  $E\text{-grlin}$  can be represented on the computer by a chain morphism between two such finite complexes, one only needs to extend these complexes to equal cohomological degrees. Again, the part of the morphism in higher cohomological degrees can be computed by an injective resolution.

**Proposition 5.8** *The construction  $\mathbf{R}$  induces two fully faithful functors  $\mathbf{R} : S\text{-grmod} \rightarrow E\text{-grlin}$  and  $\mathbf{R}^{\geq d} : S\text{-grmod}_{\geq d} \rightarrow E\text{-grlin}^{\geq d}$  for all  $d \in \mathbb{Z}$ .*

*Proof* As  $M \in S\text{-grmod}$  is finitely generated,  $\mathbf{R}(M)$  is left bounded. By definition, each  $\mathbf{R}(M)^i = \omega_E \otimes_B M_i$  is a finitely generated graded relatively free module with socle  $M_i$  concentrated in degree  $i$ . Furthermore  $\text{reg } \mathbf{R}(M) = \text{reg } M < \infty$  by Corollary 5.5.

A graded morphism  $\varphi : M \rightarrow N$  induces morphisms  $\varphi_i : M_i \rightarrow N_i$  for all  $i \in \mathbb{Z}$ . Tensoring with  $\omega_E$  yields morphisms  $\mathbf{R}(\varphi)^i : \mathbf{R}(M)^i \rightarrow \mathbf{R}(N)^i$ . These morphisms are chain morphisms, as  $x_j \circ \varphi_i = \varphi_{i+1} \circ x_j$  and the  $\mu^i$  are induced by the  $x_j$  for all  $i \in \mathbb{Z}$  and all  $0 \leq j \leq n$ .

Restricting  $\mathbf{R}$  to  $S\text{-grmod}_{\geq d}$  corestricts to  $E\text{-grlin}^{\geq d}$  by construction. These functors are obviously faithful. The fullness  $\mathbf{R}$  and  $\mathbf{R}^{\geq d}$  follows directly from the below Proposition 5.9 and Corollary 5.10, respectively.  $\square$

## 5.2 The Functor $\mathbf{R}$ Induces an Equivalence

The functor  $\mathbf{R}$  is an equivalence  $S\text{-grmod} \xrightarrow{\sim} E\text{-grlin}$  by Eisenbud et al. [11, Prop. 2.1]. In this section we explicitly construct the left adjoint quasi-inverse  $\mathbf{M}$  of  $\mathbf{R}$  and thus show *constructively* that  $\mathbf{R}$  is an adjoint equivalence.

**Proposition 5.9** *There exists a functor  $\mathbf{M} : E\text{-grlin} \rightarrow S\text{-grmod}$  such that  $\mathbf{M} \dashv \mathbf{R}$  is an adjoint equivalence of categories which sends  $S\text{-grmod}^0$  to  $E\text{-grlin}^0$ .*

*Proof* Let  $(C, \mu) \in E\text{-grlin}$ .

For a preparatory step, assume that

$$H^r(C) \text{ is the only nonvanishing cohomology (this implies that } C^{<r} = 0). \quad (\text{A})$$

Consider  $\mu^r : W \otimes_B C_r^r \rightarrow C_{r+1}^{r+1}$  and extend  $\ker(\mu^r) \xrightarrow{\kappa} W \otimes_B C_r^r$  to a map  $S \otimes_B \ker(\mu^r) \rightarrow S \otimes_B C_r^r$ . Define  $\mathbf{M}(C)$  as its cokernel (with relatively free presentation  $\pi_r : S \otimes_B C_r^r \twoheadrightarrow \mathbf{M}(C)$ ).

To justify the correctness of this preparatory step let  $M \in S\text{-grmod}$  with  $M_{<r} = 0$  and such that  $\mathbf{R}(M)$  satisfies assumption (A). The natural isomorphism  $\tilde{\delta}^{-1} : M_r \xrightarrow{\sim} \mathbf{M}(\mathbf{R}(M))_r : m \mapsto 1_S \otimes_B 1_{\omega_E} \otimes_B m$  identifies a minimal set of generators  $M$  with one of  $\mathbf{M}(\mathbf{R}(M))$ . The assumption (A) for  $\mathbf{R}(M)$  is equivalent, by Lemma 5.3, to  $M$  being generated in degree  $r$  and having a relatively free resolution which is linear in the  $x_i$ . In particular, the only relations involving the indeterminates  $x_i$  of the finite set of generators of  $M$  in  $M_r$  are linear relations. All these linear relations are encoded in the map  $\mathbf{R}(M)^r \rightarrow \mathbf{R}(M)^{r+1}$ . The construction of  $\mathbf{M}$  above just imposes these linear relations of the generators of  $M$  to the generators of  $\mathbf{M}(\mathbf{R}(M))$ . In particular,  $\tilde{\delta}$  induces an isomorphism  $\delta_M : \mathbf{M}(\mathbf{R}(M)) \rightarrow M$ . Similarly, there exists an isomorphism  $\eta_C : C \rightarrow \mathbf{R}(\mathbf{M}(C))$  for any  $C \in E\text{-grlin}$  satisfying assumption (A).

For a general  $(C, \mu) \in E\text{-grlin}$ , there is a bound  $r$  (e.g., any  $r > \text{reg}(C)$ ) such that the preparatory step applies to  $(C^{\geq r}, \mu^{\geq r})$ . Then, we inductively define  $\mathbf{M}(C)$  by decreasing the cohomological degree  $d$ . Let  $(C, \mu) \in E\text{-grlin}$  be a complex and  $d < r$  such that  $\mathbf{M}(C^{\geq d+1})$  is defined by the induction hypothesis with relatively free presentation  $\pi_{d+1} : S \otimes_B (C_{d+1}^{d+1} \oplus \dots \oplus C_r^r) \twoheadrightarrow \mathbf{M}(C^{\geq d+1})$ . We define  $\mathbf{M}(C^{\geq d})$  as a pushout of the span of  $\beta$  and  $\gamma$  defined as follows: Let  $\alpha : S \otimes_B W \rightarrow S : p \otimes x_i \rightarrow x_i p$  and  $\iota : C_{d+1}^{d+1} \hookrightarrow C_{d+1}^{d+1} \oplus \dots \oplus C_r^r$  be the embedding in the direct sum. Now set  $\beta := \alpha \otimes_B C_d^d$  and  $\gamma := \pi_{d+1} \circ (S \otimes_B (\iota \circ \mu^d))$  with common source  $S \otimes_B W \otimes C_d^d$  (recall,  $\mu^d : W \otimes_B C_d^d \rightarrow C_{d+1}^{d+1}$ ). This inductive step of the construction of  $\mathbf{M}$  is the reverse construction of  $\mathbf{R}$ .

To apply  $\mathbf{M}$  to a morphism  $\varphi : C \rightarrow D$  in  $E\text{-grlin}$  we use the identification of  $\mathbf{M}(C)_i$  with  $C_i^i$ , map  $C_i^i$  using  $\varphi^i$  to  $D_i^i$ , which we finally identify with  $\mathbf{M}(D)_i$ .

This equivalence of categories is an adjoint equivalence. We already have constructed the unit  $\eta$  and counit  $\delta$  as natural isomorphisms in the preparatory step. This unit and counit naturally extends into lower cohomological degrees using the natural  $B$ -isomorphisms  $C_i^i \xrightarrow{\sim} \mathbf{M}(C)_i : c \mapsto 1_S \otimes_B c$  and  $M_i \xrightarrow{\sim} \mathbf{R}(M)_i^i : m \mapsto 1_{\omega_E} \otimes_B m$ . The triangle identities are easily verified.  $\square$

**Corollary 5.10** *The restriction-corestriction  $\mathbf{M}_{\geq d} : E\text{-grlin}^{\geq d} \rightarrow S\text{-grmod}_{\geq d}$  of  $\mathbf{M}$  and the functor  $\mathbf{R}^{\geq d}$  form an adjoint equivalence  $\mathbf{M}_{\geq d} \dashv \mathbf{R}^{\geq d}$ , which sends  $S\text{-grmod}_{\geq d}^0$  to  $E\text{-grlin}^{\geq d,0}$ .*

### 5.3 Saturated Linear Complexes

We now give a characterization of saturated linear complexes corresponding to the one we gave for graded modules.

**Definition 5.11** The **linear regularity** of a linear complex  $C \in E\text{-grlin}$  is defined as

$$\text{linreg } C := \max\{a \in \mathbb{Z} \mid H^a(C)_{a+n+1} \neq 0 \text{ or } H^a(C)_{a+n} \neq 0\} \in \mathbb{Z} \cup \{-\infty\}.$$

We get a further characterization of  $E\text{-grlin}^0$ -saturated linear complexes.

**Corollary 5.12** A complex  $C \in E\text{-grlin}$  is  $E\text{-grlin}^0$ -saturated iff  $\text{linreg } C = -\infty$ .

*Proof* The module  $\mathbf{M}(C)$  is  $S\text{-grmod}^0$ -saturated if  $\text{Ext}_{\bullet}^j(B, \mathbf{M}(C)) = 0$  for  $j \in \{0, 1\}$  by Proposition 3.9. This is equivalent to  $H^a(\mathbf{R}(\mathbf{M}(C)))_{a+n+1-j} = 0$  for  $j \in \{0, 1\}$  by the key Lemma 5.4. The claim follows from  $C \cong \mathbf{R}(\mathbf{M}(C))$ .  $\square$

The key Lemma 5.4 also implies:

**Corollary 5.13**  $\text{linreg } C = \text{linreg } \mathbf{M}(C)$  for all  $C \in E\text{-grlin}$ .

The localizing subcategory  $S\text{-grmod}_{\geq d}^0$  of  $S\text{-grmod}_{\geq d}$  corresponds via the adjoint equivalence  $\mathbf{M} \dashv \mathbf{R}$  to the full localizing subcategory  $E\text{-grlin}^{\geq d, 0}$  of right bounded complexes in  $E\text{-grlin}^{\geq d}$ , i.e., of those complexes  $C \in E\text{-grlin}^{\geq d}$  with  $C^{\geq \ell} = 0$  for  $\ell$  large enough. A module  $M \in S\text{-grmod}_{\geq d}$  is then  $S\text{-grmod}_{\geq d}^0$ -saturated if and only if  $\mathbf{R}(M)$  is  $E\text{-grlin}^{\geq d, 0}$ -saturated, i.e., the adjoint equivalence  $\mathbf{M}_{\geq d} \dashv (\mathbf{R}^{\geq d} : S\text{-grmod}_{\geq d} \rightarrow E\text{-grlin}^{\geq d})$  restricts to an adjoint equivalence between the full subcategories of  $S\text{-grmod}_{\geq d}^0$ -saturated resp.  $E\text{-grlin}^{\geq d, 0}$ -saturated objects.

The definition of the linear regularity of complexes in  $E\text{-grlin}^{\geq d}$  and the characterization of  $E\text{-grlin}^{\geq d, 0}$ -saturated linear complexes is a little bit more subtle and is therefore deferred to the next section. The reason is that the lowest cohomology  $H^d(C)$  has to be treated separately.

## 6 Saturation of Linear Complexes

The ideal transform in Sect. 4 leads to an algorithm for the saturation of graded  $S$ -modules. In this section, we present an algorithm to saturate linear complexes. The adjoint equivalence  $\mathbf{M} \dashv \mathbf{R}$  translates this to a second algorithm for the saturation of graded  $S$ -modules.

Corollary 5.12 indicates that one has to modify a linear complex  $C$  until the conditions  $H^a(C)_{a+n+1} = 0$  and  $H^a(C)_{a+n} = 0$  hold. Our purely linear saturation is similar to that of the Tate resolution in that one truncates  $C$  in cohomological degree high enough and then computes a suitable part in lower cohomological degrees. In contrast to the Tate resolution, our approach remains in the category

of linear complexes, as we do not take free presentations of kernels to compute the part of lower cohomological degree, but so-called purely linear kernels. We can also truncate above the linear regularity, a lower bound of the Castelnuovo-Mumford regularity. For the relation between the purely linear saturation and the Tate resolution see Remark 7.4.

### 6.1 Purely Linear Kernels

Let  $C^i, C^{i+1} \in E\text{-grmod}$  be relatively free with socle concentrated in degree  $i$  and  $i + 1$ , respectively.<sup>12</sup> We call a morphism  $\varphi^i : C^i \rightarrow C^{i+1}$  **purely linear (of degree  $i$ )** if its kernel vanishes in the top degree  $i + n + 1$ .

$$\begin{array}{ccccc}
 K^{i-1} & \xrightarrow{\kappa} & C^i & \xrightarrow{\varphi^i} & C^{i+1} \\
 \uparrow \psi & \nearrow \lambda & & & \uparrow \\
 L^{i-1} & & & & \\
 & & & \searrow 0 & \\
 & & & & 
 \end{array}$$

**Definition 6.1** Let  $\varphi^i : C^i \rightarrow C^{i+1}$  be purely linear of degree  $i$ . A purely linear morphism  $\kappa : K^{i-1} \rightarrow C^i$  of degree  $i - 1$  with  $\varphi^i \circ \kappa = 0$  is called **purely linear kernel** if for any purely linear  $\lambda : L^{i-1} \rightarrow C^i$  of degree  $i - 1$  with  $\varphi^i \circ \lambda = 0$  there exists a unique morphism  $\psi : L^{i-1} \rightarrow K^{i-1}$  with  $\kappa \circ \psi = \lambda$ .

**Lemma 6.2** Each purely linear morphism has a purely linear kernel, which, by the universal property, is unique up to a unique isomorphism.

*Proof* We denote the restriction of any morphism  $\beta$  to the graded part of degree  $i + n$  by  $\beta_{i+n}$ .

Let  $\varphi^i : M^i \rightarrow M^{i+1}$  be purely linear of degree  $i$  and  $\nu : N^{i-1} \hookrightarrow M^i$  be its kernel. Denote by  $K^{i-1} := N_{i+n}^{i-1} \otimes_B E$  and by  $\lambda : K^{i-1} \rightarrow N^{i-1}$  the map induced by the identity on  $N_{i+n}^{i-1}$ . We show that  $\kappa := \nu \circ \lambda : K^{i-1} \rightarrow M^i$  is a purely linear kernel of  $\varphi^i$ .

By definition,  $K^{i-1}$  and  $M^i$  are relatively free generated in degree  $i + n$  and degree  $i + n + 1$ , respectively. As  $\varphi^i$  is purely linear,  $N^{i-1}$  lives in the degree interval  $i, \dots, i + n$ . In particular,  $\nu_{i+n}$  is a kernel of  $\varphi_{i+n}^i$ . By definition,  $\lambda_{i+n} : K_{i+n}^{i-1} \rightarrow N_{i+n}^{i-1}$  is an isomorphism and thus also  $\kappa_{i+n}$  is a kernel of  $\varphi_{i+n}^i$ . In particular, the kernel of  $\kappa$  lives in the degree interval  $i - 1, \dots, i + n - 1$ . Thus,  $\kappa$  is purely linear of degree  $i - 1$ .

The composition  $\varphi^i \circ \nu$  is zero and  $\kappa$  factors over  $\nu$  by construction. Thus,  $\varphi^i \circ \kappa = 0$ .

<sup>12</sup>Or, equivalently, freely generated in degree  $i + n + 1$  and  $i + n + 2$ , respectively.

$$\begin{array}{ccccc}
& & N^{i-1} & & \\
& & \uparrow \lambda & \searrow \nu & \\
& & K^{i-1} & \xrightarrow{\kappa} & M^i & \xrightarrow{\varphi^i} & M^{i+1} \\
& \uparrow \psi & & \nearrow \varphi^{i-1} & & & \\
& \vdots & & & & & \\
& & M^{i-1} & & & \searrow 0 & \\
& & & & & & 
\end{array}$$

To show the universal property of  $\kappa$  let  $\varphi^{i-1} : M^{i-1} \rightarrow M^i$  be purely linear with  $\varphi^i \circ \varphi^{i-1} = 0$ . From the universal property of  $\kappa_{i+n}$  as a kernel, there exists a unique  $\psi_{i+n} : M_{i+n}^{i-1} \rightarrow K_{i+n}^{i-1}$  with  $\kappa_{i+n} \circ \psi_{i+n} = \varphi_{i+n}^{i-1}$ , since  $\varphi_{i+n}^{i-1} \circ \varphi_{i+n}^i = 0$ . We define

$$\psi := \psi_{i+n} \otimes_B E : M^{i-1} \cong M_{i+n}^{i-1} \otimes_B E \longrightarrow K^{i-1} \cong K_{i+n}^{i-1} \otimes_B E,$$

which extends  $\psi_{i+n}$  to a morphism of graded  $E$ -modules. Finally,  $\varphi^{i-1} = \kappa \circ \psi$  since  $\kappa_{i+n} \circ \psi_{i+n} = \varphi_{i+n}^{i-1}$  and  $\varphi^{i-1}$  is uniquely determined by  $\varphi_{i+n}^{i-1}$ .  $\square$

Note that all steps in the proof of this last lemma are constructive.

We can now state the definition of linear regularity for complexes in  $E\text{-grlin}^{\geq d}$ .

**Definition 6.3** Define for any  $d \leq 0$  the  $d$ -th truncated linear regularity  $\text{linreg}_{\geq d} C$  of  $C \in E\text{-grlin}^{\geq d}$  as an element of  $\mathbb{Z}_{\geq d} \cup \{-\infty\}$  as follows:

If there exists an  $a \in \mathbb{Z}_{> d}$  such that  $H^a(C)_{a+n+1} \neq 0$  or  $H^a(C)_{a+n} \neq 0$  then

$$\text{linreg}_{\geq d} C := \max\{a \in \mathbb{Z}_{> d} \mid H^a(C)_{a+n+1} \neq 0 \text{ or } H^a(C)_{a+n} \neq 0\} \in \mathbb{Z}_{> d}.$$

Otherwise, if the lowest morphism  $C^d \rightarrow C^{d+1}$  is a purely linear kernel (of  $C^{d+1} \rightarrow C^{d+2}$ ) then  $\text{linreg}_{\geq d} C := -\infty$  else  $\text{linreg}_{\geq d} C := d$ .

**Corollary 6.4**  $C \in E\text{-grlin}^{\geq d}$  is  $E\text{-grlin}^{\geq d,0}$ -saturated iff  $\text{linreg}_{\geq d} C = -\infty$ .

*Proof* The claim follows from Corollaries 5.12 and 3.11 (by Remark 4.8.(4) we only need to consider the case  $d = 0$ ).  $\square$

**Corollary 6.5**  $\text{linreg}_{\geq d} C = \text{linreg}_{\geq d} \mathbf{M}_{\geq d}(C)$  for all  $C \in E\text{-grlin}^{\geq d}$ .

**Proposition 6.6** A  $(C, \mu) \in E\text{-grlin}^{\geq d}$ , is  $E\text{-grlin}^{\geq d,0}$ -saturated if and only if  $\mu^i$  is the purely linear kernel of  $\mu^{i+1}$  for all  $i \geq d$ .

*Proof* By the proof of Lemma 6.2, a morphism  $\kappa : K^{i-1} \rightarrow M^i$  is a purely linear kernel of a purely linear morphism  $\varphi^i : M^i \rightarrow M^{i+1}$  of degree  $i$  if and only if it is purely linear and  $K^{i-1} \xrightarrow{\kappa} M^i \xrightarrow{\varphi^i} M^{i+1}$  is a complex which is exact in degrees  $i+n$  and  $i+n+1$ . Now, the claim follows from the characterizations of saturated linear complexes in Corollary 6.4.  $\square$

## 6.2 Saturation of a Linear Complex

In this subsection we algorithmically saturate linear complexes by iteratively computing purely linear kernels.

Let  $(C, \mu) \in E\text{-grlin}^{\geq d}$  with regularity  $r \in \mathbb{Z}$ . Define the **purely linear saturation (truncated in degree  $d$ )** functor  $\mathbf{S}^{\geq d} : E\text{-grlin}^{\geq d} \rightarrow E\text{-grlin}^{\geq d}$  as follows. The idea is to truncate the complex above the *linear* regularity and then to “saturate” it recursively by purely linear kernels, more precisely: For cohomological degrees greater than the *linear* regularity  $r = \text{linreg}_{\geq d} C$  define  $\mathbf{S}^{\geq r+1}$  by setting  $\mathbf{S}^{\geq r+1}(C, \mu) := (C^{\geq r+1}, \mu^{\geq r+1})$ . Assume that  $(D^{\geq i}, \tau^{\geq i}) = \mathbf{S}^{\geq i}(C, \mu)$  is defined for some  $i > d$ . Let  $\tau^{i-1} : D^{i-1} \rightarrow D^i$  be the purely linear kernel of  $\tau^i$ . Define  $\mathbf{S}^{\geq i-1}(C, \mu)$  by adding  $\tau^{i-1}$  to  $(D^{\geq i}, \tau^{\geq i})$  in cohomological degree  $i - 1$ .

The morphism part  $\mathbf{S}^{\geq d}(\varphi)$  for  $\varphi : (C, \mu_C) \rightarrow (C', \mu_{C'})$  in  $E\text{-grlin}^{\geq d}$  is induced by the identity in high degrees. The universal property of the purely linear kernels implies a *unique* completion of the square and thus iteratively constructs the chain morphisms in lower degrees.

$$\begin{array}{ccc} \mathbf{S}^{\geq d}(C)^\ell & \longrightarrow & \mathbf{S}^{\geq d}(C)^{\ell+1} \\ \vdots & & \downarrow \mathbf{S}^{\geq d}(\varphi)^{\ell+1} \\ \mathbf{S}^{\geq d}(C')^\ell & \longrightarrow & \mathbf{S}^{\geq d}(C')^{\ell+1} \end{array}$$

**Theorem 6.7** *Let  $\mathcal{A} = E\text{-grlin}^{\geq d}$  and  $\mathcal{C} := E\text{-grlin}^{\geq d,0}$ . There exists a natural transformation  $\eta : \text{Id}_{\mathcal{A}} \rightarrow \mathbf{S}^{\geq d}$  such that the purely linear saturation  $\mathbf{S}^{\geq d}$  truncated in degree  $d$  together with this natural transformation  $\eta$  is a Gabriel monad of  $\mathcal{A}$  w.r.t.  $\mathcal{C}$ .*

Again, the statement of the theorem is valid for all  $d \in \mathbb{Z}$ . The statement of the following immediate corollary and the proof the theorem assume  $d \leq 0$ .

**Corollary 6.8** *The nonnegative integer  $\max\{\text{linreg}_{\geq d} C - d + 1, 0\}$  is the precise count of recursion steps needed to achieve saturation.*

Thus, the linear regularity yields a better bound for computing zeroth cohomologies and saturation than the Castelnuovo-Mumford regularity does. However, the data structure for  $E\text{-grlin}^{\geq d}$  suggested in Remark 5.7 still requires the Castelnuovo-Mumford regularity.

*Proof (of Theorem 6.7)* First, we construct the natural transformation  $\eta_C$  for the complex  $(C, \mu) \in E\text{-grlin}^{\geq d}$ . Consider the cochain-isomorphism  $\eta_C : C^{\geq r} \rightarrow \mathbf{S}^{\geq r}(C)$  induced by the identity for  $r > \text{linreg}_{\geq d} C$ . Assume that  $\eta_C$  is lifted to a cochain morphism  $C^{\geq \ell+1} \rightarrow \mathbf{S}^{\geq \ell+1}(C)$ . The universal property of the purely linear kernels implies a completion of the square by a morphism  $\eta_C^\ell : C^\ell \rightarrow \mathbf{S}^{\geq d}(C)^\ell$ . Iteratively, we get the cochain-morphism  $\eta_C : C \rightarrow \mathbf{S}^{\geq d}(C)$ .

$$\begin{array}{ccc}
C^\ell & \longrightarrow & C^{\ell+1} \\
\downarrow \eta_C^\ell & & \downarrow \eta_C^{\ell+1} \\
\mathbf{S}^{\geq d}(C)^\ell & \longrightarrow & \mathbf{S}^{\geq d}(C)^{\ell+1}
\end{array}$$

Now, we use Theorem 2.1 to show that  $\mathbf{S}^{\geq d}$  together with  $\eta$  is a Gabriel monad.

2.1.(1)  $\mathcal{C} \subset \ker \mathbf{S}^{\geq d}$ :

As  $\eta_C$  is an isomorphism in high cohomological degrees, its kernel is contained in  $\mathcal{C}$ .

2.1.(2)  $\mathbf{S}^{\geq d}(\mathcal{A}) \subset \text{Sat}_{\mathcal{C}}(\mathcal{A})$ :

$\mathbf{S}^{\geq d}(C)$  has only trivial cohomologies above the regularity of  $C$ . Below the regularity we use Proposition 6.6.

2.1.(3)  $G := \text{co-res}_{\text{Sat}_{\mathcal{C}}(\mathcal{A})} \mathbf{S}^{\geq d}$  is exact:

As  $\mathbf{S}^{\geq d}$  is the identity on objects and morphism in high cohomological degree, applying it to a short exact sequence in  $\mathcal{A}$  yields a new sequence with  $\mathcal{A}$ -defects, which are bounded by the maximum of the regularities of said short exact sequence. Thus, the  $\mathcal{A}$ -defects are contained in  $\mathcal{C}$ . In particular, this sequence is exact when considered in  $\text{Sat}_{\mathcal{A}}(\mathcal{C})$ .

2.1.(4)  $\eta \mathbf{S}^{\geq d} = \mathbf{S}^{\geq d} \eta$ :

Truncated at cohomological degree  $\ell$  above the regularity this is clear, since both natural transformations are induced by the identity. For lower degrees, this follows from the uniqueness of the universal morphism  $\psi$  in the definition of purely linear kernels.

2.1.(5)  $\eta \iota$  is a natural isomorphism:

Let  $C \in \mathcal{A}$  be  $\mathcal{C}$ -saturated. We need to show that  $\eta_C$  is a cochain isomorphism. This is clear in high cohomological degrees, as  $\mathbf{S}^{\geq d}$  is the identity on objects and morphism there. Assume that  $\eta_C$  restricted to  $C^{\geq \ell+1} \rightarrow \mathbf{S}^{\geq \ell+1}(C)$  for some  $\ell \in \mathbb{Z}$  is a cochain isomorphism. Then, the morphism  $\eta_C^\ell$  from the completion of the square is an isomorphism, since both  $C$  and  $\mathbf{S}^{\geq d}(C)$  are saturated and, by Proposition 6.6  $C^\ell$  and  $\mathbf{S}^{\geq d}(C)^\ell$  are purely linear kernels of  $\mu^{\ell+1}$  and the morphism in cohomological degree  $\ell + 1$  of  $\mathbf{S}^{\geq d}(C)$ , respectively. The uniqueness of purely linear kernels implies that  $\eta_C$  restricted to  $C^{\geq \ell} \rightarrow \mathbf{S}^{\geq \ell}(C)$  is a cochain isomorphism, and so is  $\eta_C$  by induction.  $\square$

$$\begin{array}{ccc}
C^\ell & \longrightarrow & C^{\ell+1} \\
\downarrow \eta_C^\ell & & \downarrow \eta_C^{\ell+1} \\
\mathbf{S}^{\geq d}(C)^\ell & \longrightarrow & \mathbf{S}^{\geq d}(C)^{\ell+1}
\end{array}$$

We stress that the above functors  $\mathbf{M}$ ,  $\mathbf{M}_{\geq d}$ ,  $\mathbf{R}$ ,  $\mathbf{R}^{\geq d}$ , and  $\mathbf{S}^{\geq d}$  are constructive functors between constructively Abelian categories. We furthermore note that computing the natural transformation  $\eta$  is constructive.

## 7 The Gabriel Monad of the Category of Coherent Sheaves

In this section we prove that the quotient category  $S\text{-grmod}_{\geq d}/S\text{-grmod}_{\geq d}^0$  is equivalent to the category  $\mathcal{Coh} \mathbb{P}_B^n$  for any  $d \in \mathbb{Z}$  and that the corresponding Gabriel monad computes the (truncated) module of twisted global sections.

**Proposition 7.1**  $\mathcal{Coh} \mathbb{P}_B^n \simeq S\text{-grmod}_{\geq d}/S\text{-grmod}_{\geq d}^0$  for all  $d \in \mathbb{Z}$ .

*Proof* The definitions directly imply  $S\text{-grmod}^0 \cap S\text{-grmod}_{\geq d} = S\text{-grmod}_{\geq d}^0$ . Now, a preimage of  $M \in S\text{-grmod}/S\text{-grmod}^0$  under  $S\text{-grmod}_{\geq d} \rightarrow S\text{-grmod}/S\text{-grmod}^0$  is given by  $M_{\geq d}$ , since  $M \cong M_{\geq d}$  in  $S\text{-grmod}/S\text{-grmod}^0$ . Hence, the second isomorphism theorem for Abelian categories [4, Prop. 3.2] implies the equivalence

$$S\text{-grmod}_{\geq d}/S\text{-grmod}_{\geq d}^0 \simeq S\text{-grmod}/S\text{-grmod}^0.$$

The latter category is equivalent to  $\mathcal{Coh} \mathbb{P}_B^n$  by Barakat and Lange-Hegermann [4, Coro. 4.2].  $\square$

A graded  $S$ -modules  $M$  is called quasi finitely generated if each truncation  $M_{\geq d}$  is finitely generated. We denote by  $S\text{-qfgrmod} \subset S\text{-grMod}$  the full subcategory of such modules. The functor

$$H_{\bullet}^0 : \mathcal{Coh} \mathbb{P}_B^n \rightarrow S\text{-qfgrmod} : \mathcal{F} \mapsto \bigoplus_{p \in \mathbb{Z}} H^0(\mathbb{P}_B^n, \mathcal{F}(p))$$

computing the module of twisted global sections is right adjoint to the sheafification functor  $\text{Sh} : S\text{-qfgrmod} \rightarrow \mathcal{Coh} \mathbb{P}_B^n, M \mapsto \widetilde{M}$ . This was proved by Serre in the absolute case [13, 59] and later by Grothendieck for the relative case.

Denote by  $\text{Sh}_{\geq d} : S\text{-grmod}_{\geq d} \rightarrow \mathcal{Coh} \mathbb{P}_B^n$  the restriction of  $\text{Sh}$  to  $S\text{-grmod}_{\geq d}$  and by

$$H_{\geq d}^0 : \mathcal{Coh} \mathbb{P}_B^n \rightarrow S\text{-grmod}_{\geq d} : \mathcal{F} \mapsto \bigoplus_{p \in \mathbb{Z}_{\geq d}} H^0(\mathbb{P}_B^n, \mathcal{F}(p))$$

the functor computing the truncated module of twisted global sections. It follows that  $H_{\geq d}^0$  is the right adjoint of  $\text{Sh}_{\geq d}$ .

**Proposition 7.2** *The monad  $H_{\geq d}^0(\widetilde{\phantom{x}}) = H_{\geq d}^0 \circ \text{Sh}_{\geq d}$  is a Gabriel monad of  $S\text{-grmod}_{\geq d}$  w.r.t. the localizing subcategory  $S\text{-grmod}_{\geq d}^0$ . In particular, any Gabriel monad computes the truncated module of twisted global sections.*

*Proof* Let  $\mathcal{Q}_{\geq d} : S\text{-grmod}_{\geq d} \rightarrow S\text{-grmod}_{\geq d}/S\text{-grmod}_{\geq d}^0$  be the canonical functor. The equivalence in Proposition 7.1 is constructed as a functor

$$\alpha_{\geq d} : S\text{-grmod}_{\geq d}/S\text{-grmod}_{\geq d}^0 \rightarrow \mathcal{Coh} \mathbb{P}_B^n$$



with  $\alpha_{\geq d} \circ \mathcal{Q}_{\geq d} \simeq \text{Sh}_{\geq d}$ . An easy calculation shows that a right adjoint of  $\mathcal{Q}_{\geq d}$  is given by  $\mathcal{S}_{\geq d} := H_{\geq d}^0 \circ \alpha_{\geq d}$ . In particular,  $\mathcal{S}_{\geq d} \circ \mathcal{Q}_{\geq d}$  is a Gabriel monad of  $S\text{-grmod}_{\geq d}$  w.r.t.  $S\text{-grmod}_{\geq d}^0$  by Barakat and Lange-Hegermann [3, Lemma 4.3]. Now the claim follows, as  $\mathcal{S}_{\geq d} \circ \mathcal{Q}_{\geq d} = H_{\geq d}^0 \circ \alpha_{\geq d} \circ \mathcal{Q}_{\geq d} \simeq H_{\geq d}^0 \circ \text{Sh}_{\geq d} = H_{\geq d}^0(\tilde{\cdot})$ .  $\square$

**Corollary 7.3** *There exist natural isomorphisms*

$$H_{\geq d}^0(\tilde{M}) \cong D_{m, \geq d}(M) \quad \text{and} \quad \mathbf{R}(H_{\geq d}^0(\tilde{M})) \cong \mathbf{S}^{\geq d}(\mathbf{R}(M)),$$

in particular for  $i \geq d$

$$H^0(\tilde{M}(i)) \cong (D_{m, \geq d}(M))_i \cong (\mathbf{S}^{\geq d}(\mathbf{R}(M)))_i^i.$$

*Remark 7.4* In the absolute case, i.e., when  $B = k$  is a field, the (objects of the truncated) Tate resolution  $\mathbf{T}^{\geq d}(M)$  relate to the higher cohomology modules  $H_{\geq d}^q(\tilde{M})$  by Eisenbud et al. [11]

$$\mathbf{T}^{\geq d}(M)^i = \bigoplus_{q=0}^{\min\{n, i-d\}} \omega_E \otimes_k H^q(\tilde{M}(i-q)), \quad (*)$$

while the (truncated) purely linear saturation directly extracts  $H^0$ :

$$\mathbf{S}^{\geq d}(\mathbf{R}(M))^i = \omega_E \otimes_B H^0(\tilde{M}(i)).$$

In the relative case the analogue of (\*) is more subtle: The Tate resolution  $\mathbf{T}^{\geq d}(M)$  is by its bi-graded structure in fact a multi-complex  $\mathbf{T}^{\geq d, \bullet}(M)$ . Since each multi-complex is a filtered complex and hence induces a spectral sequence<sup>13</sup>  $E^{p,q}(\mathbf{T}^{\geq d, \bullet}(M)) \implies 0$ , where for each row-complex on the first page the following isomorphism holds

$$E_1^{\geq d, q}(\mathbf{T}^{\geq d, \bullet}(M)) \cong \mathbf{R}(H_{\geq d}^q(\tilde{M})).$$

This is implicit in [10], see also [11, Corollary 3.6]. Thus, the relation between the purely linear saturation and the Tate resolution is given by

$$E_1^{\geq d, 0}(\mathbf{T}^{\geq d, \bullet}(M)) \cong \mathbf{S}^{\geq d}(\mathbf{R}(M)) \cong \mathbf{R}(H_{\geq d}^0(\tilde{M})).$$

<sup>13</sup>The vertical morphisms of this multi-complex are the morphisms between the graded summands represented by scalar matrices (i.e., degree zero in  $E$ ). This differs from the MACAULAY2 convention used in [10], where these morphisms are arranged “diagonally up and to the right” (cf. [10, Chapter 3]). Hence, we do not arrange the direct summands of the modules in the Tate resolution vertically, but diagonally up and to the left.

The first isomorphism is not a priori obvious in the relative case and gives a direct way to compute  $\mathbf{R}(H_{\geq d}^0(\widetilde{M}))$  via  $\mathbf{S}^{\geq d}(\mathbf{R}(M))$  without computing (most of) the Tate resolution.<sup>14</sup>

## References

1. W.W. Adams, P. Lounstaunau, *An Introduction to Gröbner Bases*. Graduate Studies in Mathematics, vol. 3 (American Mathematical Society, Providence, RI, 1994). MR 1287608 (95g:13025)
2. M. Barakat, M. Lange-Hegermann, An axiomatic setup for algorithmic homological algebra and an alternative approach to localization. *J. Algebra Appl.* **10**(2), 269–293 (2011). [arXiv:1003.1943](#). MR 2795737 (2012f:18022)
3. M. Barakat, M. Lange-Hegermann, On monads of exact reflective localizations of Abelian categories. *Homology Homotopy Appl.* **15**(2), 145–151 (2013). [arXiv:1202.3337](#). MR 3138372
4. M. Barakat, M. Lange-Hegermann, Characterizing Serre quotients with no section functor and applications to coherent sheaves. *Appl. Categ. Struct.* **22**(3), 457–466 (2014). [arXiv:1210.1425](#). MR 3200455
5. M. Barakat, M. Lange-Hegermann, Gabriel morphisms and the computability of Serre quotients with applications to coherent sheaves (2014). [arXiv:1409.2028](#)
6. M. Barakat, M. Lange-Hegermann, On the Ext-computability of Serre quotient categories. *J. Algebra* **420**, 333–349 (2014). [arXiv:1212.4068](#). MR 3261464
7. I.N. Bernšteĭn, I.M. Gel’fand, S.I. Gel’fand, Algebraic vector bundles on  $\mathbf{P}^n$  and problems of linear algebra. *Funktional. Anal. i Prilozhen.* **12**(3), 66–67 (1978). MR 509387 (80c:14010a)
8. M.P. Brodmann, R.Y. Sharp, *Local Cohomology: An Algebraic Introduction with Geometric Applications*. Cambridge Studies in Advanced Mathematics, vol. 60 (Cambridge University Press, Cambridge, 1998). MR 1613627 (99h:13020)
9. W. Decker, D. Eisenbud, Sheaf algorithms using the exterior algebra, in *Computations in Algebraic Geometry with Macaulay 2*. Algorithms and Computation in Mathematics, vol. 8 (Springer, Berlin, 2002), pp. 215–249. MR 1949553
10. D. Eisenbud, F.-O. Schreyer, Relative Beilinson monad and direct image for families of coherent sheaves. *Trans. Am. Math. Soc.* **360**(10), 5367–5396 (2008). [arXiv:math/0506391](#). MR 2415078 (2009f:14030)
11. D. Eisenbud, G. Fløystad, F.-O. Schreyer, Sheaf cohomology and free resolutions over exterior algebras. *Trans. Am. Math. Soc.* **355**(11), 4397–4426 (2003). (electronic). MR 1990756 (2004f:14031)
12. T.Y. Lam, *Lectures on Modules and Rings*. Graduate Texts in Mathematics, vol. 189 (Springer, New York, 1999). MR 1653294 (99i:16001)
13. J.-P. Serre, Faisceaux algébriques cohérents. *Ann. Math. (2)* **61**, 197–278 (1955). MR 0068874 (16,953c)
14. W.V. Vasconcelos, *Computational Methods in Commutative Algebra and Algebraic Geometry*. Algorithms and Computation in Mathematics, vol. 2 (Springer, Berlin, 1998). With chapters by David Eisenbud, Daniel R. Grayson, Jürgen Herzog and Michael Stillman. MR 1484973 (99c:13048)

---

<sup>14</sup>In absolute case  $B = k$  the isomorphism easily follows from the fact that the bottom complex  $E_1^{\geq d, 0}(\mathbf{T}^{\geq d, \bullet}(M))$  of the first spectral sequence is already the subcomplex of the Tate resolution consisting of those direct summands of the objects in  $\mathbf{T}^{\geq d}(M)$  where the degree of the socle equals the cohomological degree.

# Local to Global Algorithms for the Gorenstein Adjoint Ideal of a Curve



Janko Böhm, Wolfram Decker, Santiago Laplagne, and Gerhard Pfister

**Abstract** We present new algorithms for computing adjoint ideals of curves and thus, in the planar case, adjoint curves. With regard to terminology, we follow Gorenstein who states the adjoint condition in terms of conductors. Our main algorithm yields the Gorenstein adjoint ideal  $\mathfrak{G}$  of a given curve as the intersection of what we call local Gorenstein adjoint ideals. Since the respective local computations do not depend on each other, our approach is inherently parallel. Over the rationals, further parallelization is achieved by a modular version of the algorithm which first computes a number of the characteristic  $p$  counterparts of  $\mathfrak{G}$  and then lifts these to characteristic zero. As a key ingredient, we establish an efficient criterion to verify the correctness of the lift. Well-known applications are the computation of Riemann-Roch spaces, the construction of points in moduli spaces, and the parametrization of rational curves. We have implemented different variants of our algorithms together with MnuK's approach in the computer algebra system SINGULAR and give timings to compare the performance.

**Keywords** Adjoint ideals • Singularities • Algebraic curves

**2010 Mathematics Subject Classification** Primary 14Q05; Secondary 14H20, 14H50, 68W10

---

J. Böhm (✉) • W. Decker • G. Pfister

Fachbereich Mathematik, Technische Universität Kaiserslautern, Postfach 3049, 67653  
Kaiserslautern, Germany

e-mail: [boehm@mathematik.uni-kl.de](mailto:boehm@mathematik.uni-kl.de); [decker@mathematik.uni-kl.de](mailto:decker@mathematik.uni-kl.de);  
[pfister@mathematik.uni-kl.de](mailto:pfister@mathematik.uni-kl.de)

S. Laplagne

Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Ciudad Universitaria,  
1428 Pabellón I, Buenos Aires, Argentina

e-mail: [slaplagn@dm.uba.ar](mailto:slaplagn@dm.uba.ar)

## 1 Introduction

In classical algebraic geometry, starting from Riemann’s paper on abelian functions [53], the adjoint curves of an irreducible plane curve  $\Gamma$  have been used as an essential tool in the study of the geometry of  $\Gamma$ . The defining property of an *adjoint curve* is that it passes with “sufficiently high” multiplicity through the singularities of  $\Gamma$ . There are several ways of making this precise, developed in classical papers by Brill and Noether [14], Castelnuovo [16, 17], and Petri [52], and in more recent work by Gröbner [36], Gorenstein [27] and van der Waerden [59], Keller [42]. We refer to [29, 30, 41], and [19] for results comparing the different notions: whereas the adjoint condition given by Brill and Noether is more restrictive, the notions of adjoint curves given by the other authors above coincide.

In this paper, we always consider adjoint curves in the less restrictive sense. In fact, we rely on Gorenstein’s algebraic definition which states the adjoint condition at a singular point  $P \in \Gamma$  by considering the conductor of the local ring  $\mathcal{O}_{\Gamma,P}$  in its normalization. It is a well-known consequence of Max Noether’s Fundamentalsatz that the adjoint curves of any given degree  $m$  cut out, residual to a fixed divisor supported on the singular locus of  $\Gamma$ , a complete linear series. Of fundamental importance is the case  $m = \deg \Gamma - 3$  which, as shown by Gorenstein, yields the canonical series.

The ideal generated by the defining forms of the adjoint curves of  $\Gamma$  is called the *adjoint ideal* of  $\Gamma$ . In [1], the concept of adjoint ideals is extended to the non-planar case: consider a non-degenerate integral curve  $\Gamma \subset \mathbb{P}_k^r = \text{Proj}(S)$ , and let  $I$  be a saturated homogeneous ideal of  $S$  properly containing the ideal of  $\Gamma$ . Then  $I$  is an adjoint ideal of  $\Gamma$  if its homogeneous elements of degree  $m \gg 0$  cut out, residual to a fixed divisor whose support contains the singular locus, a complete linear series. As pointed out in [1], the existence of adjoint ideals is implicit in classical papers: examples are the Castelnuovo adjoint ideal and the Petri adjoint ideal. In [19], it is shown that Gorenstein’s condition leads to the largest possible adjoint ideal, supported on the singular locus and containing all other adjoint ideals, and referred to as the *Gorenstein adjoint ideal*  $\mathfrak{G} = \mathfrak{G}(\Gamma)$ . See [19] for some remarks on how the different concepts of adjoint ideals compare in the non-planar case.

With regard to practical applications, adjoint curves enter center stage in the classical Brill-Noether algorithm for computing Riemann-Roch spaces, which in turn can be used to construct Goppa codes (see [43]). Furthermore, linear series cut out by adjoint curves allow us to construct explicit examples of smooth curves via singular plane models; a typical application is the experimental study of moduli spaces of curves. If the geometric genus of a plane curve  $\Gamma$  is zero, then the adjoint curves of degree  $\deg \Gamma - 2$  specify a birational map to a rational normal curve. Based on this, we can find an explicit parametrization of  $\Gamma$  over its field of definition, starting either from the projective line or a conic. See [5] and the implementation in the SINGULAR library [8]. Algorithms for parametrization, in turn, have applications in computer aided design, for example, to compute intersections of curves with other algebraic varieties. See also [55].

A well-known algorithm for computing the Gorenstein adjoint ideal  $\mathfrak{G} = \mathfrak{G}(\Gamma)$  in the planar case is due to Mnuk [50]. This algorithm makes use of linear algebra to obtain  $\mathfrak{G}$  from an integral basis for the normalization  $\overline{k[C]}$ , where  $C$  is an affine part of  $\Gamma$  containing all singularities of  $\Gamma$ . Efficient ways of finding integral bases rely on Puiseux series techniques (see [10, 60]). This somewhat limits Mnuk’s approach to characteristic zero. The same applies to the algorithm of El Kahoui and Moussa [26], which also computes the Gorenstein adjoint ideal of a plane curve from an integral basis of  $\overline{k[C]}$ . The approach of Orecchia and Ramella [51], on the other hand, is limited to curves with ordinary multiple points only.

In this paper, we present a new algorithm for computing  $\mathfrak{G}$ . This algorithm is highly efficient and not restricted to the planar case, special types of singularities, or to characteristic zero. The basic idea is to compute  $\mathfrak{G}$  as the intersection of “local Gorenstein ideals”, one for each singular point of  $\Gamma$ . Each local ideal is obtained via Gröbner bases, starting from a “local contribution” to the normalization  $\overline{k[C]}$  at the respective singular point. To find these contributions, we use the algorithm from [6] which is a local variant of the normalization algorithm designed in [34, 35].

Our approach is already faster per se. In addition, it can take advantage of handling special classes of singularities in an ad hoc way. Above all, it is inherently parallel. For input over the rationals, further parallelization is achieved by a modular version of the algorithm which first computes a number of characteristic  $p$  counterparts of  $\mathfrak{G}$  and then lifts these to characteristic zero. This allows us, in addition, to avoid intermediate coefficient growth over the rationals. To apply the general rational reconstruction scheme from [11], we prove an efficient criterion to verify the correctness of the lift. Note that the local-to-global approach is particularly useful when combined with modular methods: By Chebotarev’s density theorem [58], the primes  $p$  for which the singular locus decomposes over  $\mathbb{F}_p$  have positive density among all primes, provided the singular locus is decomposable over  $\overline{\mathbb{Q}}$ .

Our paper is organized as follows: In Sect. 2, we discuss algorithmic normalization. In Sect. 3, we review the definition of adjoint ideals and some related facts. In Sect. 4, we describe global algorithmic approaches to obtain  $\mathfrak{G}$ . We first discuss Mnuk’s approach. Then we describe a global approach which relies on normalization and Gröbner bases. In Sects. 5 and 6, we present our local to global algorithm for finding  $\mathfrak{G}$  via normalization and Gröbner bases. Section 7 pays particular attention to the planar case, commenting on the direct treatment of special types of singularities. In Sects. 8 and 9, we discuss the modular version of our algorithm. Finally, in Sect. 10, we compare the performance of the different approaches, relying on our implementations in the computer algebra system SINGULAR [21], and running various examples coming from algebraic geometry.

## 2 Algorithms for Normalization

We begin with some general remarks on normalization and the role played by the conductor. For these, let  $A$  be any reduced Noetherian ring, and let  $Q(A)$  be its total ring of fractions. Then  $Q(A)$  is again a reduced Noetherian ring. We write

$$\text{Spec}(A) = \{P \subset A \mid P \text{ prime ideal}\}$$

for the *spectrum* of  $A$ . The *vanishing locus* of an ideal  $J$  of  $A$  is the set  $V(J) = \{P \in \text{Spec}(A) \mid P \supset J\}$ .

The *normalization* of  $A$ , written  $\bar{A}$ , is the integral closure of  $A$  in  $Q(A)$ . We call  $A$  *normalization-finite* if  $\bar{A}$  is a finite  $A$ -module, and we call  $A$  *normal* if  $A = \bar{A}$ .

We denote by

$$N(A) = \{P \in \text{Spec}(A) \mid A_P \text{ is not normal}\}$$

the *non-normal locus* of  $A$ , and by

$$\text{Sing}(A) = \{P \in \text{Spec}(A) \mid A_P \text{ is not regular}\}$$

the *singular locus* of  $A$ .

**Remark 2.1** Note that  $N(A) \subset \text{Sing}(A)$ . Equality holds if  $A$  is of pure dimension one. Indeed, a Noetherian local ring of dimension one is normal iff it is regular (see [23, Thm. 4.4.9]).

**Definition 2.2** If  $R \subset S$  is an extension of rings, the *conductor* of  $A$  in  $B$  is

$$\mathcal{C}_{S/R} = \text{Ann}_R(S/R) = \{r \in R \mid rS \subset R\}.$$

Note that  $\mathcal{C}_{S/R}$  is the largest ideal of  $R$  which is also an ideal of  $S$ .

**Remark 2.3** Specializing to the normalization, we write

$$\mathcal{C}_A = \mathcal{C}_{\bar{A}/A} = \text{Ann}_A(\bar{A}/A) = \{a \in A \mid a\bar{A} \subset A\}.$$

Note that  $\mathcal{C}_A$  can be naturally identified with  $\text{Hom}_A(\bar{A}, A)$  (see [57, Lemma 2.4.2]).

**Lemma 2.4** We have  $N(A) \subset V(\mathcal{C}_A)$ . Furthermore,  $A$  is normalization-finite iff  $\mathcal{C}_A$  contains a non-zerodivisor of  $A$ . In this case,  $N(A) = V(\mathcal{C}_A)$ .

*Proof* See [32, Lemmas 3.6.1, 3.6.3].

**Remark 2.5 (Splitting of Normalization)** Finding the normalization can be reduced to the case of integral domains: If  $P_1, \dots, P_s$  are the minimal primes of  $A$ , then

$$\overline{A} \cong \overline{A/P_1} \times \cdots \times \overline{A/P_s}$$

(see [23, Thm. 1.5.20]).

**Remark 2.6** Let  $k$  be a field. An *affine  $k$ -domain* is a finitely generated  $k$ -algebra which is an integral domain. By Emmy Noether's finiteness theorem (see [25, Cor. 13.13]), any such domain is normalization-finite, and its normalization is an affine  $k$ -domain as well. Geometrically, by gluing, this implies that any integral algebraic variety  $X$  over  $k$  admits a (unique) *normalization map*  $\pi : \overline{X} \rightarrow X$ , where  $\pi$  is a finite morphism and, hence, the normal scheme  $\overline{X}$  is an algebraic variety over  $k$  as well (see, for example, [45, Sec. 4.1.2]). Specifically, by Remark 2.1, if  $\Gamma$  is an integral algebraic curve over  $k$ , we get the *nonsingular model*  $\pi : \overline{\Gamma} \rightarrow \Gamma$ .

**Definition 2.7** A homomorphism  $A \rightarrow B$  of reduced Noetherian rings is called *normal* if it is flat and if for every  $P \in \text{Spec}(A)$  and every field extension  $L$  of  $A_P/PA_P$ , the ring  $B \otimes_A L$  is normal.

**Remark 2.8 (Base Change)** Let  $\ell \subset k$  be a separable field extension, and let  $A$  be a finitely generated reduced  $\ell$ -algebra. Then  $A \rightarrow A \otimes_{\ell} k$  is a normal homomorphism, so that  $\overline{A} \otimes_{\ell} k$  is a normal ring (see [57, Propositions 19.1.1, 19.1.2, Thm. 19.4.2]). On the other hand, by Swanson and Huneke [57, Thm. 19.5.1], we may identify  $\overline{A} \otimes_{\ell} k$  with the integral closure of  $A \otimes_{\ell} k$  in  $\mathbb{Q}(A) \otimes_{\ell} k$ . In turn, since every non-zero-divisor of  $A$  is a non-zero-divisor of  $A \otimes_{\ell} k$ , we may regard  $\mathbb{Q}(A) \otimes_{\ell} k$  as a subring of  $\mathbb{Q}(A \otimes_{\ell} k)$ , and thus  $\overline{A} \otimes_{\ell} k$  as a subring of  $\overline{A \otimes_{\ell} k}$ . Since  $\overline{A} \otimes_{\ell} k$  is already normal, we conclude that  $\overline{A \otimes_{\ell} k} = \overline{A} \otimes_{\ell} k$ . In particular,

$$\begin{aligned} \mathcal{C}_{A \otimes_{\ell} k} &\cong \text{Hom}_{A \otimes_{\ell} k}(\overline{A \otimes_{\ell} k}, A \otimes_{\ell} k) = \text{Hom}_{A \otimes_{\ell} k}(\overline{A} \otimes_{\ell} k, A \otimes_{\ell} k) \\ &= \text{Hom}_{A \otimes_{\ell} k}(\overline{A} \otimes_A (A \otimes_{\ell} k), A \otimes_A (A \otimes_{\ell} k)) \cong \text{Hom}_A(\overline{A}, A) \otimes_{\ell} k \\ &\cong \mathcal{C}_A \otimes_{\ell} k \end{aligned}$$

(see [25, Prop. 2.10] for the second to last identity).

Now, we briefly discuss algorithmic normalization. We begin by recalling the normalization algorithm of Greuel et al. [34], which is an improvement of de Jong's algorithm (see [20, 22]). This algorithm, to which we refer as the GLS Algorithm, is based on the normality criterion of Grauert and Remmert. To state this criterion, we need:

**Lemma 2.9** *Let  $A$  be a reduced Noetherian ring, and let  $J \subset A$  be an ideal which contains a non-zero-divisor  $g$  of  $A$ . Then:*

1. *If  $\varphi \in \text{Hom}_A(J, J)$ , the fraction  $\varphi(g)/g \in \bar{A}$  is independent of the choice of  $g$ , and  $\varphi$  is multiplication by  $\varphi(g)/g$ .*
2. *There are natural inclusions of rings*

$$A \subset \text{Hom}_A(J, J) \cong \frac{1}{g}(gJ :_A J) \subset \bar{A} \subset Q(A), \quad a \mapsto \varphi_a, \quad \varphi \mapsto \frac{\varphi(g)}{g},$$

where  $\varphi_a$  is multiplication by  $a$ .

*Proof* See [32, Lemmas 3.6.1, 3.6.3].

**Proposition 2.10 (Grauert and Remmert Criterion)** *Let  $A$  be a reduced Noetherian ring, and let  $J \subset A$  be a radical ideal which contains a non-zero-divisor  $g$  of  $A$  and satisfies  $V(\mathcal{C}_A) \subset V(J)$ . Then  $A$  is normal iff  $A \cong \text{Hom}_A(J, J)$  via the map which sends  $a$  to multiplication by  $a$ .*

*Proof* See [28], [32, Prop. 3.6.5].

**Definition 2.11** A pair  $(J, g)$  as in the proposition is called a *test pair* for  $A$ , and  $J$  is called a *test ideal* for  $A$ .

If  $k$  is a field and  $A$  is an affine  $k$ -domain, then test pairs exist by Lemma 2.4 and Emmy Noether’s finiteness theorem. If, in addition,  $k$  is perfect, an explicit test pair can be found by applying the Jacobian criterion (see [25, Thm. 16.19] for this criterion). In fact, in this case, we may choose the radical of the Jacobian ideal  $M$  together with any non-zero element  $g$  of  $M$  as a test pair. Given a test pair  $(J, g)$ , the basic idea of finding  $\bar{A}$  is to enlarge  $A$  by a sequence of finite extensions of affine  $k$ -domains

$$A_{i+1} = \text{Hom}_{A_i}(J_i, J_i) \cong \frac{1}{g}(gJ_i :_{A_i} J_i) \subset \bar{A} \subset Q(A),$$

with  $A_0 = A$  and test ideals  $J_i = \sqrt{JA_i}$ , until the Grauert and Remmert criterion allows one to stop. According to [34], each  $A_i$  can be represented as a quotient  $\frac{1}{d_i}U_i \subset Q(A)$ , where  $U_i \subset A$  is an ideal and  $d_i \in U_i$  is non-zero. In this way, all computations except those of the radicals  $J_i$  may be carried through in  $A$ .

*Example 2.12* For

$$A = \mathbb{C}[x, y] = \mathbb{C}[X, Y]/\langle X^5 - Y^2(Y - 1)^3 \rangle,$$

the radical of the Jacobian ideal is

$$J := \langle x, y(y - 1) \rangle_A,$$



so that we can take  $(J, x)$  as a test pair. Then, in its first step, the normalization algorithm yields

$$A_1 = \frac{1}{x}U_1 = \frac{1}{x}\langle x, y(y-1)^2 \rangle_A.$$

In the next steps, we get

$$A_2 = \frac{1}{x^2}U_2 = \frac{1}{x^2}\langle x^2, xy(y-1), y(y-1)^2 \rangle_A$$

and

$$A_3 = \frac{1}{x^3}U_3 = \frac{1}{x^3}\langle x^3, x^2y(y-1), xy(y-1)^2, y^2(y-1)^2 \rangle_A.$$

In the final step, we find that  $A_3$  is normal and, hence, equal to  $\bar{A}$ .

Next, we describe the local to global variant of the GLS algorithm given in [6]. This variant is a considerable enhancement of the algorithm which serves as a motivation for our local to global approach to compute the Gorenstein adjoint ideal. It is based on the following two observations from [6]: First, the normalization  $\bar{A}$  can be computed as the sum of local contributions  $A \subset A^{(i)} \subset \bar{A}$ , and second, local contributions can be obtained efficiently by a local variant of the GLS algorithm. For our purposes, it is enough to present the relevant results in a special case. Here, as usual, if  $P$  is a prime of a ring  $R$ , and  $M$  is an  $R$ -module, we write  $M_P$  for the localization of  $M$  at  $R \setminus P$ .

**Proposition 2.13** *Let  $A$  be an affine domain of dimension one over a field  $k$ , and let  $\text{Sing}(A) = \{P_1, \dots, P_s\}$  be its singular locus. For  $i = 1, \dots, s$ , let an intermediate ring  $A \subset A^{(i)} \subset \bar{A}$  be given such that  $A_{P_i}^{(i)} = \bar{A}_{P_i}$ . Then*

$$\sum_{i=1}^s A^{(i)} = \bar{A}.$$

*Proof* See [6, Prop. 15].

**Definition 2.14** A ring  $A^{(i)}$  as above is called a *local contribution* to  $\bar{A}$  at  $P_i$ . It is called a *minimal local contribution* if  $A_{P_j}^{(i)} = A_{P_j}$  for  $j \neq i$ .

The computation of local contributions is based on the modified version of the Grauert and Remmert criterion below:

**Proposition 2.15** *Let  $A$  be an affine domain of dimension one over a field  $k$ , let  $A \subset A'$  be a finite ring extension, let  $P \in \text{Sing}(A)$ , and let  $J' = \sqrt{PA'}$ . If*

$$A' \cong \text{Hom}_{A'}(J', J')$$

*via the map which sends  $a'$  to multiplication by  $a'$ , then  $A'_P$  is normal.*

*Proof* See [6, Prop. 16].

Considering an affine domain  $A$  of dimension one over a perfect field  $k$ , let  $P \in \text{Sing}(A)$ . Choose  $P$  together with a non-zero element  $g \in P$  instead of a test pair as in Definition 2.11. Then, proceeding as before, we get a chain of affine  $k$ -domains

$$A \subset A_1 \subset \cdots \subset A_m \subset \bar{A}$$

such that  $A_m$  is a local contribution to  $\bar{A}$  at  $P$ .

*Remark 2.16* Given  $A$  as above, a finite ring extension  $A \subset A'$ , and a prime  $P \in \text{Sing}(A)$ , let  $Q \in \text{Sing}(A)$  be a prime different from  $P$ , and let  $J' = \sqrt{PA'}$ . Then

$$\begin{aligned} \text{Hom}_{A'}(J', J')_Q &\cong \text{Hom}_{A'_Q}(J'_Q, J'_Q) \\ &\cong \text{Hom}_{A'_Q}(A'_Q, A'_Q) \cong A'_Q \end{aligned}$$

(see [25, Proposition 2.10] for the first identity). Inductively, this shows that the algorithm outlined above computes a minimal local contribution to  $\bar{A}$  at  $P$ . Note that such a contribution is uniquely determined since, by definition, its localization at each  $Q \in \text{Spec}(A)$  is determined.

*Example 2.17* In Example 2.12, there are two singularities, namely  $P_1 = \langle x, y \rangle$  and  $P_2 = \langle x, y - 1 \rangle$ . Geometrically, these are a singularity of type  $A_4$  at  $(0, 0)$  and a threefold point of type  $E_8$  at  $(0, 1)$ . For  $P_1$ , the local normalization algorithm yields  $\bar{A}_{P_1} = (\frac{1}{d_1}U_1)_{P_1}$ , where

$$d_1 = x^2 \quad \text{and} \quad U_1 = \langle x^2, y(y-1)^3 \rangle_A.$$

For  $P_2$ , we get  $\bar{A}_{P_2} = (\frac{1}{d_2}U_2)_{P_2}$ , where

$$d_2 = x^3 \quad \text{and} \quad U_2 = \langle x^3, x^2y^2(y-1), y^2(y-1)^2 \rangle_A.$$

Combining the local contributions, we get

$$\frac{1}{d}U = \frac{1}{d_1}U_1 + \frac{1}{d_2}U_2,$$

with  $d = x^3$  and

$$U = \langle x^3, xy(y-1)^3, x^2y^2(y-1), y^2(y-1)^2 \rangle_A.$$

Note that  $U$  coincides with the ideal  $U_3$  computed in Example 2.12.

In the following sections we will use the notation below:

**Notation 2.18** Given an affine algebraic curve  $C \subset \mathbb{A}_k^r$  over a field  $k$  with vanishing ideal  $I(C)$  and a point<sup>1</sup>  $P \in C$ , if  $I \subset k[X_1, \dots, X_r]$  is an ideal properly containing  $I(C)$ , we will write  $I_P = I_{\mathcal{O}_{C,P}}$  for the *local ideal* of  $I$  at  $P$ . Similarly for a projective algebraic curve  $\Gamma \subset \mathbb{P}_k^r$  and a homogeneous ideal  $I \subset k[X_0, \dots, X_r]$ .

### 3 Adjoint Ideals

Let  $k$  be a field, and let  $\Gamma \subset \mathbb{P}_k^r$  be an integral non-degenerate projective algebraic curve. Write  $S = k[X_0, \dots, X_r]$  for the homogeneous coordinate ring of  $\mathbb{P}_k^r$ ,  $I(\Gamma) \subset S$  for the homogeneous vanishing ideal of  $\Gamma$ ,  $k[\Gamma] = S/I(\Gamma)$  for the homogeneous coordinate ring of  $\Gamma$ , and  $\text{Sing}(\Gamma)$  for the singular locus of  $\Gamma$ .

Let  $\pi : \overline{\Gamma} \rightarrow \Gamma$  be the normalization map, let  $P$  be a point of  $\Gamma$ , and let  $\mathcal{O}_{\Gamma,P}$  be the local ring of  $\Gamma$  at  $P$ . Then the normalization  $\overline{\mathcal{O}_{\Gamma,P}}$  is a semi-local ring whose maximal ideals correspond to the points of  $\overline{\Gamma}$  lying over  $P$ . Furthermore,  $\overline{\mathcal{O}_{\Gamma,P}}$  is finite over  $\mathcal{O}_{\Gamma,P}$ , so that  $\overline{\mathcal{O}_{\Gamma,P}}/\mathcal{O}_{\Gamma,P}$  is a finite-dimensional  $k$ -vector space. The dimension

$$\delta_P(\Gamma) = \delta(\mathcal{O}_{\Gamma,P}) = \dim_k \overline{\mathcal{O}_{\Gamma,P}}/\mathcal{O}_{\Gamma,P}$$

is called the *delta invariant* of  $\Gamma$  at  $P$ . The *arithmetic genus* of  $\Gamma$  is  $p_a(\Gamma) = 1 - P_\Gamma(0)$ , where  $P_\Gamma$  is the Hilbert polynomial of  $k[\Gamma]$ . Making use of the (global) *delta invariant*

$$\delta(\Gamma) = \sum_{P \in \text{Sing}(\Gamma)} \delta_P(\Gamma)$$

of  $\Gamma$ , the *geometric genus*  $p(\Gamma)$  of  $\Gamma$  is given by

$$p(\Gamma) = p(\overline{\Gamma}) = p_a(\Gamma) - \delta(\Gamma)$$

(see [39]). If  $\Gamma$  is a plane curve of degree  $n$ , we have  $p_a(\Gamma) = \binom{n-1}{2}$ .

Following the presentation in [18], we now recall the definition and characterization of adjoint ideals due to [1] and [19]. Let  $I = \bigoplus_{m \geq 0} I_m \subset S = k[X_0, \dots, X_r]$  be a saturated homogeneous ideal properly containing  $I(\overline{\Gamma})$ . Pulling back  $\text{Proj}(S/I)$  via  $\pi$ , we get an effective divisor  $\Delta(I)$  on  $\overline{\Gamma}$ . Let  $H$  be a divisor on  $\overline{\Gamma}$  given as the pullback of a hyperplane in  $\mathbb{P}_k^r$ . Then, since any divisor on  $\overline{\Gamma}$  cut out by a homogeneous polynomial in  $I$  is of the form  $D + \Delta(I)$  for some effective divisor  $D$ ,

---

<sup>1</sup>The term *point* will always refer to a closed point.

we have natural linear maps

$$\varrho_m : I_m \rightarrow H^0(\overline{\Gamma}, \mathcal{O}_{\overline{\Gamma}}(mH - \Delta(I))),$$

for all  $m \geq 0$ .

*Remark 3.1* Consider the exact sequence

$$0 \rightarrow \tilde{I}\mathcal{O}_{\Gamma} \rightarrow \pi_*(\tilde{I}\mathcal{O}_{\overline{\Gamma}}) \rightarrow \mathcal{F} \rightarrow 0,$$

where  $\tilde{I}$  is the ideal sheaf on  $\mathbb{P}_k^r$  associated to  $I$ , and  $\mathcal{F}$  is the cokernel. Twisting by  $m \gg 0$  and taking global sections, we get the exact sequence

$$0 \rightarrow H^0(\Gamma, \tilde{I}\mathcal{O}_{\Gamma}(m)) \rightarrow H^0(\overline{\Gamma}, \tilde{I}\mathcal{O}_{\overline{\Gamma}}(mH)) \rightarrow H^0(\Gamma, \mathcal{F}) \rightarrow 0.$$

Indeed,  $\mathcal{F}$  has finite support and, since the normalization map  $\pi$  is finite, we have  $H^0(\overline{\Gamma}, \tilde{I}\mathcal{O}_{\overline{\Gamma}}(mH)) \cong H^0(\Gamma, \pi_*(\tilde{I}\mathcal{O}_{\overline{\Gamma}})(m))$ . Since  $\tilde{I}\mathcal{O}_{\overline{\Gamma}}(mH) = \mathcal{O}_{\overline{\Gamma}}(mH - \Delta(I))$  and, for  $m \gg 0$ ,  $H^0(\Gamma, \tilde{I}\mathcal{O}_{\Gamma}(m)) = I_m/I(\Gamma)_m$ , we get, for  $m \gg 0$ , the exact sequence

$$0 \rightarrow I_m/I(\Gamma)_m \xrightarrow{\overline{\varrho}_m} H^0(\overline{\Gamma}, \mathcal{O}_{\overline{\Gamma}}(mH - \Delta(I))) \rightarrow H^0(\Gamma, \mathcal{F}) \rightarrow 0.$$

In particular, for  $m \gg 0$ ,

$$\ker(\varrho_m) = I(\Gamma)_m.$$

**Definition 3.2** With notation and assumptions as above, the ideal  $I$  is called an *adjoint ideal* of  $\Gamma$  if the maps

$$\varrho_m : I_m \rightarrow H^0(\overline{\Gamma}, \mathcal{O}_{\overline{\Gamma}}(mH - \Delta(I)))$$

are surjective for  $m \gg 0$ .

As already remarked in the introduction, the existence of adjoint ideals is classical. Locally, adjoint ideals are characterized by the following criterion:

**Theorem 3.3** *The ideal  $I$  is an adjoint ideal of  $\Gamma$  iff  $I_P = \overline{I_P\mathcal{O}_{\Gamma,P}}$  for all  $P \in \text{Sing}(\Gamma)$ .*

*Proof* Using the notation from Remark 3.1, we have, for  $m \gg 0$ ,

$$\dim_k \text{coker } \varrho_m = h^0(\Gamma, \mathcal{F}) = \sum_{P \in \text{Sing}(\Gamma)} \ell(I_P\overline{\mathcal{O}_{\Gamma,P}}/I_P).$$

Hence,  $\varrho_m$  is surjective iff  $I_P\overline{\mathcal{O}_{\Gamma,P}} = I_P$  for all  $P \in \text{Sing}(\Gamma)$ .

**Corollary 3.4** *If  $I$  is an adjoint ideal of  $\Gamma$  and  $P \in \text{Sing}(\Gamma)$ , then  $I_P \subsetneq \overline{I_P\mathcal{O}_{\Gamma,P}}$ .*

*Proof* Suppose  $I_P = \mathcal{O}_{\Gamma, P}$ . Then  $I_P \subsetneq \overline{I_P \mathcal{O}_{\Gamma, P}}$ , a contradiction to Theorem 3.3.

**Corollary 3.5** *The support of  $\text{Proj}(S/I)$  contains  $\text{Sing}(\Gamma)$ .*

*Proof* Follows immediately from Corollary 3.4.

**Theorem 3.6** *There is a unique largest homogeneous ideal  $\mathfrak{G} \subset S$  which satisfies*

$$\mathfrak{G}_P = \mathcal{C}_{\mathcal{O}_{\Gamma, P}} \text{ for all } P \in \text{Sing}(\Gamma).$$

*The ideal  $\mathfrak{G}$  is an adjoint ideal of  $\Gamma$  containing all other adjoint ideals of  $\Gamma$ . In particular,  $\mathfrak{G}$  is saturated and  $\text{Proj}(S/\mathfrak{G})$  is supported on  $\text{Sing}(\Gamma)$ .*

*Proof* For the conductor ideal sheaf  $\mathcal{C} = \text{Ann}_{\mathcal{O}_\Gamma}(\pi_* \mathcal{O}_{\overline{\Gamma}} / \mathcal{O}_\Gamma)$  on  $\Gamma$ , we have  $\mathcal{C}_P = \mathcal{C}_{\mathcal{O}_{\Gamma, P}}$  for all  $P \in \Gamma$ . If  $j : \Gamma \rightarrow \mathbb{P}_k^r$  is the inclusion, then the graded  $S$ -module  $\mathfrak{G} = \bigoplus_{n \in \mathbb{Z}} H^0(\mathbb{P}_k^r, j_* \mathcal{C}(n))$  associated to  $j_* \mathcal{C}$  is the unique largest homogeneous ideal with  $\mathfrak{G}_P = \mathcal{C}_{\mathcal{O}_{\Gamma, P}}$  for all  $P \in \text{Sing}(\Gamma)$ . By Theorem 3.3 and the properties of the conductor,  $\mathfrak{G}$  is an adjoint ideal. Moreover, if  $I$  is any other adjoint ideal, then  $I_P \subset \mathfrak{G}_P$  for all  $P \in \Gamma$ , hence  $I \subset \mathfrak{G}$ .

**Definition 3.7** With notation as in Theorem 3.6, the ideal  $\mathfrak{G}$  is called the *Gorenstein adjoint ideal* of  $\Gamma$ . We also write  $\mathfrak{G}(\Gamma) = \mathfrak{G}$ .

For repeated subsequent use, we introduce the following notation:

**Notation 3.8** Given an integral non-degenerate projective algebraic curve  $\Gamma \subset \mathbb{P}_k^r$ , let  $C$  be the affine part of  $\Gamma$  with respect to the chart

$$\mathbb{A}_k^r \hookrightarrow \mathbb{P}_k^r, (X_1, \dots, X_r) \mapsto (1 : X_1 : \dots : X_r).$$

Let  $I(C) \subset k[X_1, \dots, X_r]$  be the vanishing ideal of  $C$ , let

$$k[C] = k[x_1, \dots, x_r] = k[X_1, \dots, X_r]/I(C)$$

be its coordinate ring, and let  $\text{Sing}(C)$  be the set of singular points of  $C$ .

**Proposition 3.9** *Let  $C$  be the affine part of  $\Gamma$  in the chart  $X_0 \neq 0$  as in Notation 3.8, and let  $\overline{\mathfrak{G}}$  be the ideal of  $k[C]$  obtained by dehomogenizing  $\mathfrak{G}$  with respect to  $X_0$  and mapping the result to  $k[C]$ . Then*

$$\overline{\mathfrak{G}} = \mathcal{C}_{k[C]}.$$

*If  $\Gamma$  has no singularities at infinity and  $\mathcal{C}_{k[C]} = \langle g_i(x_1, \dots, x_r) \mid i = 1, \dots, m \rangle_{k[C]}$ , with polynomials  $g_i \in k[X_1, \dots, X_r]$ , then  $\mathfrak{G}$  is the homogenization of the ideal*

$$\langle g_i(X_1, \dots, X_r) \mid i = 1, \dots, m \rangle_{k[X_1, \dots, X_r]} + I(C)$$

*with respect to  $X_0$ .*

*Proof* The first assertion is obtained by localizing at the points of  $C$ :

$$\overline{\mathfrak{G}}_P = \mathcal{C}_{\mathfrak{G}_C, P} = (\mathcal{C}_{k[C]})_P \text{ for each } P \in C.$$

Here, the first equality is clear from the definition of  $\mathfrak{G}$  (see Theorem 3.6). The second equality holds since forming the conductor commutes with localization since  $k[C]$  is normalization-finite (see [61, Ch. V, § 5]).

The second assertion follows from the first one since there are no singularities at infinity,  $\mathfrak{G}$  is saturated, and the support of  $\mathfrak{G}$  is contained in  $C$ .

*Remark 3.10 (Base Change)* Suppose that  $\Gamma$  is defined over a subfield  $\ell$  of  $k$  such that  $\ell \subset k$  is separable, and let  $\Gamma(\ell) \subset \mathbb{P}_\ell^r$  be the set of  $\ell$ -rational points of  $\Gamma$ . Then it follows from Remark 2.8 and Proposition 3.9 that

$$\delta(\Gamma(\ell)) = \delta(\Gamma) \text{ and } \mathfrak{G}(\Gamma(\ell))k[X_0, \dots, X_n] = \mathfrak{G}(\Gamma).$$

We now take a moment to specialize to plane curves.

*Remark 3.11* Assume  $\Gamma$  is a plane curve. Then, by Max Noether's Fundamental-satz, the maps  $\varrho_m : \mathfrak{G}_m \rightarrow H^0(\overline{\Gamma}, \mathcal{O}_{\overline{\Gamma}}(mH - \Delta(\mathfrak{G})))$  are surjective for all  $m$ . Referring to each homogeneous polynomial in  $\mathfrak{G}$  not contained in  $I(\Gamma)$  as an *adjoint curve* to  $\Gamma$ , this means that residual to  $\Delta(\mathfrak{G})$ , the adjoint curves of any degree  $m$  cut out the complete linear series  $\mathcal{A}_m = |mH - \Delta(\mathfrak{G})|$ . See [59, § 49].

**Theorem 3.12** *Assume  $\Gamma$  is a plane curve of degree  $n$ . Then, residual to  $\Delta(\mathfrak{G})$ , the elements of  $\mathfrak{G}_{n-3}$  cut out the complete canonical linear series. Equivalently,*

$$\deg \Delta(\mathfrak{G}) = 2\delta(\Gamma). \tag{1}$$

*Proof* See [27, Thm. 9].

Recall that the dimension of the canonical linear series is  $\dim \mathcal{A}_{n-3} = p(\Gamma) - 1$ .

*Remark 3.13* Assume  $\Gamma$  is a plane curve of degree  $n$ . If  $p(\Gamma) = 0$ , that is,  $\Gamma$  is rational, then  $\dim \mathcal{A}_{n-2} = \deg \mathcal{A}_{n-2} = n - 2$ , and the image of  $\Gamma$  under  $\mathcal{A}_{n-2}$  is a rational normal curve  $\Gamma_{n-2} \subset \mathbb{P}_k^{n-2}$  of degree  $n - 2$ . Via the birational morphism  $\Gamma_{n-2} \rightarrow \Gamma$ , the problem of parametrizing  $\Gamma$  is reduced to parametrizing the smooth curve  $\Gamma_{n-2}$ . For the latter, we may successively decrease the degree of the rational normal curve by 2 via the anti-canonical linear series. This yields an isomorphism from  $\Gamma_{n-2}$  either to  $\mathbb{P}^1$  or to a plane conic, depending on whether  $n$  is odd or even.

We will now return to the general case and discuss a version of Eq. (1) which is also valid if  $\Gamma$  is not necessarily planar. In fact, this equation characterizes adjoint ideals. We use the following notation: If  $I \subset S$  is a homogeneous ideal, write  $\deg I = \deg \text{Proj}(S/I)$ . That is,  $\deg I$  is  $(\dim I - 1)!$  times the leading coefficient of the Hilbert polynomial of  $S/I$ .

**Lemma 3.14** *Let  $I \subset S$  be a saturated homogeneous ideal with  $I(\Gamma) \subsetneq I$ . Then*

$$\deg \Delta(I) \leq \deg I + \delta(\Gamma),$$

*and  $I$  is an adjoint ideal of  $\Gamma$  iff*

$$\deg \Delta(I) = \deg I + \delta(\Gamma).$$

*Proof* Let  $P_\Gamma(t) = (\deg \Gamma) \cdot t - p_a(\Gamma) + 1$  be the Hilbert polynomial of  $k[\Gamma]$ . Denote by  $I_\Gamma$  the image of  $I$  in  $k[\Gamma]$ . Then, for  $m \gg 0$ ,

$$\deg I = \dim_k(S_m/I_m) = \dim_k(k[\Gamma]_m/(I_\Gamma)_m) = P_\Gamma(m) - \dim_k(I_\Gamma)_m.$$

Moreover, by Remark 3.1 and with notation as in that remark,

$$h^0(\overline{\Gamma}, \mathcal{O}_{\overline{\Gamma}}(mH - \Delta(I))) = \dim_k(I_\Gamma)_m + h^0(\Gamma, \mathcal{F}) \geq \dim_k(I_\Gamma)_m$$

for  $m \gg 0$ . Hence, by Riemann-Roch, we have

$$\begin{aligned} (\deg \Gamma) \cdot m - \deg \Delta(I) &= \deg |mH - \Delta(I)| = \dim |mH - \Delta(I)| + p(\Gamma) \\ &\geq \dim_k(I_\Gamma)_m - 1 + p(\Gamma) \\ &= P_\Gamma(m) - \deg I - 1 + p(\Gamma) \\ &= (\deg \Gamma) \cdot m - \delta(\Gamma) - \deg I \end{aligned}$$

for  $m \gg 0$  since  $|mH - \Delta(I)|$  is non-special for large  $m$  by reason of its degree. For such  $m$ , equality holds iff  $\varrho_m$  is surjective.

*Remark 3.15* In the case where  $\Gamma$  is a plane curve and  $I = \mathfrak{G}$  is its Gorenstein adjoint ideal, Lemma 3.14 shows that Eq. (1) may be rewritten as

$$\deg \mathfrak{G} = \delta(\Gamma). \quad (2)$$

Note that (1) and (2) may not hold in the non-planar case:

*Example 3.16* ([23, Example 5.2.5]) Let  $\Gamma \subset \mathbb{P}_{\mathbb{C}}^3$  be the image of the parametrization

$$\mathbb{P}_{\mathbb{C}}^1 \longrightarrow \mathbb{P}_{\mathbb{C}}^3, (s : t) \mapsto (s^5 : t^3 s^2 : t^4 s : t^5).$$

Then  $\Gamma$  has exactly one singularity at  $(1 : 0 : 0 : 0)$ . Furthermore,  $p(\Gamma) = 0$  and  $p_a(\Gamma) = 2$ , hence  $\delta(\Gamma) = 2$ . However,  $\mathfrak{G} = \langle X_1, X_2, X_3 \rangle \subset \mathbb{C}[X_0, \dots, X_3]$ , hence  $\deg \mathfrak{G} = 1$ .

*Remark 3.17* If  $\Gamma \subset \mathbb{P}_k^r$  is any curve as in Notation 3.8, with affine part  $C$  and no singularities at infinity, then it follows from Proposition 3.9 that

$$\deg \mathfrak{G} = \dim_k (k[C]/\mathcal{C}_{k[C]}) = \sum_{P \in \text{Sing}(C)} \dim_k (\mathcal{O}_{C,P}/\mathcal{C}_{\mathcal{O}_{C,P}}).$$

**Lemma 3.18** *If  $\text{char } k = 0$ , then  $\dim_k(\mathcal{O}_{\Gamma,P}/\mathcal{C}_{\mathcal{O}_{\Gamma,P}}) \leq \delta_P(\Gamma)$  for any point  $P \in \Gamma$ .*

*Proof* This follows from the case  $k = \mathbb{C}$  proved in [31, 2.4] by base change (see Remark 2.8).

Now recall that a point  $P \in \text{Sing}(\Gamma)$  is called a *Gorenstein singularity* if

$$\dim_k(\mathcal{O}_{\Gamma,P}/\mathcal{C}_{\mathcal{O}_{\Gamma,P}}) = \delta_P(\Gamma).$$

*Example 3.19* Plane curve singularities are Gorenstein (see, for example, [23, Corollary 5.2.9]).

**Corollary 3.20** *We have:*

1. *If  $\text{char } k = 0$ , then  $\deg \mathfrak{G} \leq \delta(\Gamma)$ .*
2. *If  $\Gamma$  has only Gorenstein singularities, then*

$$\deg \mathfrak{G} = \delta(\Gamma) \text{ and } \deg \Delta(\mathfrak{G}) = 2\delta(\Gamma).$$

*Proof* This is clear from the discussion above.

We now begin with the discussion of how to compute the Gorenstein adjoint ideal. One possible way of finding  $\mathfrak{G}$  is to apply the global algorithms presented in Sect. 4.2 below, starting from the normalization  $\overline{k[C]}$ , and relying on Proposition 3.9. To compute  $\overline{k[C]}$ , in turn, we may use the local to global approach outlined in Sect. 2. As we will see, however, it is more efficient to directly proceed with a local to global approach for finding  $\mathfrak{G}$ , computing local Gorenstein adjoint ideals at the singular points, and obtaining  $\mathfrak{G}$  as their intersection. This will be the theme of Sects. 5 and 6, while in Sect. 7, focusing on the case of plane curves, we will write down explicit generators for the local Gorenstein adjoint ideals at various types of singularities.

*Remark 3.21* With regard to implementing Proposition 3.9 as a part of the global algorithms, we note that if  $k$  is infinite, then the assumption on the singularities can always be achieved by a projective change of coordinates defined over  $k$ . If  $k$  is finite, however, we may have to replace  $k$  by an algebraic extension field of  $k$ . Our local to global algorithm, on the other hand, does not require a coordinate change. If  $\Gamma$  is defined over a subfield  $\ell$  of  $k$  such that  $\ell \subset k$  is separable, then it follows from Remark 3.10 that we may find the Gorenstein ideal by computations over  $\ell$ .

If  $\Gamma$  is defined over  $\mathbb{Q}$ , we will use the equality

$$\deg \mathfrak{G} = \deg \Delta(\mathfrak{G}) - \delta(\Gamma)$$



from Lemma 3.14 to compute  $\deg \mathfrak{G}$  without actually knowing  $\mathfrak{G}$ , and apply this in the final verification step of our modularized adjoint ideal algorithm (see Sects. 8 and 9). In fact, we will present a modular approach to computing  $\deg \Delta(\mathfrak{G})$ , and we will use standard techniques to compute  $\delta(\Gamma)$ . For the latter, first note that the delta invariant of  $\Gamma$  differs from that of a plane model of  $\Gamma$  by the quantity  $p_a(\Gamma) - \binom{\deg \Gamma - 1}{2}$ . The delta invariant of a plane curve, in turn, can be computed locally at the singular points, either from the semigroups of values of the analytic branches of the singularity (see [23, 33]), or from a formula relating the local delta invariant to the Milnor number (see Remark 7.3 in Sect. 7).

*Remark 3.22* Let  $\Gamma \subset \mathbb{P}_k^r$  be a curve with affine part  $C$  as in Notation 3.8 and no singularities at infinity. Then computing  $\deg \Delta(\mathfrak{G})$  also means to compute the dimension  $\dim_k \overline{k[C]}/\mathcal{C}_{k[C]}$ :

$$\begin{aligned} \deg \Delta(\mathfrak{G}) &= \delta(\Gamma) + \deg \mathfrak{G} \\ &= \dim_k \overline{k[C]}/k[C] + \dim_k (k[C]/\mathcal{C}_{k[C]}) \\ &= \dim_k \overline{k[C]}/\mathcal{C}_{k[C]}. \end{aligned}$$

## 4 Global Approaches

### 4.1 Computing the Conductor via the Trace Matrix

We will require some facts from classical ideal theory (see [61, Ch. V] for details and proofs): Let  $R$  be an integral domain, and let  $K = \mathbb{Q}(R)$  be its quotient field. A *fractionary ideal* of  $R$  is an  $R$ -submodule  $\mathfrak{b}$  of  $K$  admitting a common denominator: there is an element  $0 \neq d \in R$  such that  $d\mathfrak{b} \subset R$ .

*Example 4.1* The extensions  $A_i$  computed by the normalization algorithms from Sect. 2 are fractionary ideals of the given affine domain  $A$ .

If  $\mathfrak{b}, \mathfrak{b}'$  are two fractionary ideals of  $R$ , with  $\mathfrak{b}'$  non-zero, then  $\mathfrak{b} : \mathfrak{b}' = \{z \in K \mid z\mathfrak{b}' \subset \mathfrak{b}\}$  is a fractionary ideal of  $R$  as well. A fractionary ideal  $\mathfrak{b}$  of  $R$  is *invertible* if there is a fractionary ideal  $\mathfrak{b}'$  of  $R$  such that  $\mathfrak{b} \cdot \mathfrak{b}' = R$ . In this case,  $\mathfrak{b}'$  is uniquely determined and equal to  $R : \mathfrak{b}$ .

Suppose in addition that  $R$  is normal. Let  $K'$  be a finite separable extension of  $K$ , and let  $R'$  be an integral extension of  $R$  such that  $K' = \mathbb{Q}(R')$ . Moreover, let

$$\mathrm{Tr}_{K'/K} : K' \rightarrow K, z \mapsto \sum_{g \in \mathrm{Gal}(K'/K)} g(z),$$

be the corresponding *trace map*. Then the *complementary module*

$$\mathfrak{C}_{R'/R} := \{z \in K' \mid \mathrm{Tr}_{K'/K}(zR') \subset R\}$$

of  $R'$  with respect to  $R$  is a *fractionary ideal* of  $R'$  containing  $R'$ . Hence, the *different*

$$\begin{aligned} \mathfrak{D}_{R'/R} &= R' : \mathfrak{C}_{R'/R} = \{z \in K' \mid z\mathfrak{C}_{R'/R} \subset R'\} \\ &= \{z \in K' \mid zX \in R' \text{ for all } x \in K' \text{ with } \text{Tr}_{K'/K}(xR') \subset R\} \end{aligned}$$

of  $R'$  over  $R$  is a non-zero ideal of  $R'$ .

Now, keeping our assumptions, we focus on the case where  $R$  is a Dedekind domain, and where  $R'$  is the integral closure of  $R$  in  $K'$ . Then  $R'$  is a Dedekind domain as well, which implies that every non-zero fractionary ideal of  $R'$  is invertible. On the other hand, by the primitive element theorem, there is an element  $y \in R'$  with  $K' = K(y)$ . Denote by  $f(Y) \in K[Y]$  the minimal polynomial of  $y$  over  $K$ . Then, as shown in [61, Ch. V],

$$f'(y)R' = \mathcal{C}_{R'/R[y]}\mathfrak{D}_{R'/R},$$

hence

$$\mathcal{C}_{R'/R[y]} = f'(y)\mathfrak{C}_{R'/R}. \quad (3)$$

We now fix the following setup:

**Notation 4.2** Let  $k$  be a field, and let  $\Gamma \subset \mathbb{P}_k^2$  be a plane curve of degree  $n$  defined by an irreducible polynomial  $F \in k[X, Y, Z]$ . Suppose that the equation  $f \in k[X, Y]$  of the affine part  $C$  of  $\Gamma$  in the chart

$$\mathbb{A}_k^2 \hookrightarrow \mathbb{P}_k^2, (X, Y) \mapsto (1 : X : Y),$$

is monic in  $Y$ .

Write  $k[C] = k[x, y] = k[X, Y]/\langle f(X, Y) \rangle$  for the affine coordinate ring of  $C$  and

$$k(C) = k(x, y) = k(X)[Y]/\langle f(X, Y) \rangle$$

for its function field. Then  $x$  is a separating transcendence basis of  $k(C)$  over  $k$ , and  $y$  is integral over  $k[x]$ , with integral equation  $f(x, y) = 0$ . In particular,  $k[C]$  is integral over  $k[x]$ , which implies that  $\overline{k[C]}$  coincides with the integral closure  $\overline{k[x]}$  of  $k[x]$  in  $k(C)$ . Furthermore,  $\overline{k[C]}$  is a free  $k[x]$ -module of rank

$$n := \deg_y(f) = [k(C) : k(x)].$$

**Definition 4.3** An *integral basis* for  $\overline{k[C]}$  is a set  $b_0, \dots, b_{n-1}$  of free generators for  $\overline{k[C]}$  over  $k[x]$ :

$$\overline{k[C]} = k[x]b_0 \oplus \dots \oplus k[x]b_{n-1}.$$

*Remark 4.4* Since  $k(C) = k(x, y) = k(X)[Y]/\langle f \rangle$ , any element  $\alpha \in k(C)$  can be represented as a polynomial in  $k(X)[Y]$  of degree less than  $n = \deg f$ . Hence, we

may associate to  $\alpha$  a well-defined degree  $\deg_y(\alpha)$  in  $y$  and a smallest common denominator in  $k[x]$  of the coefficients of  $\alpha$ . In particular,  $\overline{k[C]}$  has an integral basis  $(b_i)$  in triangular form, that is, with  $\deg_y(b_i) = i$ , for  $i = 0, \dots, n - 1$  (see [10, Remark 1.4]). If not stated otherwise, all integral bases considered here will be of this form. In principle, such a basis can be found by applying one of the normalization algorithms discussed earlier (see [10, Remark 1.5]). However, in the characteristic zero case, methods relying on Puiseux series techniques are much more efficient (see [10] and [60]). Note that when using these methods, we temporarily may have to pass to an algebraic extension field of  $k$ .

*Example 4.5* An integral basis for the curve considered in Examples 2.12, 2.17 is given below:

$$1, y, \frac{y(y-1)}{x}, \frac{y(y-1)^2}{x^2}, \frac{y^2(y-1)^2}{x^3}.$$

Using Proposition 3.9 and Eq. (3), with  $R = k[x]$ ,  $R' = \overline{k[C]}$ ,  $K = k(x)$ , and  $K' = k(C)$ , we get Algorithm 1.

---

**Algorithm 1** Gorenstein adjoint ideal via linear algebra (see Mruk [50])

---

**Input:** A plane curve  $\Gamma$  over a perfect field  $k$  with affine part  $C$  as in Notation 4.2 and no singularities at infinity.

**Output:** The Gorenstein adjoint ideal  $\mathfrak{G}$  of  $\Gamma$ .

- 1: Compute an integral basis  $(b_i)_{i=0, \dots, n-1}$  for  $\overline{k[C]}$ .
- 2: Compute the (symmetric and invertible) trace matrix

$$T = (\text{Tr}_{k(C)/k(x)}(b_i b_j))_{i,j=0, \dots, n-1} \in k(x)^{n \times n}.$$

- 3: Compute a decomposition  $L \cdot R = P \cdot T$ , where  $L$  is left triangular matrix with diagonal entries equal to one,  $R$  is a right triangular matrix, and  $P$  is a permutation matrix.
- 4: For  $j = 0, \dots, n - 1$ , use forward and backward substitution to compute

$$\eta_j = \sum_{i=0}^{n-1} s_{ij} b_i,$$

where  $(s_{ij}) = T^{-1}$ . The  $\eta_j$  are  $k[x]$ -module generators for  $\mathfrak{C}_{\overline{k[C]}/k[x]}$ . By Eq. (3),  $\mathfrak{C}_{k[C]} = \langle \frac{\partial f}{\partial Y}(x, y) \eta_j \mid j = 0, \dots, n - 1 \rangle$ .

- 5: Let  $\mathcal{C}$  be the ideal of  $k[X, Y]$  generated by representatives of minimal  $y$ -degree of the  $\frac{\partial f}{\partial Y}(x, y) \eta_j, j = 0, \dots, n - 1$ .
  - 6: **return** the homogenization of  $\mathcal{C}$  with respect to  $X_0$ .
-

*Example 4.6* Let  $\Gamma \subset \mathbb{P}_{\mathbb{C}}^2$  be the projective closure of the curve  $C$  with affine equation

$$X^5 - Y^2(1 - Y)^3 = 0$$

as in Examples 2.12, 2.17, and 4.5. From the integral basis

$$1, y, \frac{y(y-1)}{x}, \frac{y(y-1)^2}{x^2}, \frac{y^2(y-1)^2}{x^3}.$$

given in Example 4.5, we compute the trace matrix

$$T = \begin{pmatrix} 5 & 3 & 0 & 0 & 0 \\ 3 & 3 & 0 & 0 & -5x^2 \\ 0 & 0 & 0 & -5x^2 & -3x \\ 0 & 0 & -5x^2 & -3x & 0 \\ 0 & -5x^2 & -3x & 0 & 0 \end{pmatrix},$$

which yields by forward and backward substitution

$$\mathcal{C}_{\mathbb{C}[C]} = \left\langle x^3, x^2(y-1), xy(x-1), y(y-1)^2 \right\rangle_{\mathbb{C}[C]}.$$

Homogenization gives the Gorenstein ideal  $\mathfrak{G}$  which can be decomposed using primary decomposition:

$$\mathfrak{G} = \langle X^2, Y \rangle \cap \langle X^3, X(Y-Z), (Y-Z)^2 \rangle.$$

Note the two ideals on the right hand side correspond to the two singularities of  $C$ . This somewhat motivates the local to global algorithm discussed in Sects. 5 and 6 below, where  $\mathfrak{G}$  will be found as the intersection of local Gorenstein ideals.

## 4.2 Computing the Adjoint Ideal via Ideal Quotients

The algorithm presented in what follows relies on normalization and ideal quotients. It is not limited to plane curves.

**Proposition 4.7** *Let  $\Gamma \subset \mathbb{P}_k^r$  be a curve with affine part  $C$  as in Notation 3.8. Write  $\overline{k[C]} = \frac{1}{d}U$ , where  $U \subset k[C]$  is an ideal and  $d \in U$  is non-zero. Then the conductor is*

$$\mathcal{C}_{k[C]} = \langle d \rangle_{k[C]} : U.$$

**Algorithm 2** Gorenstein adjoint ideal via ideal quotients

**Input:** A curve  $\Gamma \subset \mathbb{P}_k^r$  over a perfect field  $k$  with affine part  $C$  as in Notation 3.8 and no singularities at infinity.

**Output:** The Gorenstein adjoint ideal  $\mathfrak{G}$  of  $\Gamma$ .

- 1: Normalization: Compute polynomials  $d, a_0, \dots, a_s \in k[X_1, \dots, X_r]$  such that the fractions  $\frac{a_i(x_1, \dots, x_r)}{d(x_1, \dots, x_r)}$  generate  $\overline{k[C]}$  as a  $k[C]$ -module.
- 2: Compute the ideal quotient

$$\mathcal{C} = (\langle d \rangle + I(C)) : (\langle a_0, \dots, a_s \rangle + I(C)) \subset k[X_1, \dots, X_r].$$

- 3: **return** the homogenization of  $\mathcal{C}$  with respect to  $X_0$ .

*Proof* By definition,

$$\begin{aligned} \mathcal{C}_{k[C]} &= \left\{ s \in k[C] \mid s \cdot \overline{k[C]} \subset k[C] \right\} \\ &= \left\{ s \in k[C] \mid s \cdot g \in \langle d \rangle_{k[C]} \text{ for all } g \in U \right\} \\ &= \langle d \rangle_{k[C]} : U. \end{aligned}$$

Using once more Proposition 3.9, we get Algorithm 2.

*Example 4.8* In Example 4.6,

$$a_0 = X^3, a_1 = X^2Y(Y-1), a_2 = XY(Y-1)^2, a_3 = Y^2(Y-1)^2,$$

and  $d = X^3$ . Hence,

$$\langle d, f \rangle : \langle a_0, \dots, a_3, f \rangle = \left\langle X^3, X^2(Y-1), XY(Y-1), Y(Y-1)^2 \right\rangle.$$

## 5 A Local to Global Approach

In this section, motivated by the local to global approach for normalization, we introduce local Gorenstein adjoint ideals of a given curve  $\Gamma$  and show how to find the Gorenstein adjoint ideal  $\mathfrak{G}$  of  $\Gamma$  as their intersection. Together with the algorithm presented in the next section, which computes the local ideals, this yields a local to global approach for finding  $\mathfrak{G}$ . As we will see in Sect. 10, this approach is per se faster than the algorithms discussed so far. In addition, it is well-suited for parallel computations.

We consider a curve  $\Gamma \subset \mathbb{P}_k^r$  as in Notation 3.8.

**Definition 5.1** Let  $W \subset \text{Sing}(\Gamma)$  be any set of singular points of  $\Gamma$ . The *local Gorenstein adjoint ideal* of  $\Gamma$  at  $W$  is defined to be the largest homogeneous ideal  $\mathfrak{G}(W) \subset S$  which satisfies

$$\mathfrak{G}(W)_P = \mathcal{C}_{\mathcal{O}_{\Gamma,P}} \text{ for all } P \in W. \quad (4)$$

For a single point  $P \in \text{Sing}(\Gamma)$ , we write  $\mathfrak{G}(P) := \mathfrak{G}(\{P\})$ .

*Remark 5.2* Since  $\mathfrak{G}(W)$  is the largest homogeneous ideal satisfying (4), it is saturated and  $\text{Proj}(S/\mathfrak{G}(W))$  is supported on  $W$ .

**Proposition 5.3** Let  $W \subset \text{Sing}(\Gamma)$ . Then

$$\mathfrak{G}(W) = \bigcap_{P \in W} \mathfrak{G}(P).$$

*Proof* This is immediate from the definition: If  $\mathfrak{G}' := \bigcap_{P \in W} \mathfrak{G}(P)$ , then  $\text{Proj}(S/\mathfrak{G}')$  and  $\text{Proj}(S/\mathfrak{G}(W))$  have the same support  $W$ , and

$$\mathfrak{G}'_Q = \mathfrak{G}(Q)_Q = \mathcal{C}_{\mathcal{O}_{\Gamma,Q}} = \mathfrak{G}(W)_Q$$

for all  $Q \in W$ , hence  $\mathfrak{G}(W) = \mathfrak{G}'$ .

Proposition 5.3 yields Algorithm 3.

*Remark 5.4* It is clear from Proposition 5.3 that we may choose any partition  $\text{Sing}(\Gamma) = \bigcup_{i=1}^s W_i$  of  $\text{Sing}(\Gamma)$  and have

$$\mathfrak{G} = \bigcap_{i=1}^s \mathfrak{G}(W_i).$$

This is useful in that for some subsets  $W_i$ , specialized approaches or a priori knowledge may ease the computation of  $\mathfrak{G}(W_i)$ . In Sect. 7, focusing on plane curves, we will present some ideas in this direction.

---

**Algorithm 3** Gorenstein adjoint ideal, local to global

---

**Input:** A curve  $\Gamma \subset \mathbb{P}_k^r$  over a perfect field  $k$  as in Notation 3.8.

**Output:** The Gorenstein adjoint ideal  $\mathfrak{G}$  of  $\Gamma$ .

- 1: Compute  $\text{Sing}(\Gamma) = \{P_1, \dots, P_s\}$ .
  - 2: Apply Algorithm 4 in Sect. 6 below to compute  $\mathfrak{G}(P_i)$  for all  $i$ .
  - 3: **return**  $\bigcap_{i=1}^s \mathfrak{G}(P_i)$ .
-

## 6 Computing Local Adjoint Ideals

In this section, we modify Algorithm 2 so that it computes the local Gorenstein adjoint ideal of  $\Gamma$  at a point  $P$  from a minimal local contribution to  $\overline{k[C]}$  at  $P$  via ideal quotients.

Fix a curve  $\Gamma \subset \mathbb{P}_k^r$  as in Notation 3.8, a point  $P \in \text{Sing}(\Gamma)$ , and an affine chart containing  $P$ . For simplicity of the presentation, we stick with the chart  $X_0 \neq 0$ , and let  $C$  be the affine part of  $\Gamma$  as before. Consider an ideal  $U \subset k[C]$  and a non-zero element  $d \in U$  such that  $\frac{1}{d}U$  is the minimal local contribution to  $\overline{k[C]}$  at  $P$ .

**Proposition 6.1** *With notation as above, and given  $Q \in C$ , we have*

$$(\langle d \rangle_{k[C]} : U)_Q = \begin{cases} \mathcal{C}_{\mathcal{O}_{C,Q}} & \text{if } Q = P, \\ \mathcal{O}_{C,Q} & \text{if } Q \neq P. \end{cases}$$

*Proof* By the minimality assumption, we have

$$\left(\frac{1}{d}U\right)_Q = \begin{cases} \overline{\mathcal{O}_{C,Q}} & \text{if } Q = P, \\ \mathcal{O}_{C,Q} & \text{if } Q \neq P. \end{cases}$$

The claim follows since localization commutes with forming the conductor:

$$(\langle d \rangle_{k[C]} : U)_Q = \left(\mathcal{C}_{\left(\frac{1}{d}U\right) / k[C]}\right)_Q = \mathcal{C}_{\left(\frac{1}{d}U\right)_Q / k[C]_Q}.$$

Now, we argue as in the proof of Proposition 3.9: From Proposition 6.1 and Remark 5.2, it follows that  $\langle d \rangle_{k[C]} : U$  coincides with the ideal obtained by dehomogenizing  $\mathfrak{G}(P)$  with respect to  $X_0$  and mapping the result to  $k[C]$ . Hence, since  $\mathfrak{G}(P)$  is saturated, Algorithm 4 below indeed computes  $\mathfrak{G}(P)$ .

---

### Algorithm 4 Local Gorenstein adjoint ideal from a local contribution

---

**Input:** A curve  $\Gamma \subset \mathbb{P}_k^r$  over a perfect field  $k$  with affine part  $C$  as in Notation 3.8 and a point  $P \in \text{Sing}(C) \subset \text{Sing}(\Gamma)$ .

**Output:** The local Gorenstein adjoint ideal  $\mathfrak{G}(P)$  of  $\Gamma$ .

- 1: Compute polynomials  $d, a_0, \dots, a_s \in k[X_1, \dots, X_r]$  such that the fractions  $\frac{a_i(x_1, \dots, x_r)}{d(x_1, \dots, x_r)}$  generate the minimal local contribution to  $\overline{k[C]}$  at  $P$  as a  $k[C]$ -module.
- 2: Compute the ideal quotient

$$\mathcal{C} = (\langle d \rangle + I(C)) : (\langle a_0, \dots, a_s \rangle + I(C)) \subset k[X_1, \dots, X_r].$$

- 3: **return** the homogenization of  $\mathcal{C}$  with respect to  $X_0$ .
-

*Example 6.2* Let  $\Gamma \subset \mathbb{P}_{\mathbb{C}}^2$  be the projective closure of the curve  $C$  with affine equation

$$X^5 - Y^2(1 - Y)^3 = 0$$

as in Examples 2.12, 2.17, 4.5, and 4.8. We compute the local Gorenstein adjoint ideals. For the  $A_4$ -singularity  $P_1$ , we know from Example 2.17 that

$$d_1 = x^2 \text{ and } U_1 = \langle x^2, y(y-1)^3 \rangle_{\mathbb{C}[C]},$$

so that

$$\mathfrak{G}(P_1) = \langle X^2, Y \rangle.$$

For the  $E_8$  singularity  $P_2$ , in turn, we have

$$d_2 = x^3 \text{ and } U_2 = \langle x^3, x^2y^2(y-1), y^2(y-1)^2 \rangle_{\mathbb{C}[C]},$$

leading to

$$\mathfrak{G}(P_2) = \langle X^3, X(Y-Z), (Y-Z)^2 \rangle.$$

Note that  $\mathfrak{G}(P_1)$  and  $\mathfrak{G}(P_2)$  are the ideals already obtained in Example 4.6.

## 7 Improvements to the Local Strategy for Plane Curves

In this section, we focus on the case of a plane curve  $\Gamma$  with affine part  $C = V(f)$  and  $\text{Sing}(\Gamma) = \text{Sing}(C)$  as in Notation 4.2. *For simplicity of the presentation, we suppose throughout the section that our ground field  $k = \mathbb{C}$ .*

As explained in Sect. 5, the Gorenstein adjoint ideal  $\mathfrak{G}$  can be computed as the intersection of local Gorenstein ideals via a partition of  $\text{Sing}(C)$ . To begin with, consider the following partition:

$$\text{Sing}(C) = W_2 \cup W_3 \cup \dots \cup W_r \cup W', \tag{5}$$

where, for each  $i$ ,  $W_i$  denotes the locus of ordinary  $i$ -fold points (ordinary multiple points of multiplicity  $i$ )<sup>2</sup> and where  $W'$  collects the remaining singularities of  $C$ . In particular,  $W_2$  is the set of nodes of  $C$ . Note that in many practical examples  $W' = \emptyset$ .

---

<sup>2</sup>Recall that an ordinary multiple point of multiplicity  $i$  is a singularity where the lowest non-vanishing jet of  $f$  factors into  $i$  distinct linear factors.



**Lemma 7.1** *Let  $P \in \text{Sing}(C)$ , and let  $\mathfrak{m}_P \subset k[X, Y]$  be the corresponding maximal ideal. If  $P$  is an ordinary  $i$ -fold point of  $C$ , then*

$$\mathfrak{G}(P) = \mathfrak{m}_P^{i-1}.$$

*Proof* Since  $C$  is a plane curve and  $P$  is an ordinary  $i$ -fold point of  $C$ , the conductor  $\mathcal{C}_{\mathcal{O}_{C,P}} = \mathfrak{m}_{C,P}^{i-1}$ , where  $\mathfrak{m}_{C,P}$  is the maximal ideal of  $\mathcal{O}_{C,P}$  (see [29, 47]). The result follows from the very definition of  $\mathfrak{G}(P)$ .

Applying the lemma to the partition (5), we get the intersection of ideals

$$\mathfrak{G} = I(W_2) \cap I(W_3)^2 \cap \cdots \cap I(W_r)^r \cap \mathfrak{G}(W'). \quad (6)$$

Hence, in the case where  $\Gamma$  is known to have ordinary multiple points as singularities only (that is,  $W' = \emptyset$ ), we can compute  $\mathfrak{G}$  in a very efficient way by using Algorithm 5 below (see [5]).

In the general case, Eq. (6) allows us to reduce the computation of  $\mathfrak{G}$  to the less involved task of computing  $\mathfrak{G}(W')$  as soon as we have detected the ordinary  $i$ -fold points. To begin with treating these, here is how to find the nodes:

*Remark 7.2* We know how to find *all* singularities:  $\text{Sing}(C)$  is given by the ideal

$$J = \left\langle f, \frac{\partial f}{\partial X}, \frac{\partial f}{\partial Y} \right\rangle.$$

Now consider the Hessian matrix  $\text{Hess}(f)$  formed by the second partial derivatives of  $f$ . By the Morse lemma (see [49]), a point  $P \in \text{Sing}(C)$  is a node iff  $\text{Hess}(f)$  is

---

**Algorithm 5** Gorenstein adjoint ideal, ordinary multiple points only

---

**Input:** A plane curve  $\Gamma$  of degree  $n$  with defining polynomial  $F$  as in Notation 4.2 with only ordinary multiple points as singularities.

**Output:** The Gorenstein adjoint ideal  $\mathfrak{G}$  of  $\Gamma$ .

- 1:  $J_1 = \left\langle \frac{\partial F}{\partial X}, \frac{\partial F}{\partial Y}, \frac{\partial F}{\partial Z} \right\rangle$  (the ideal defining  $\text{Sing}(\Gamma)$ )
  - 2:  $i = 1$
  - 3: **while**  $(J_i : \langle X, Y, Z \rangle^\infty) \neq \langle 1 \rangle$  **do**
  - 4:    $i = i + 1$
  - 5:    $J_i = \left\langle \frac{\partial^{i+l+m} F}{\partial X^j \partial Y^l \partial Z^m} \mid j+l+m = i; j, l, m \in \mathbb{N} \right\rangle$
  - 6:  $B = \langle X, Y, Z \rangle^{n-i}$
  - 7: **while**  $i > 0$  **do**
  - 8:    $I_i = (J_{i-1} : B^\infty)$  (the ideal of the  $i$ -fold points of  $\Gamma$ )
  - 9:    $B = ((B \cap I_i^{i-1}) : \langle X, Y, Z \rangle^\infty)$
  - 10:    $i = i - 1$
  - 11: **return**  $B$
-

non-degenerate at  $P$ . That is,  $P$  is a node iff

$$I(P) + \langle \det(\text{Hess}(f)) \rangle = k[X, Y].$$

This gives us a fast way of computing  $W_2$ .

Carrying our efforts one step further, we discuss the local analysis of the singularities via invariants. This yields an efficient method not only for finding the delta invariant, but also for detecting the ordinary  $i$ -fold points, for each  $i$ :

*Remark 7.3* Let  $P \in \text{Sing}(C)$ . After a translation, we may assume that  $P = (0, 0)$  is the origin. Write  $m_P$  for the multiplicity and

$$\mu_P = \dim_k \left( k[[X, Y]] / \left\langle \frac{\partial f}{\partial X}, \frac{\partial f}{\partial Y} \right\rangle \right)$$

for the *Milnor number* of  $C$  at  $P$ . Then  $m_P = \deg h_P$ , where  $h_P$  is the lowest degree homogeneous summand of the Taylor expansion of  $f$  at  $P$ . Recall that  $\mu_P$  can be computed via standard bases (see [32]). Furthermore, if the Newton polygon of  $f$  is non-degenerate (otherwise, successively blow up), the *number  $r_P$  of branches* of  $f$  at  $P$  can be computed as

$$r_P = \sum_{j=1}^{s-1} \gcd \left( V_X^{(j+1)} - V_X^{(j)}, V_Y^{(j+1)} - V_Y^{(j)} \right),$$

where  $V^{(1)}, \dots, V^{(s)}$  are the (ordered) vertices of the Newton polygon (and  $X$  and  $Y$  refer to their respective coordinates). This is immediate from [13, Section 8.4, Lemma 3]. The delta invariant of  $C$  at  $P$  is then obtained as

$$\delta_P = \frac{1}{2}(\mu_P + r_P - 1)$$

(see, for example, [33, Chapter 1, Proposition 3.34]). Furthermore,  $P$  is an ordinary  $i$ -fold point iff  $h_P$  is square-free and  $m_P = i$ . Equivalently,

$$(m_P, r_P, \delta_P) = \left( i, i, \binom{i}{2} \right).$$

See [33, Chapter 1, Proposition 3.33].

The local analysis of the singularities may be used to further refine our partition of  $\text{Sing}(C)$ . For example, singularities of type *ADE* can be identified as follows:

*Remark 7.4* With notation as in Remark 7.3, the point  $P = (0, 0) \in \text{Sing}(C)$  is

1. of type  $A_n$ ,  $n \geq 2$ , iff  $h_P = l_1^2$ , with  $l_1 \in k[X, Y]$  linear, and  $\mu_P = n$ ,
2. of type  $D_n$ ,  $n \geq 4$ , iff  $h_P = l_1 l_2 l_3$  or  $h_P = l_1^2 l_2$ , with pairwise different linear polynomials  $l_j \in k[X, Y]$ , and  $\mu_P = n$ , and

3. of type  $E_n$ ,  $n = 6, 7, 8$ , iff  $h_P = l_1^3$ , with  $l_1 \in k[X, Y]$  linear, and  $\mu_P = n$ .

Here, in (2),  $h_P$  splits into three different linear factors iff  $P$  is of type  $D_4$ . See, for example, [33, Chapter 1, Theorems 2.48, 2.51, 2.54].

To describe the local Gorenstein adjoint ideal at a singularity of type  $A$ ,  $D$ , or  $E$ , we use the following notation:

**Notation 7.5** For any element  $g \in k[[X, Y]]$ , let  $g_j = \text{taylor}(g, j) \in k[X, Y]$  be the Taylor expansion of  $g$  at  $P = (0, 0)$  modulo  $O(j + 1)$ .<sup>3</sup>

If  $C$  has a singularity of type  $A_n$  at  $P = (0, 0)$ , we may write  $f$  in the form  $f = T^2 + W^{n+1}$ , where  $T, W \in k[[X, Y]]$  is a regular system of parameters. Let  $s = \lfloor \frac{n+1}{2} \rfloor$  (the meaning of  $s$  will become clear in the proof of Lemma 7.6). We may compute the Taylor expansion  $T_{s-1} \in k[X, Y]$  as follows. If  $n$  and thus  $s$  is equal to 1, set  $T_0 = 0$ . Otherwise, inductively solve  $f$  for  $T$ : Start by choosing a linear form  $T_1 \in k[X, Y]$  such that  $\text{taylor}(f, 2) = T_1^2$ . Supposing that  $1 < j < s - 1$  and  $T_j = T + O(j + 1)$  has already been computed, write

$$\text{taylor}(f - T_j^2, j + 2) = 2T_1 \cdot m,$$

with  $m \in k[X, Y]$  homogeneous of degree  $j + 1$ , and set  $T_{j+1} = T_j + m$ .

**Lemma 7.6** *Let  $C$  have a singularity of type  $A_n$ ,  $n \geq 1$ , at  $P = (0, 0)$ . Set  $s = \lfloor \frac{n+1}{2} \rfloor$ , and let  $T_{s-1}$  be defined as above. Then  $\mathfrak{G}(P)$  is the homogenization of*

$$\langle X^s, T_{s-1}, Y^s \rangle \subset k[X, Y]$$

with respect to  $Z$ .

*Proof* The case  $n = 1$  is clear, so we may suppose  $n \geq 2$ . If  $\mathfrak{G}' = \langle X^s, T_{s-1}, Y^s \rangle \subset k[X, Y]$ , then  $\mathfrak{G}'_Q = \mathcal{O}_{C,Q}$  for all  $Q \in C \setminus \{P\}$ , so it suffices to show that  $\mathfrak{G}'_P = \mathcal{C}_B$ , where  $B = \mathcal{O}_{C,P}$ . For this, we pass to the completion

$$\widehat{B} = k[[x, y]] = k[[X, Y]] / \langle f(X, Y) \rangle,$$

and consider the isomorphism

$$A = k[[t, w]] = k[[T, W]] / \langle T^2 + W^{n+1} \rangle \rightarrow \widehat{B}, t \mapsto T(x, y), w \mapsto W(x, y).$$

An analysis of the normalization algorithm applied to  $A$  shows that

$$\bar{A} = \sum_{i=0}^{n-s} k[[t]] \cdot w^i + \sum_{i=n-s+1}^n k[[t]] \cdot \frac{w^i}{t},$$

<sup>3</sup>The notation  $O(m)$  stands for terms of degree  $\geq m$ .

and that it takes  $s = \lfloor \frac{n+1}{2} \rfloor$  steps to reach  $\bar{A}$  (see [9, Sect. 4]). Hence,

$$\mathcal{C}_A = \langle t, w^s \rangle_A, \text{ so that } \mathcal{C}_{\bar{B}} = \langle T(x, y), W(x, y)^s \rangle_{\bar{B}}.$$

Working in  $k[[X, Y]]$ , we write

$$T = aX + bY \text{ and } W = cX + dY,$$

where  $a, b, c, d \in k[[X, Y]]$  are such that  $ad - bc$  is a unit in  $k[[X, Y]]$ . Since  $\langle X, Y \rangle = \langle T, W \rangle$ , it follows that  $\langle X, Y \rangle^s = \langle T, W \rangle^s \subset \langle T, W^s \rangle$ . Since  $\langle X, Y \rangle = \langle X, T \rangle$  or  $\langle X, Y \rangle = \langle T, Y \rangle$ , we have  $W^s \in \langle X, Y \rangle^s \subset \langle X^s, T, Y^s \rangle$ . We conclude that

$$\langle X^s, T, Y^s \rangle = \langle T, W^s \rangle.$$

If  $s > 1$ , then  $\langle X, Y \rangle = \langle X, T_{s-1} \rangle$  or  $\langle X, Y \rangle = \langle T_{s-1}, Y \rangle$ . Hence, for any  $s$ , we have  $\langle X, Y \rangle^s \subset \langle X^s, T_{s-1}, Y^s \rangle$ . We conclude that

$$\langle X^s, T_{s-1}, Y^s \rangle = \langle X^s, T, Y^s \rangle.$$

Now recall that  $B$  is an excellent ring, which implies that  $\bar{\bar{B}} = \widehat{\bar{B}}$  (see, for example, [9, Sect. 1]). It follows that

$$\widehat{\mathcal{C}_B} = \text{Hom}_{\widehat{B}}(\widehat{\bar{B}}, B) = \text{Hom}_B(\bar{B}, B) \otimes_B \widehat{B} = \mathcal{C}_B \otimes_B \widehat{B}. \quad (7)$$

Since completion is faithfully flat in the case considered here, we conclude that

$$\mathcal{C}_B = \langle x^s, T_{s-1}(x, y), y^s \rangle_B.$$

*Remark 7.7* In particular, if  $P$  is a cusp, then  $\mathfrak{G}(P) = \langle X, Y \rangle$ . So in Eq. (6), nodes and cusps may be treated simultaneously.

If  $C$  has a singularity of type  $D_n$  at  $P = (0, 0)$ , we may write  $f$  in the form  $f = W \cdot (T^2 + W^{n-2})$ , where  $T, W \in k[[X, Y]]$  is a regular system of parameters. Let  $s = \lfloor \frac{n}{2} \rfloor$ . We may compute the Taylor expansion  $T_{s-2} \in k[X, Y]$  as follows. If  $n = 4$ , set  $T_0 = 0$ . If  $n \geq 5$ , choose linear forms  $T_1, W_1 \in k[X, Y]$  such that  $\text{taylor}(f, 3) = T_1^2 \cdot W_1$ . For  $j \leq s-2$ , determine  $W_j = W + O(j+1)$  as the Puiseux expansion up to order  $j$  of  $f$  corresponding to  $W_1$ . Supposing that  $1 < j < s-2$  and  $T_j = T + O(j+1)$  has already been computed, write

$$\text{taylor}(f - T_j^2 \cdot W_{j+1}, j+3) = 2Z_1 \cdot W_1 \cdot m,$$

with  $m \in k[X, Y]$  homogeneous of degree  $j+1$ , and set  $T_{j+1} = T_j + m$ .

**Lemma 7.8** *Let  $C$  have a singularity of type  $D_n$ ,  $n \geq 4$ , at  $P = (0, 0)$ . Set  $s = \lfloor \frac{n}{2} \rfloor$ , and let  $T_{s-2}$  be defined as above. Then  $\mathfrak{G}(P)$  is the homogenization of*

$$\langle X, Y \rangle \cdot \langle X^{s-1}, T_{s-2}, Y^{s-1} \rangle \subset k[X, Y]$$

with respect to  $Z$ .

*Proof* We have an isomorphism

$$A \rightarrow \widehat{B}, q \mapsto T(x, y), w \mapsto W(x, y),$$

where  $B = \mathcal{O}_{C,P}$  and

$$A = k[[t, w]] = k[[T, W]] / \langle W \cdot (T^2 + W^{n-2}) \rangle.$$

This time, the normalization is

$$\bar{A} = \sum_{i=0}^{n-2-s} k[[t]] \cdot w^i + \sum_{i=n-1-s}^{n-3} k[[t]] \cdot \frac{w^i}{t} + k[[t]] \cdot \frac{w^{n-2}}{t^2},$$

and it takes  $s = \lfloor \frac{n}{2} \rfloor$  steps to reach  $\bar{A}$  (see again [9, Sect. 4]). Hence,

$$\mathcal{C}_A = \langle t^2, tw, w^s \rangle.$$

Write

$$T = aX + bY \text{ and } W = cX + dY,$$

where  $a, b, c, d \in k[[X, Y]]$  are such that  $ad - bc$  is a unit in  $k[[X, Y]]$ . Since  $\langle X, Y \rangle = \langle T, W \rangle$ , we have  $\langle XT, YT \rangle = \langle T^2, TW \rangle$  and  $\langle X, Y \rangle^s = \langle T, W \rangle^s \subset \langle T^2, TW, W^s \rangle$ . Hence,

$$\langle X, Y \rangle \cdot \langle X^{s-1}, T, Y^{s-1} \rangle \subset \langle T^2, TW, W^s \rangle.$$

For the other inclusion, observe that  $\langle X, Y \rangle = \langle X, T \rangle$  or  $\langle X, Y \rangle = \langle T, Y \rangle$ , so that  $\langle X, Y \rangle^{s-1} \subset \langle X^{s-1}, T, Y^{s-1} \rangle$ . Hence,

$$W^s \in \langle X, Y \rangle^s \subset \langle X, Y \rangle \cdot \langle X^{s-1}, T, Y^{s-1} \rangle.$$

If  $s > 2$ , then  $\langle X, Y \rangle = \langle X, T_{s-2} \rangle$  or  $\langle X, Y \rangle = \langle T_{s-2}, Y \rangle$ . Hence, for any  $s$ , we have  $\langle X, Y \rangle^{s-1} \subset \langle X^{s-1}, T_{s-2}, Y^{s-1} \rangle$ . We conclude that

$$\langle X^{s-1}, T_{s-2}, Y^{s-1} \rangle = \langle X^{s-1}, T, Y^{s-1} \rangle.$$

To summarize,

$$\langle T^2, TW, W^s \rangle = \langle X, Y \rangle \cdot \langle X^{s-1}, T, Y^{s-1} \rangle = \langle X, Y \rangle \cdot \langle X^{s-1}, T_{s-2}, Y^{s-1} \rangle,$$

so that

$$\mathcal{C}_{\widehat{B}} = \langle x, y \rangle \cdot \langle x^{s-1}, T_{s-2}(x, y), y^{s-1} \rangle \subset \widehat{B}.$$

Then the claim follows as before.

**Lemma 7.9** *Let  $C$  have a singularity of type  $E_n$ ,  $n = 6, 7, 8$ , at  $P = (0, 0)$ . Set  $s = \lfloor \frac{n-1}{2} \rfloor$ , and let  $l_1$  be as in Remark 7.4. Then  $\mathfrak{G}(P)$  is the homogenization of*

$$\langle X, Y \rangle \cdot \langle X^{s-1}, l_1, Y^{s-1} \rangle \subset k[X, Y]$$

with respect to  $Z$ .

*Proof* Depending on  $n \in \{6, 7, 8\}$ , we have an isomorphism

$$A \rightarrow \widehat{B}, q \mapsto T(x, y), w \mapsto W(x, y),$$

where  $B = \mathcal{O}_{C,P}$  and

$$A = k[[t, w]] = k[[T, W]] / \langle T^3 + W^4 \rangle \text{ or}$$

$$A = k[[t, w]] = k[[T, W]] / \langle T(T^2 + W^3) \rangle \text{ or}$$

$$A = k[[t, w]] = k[[T, W]] / \langle T^3 + W^5 \rangle.$$

In each case, by Böhm et al. [9, Sect. 4],

$$\overline{A} = k[[w]] \cdot 1 + k[[w]] \cdot \frac{t}{w} + k[[w]] \cdot \frac{t^2}{w^s},$$

which implies that

$$\mathcal{C}_A = \langle t^2, tw, w^s \rangle.$$

The same argument as in the proof of Lemma 7.8 shows that

$$\mathcal{C}_{\widehat{B}} = \langle x, y \rangle \cdot \langle x^{s-1}, T_{s-2}(x, y), y^{s-1} \rangle \subset \widehat{B},$$

and the claim follows as before. Note that  $T_{s-2} = 0$  if  $s = 2$ , and  $T_{s-2} = l_1$  if  $s = 3$ .

In principle, we could pursue a similar strategy for all singularities classified by Arnold in [4]. However, in [10], we give an algorithm which, for plane curves in characteristic zero, allows us to compute the local contributions to the normalization

for a broad class of singularities in a direct way. Combining the approach of Sect. 6 with this algorithm or with modular techniques and normalization as described in Sect. 8 below, we already get a very efficient algorithm for computing  $\mathfrak{O}$ .

## 8 Parallel Computation Using Modular Techniques

Algorithm 3 is parallel in nature since the computations of the local adjoint ideals do not depend on each other. In this section, in the case where the given curve is defined over  $\mathbb{Q}$ , we describe a modular way of parallelizing Algorithm 3 even further. One possible approach is to replace the computations of the Gröbner bases involved, the computation of the (minimal) associated primes in the singular locus, and the computations yielding the normalizations by their modular variants as introduced in [3, 40], and [6]. These variants are either probabilistic or require expensive tests to verify the results at the end. To reduce the number and complexity of the verification tests, we provide a direct modularization for the adjoint ideal algorithm. The approach we propose requires only the verification of the final result: In the next section, we give efficient conditions for checking whether the result obtained is indeed the Gorenstein adjoint ideal.

Our approach relies on the general scheme for modular computations presented in Böhm et al. [11] and provided, in fact, motivation for developing the scheme. This is based on error tolerant rational reconstruction (a short account of which will be given in Remark 8.8 below) and can handle *bad primes*,<sup>4</sup> provided there are only finitely many such primes. Referring to [11] for details, we will now outline the main ideas behind the scheme.

Fix a global monomial ordering  $>$  on the monoid of monomials in the variables  $X = \{X_0, \dots, X_r\}$ . Consider the polynomial rings  $R = \mathbb{Q}[X]$  and, given an integer  $N \geq 2$ ,  $R_N = (\mathbb{Z}/N\mathbb{Z})[X]$ . If  $H \subset R$  or  $H \subset R_N$  is a set of polynomials, then denote by  $\text{LM}(H) := \{\text{LM}(h) \mid h \in H\}$  its set of leading monomials.

If  $\frac{a}{b} \in \mathbb{Q}$  with  $\text{gcd}(a, b) = 1$  and  $\text{gcd}(b, N) = 1$ , set  $\left(\frac{a}{b}\right)_N := (a + N\mathbb{Z})(b + N\mathbb{Z})^{-1} \in \mathbb{Z}/N\mathbb{Z}$ . If  $f \in R$  is a polynomial such that  $N$  is coprime to any denominator of a coefficient of  $f$ , then its *reduction modulo  $N$*  is the polynomial  $f_N \in R_N$  obtained by mapping each coefficient  $c$  of  $f$  to  $c_N$ . If  $H = \{h_1, \dots, h_s\} \subset R$  is a set of polynomials such that  $N$  is coprime to any denominator of a coefficient of any  $h_i$ , set  $H_N = \{(h_1)_N, \dots, (h_s)_N\}$ . If  $J \subset R$  is an ideal, we write

$$J_0 = J \cap \mathbb{Z}[X] \quad \text{and} \quad J_N = \langle f_N \mid f \in J \rangle \subset R_N,$$

and call  $J_N$  the *reduction of  $J$  modulo  $N$* . We also write  $(R/J)_N = R_N/J_N$ .

---

<sup>4</sup>In our context, a prime  $p$  is *bad* if Algorithm 3, applied to the modulo  $p$  values of the input over the rationals, does not return the reduction of the characteristic zero result.

As a first step towards the modular algorithm, we explain how to compute the reduction of a given ideal  $J \subset R$  modulo a prime, supposing that a Gröbner basis for  $J$  is already known.

**Lemma 8.1** *With notation as above, let  $J \subset R$  be an ideal, let  $H = \{h_1, \dots, h_s\}$  be a Gröbner basis for  $J$  with elements  $h_i \in \mathbb{Z}[X]$ , and let  $p$  be a prime not dividing any of the leading coefficients  $\text{LC}(h_i)$ . Then for every  $f \in J \cap \mathbb{Z}[X]$ , there exists an integer  $d \in \mathbb{Z}$  not divisible by  $p$ , and such that  $df \in \langle H \rangle_{\mathbb{Z}[X]}$ .*

*Proof* Let  $f \in J \cap \mathbb{Z}[X]$ . Then, since  $H$  is a Gröbner basis for  $J$ , there exists an  $h_i \in H$  such that  $\text{LM}(f)$  is divisible by  $\text{LM}(h_i)$ . We hence have a representation  $\text{LC}(h_i) \cdot f = m \cdot h_i + f^{(1)}$  with  $f^{(1)} \in J \cap \mathbb{Z}[X]$ , and such that  $\text{LM}(f) > \text{LM}(f^{(1)})$ . Proceeding with  $f^{(1)}$  instead of  $f$  and continuing that way, we get an integer  $d \in \mathbb{Z}$  and a representation  $df = \sum_{i=1}^s \xi_i h_i$  as desired.

**Corollary 8.2** *If  $J, H$ , and  $p$  are as above, then  $J_p = \langle H_p \rangle_{\mathbb{F}_p[X]}$ .*

*Proof* Given  $f \in J \cap \mathbb{Z}[X]$ , let  $df = \sum_{i=1}^s \xi_i h_i$  be a representation as above. Then  $d_p f_p = \sum_{i=1}^s (\xi_i)_p (h_i)_p$ . We conclude that  $f_p \in \langle H_p \rangle_{\mathbb{F}_p[X]}$ .

**Corollary 8.3** *With  $J$  and  $H$  as above, let  $p$  be a prime such that  $H_p$  is a Gröbner basis with  $\text{LM}(H) = \text{LM}(H_p)$ . Then  $J_p = \langle H_p \rangle_{\mathbb{F}_p[X]}$ .*

We now fix the following setup for the rest of this section:

**Notation 8.4** Let  $\Gamma \subset \mathbb{P}_{\mathbb{Q}}^r$  be an integral non-degenerate projective algebraic curve, let  $I(\Gamma)$  be the ideal of  $\Gamma$  in  $R$ , and let  $G(0) \subset R$  be the reduced Gröbner basis of  $\mathfrak{G}(\Gamma)$ . If  $p$  is a prime such that  $\text{LM}(I(\Gamma)) = \text{LM}(I(\Gamma)_p)$ , and  $I(\Gamma)_p$  is radical and defines an integral non-degenerate projective algebraic curve in  $\mathbb{P}_{\mathbb{F}_p}^r$ , then write  $\Gamma_p$  for this curve and  $G(p) \subset R_p$  for the reduced Gröbner basis of  $\mathfrak{G}(\Gamma_p)$ .

*Remark 8.5* There are only finitely many primes  $p$  for which the desired conditions on  $I(\Gamma)_p$  in Notation 8.4 are not satisfied. Since these conditions can be checked using Gröbner bases and square-free decomposition, we may reject such a prime if we encounter it in the modular algorithm. In the following discussion, we will ignore these bad primes for simplicity of the presentation. In particular, we will assume that the Gröbner bases  $G(p)$  are defined for all primes  $p$ .

The basic idea of the modular adjoint ideal algorithm can now be described as follows: First, choose a set of primes  $\mathcal{P}$  and compute  $G(p)$  for each  $p \in \mathcal{P}$ . Second, lift the  $G(p)$  coefficientwise to a set of polynomials  $G \subset R$ . Provided that  $\mathfrak{G}(\Gamma)_p = \mathfrak{G}(\Gamma_p)$  for each  $p \in \mathcal{P}$ , we then expect that  $G$  is a Gröbner basis which coincides with our target Gröbner basis  $G(0)$ .

The lifting process consists of two steps. First, use Chinese remaindering to lift the  $G(p) \subset R_p$  to a set of polynomials  $G(N) \subset R_N$ , with  $N := \prod_{p \in \mathcal{P}} p$ . Second, compute a set of polynomials  $G \subset R$  by lifting the coefficients occurring in  $G(N)$  to rational coefficients. Here, to identify Gröbner basis elements corresponding to each other, we require that  $\text{LM}(G(p)) = \text{LM}(G(q))$  for all  $p, q \in \mathcal{P}$ . This leads to



condition (L2) in the definition below:

**Definition 8.6** With notation as above, a prime  $p$  is called *lucky* if

$$(L1) \quad \mathfrak{G}(\Gamma)_p = \mathfrak{G}(\Gamma_p) \text{ and}$$

$$(L2) \quad \text{LM}(G(0)) = \text{LM}(G(p)).$$

Otherwise  $p$  is called *unlucky*.

**Lemma 8.7** *All but finitely many primes are lucky.*

*Proof* As is clear from the proof of [11, Lemma 5.5], it is enough to show that condition (L1) is true for all but finitely many primes. For this, we may assume that  $\Gamma$  does not have singularities at  $X_0 = 0$ . Then for all but finitely many primes  $p$ , the curve  $\Gamma_p$  does not have singularities at  $X_0 = 0$ .

Let  $C$  be the affine part of  $\Gamma$  in the chart  $X_0 \neq 0$ . Write  $A = \mathbb{Q}[X_1, \dots, X_r]/I(C)$ . Using a Gröbner basis argument as summarized in [11, Remark 5.3], it is shown in [6, Section 4] that  $(\overline{A})_p = \overline{A}_p$  for all but finitely many primes  $p$ . So if we write  $\overline{A} = \frac{1}{d}U$ , with an ideal  $U \subset A$  and an element  $0 \neq d \in A$ , and  $\overline{A}_p = \frac{1}{d(p)}U(p)$ , with an ideal  $U(p) \subset A_p$  and an element  $0 \neq d(p) \in A_p$ , then

$$(d_p : U_p) = (d(p) : U(p))$$

for all but finitely many primes  $p$ .

Computing an ideal quotient amounts to another Gröbner basis computation. Hence, we may again apply [11, Remark 5.3] to conclude that

$$(d : U)_p = (d_p : U_p)$$

for all but finitely many primes  $p$ .

Summing up, the result follows from Propositions 3.9 and 4.7.

When performing the modular algorithm, condition (L1) can only be checked a posteriori: We compute  $G(p)$  and, thus,  $\mathfrak{G}(\Gamma_p)$  on our way, but  $\mathfrak{G}(\Gamma)_p$  is only known to us after  $G(0)$  and, thus,  $\mathfrak{G}(\Gamma)$  has been computed. This is not a problem, however, since the finitely many primes where  $\mathfrak{G}(\Gamma)_p \neq \mathfrak{G}(\Gamma_p)$  will not influence the final result if we apply error tolerant rational reconstruction as discussed now.

*Remark 8.8* Let  $N'$  and  $M$  be integers with  $\gcd(N', M) = 1$ , let  $N = N' \cdot M$ , and let  $\frac{a}{b} \in \mathbb{Q}$  with  $\gcd(a, b) = \gcd(N', b) = 1$ . Set  $\bar{r}_1 := (\frac{a}{b})_{N'} \in \mathbb{Z}/N'\mathbb{Z}$ , let  $\bar{r}_2 \in \mathbb{Z}/M\mathbb{Z}$  be arbitrary, and denote by  $\bar{r}$ , with  $0 \leq r \leq N - 1$ , the image of  $(\bar{r}_1, \bar{r}_2)$  under the isomorphism

$$\mathbb{Z}/N'\mathbb{Z} \times \mathbb{Z}/M\mathbb{Z} \rightarrow \mathbb{Z}/N\mathbb{Z}.$$

Lifting  $\bar{r}$  to a rational number by Gaussian reduction, starting from  $(a_0, b_0) = (N'M, 0)$  and  $(a_1, b_1) = (r, 1)$ , we create the sequence  $(a_i, b_i)$  obtained by

$$(a_{i+2}, b_{i+2}) = (a_i, b_i) - q_i(a_{i+1}, b_{i+1}),$$

with

$$q_i = \left\lfloor \frac{\langle (a_i, b_i), (a_{i+1}, b_{i+1}) \rangle}{\|(a_{i+1}, b_{i+1})\|^2} \right\rfloor.$$

Computing this sequence until  $\|(a_{i+2}, b_{i+2})\| \geq \|(a_{i+1}, b_{i+1})\|$ , we return false if  $\|(a_{i+1}, b_{i+1})\|^2 \geq N$ , and  $\frac{a_{i+1}}{b_{i+1}}$ , otherwise. By Böhm et al. [11, Lemma 4.3], this algorithm will return  $\frac{a_{i+1}}{b_{i+1}} = \frac{a}{b}$ , provided that  $N$  is large enough and  $M \ll N'$ . More precisely, we ask that  $N' > (a^2 + b^2) \cdot M$ .

**Definition 8.9** If  $\mathcal{P}$  is a finite set of primes, set

$$N' = \prod_{p \in \mathcal{P} \text{ lucky}} p \quad \text{and} \quad M = \prod_{p \in \mathcal{P} \text{ unlucky}} p.$$

Then  $\mathcal{P}$  is called *sufficiently large* if

$$N' > (a^2 + b^2) \cdot M$$

for any coefficient  $\frac{a}{b}$  of any polynomial in  $G(0)$  (assume  $\gcd(a, b) = 1$ ).

**Lemma 8.10** *If  $\mathcal{P}$  is a sufficiently large set of primes satisfying condition (L2), then the reduced Gröbner bases  $G(p)$ ,  $p \in \mathcal{P}$ , lift via Chinese remaindering and error tolerant rational reconstruction to the reduced Gröbner basis  $G(0)$ .*

*Proof* By Lemma 8.7, condition (L1) holds for all but finitely many primes  $p$ . Hence, since  $\mathcal{P}$  is sufficiently large, the result follows as in the proof of [11, Lemma 5.6] from [11, Lemma 4.3].

Lemma 8.7 guarantees, in particular, that a sufficiently large set  $\mathcal{P}$  of primes satisfying condition (L2) exists. So from a theoretical point of view, the idea of finding  $G(0)$  is now as follows: Consider such a set  $\mathcal{P}$ , compute the reduced Gröbner bases  $G(p)$ ,  $p \in \mathcal{P}$ , and lift the results to  $G(0)$ .

From a practical point of view, however, we face the problem that condition (L2) can only be checked a posteriori. On the other hand, as already pointed out, we need that the  $G(p)$ ,  $p \in \mathcal{P}$ , have the same set of leading monomials in order to identify corresponding Gröbner basis elements in the lifting process. To remedy this situation, we suggest to proceed in a randomized way: First, fix an integer  $t \geq 1$  and choose a set of  $t$  primes  $\mathcal{P}$  at random. Second, compute  $\mathcal{G} = \{G(p) \mid p \in \mathcal{P}\}$ , and use a majority vote on the set of lead monomials to choose a subset of  $\mathcal{G}$  such that all Gröbner bases in the subset have the same set of lead monomials:

`DELETEBYMAJORITYVOTE`: Define an equivalence relation on  $\mathcal{P}$  by setting  $p \sim q : \iff \text{LM}(G(p)) = \text{LM}(G(q))$ . Then replace  $\mathcal{P}$  by the equivalence class of largest cardinality,<sup>5</sup> and change  $\mathcal{G}$  accordingly.

---

<sup>5</sup>We have to use a weighted cardinality count: when enlarging  $\mathcal{P}$ , the total weight of the elements already present must be strictly smaller than the total weight of the new elements. Otherwise, though highly unlikely in practical terms, it may happen that only unlucky primes are accumulated.

Now, all  $G(p)$ ,  $p \in \mathcal{P}$ , have the same set of leading monomials. Hence, we can apply the error tolerant lifting algorithm to the coefficients of the Gröbner bases in  $\mathcal{G}$ . If this algorithm returns `false` at some point, we enlarge the set  $\mathcal{P}$  by  $t$  primes not used so far, and repeat the whole process. Otherwise, the lifting yields a set of polynomials  $G \subset R$ . Furthermore, if  $\mathcal{P}$  is sufficiently large, all primes in  $\mathcal{P}$  satisfy condition (L2). Since we cannot check, however, whether  $\mathcal{P}$  is sufficiently large, a final verification step over  $\mathbb{Q}$  is required. We will establish such a test in Sect. 9 below. Since this test is particularly expensive if  $G \neq G(0)$ , we first perform a test in positive characteristic in order to increase our chances that the two sets are equal:

**PTEST:** *Randomly choose a prime  $p \notin \mathcal{P}$  which does neither divide the numerator nor the denominator of any coefficient occurring in any polynomial in  $G$ . Return `true` if  $G_p = G(p)$ , and `false` otherwise.*

If PTEST returns `false`, then  $\mathcal{P}$  is not sufficiently large (or the extra prime chosen in PTEST is bad). In this case, we enlarge  $\mathcal{P}$  as above and repeat the process. If PTEST returns `true`, however, then most likely  $G = G(0)$ . Only now, we verify the result over  $\mathbb{Q}$ . If the verification fails, we again enlarge  $\mathcal{P}$  and repeat the process.

## 9 Verification

Throughout this section, we consider a curve  $\Gamma \subset \mathbb{P}_{\mathbb{Q}}^r$  with Gorenstein adjoint ideal  $\mathfrak{G} = \mathfrak{G}(\Gamma) \subset R = \mathbb{Q}[X]$  as in Notation 8.4. Our goal is to derive a criterion which, in combination with the procedure PTEST from the previous section, provides an effective way of checking whether the result of our modular algorithm is correct. The verification is based on the following observation obtained from Lemma 3.14:

If  $I \subset R$  is a homogeneous ideal, then  $I = \mathfrak{G}$  iff the following hold:

1.  $I$  is saturated and  $I(\Gamma) \subsetneq I$ ,
2.  $\deg \Delta(I) = \deg I + \delta(\Gamma)$ , and
3.  $\deg I = \deg \mathfrak{G}$ .

To turn this into an algorithmic criterion, we need some preparations. If  $A$  is any reduced Noetherian algebra over a field  $k$ , then, as a direct generalization of the definition from Sect. 3, we can associate to  $A$  the delta invariant

$$\delta_k(A) = \dim_k \bar{A}/A.$$

**Proposition 9.1** *Let  $B \rightarrow A$  be a homomorphism of reduced Noetherian rings. Suppose:*

1.  $(B, \mathfrak{m})$  is a normal local domain with perfect residue class field  $k$ ;
2. the natural homomorphism  $B \rightarrow \widehat{B}$  from  $B$  to its completion  $\widehat{B}$  is normal;

3.  $A$  is a formally equidimensional Nagata ring;
4.  $A$  is a flat  $B$ -algebra,  $\mathfrak{m}A$  is contained in every maximal ideal of  $A$ , the ring  $A/\mathfrak{m}A$  is reduced, and  $\delta_k(A/\mathfrak{m}A) < \infty$ ;
5.  $\bar{A}/A$  is a finite  $B$ -module;
6. the unique map  $\overline{A/\mathfrak{m}A} \rightarrow \overline{A/\mathfrak{m}A}$  which factorizes the normalization map  $A/\mathfrak{m}A \rightarrow \bar{A}/\mathfrak{m}\bar{A}$  as

$$A/\mathfrak{m}A \rightarrow \overline{A/\mathfrak{m}A} \rightarrow \overline{A/\mathfrak{m}A}$$

is injective.

Then

$$\delta_{Q(B)}(A \otimes_B Q(B)) \leq \delta_k(A/\mathfrak{m}A).$$

*Proof* See [44, Prop. 2.1.1(i)] for the factorization in (6) and [44, Prop. 3.3] for the proof of the proposition.

**Corollary 9.2 (see also [15])** *With notation as above, let  $p$  be a prime such that  $I(\Gamma)_p$  is radical and defines an integral non-degenerate curve  $\Gamma_p \subset \mathbb{P}_{\mathbb{F}_p}^r$ . Then*

$$\delta(\Gamma) \leq \delta(\Gamma_p).$$

*Proof* Write  $X' = \{X_1, \dots, X_r\}$ . We may assume that  $\Gamma$  has no singularities at  $X_0 = 0$ . Let  $C$  be the affine part of  $\Gamma$  in the chart  $X_0 \neq 0$  as before. Write

$$J = I(C) \cap \mathbb{Z}[X'] \text{ and } I(C)_p = \langle f_p \mid f \in J \rangle \subset \mathbb{F}_p[X'].$$

Then  $J$  is a prime ideal of height  $r - 1$ ,  $\langle p, J \rangle$  is a prime ideal, and  $J \cap \mathbb{Z} = \langle 0 \rangle$ . The claim follows by applying Proposition 9.1 to  $(B, \mathfrak{m}) = (\mathbb{Z}_{\langle p \rangle}, \langle p \rangle)$  and  $A = \mathbb{Z}_{\langle p \rangle}[X']/J \mathbb{Z}_{\langle p \rangle}[X']$  since, in this case,  $A \otimes_B Q(B) = \mathbb{Q}[X']/I(C)$ ,  $A/\mathfrak{m}A = \mathbb{F}_p[X']/I(C)_p$ , and conditions (1) through (6) of the proposition are satisfied. Indeed, this is clear for (1), while (2) holds since  $B$  is excellent. We have (3) since  $A$  is of finite type over  $B$  and  $J \mathbb{Z}_{\langle p \rangle}[X']$  is a prime ideal. Moreover, (4) is satisfied since  $A$  is a torsion-free  $B$ -module,  $\langle p, J \rangle$  is a prime ideal, and  $\text{Spec}(A/\mathfrak{m}A)$  is a curve. We get (5) since  $A/\mathcal{C}_A$  is a finite  $B$ -module and  $\bar{A}/\mathcal{C}_A$  is a finite  $A/\mathcal{C}_A$ -module. Finally, condition (6) holds by Lemma 9.4 below: Taking into account that  $Q(A) = Q(\mathbb{Z}[X']/J)$ , the lemma gives us a canonical map

$$\bar{A} \rightarrow \overline{A/\mathfrak{m}A}, \alpha = \frac{\bar{a}}{\bar{b}} \mapsto \frac{a \bmod \langle p, J \rangle}{b \bmod \langle p, J \rangle},$$

where  $a, b \in \mathbb{Z}[X']$ , with  $b \notin \langle p, J \rangle$ , and where  $\bar{a}, \bar{b}$  denote the images of  $a, b$  in  $A$ . Since  $\alpha = \frac{\bar{a}}{\bar{b}}$  is in the kernel of this map iff  $a \in \langle p, J \rangle$ , we get an injective map  $\overline{A/\mathfrak{m}A} \rightarrow \overline{A/\mathfrak{m}A}$  which factorizes the normalization map as desired.

Before deriving Lemma 9.4, we illustrate condition (6) by an example.

*Example 9.3* Let  $(B, \mathfrak{m}) = (\mathbb{Z}_{(3)}, \langle 3 \rangle)$  and  $A = \mathbb{Z}_{(3)}[X, Y]/\langle X^3 + Y^3 + Y^5 \rangle$ . Then  $\overline{A/\mathfrak{m}A} = \left\langle 1, \frac{x}{y}, \frac{(x+y)^2}{y^3} \right\rangle_{A/\mathfrak{m}A}$  and  $\overline{A} = \left\langle 1, \frac{x}{y}, \frac{x^2}{y^2} \right\rangle_A$ . We compute  $\delta_{\mathbb{Q}}(A \otimes_B \mathbb{Q}) = 3$  and  $\delta_{\mathbb{F}_3}(A/\mathfrak{m}A) = 4$ , and find that

$$\overline{A/\mathfrak{m}A} = \left\langle 1, \frac{x}{y}, \frac{x^2}{y^2} \right\rangle_{A/\mathfrak{m}A} \subsetneq \left\langle 1, \frac{x}{y}, \frac{(x+y)^2}{y^3} \right\rangle_{A/\mathfrak{m}A} = \overline{A/\mathfrak{m}A}.$$

**Lemma 9.4** *With notation as in the proof of Corollary 9.2, for any  $\alpha \in \overline{A}$ , there exist  $a, b \in \mathbb{Z}[X']$  with  $b \notin \langle p, J \rangle$  and  $\alpha = \frac{a}{b} \in \mathbb{Q}(A) = \mathbb{Q}(\mathbb{Z}[X']/J)$ .*

*Proof* For  $\alpha \in \overline{A}$ , there are  $a, b \in \mathbb{Z}[X']$  with  $b \notin J$  and  $\alpha = \frac{a \bmod J}{b \bmod J}$ , and there are  $a_0, \dots, a_{m-1} \in \mathbb{Z}[X']$  and  $d \in \mathbb{Z}$  with  $p \nmid d$  and

$$\alpha^m + \frac{a_{m-1} \bmod J}{d} \alpha^{m-1} + \dots + \frac{a_0 \bmod J}{d} = 0.$$

Then  $d \cdot a^m + a_{m-1} \cdot ba^{m-1} + \dots + a_0 \cdot b^m \in J$ .

If  $b_0 = b \in \langle p, J \rangle$ , then  $d \cdot a^m \in \langle p, J \rangle$ . Hence, since  $(p, J)$  is prime,  $a \in \langle p, J \rangle$ . Then  $a = pa_1 + c_1$  and  $b = pb_1 + d_1$  for some  $a_1, b_1 \in \mathbb{Z}[X']$  and some  $c_1, d_1 \in J$ . If  $b_1 \in \langle p, J \rangle$ , we can iterate the process. Inductively, as long as  $b_{s-1} \in \langle p, J \rangle$ , we obtain  $a_s, b_s \in \mathbb{Z}[X']$  and  $c_s, d_s \in J$  with  $a = p^s a_s + c_s$  and  $b = p^s b_s + d_s$ . If  $b_s$  were in  $\langle p, J \rangle$  for all  $s$ , then  $b \in \bigcap_s \langle p^s, J \rangle = J$ , contradicting our assumption on  $b$ . Thus, there is an  $s$  with  $b_s \notin \langle p, J \rangle$ . Then

$$\alpha = \frac{a \bmod J}{b \bmod J} = \frac{p^s a_s \bmod J}{p^s b_s \bmod J} = \frac{a_s \bmod J}{b_s \bmod J}.$$

**Notation 9.5** Let  $I \subset R$  be a saturated homogeneous ideal such that  $I(\Gamma) \subsetneq I$ , let  $m$  be an integer, and let  $g \in I$  be a homogeneous polynomial of degree  $m$  not contained in  $I(\Gamma)$ . Let  $\text{div}(g)$  be the divisor cut out by  $g$  on  $\overline{\Gamma}$ , let  $D(g) = \text{div}(g) - \Delta(I)$  be the corresponding divisor in  $|mH - \Delta(I)|$ , and let  $d(g) = \deg D(g)$ . Furthermore, write  $\tilde{d}(g)$  for the degree of the part of  $D(g)$  away from  $\text{Sing}(\Gamma)$ .

Given a prime  $p$ , use the same notation for  $\Gamma_p$  and  $g_p$  if these are defined.

Note that  $\deg \text{div}(g) = m \cdot \deg \Gamma$  and  $\tilde{d}(g) \leq d(g)$ .

**Theorem 9.6** *Let  $I \subset R$  be a saturated homogeneous ideal such that  $I(\Gamma) \subsetneq I$ , let  $m$  be an integer, let  $g \in I$  be a homogeneous polynomial of degree  $m$  not contained in  $I(\Gamma)$ , and let  $p$  be a prime. With notation as above, suppose:*

1.  $I(\Gamma)_p$  is radical and defines an integral non-degenerate curve  $\Gamma_p \subset \mathbb{P}_{\mathbb{F}_p}^r$ ;
2.  $g_p$  is defined and non-zero;
3.  $\Gamma$  and  $\Gamma_p$  have the same Hilbert polynomial;
4.  $I_p$  is an adjoint ideal of  $\Gamma_p$ ;

5.  $\deg I = \deg I_p$ ;
6.  $m$  is large enough to ensure that  $|mH_p - \Delta(I_p)|$  is nonspecial;
7.  $\tilde{d}(g_p) = (\deg \Gamma) \cdot m - \deg I_p - \delta(\Gamma)$ .

Then

$$\delta(\Gamma) = \delta(\Gamma_p), \tilde{d}(g) = d(g), \text{ and } \deg \Delta(I) = \deg \Delta(I_p).$$

Moreover,  $I$  is an adjoint ideal of  $\Gamma$ , and we have

$$\deg \Delta(I) = (\deg \Gamma) \cdot m - \tilde{d}(g_p).$$

*Proof* By (3), we have

$$\deg(\Gamma) = \deg(\Gamma_p) \text{ and } p_a(\Gamma) = p_a(\Gamma_p). \quad (8)$$

Moreover, since (1) holds, it follows from Corollary 9.2 that  $\delta(\Gamma) \leq \delta(\Gamma_p)$ . Hence, taking (4) and Lemma 3.14 into account, we get

$$\begin{aligned} \tilde{d}(g_p) &\leq d(g_p) = (\deg \Gamma_p) \cdot m - \deg \Delta(I_p) \\ &= (\deg \Gamma) \cdot m - \deg I_p - \delta(\Gamma_p) \\ &\leq (\deg \Gamma) \cdot m - \deg I_p - \delta(\Gamma). \end{aligned}$$

By (6), this chain of inequalities is an equality, so that

$$\tilde{d}(g_p) = d(g_p) = (\deg \Gamma_p) \cdot m - \deg \Delta(I_p) \quad (9)$$

and

$$\delta(\Gamma) = \delta(\Gamma_p). \quad (10)$$

Together with Lemma 3.14 and conditions (4) and (5), this implies that

$$\deg \Delta(I_p) = \deg I_p + \delta(\Gamma_p) = \deg I + \delta(\Gamma) \geq \deg \Delta(I), \quad (11)$$

or equivalently that

$$d(g_p) \leq d(g). \quad (12)$$

Next, in the main part of the proof, we show equality in (12). For this, we consider the closed subscheme

$$X = V(I(\Gamma)_0) \subset \mathbb{P}'_{\mathbb{Z}} \xrightarrow{\pi} \text{Spec } \mathbb{Z}$$

with projection  $\Pi$  and fibers  $X_q = X \times_{\text{Spec } \mathbb{Z}} \text{Spec } \kappa(\langle q \rangle)$ . Then the fiber over the generic point  $\langle 0 \rangle \in \text{Spec } \mathbb{Z}$  is  $X_0 = \Gamma$ , while over  $\langle p \rangle$  we have  $X_p = \Gamma_p$ . Since  $\Gamma$  and  $\Gamma_p$  have the same Hilbert polynomial by (3), there is a Zariski open subset  $V \subset \text{Spec } \mathbb{Z}$  containing  $p$  and such that the Hilbert polynomial is constant on  $V$ . It follows that the restriction map  $\Pi_V : X_V = \Pi^{-1}(V) \rightarrow V$  constitutes a flat family (see [37, Ch. III, Thm. 9.9]).

Since  $\delta(\Gamma_p) = \delta(\Gamma)$ , the  $\delta$ -constant criterion for simultaneous normalization (see [44, Cor. 3.3.1]) implies that there is a Zariski open subset  $U \subset V \subset \text{Spec } \mathbb{Z}$  containing  $p$  and such that  $\pi_U : X_U = \Pi^{-1}(U) \rightarrow U$  is equinormalizable. That is, there is a finite map  $\nu : \bar{X} \rightarrow X_U$  such that  $\bar{\Pi} := \pi_U \circ \nu$  is flat with non-empty geometrically normal fibers, and such that for each  $\langle q \rangle \in U$  the induced map of fibers  $\nu_q : \bar{X}_q \rightarrow X_q$  is the normalization map.

By construction of  $I$ , the family of sheaves defined by  $I_0 = I \cap \mathbb{Z}[X]$  is flat over a Zariski open subset of  $U$  containing both  $\langle 0 \rangle$  and  $\langle p \rangle$ . Hence, the semi-continuity theorem (see, for example, [45, Ch. 5, Thm. 3.20]) implies that the dimensions of the linear series induced by  $I$  on  $\bar{\Gamma}$  and  $I_p$  on  $\bar{\Gamma}_p$  satisfy

$$h^0(\bar{\Gamma}_p, \mathcal{O}_{\bar{\Gamma}_p}(mH_p - \Delta(I_p))) \geq h^0(\bar{\Gamma}, \mathcal{O}_{\bar{\Gamma}}(mH - \Delta(I))).$$

Hence  $d(g_p) \geq d(g)$  by condition (7) and Riemann-Roch, and since  $p(\Gamma_p) = p(\Gamma)$  by (8) and (10). Taking (9) and (12) into account, it follows that

$$\tilde{d}(g_p) = d(g_p) = d(g).$$

The second equality translates into  $\deg \Delta(I_p) = \deg \Delta(I)$ , so that  $I$  is an adjoint ideal by Lemma 3.14 and (11). Finally,

$$(\deg \Gamma) \cdot m - \deg \Delta(I) = (\deg \Gamma_p) \cdot m - \deg \Delta(I_p) = \tilde{d}(g_p).$$

**Corollary 9.7** *In the situation of Theorem 9.6, suppose that all assumptions of the theorem are satisfied. Suppose in addition that  $I_p$  is the Gorenstein adjoint ideal of  $\Gamma_p$ . Then  $I$  is the Gorenstein adjoint ideal of  $\Gamma$ .*

*Proof* Theorem 9.6 already tells us that  $I$  is an adjoint ideal of  $\Gamma$ . In particular,  $I \subset \mathfrak{G}$  and, thus,  $\deg I \geq \deg \mathfrak{G}$ . This implies that

$$\deg \Delta(\mathfrak{G}) = \deg \mathfrak{G} + \delta(\Gamma) \leq \deg I + \delta(\Gamma) = \deg \Delta(I) = \deg \Delta(I_p), \quad (13)$$

where the last equality holds by the theorem. On the other hand, since we suppose that  $I_p$  is the Gorenstein adjoint ideal of  $\Gamma_p$ , we have

$$\dim |mH_p - \Delta(I_p)| \geq \dim |mH - \Delta(\mathfrak{G})|$$

for  $m$  large enough by semi-continuity. Hence, by Riemann-Roch and since  $\delta(\Gamma_p) = \delta(\Gamma)$  by the theorem, we have

$$\deg \Delta(I_p) \leq \deg \Delta(\mathfrak{G})$$

(see Lemma 3.14 and its proof). This shows that (13) is an equality. In particular,

$$\deg I = \deg \mathfrak{G}.$$

We conclude that  $I = \mathfrak{G}$ .

*Remark 9.8* In the final verification step of our modular algorithm for computing  $\mathfrak{G}$ , if  $G$  denotes the result of the lifting process as in the previous section, we randomly choose one of the primes  $p \in \mathcal{P}$  already used in the lifting process, and apply Theorem 9.6 to the ideal  $I = \langle G \rangle$ . For this, we need to know whether the assumptions of the theorem hold. Checking condition (2) is trivial, while conditions (1) and (3) are fulfilled by construction (see Remark 8.5 and step 6 of Algorithm 6 below, where we in particular check that  $\text{LM}(I(\Gamma)) = \text{LM}(I(\Gamma)_p)$ ). Similarly conditions (4) and (5) are fulfilled since by construction  $G_p = G(p)$  and thus  $I_p = \mathfrak{G}(\Gamma_p)$ . With respect to condition (6), we will comment on how to choose  $m$  in Lemma 9.9 below. Finally, since we know how to compute  $\delta(\Gamma)$ , we can also check condition (7) (see step 18 of Algorithm 6).

In the situation above, if all assumptions of Theorem 9.6 are fulfilled, then by the assertions of the theorem, we may rewrite the formula in condition (7) as

$$\tilde{d}(g) = \deg(\Gamma) \cdot m - \deg \Delta(\mathfrak{G}). \quad (14)$$

So in order to expect that condition (7) holds for a given  $m$  and randomly chosen  $g \in I_m$  and  $p \in \mathcal{P}$ , the degree  $m$  needs to be large enough so that Eq. (14) is satisfied. The following lemma specifies an appropriate bound for  $m$ , which is also sufficient to guarantee that condition (6) is fulfilled.

**Lemma 9.9** *Consider an integer  $m$  such that  $P_\Gamma(m) - 1 \geq p_a(\Gamma)$ , and suppose that  $g \in \mathfrak{G}_m$  is generic. Then (14) is satisfied, and  $|mH - \Delta(\mathfrak{G})|$  is nonspecial.*

*Proof* By assumption and since  $P_\Gamma(m) = (\deg \Gamma) \cdot m - p_a(\Gamma) + 1$ , we have

$$(\deg \Gamma) \cdot m \geq 2p_a(\Gamma).$$

On the other hand, by Lemma 3.14 and Corollary 3.20,

$$\deg \Delta(\mathfrak{G}) \leq 2\delta(\Gamma).$$

Putting these inequalities together, we get

$$(\deg \Gamma) \cdot m - \deg \Delta(\mathfrak{G}) \geq 2p_a(\Gamma) - 2\delta(\Gamma) = 2p(\Gamma).$$



**Algorithm 6** Modular adjoint ideal**Input:** A curve  $\Gamma \subset \mathbb{P}_{\mathbb{Q}}^r$  satisfying the conditions of Notation 8.4.**Output:** The Gorenstein adjoint ideal  $\mathfrak{G}(\Gamma)$ .

---

```

1: choose an integer  $t \geq 1$ 
2:  $\mathcal{P} = \mathcal{G} = \emptyset$ 
3: loop
4:   choose a list  $\mathcal{Q}$  of  $t$  random primes not used so far
5:   for all  $p \in \mathcal{Q}$  do
6:     if  $I(\Gamma)_p$  satisfies the conditions of Notation 8.4 then
7:       compute the reduced Gröbner basis  $G(p)$  of  $\mathfrak{G}(\Gamma_p) \subset R_p$  (via Alg. 3)
8:        $\mathcal{P} = \mathcal{P} \cup \{p\}$ ,  $\mathcal{G} = \mathcal{G} \cup \{G(p)\}$ 
9:      $(\mathcal{G}, \mathcal{P}) = \text{DELETEBYMAJORITYVOTE}(\mathcal{G}, \mathcal{P})$ 
10:    lift  $(\mathcal{G}, \mathcal{P})$  to a set of polynomials  $G \subset R$  via the Chinese remainder theorem
    and Gaussian reduction
11:    if the lifting succeeds and  $\text{PTTEST}(I(\Gamma), G, \mathcal{P})$  then
12:      if  $G$  is a Gröbner basis and  $\langle G \rangle$  is saturated then
13:        choose  $m$  such that  $P_{\Gamma}(m) - 1 \geq p_a(\Gamma)$ 
14:        choose  $g \in \langle G \rangle_m$  at random
15:        choose a prime  $p \in \mathcal{P}$ 
16:        if  $g_p$  is defined and non-zero then
17:           $M_p = \text{Jacobian ideal of } I(\Gamma)_p$ 
18:          compute  $\tilde{d}(g_p) = \deg((I(\Gamma_p) + \langle g_p \rangle) : M_p^{\infty})$ 
19:          compute  $\delta(\Gamma)$ 
20:          if  $\tilde{d}(g_p) = \deg(\Gamma) \cdot m - \deg \langle G(p) \rangle - \delta(\Gamma)$  then
21:            return  $\langle G \rangle$ 

```

---

In particular,  $|mH - \Delta(\mathfrak{G})|$  is base-point free, which implies that  $d(g) = \tilde{d}(g)$  since  $g$  is generic. Furthermore, by reason of its degree,  $|mH - \Delta(\mathfrak{G})|$  is nonspecial.

*Remark 9.10* For a plane curve  $\Gamma$  of degree  $n$  the condition  $P_{\Gamma}(m) - 1 \geq p_a(\Gamma)$  means that  $n \cdot m \geq (n-1)(n-2)$ , which is satisfied for  $m \geq n-2$ .

We summarize our approach in Algorithm 6.

*Remark 9.11* In Algorithm 6, the  $G(p)$ ,  $p \in \mathcal{P}$ , can be computed in parallel. Each individual computation, in turn, can be parallelized by partitioning the singular loci.

*Remark 9.12* The most expensive step of the verification is the computation of  $\delta(\Gamma)$ . If we skip the verification, the algorithm will become probabilistic. That is, the output can only be expected to be the Gorenstein adjoint ideal, with high probability. Skipping the verification usually accelerates the algorithm considerably. This gives us, in particular, a fast probabilistic way to compute both the geometric genus  $\rho(\Gamma)$  and  $\deg \Delta(\mathfrak{G}) = \dim_{\mathbb{Q}}(\overline{\mathbb{Q}[C]}/\mathcal{C}_{\mathbb{Q}[C]})$ .

## 10 Timings

The algorithms for adjoint ideals presented in this paper are implemented in the SINGULAR library `adjointideal.lib` (see [12]). They make use of the normalization algorithm of Sect. 2 either in its local or local to global variant, as appropriate. These variants, in turn, are part of the SINGULAR library `locnormal.lib` (see [7]).

In this section, we compare the performance of the different algorithms. Specifically, we consider

LA	Mnuk's global linear algebra approach (Algorithm 1),
IQ	the global ideal quotient approach (Algorithm 2),
locIQ	the local ideal quotient approach (Algorithm 3 using Algorithm 4),
locIQP2	the local ideal quotient approach for plane curves with the improvements of Sect. 7 concerning ordinary multiple points and singularities of type <i>ADE</i> , and
modLocIQ	the modular local ideal quotient strategy (Algorithm 6).

For the modular approach, we do not make use of a local analysis of the singular locus except for computing the invariants needed in the verification step.

To quantify the improvement in computation time obtained by omitting the verification step in the modular approach, we give timings for the *resulting, now probabilistic, version of Algorithm 6* (denoted by `modLocIQ'` in the tables). Note that in all examples where we could check the output of the modular algorithm by computing the desired Gröbner basis also directly over  $\mathbb{Q}$ , the result was indeed correct.

To quantify the contributions of the different normalization algorithms and to provide a lower bound for any adjoint ideal algorithm using them, we also specify the computation times for the normalization step in SINGULAR via the local to global approach outlined in Sect. 2 (denoted by `locNormal`), and for finding an integral basis in MAPLE [46] via the algorithm of van Hoeij (denoted by `Maple-IB`). Once being fully implemented in SINGULAR, we expect further improvements of the performance by computing the local contribution or just an integral basis of the local ring by the algorithm discussed in [10]. Since this algorithm and van Hoeij's algorithm rely on Puiseux series, they work in characteristic zero only.

All timings are in seconds on an AMD Opteron 6174 machine with 48 cores, 2.2 GHz, and 128 GB of RAM running a Linux operating system. A dash indicates that the computation did not finish within 10,000s. The *timings for parallel computations are marked by the symbol \* and the maximum number of cores used in parallel is indicated in brackets*.

*Remark 10.1* All examples are defined over the field of rationals. For `locIQ*`, the number of cores used corresponds to the number of components of the

decomposition of the singular locus over  $\mathbb{Q}$ . For  $\text{modLocIQ}^*$ , the number of cores used in a given iteration of the algorithm is obtained by summing up the number of components modulo  $p$  over all primes  $p \in \mathcal{Q}$  chosen in Step 4 of Algorithm 6.

To show the power of the modular algorithm, we give simulated parallel timings even if the number of processes exceeds the number of cores available on our machine (which is a valid approach since the algorithm has basically zero communication overhead). For the single-core timings of  $\text{modLocIQ}$ , we indicate in square brackets the number of primes used by the algorithm.

Now we turn to explicit examples. First we consider rational plane curves defined by a random parametrization of degree  $n$ . These curves have  $\binom{n-1}{2}$  ordinary double points. Their defining equations  $f_{1,n}$  were generated by the function `randomRatCurve` from the SINGULAR library `paraplanecurves.lib` (see [8]), using the random seed 1 and a random parametrization with coefficients of bitlength 15. For the resulting timings, see Table 1.

We observe that the detection of special types of singularities is fast and yields the best performance among the non-probabilistic algorithms, while the modular local strategy provides a very fast probabilistic algorithm.

To compare the algorithms at a single singularity, we consider plane curves with exactly one  $A_n$  respectively  $D_n$  singularity at the origin of the affine chart  $\{Z \neq 0\}$  (ignoring singularities at infinity). For the modular approach, we omit verification since this step relies on global properties of the curve. The curves with affine equation  $f_{2,n,d} = Y^2 + X^{n+1} + Y^d$ ,  $n \geq 1, d \geq 3$ , have precisely one singularity of type  $A_n$  at the origin. The curves with affine equation  $f_{3,n,d} = X(X^{n-1} + Y^2) + Y^d$ ,  $n \geq 2, d \geq 3$ , have exactly one singularity of type  $D_n$  at the origin. For timings, see Tables 2 and 3, respectively.

In both examples, the best strategy is IQ since we consider only one singularity and since no coefficients of large bitlength occur.

**Table 1** Timings for curves given by a random parametrization

	$f_{1,5}$	$f_{1,6}$	$f_{1,7}$
deg	5	6	7
locNormal	2.1	56	–
Maple-IB	5.1	47	318
LA	98	4400	–
IQ	2.1	56	–
locIQ	1.3	54	3800
locIQ*	1.3 (1)	54 (1)	3800 (1)
locIQP2	0.18	1.2	49
locIQP2*	0.18 (1)	1.2 (1)	49 (1)
modLocIQ	6.4 [33]	19 [53]	150 [75]
modLocIQ'	6.2 [33]	18 [53]	104 [75]
modLocIQ*	0.36 (74)	1.6(153)	51(230)
modLocIQ' *	0.21 (74)	0.48(153)	5.2 (230)

**Table 2** Timings for curves with one singularity of type  $A_n$

	$f_{2.5,10}$	$f_{2.5,100}$	$f_{2.5,500}$	$f_{2.50,100}$	$f_{2.50,500}$	$f_{2,400,500}$
deg	10	100	500	100	500	500
locNormal	0.12	0.12	0.12	0.51	0.51	3.6
Maple-IB	0.08	1.5	96	4.7	150	630
LA	0.18	140	–	150	–	–
IQ	0.12	0.12	0.12	0.51	0.51	3.6
modLocIQ'	0.20 [2]	0.22 [2]	0.96 [2]	1.1 [2]	2.0 [2]	11 [2]
modLocIQ'*	0.10 (2)	0.13 (2)	0.48 (2)	0.54 (2)	1.2 (2)	5.8 (2)

**Table 3** Timings for curves with one singularity of type  $D_n$

	$f_{3.5,10}$	$f_{3.5,100}$	$f_{3.5,500}$	$f_{3,50,100}$	$f_{3,50,500}$	$f_{3,400,500}$
deg	10	100	500	100	50	500
locNormal	0.15	0.15	0.15	0.67	0.67	4.9
Maple-IB	0.05	1.7	100	34	1830	–
LA	0.20	140	–	140	–	–
IQ	0.15	0.15	0.15	0.67	0.67	5.0
modLocIQ'	0.22 [2]	0.23 [2]	0.23 [2]	1.5 [2]	1.5 [2]	24 [2]
modLocIQ'*	0.09 (2)	0.10 (2)	0.10 (2)	0.74 (2)	0.77 (2)	17 (2)

**Table 4** Timings for curves with many  $A_n$ -singularities

	$f_{4,4}$	$f_{4,6}$	$f_{4,8}$
deg	10	14	18
locNormal	1.6	–	–
Maple-IB	2.2	14	70
LA	89	–	–
IQ	2.5	–	–
locIQ	0.96	–	–
locIQ*	0.36 (6)	–	–
locIQP2	1.0	–	–
locIQP2*	0.38 (6)	–	–
modLocIQ	3.7 [3]	23 [4]	190 [4]
modLocIQ'	3.3 [3]	20 [4]	170 [4]
modLocIQ*	0.63 (27)	4.4 (48)	50 (48)
modLocIQ'*	0.38 (27)	2.2 (48)	30 (48)

The plane curves with defining equations

$$f_{4,n} = (X^{n+1} + Y^{n+1} + Z^{n+1})^2 - 4(X^{n+1}Y^{n+1} + Y^{n+1}Z^{n+1} + Z^{n+1}X^{n+1})$$

were given in [38] and have  $3(n + 1)$  singularities of type  $A_n$  if  $n$  is even. To ensure that all singularities of the curves are in the affine chart  $\{Z \neq 0\}$ , we substitute  $Z = 2X - 3Y + 1$ . For timings, see Table 4.

**Table 5** Timings for non-planar curves

	$L_{25}$	$L_{50}$	$I_4$	$I_6$
deg	25	50	20	28
locNormal	3.9	84	21	–
IQ	3.9	84	30	–
locIQ	3.9	84	18	–
locIQ*	3.9 (1)	84 (1)	7.5 (6)	–
modLocIQ'	6.5 [2]	220 [2]	74 [5]	2600 [5]
modLocIQ'*	3.3 (2)	140 (2)	4.0 (45)	59 (69)

To conclude this section, we present examples of curves in higher-dimensional projective space. As above, we first consider curves with only one singularity in a given affine chart: let  $L_n$  be the ideal of the image of

$$\mathbb{A}^1 \longrightarrow \mathbb{A}^3, t \mapsto (t^{n-2}, t^{n-1}, t^n).$$

Second, denote by  $I_n$  the ideal of the image in  $\mathbb{P}^5$  under the degree-2 Veronese embedding of the curve  $\{f_{4,n} = 0\}$ . For the resulting timings, see Table 5.

To summarize, we observe that the ideal quotient approach is faster than the linear algebra one. To some extent, this is due to the lack of efficiency of the rational function arithmetic in SINGULAR. The local strategy is faster than the global one if there is more than one component in the decomposition of the singular locus over  $\mathbb{Q}$ . In addition, the local algorithm can be run in parallel and is, then, even faster. In most examples, especially when the coefficients have large bitlength, the fastest approach is the modular local strategy, which parallelizes in a twofold way, via localization and modularization. Note that, even if the singular locus of the curve is irreducible over the rationals, by Chebotarev’s density theorem the singular locus is likely to decompose when passing to a finite field (see, for example,  $f_{1,7}$ ). In contrast to other modular algorithms (such as modular normalization), the verification step is usually very fast.

**Acknowledgements** We would like to thank Gert-Martin Greuel, Christoph Lossen, Thomas Markwig, Mathias Schulze, and Frank Seelisch for helpful discussions.

## References

1. E. Arbarello, C. Ciliberto, Adjoint hypersurfaces to curves in  $\mathbb{P}^r$  following Petri, in *Commutative Algebra*. Lecture Notes in Pure and Applied Mathematics, vol. 84 (Dekker, New York, 1983), pp. 1–21
2. E. Arbarello, M. Cornalba, P.A. Griffiths, J. Harris, *Geometry of Algebraic Curves*, vol. I (Springer, Berlin, 1985)
3. E.A. Arnold, Modular algorithms for computing Gröbner bases. *J. Symb. Comput.* **35**, 403–419 (2003)

4. V.I. Arnold, S.M. Gusein-Zade, A.N. Varchenko, *Singularities of Differential Maps*, vol. I (Birkhäuser, Basel, 1995)
5. J. Böhm, Parametrisierung rationaler Kurven. Diploma thesis, Institut für Mathematik und Physik der Universität Bayreuth, 1999
6. J. Böhm, W. Decker, S. Laplagne, G. Pfister, A. Steenpaß, S. Steidel, Parallel algorithms for normalization. *J. Symb. Comput.* **51**, 99–114 (2013)
7. J. Böhm, W. Decker, S. Laplagne, G. Pfister, A. Steenpaß, S. Steidel, locnormal.lib - a SINGULAR 4-1-0 library for computing integral bases of algebraic function fields. SINGULAR distribution. <http://www.singular.uni-kl.de>
8. J. Böhm, W. Decker, S. Laplagne, F. Seelisch, paraplancurves.lib - a SINGULAR 4-1-0 library for computing parametrizations of rational curves. SINGULAR distribution. <http://www.singular.uni-kl.de>
9. J. Böhm, W. Decker, M. Schulze, Local analysis of Grauert-Remmert-type normalization algorithms. *Int. J. Algebra Comput.* **24**(1), 69–94 (2014)
10. J. Böhm, W. Decker, S. Laplagne, G. Pfister, Computing integral bases via localization and Hensel lifting (2015). <http://arxiv.org/abs/1505.05054>
11. J. Böhm, W. Decker, C. Fieker, G. Pfister, The use of bad primes in rational reconstruction. *Math. Comput.* **84**, 3013–3027 (2015)
12. J. Böhm, W. Decker, S. Laplagne, F. Seelisch, adjointideal.lib - a SINGULAR 4-1-0 library for computing adjoint ideals of curves. SINGULAR distribution. <http://www.singular.uni-kl.de>
13. N. Brieskorn, *Plane Algebraic Curves* (Birkhäuser, Basel, 1986)
14. A. Brill, M. Noether, Über die algebraischen Functionen und ihre Anwendung in der Geometrie. *Math. Ann.* **7**, 269–310 (1874)
15. R. Buchweitz, G.-M. Greuel, The Milnor number and deformations of complex curve singularities. *Invent. Math.* **58**, 241–281 (1980)
16. G. Castelnuovo, Massima dimensione dei sistemi lineari di curve piane di dato genere. *Ann. Mat. (2)* **18**, 119–128 (1890)
17. G. Castelnuovo, Sui multipli di una serie lineare di gruppi di punti appartenenti ad una curva algebrica. *Rend. Circ. Mat. Palermo* **7**, 89–110 (1893)
18. N. Chiarli, Deficiency of linear series on the normalization of a space curve. *Commun. Algebra* **12**, 2231–2242 (1984)
19. C. Ciliberto, F. Orecchia, Adjoint ideals to projective curves are locally extended ideals. *Bollettino U.M.I. (6)* **3-B**, 39–52 (1984)
20. W. Decker, G.-M. Greuel, G. Pfister, T. de Jong, The normalization: a new algorithm, implementation and comparisons, in *Computational Methods for Representations of Groups and Algebras (Essen, 1997)* (Birkhäuser, Basel, 1999)
21. W. Decker, G.-M. Greuel, G. Pfister, H. Schönemann, SINGULAR 4-1-0 — a computer algebra system for polynomial computations. <http://www.singular.uni-kl.de>
22. T. De Jong, An algorithm for computing the integral closure. *J. Symb. Comput.* **26**, 273–277 (1998)
23. T. De Jong, G. Pfister, *Local Analytic Geometry* (Vieweg, Braunschweig, 2000)
24. J. Dieudonné, *Topics in Local Algebra*. Notre Dame Mathematical Lectures (University of Notre Dame Press, Notre Dame, 1967)
25. D. Eisenbud, *Commutative Algebra with a View Toward Algebraic Geometry* (Springer, Berlin, 1995)
26. M. El Kahoui, Z.Y. Moussa, An algorithm to compute the adjoint ideal of an affine plane curve. *Math. Comput. Sci.* **8**, 289–298 (2014)
27. D. Gorenstein, An arithmetic theory of adjoint plane curves. *Trans. Am. Math. Soc.* **72**, 414–436 (1952)
28. H. Grauert, R. Remmert, *Analytische Stellenalgebren*. Unter Mitarbeit von O. Riemenschneider, Die Grundlehren der mathematischen Wissenschaften, Band 176 (Springer, Berlin, 1971)
29. S. Greco, P. Valabrega, On the theory of adjoints, in *Algebraic Geometry*. Lecture Notes in Mathematics, vol. 732 (Springer, Berlin, 1979), pp. 99–123

30. S. Greco, P. Valabrega, On the theory of adjoints II. Rendiconti del Circolo Matematico di Palermo, Serie II, Tomo **XXXI**, 5–15 (1982)
31. G.-M. Greuel, *On Deformations of Curves and a Formula of Deligne*. Algebraic Geometry (La Rábida 1981). Lecture Notes in Mathematics, vol. 961 (Springer, Berlin, 1982)
32. G.-M. Greuel, G. Pfister, *A Singular Introduction to Commutative Algebra* (Springer, Berlin, 2008)
33. G.-M. Greuel, C. Lossen, E. Shustin, *Introduction to Singularities and Deformations* (Springer, Berlin, 2007)
34. G.-M. Greuel, S. Laplagne, F. Seelisch, Normalization of rings. J. Symb. Comput. **45**(9), 887–901 (2010)
35. G.-M. Greuel, S. Laplagne, G. Pfister, *normal.lib – a SINGULAR library for computing the normalization of affine rings*. SINGULAR distribution, <http://www.singular.uni-kl.de>
36. W. Gröbner, *Idealtheoretischer Aufbau der algebraischen Geometrie, Teil I* (Teubner, Leipzig, 1941)
37. R. Hartshorne, *Algebraic Geometry* (Springer, Berlin, 1977)
38. A. Hirano, Construction of plane curves with cusps. Saitama Math. J. **10**, 21–24 (1992)
39. H. Hironaka, On the arithmetic genera and the effective genera of algebraic curves. Mem. College Sci. Univ. Kyoto Ser. A Math. **30**(2), 177–195 (1957)
40. N. Idrees, G. Pfister, S. Steidel, Parallelization of modular algorithms. J. Symb. Comput. **46**, 672–684 (2011)
41. O. Keller, Die verschiedenen Definitionen des adjungierten Ideals einer ebenen algebraischen Kurve. Math. Ann. **159**, 130–144 (1965)
42. O. Keller, *Vorlesungen über algebraische Geometrie* (Akademische Verlagsgesellschaft, Leipzig, 1974)
43. D. Le Brigand, J.J. Risler, Algorithme de Brill-Noether et codes de Goppa. Bulletin de la S. M. F. **116**, 231–253 (1988)
44. J. Lipman, A numerical criterion for simultaneous normalization. Duke Math. J. **133**(2), 347–390 (2006)
45. Q. Liu, *Algebraic Geometry and Arithmetic Curves* (Oxford University Press, Oxford, 2002)
46. MAPLE (Waterloo MAPLE Inc.), MAPLE (2012). <http://www.maplesoft.com/>
47. E. Matlis, *1-Dimensional Cohen-Macaulay Rings*. Lecture Notes in Mathematics, vol. 327 (Springer, Berlin, 1970)
48. J.S. Milne, *Étale Cohomology* (Princeton University Press, Princeton, NJ, 1980)
49. T. Milnor, *Singular Points of Complex Hypersurfaces*. Annals of Mathematics Studies, vol. 61 (Princeton University Press, Princeton, NJ, 1968)
50. M. Mnuk, An algebraic approach to computing adjoint curves. J. Symb. Comput. **23**(2–3), 229–240 (1997)
51. F. Orecchia, I. Ramella, On the computation of the adjoint ideal of curves with ordinary singularities. Appl. Math. Sci. **8**(136), 6805–6812 (2014)
52. K. Petri, Über Spezialkurven I. Math. Ann. **93**, 182–209 (1924)
53. R. Riemann, Theorie der Abel’schen Functionen. J. Reine Angew. Math. **54**(14), 115–155 (1857)
54. J.R. Sendra, F. Winkler, Parametrization of algebraic curves over optimal field extensions. Parametric algebraic curves and applications (Albuquerque, NM, 1995). J. Symb. Comput. **23**(2–3), 191–207 (1997)
55. J.R. Sendra, F. Winkler, S. Perez-Díaz, *Rational Algebraic Curves*. Algorithms and Computation in Mathematics, vol. 22 (Springer, Berlin, 2008)
56. I.R. Shafarevich, *Algebraic Geometry I* (Springer, Berlin, 1994)
57. I. Swanson, C. Huneke, *Integral Closure of Ideals, Rings, and Modules* (Cambridge University Press, Cambridge, 2006)
58. N. Tschebotareff (Chebotarev), Die Bestimmung der Dichtigkeit einer Menge von Primzahlen, welche zu einer gegebenen Substitutionsklasse gehören. Math. Ann. **95**, 191–228 (1925)
59. B.L. van der Waerden, *Einführung in die algebraische Geometrie*. Die Grundlehren der Mathematischen Wissenschaften (Vieweg, Braunschweig, 1939)

60. M. van Hoeij, An algorithm for computing an integral basis in an algebraic function field. *J. Symb. Comput.* **18**(4), 353–363 (1994)
61. O. Zariski, P. Samuel, *Commutative Algebra I* (Springer, Berlin, 1975)



# Picard Curves with Small Conductor



Michel Börner, Irene I. Bouw, and Stefan Wewers

**Abstract** We study the conductor of Picard curves over  $\mathbb{Q}$ , which is a product of local factors. Our results are based on previous results on stable reduction of superelliptic curves that allow one to compute the conductor exponent  $f_p$  at the primes  $p$  of bad reduction. A careful analysis of the possibilities of the stable reduction at  $p$  yields restrictions on the conductor exponent  $f_p$ . We prove that Picard curves over  $\mathbb{Q}$  always have bad reduction at  $p = 3$ , with  $f_3 \geq 4$ . As an application we discuss the question of finding Picard curves with small conductor.

**Keywords** Picard curves • Conductor • Semistable reduction

**Subject Classifications** Primary 14H25. Secondary: 11G30, 14H45

## 1 Introduction

Let  $Y$  be a smooth projective curve of genus  $g$  over a number field  $K$ . To simplify the exposition, let us assume that  $K = \mathbb{Q}$ . With  $Y$  we can associate an  $L$ -function  $L(Y, s)$  and a conductor  $N_Y \in \mathbb{N}$ . Conjecturally, the  $L$ -function satisfies a functional equation of the form

$$\Lambda(Y, s) = \pm \Lambda(Y, 2 - s),$$

where

$$\Lambda(Y, s) := \sqrt{N_Y}^{-s} \cdot (2\pi)^{-gs} \cdot \Gamma(s)^g \cdot L(Y, s).$$

---

M. Börner • I.I. Bouw • S. Wewers (✉)  
Institut für Reine Mathematik, Universität Ulm, Germany  
e-mail: [irene.bouw@uni-ulm.de](mailto:irene.bouw@uni-ulm.de); [stefan.wewers@uni-ulm.de](mailto:stefan.wewers@uni-ulm.de)

© Springer International Publishing AG, part of Springer Nature 2017  
G. Böckle et al. (eds.), *Algorithmic and Experimental Methods  
in Algebra, Geometry, and Number Theory*,  
[https://doi.org/10.1007/978-3-319-70566-8\\_4](https://doi.org/10.1007/978-3-319-70566-8_4)

By definition, both  $L(Y, s)$  and  $N_Y$  are a product of local factors. In this paper we are really only concerned with the conductor, which can be written as

$$N_Y = \prod_p p^{f_p}.$$

The exponent  $f_p$  is called the *conductor exponent* of  $Y$  at  $p$ . It is known that  $f_p$  only depends on the ramification of the local Galois representation associated with  $Y$ . In particular, if  $Y$  has good reduction at  $p$  then  $f_p = 0$ . If  $Y$  has bad reduction at  $p$  then the computation of  $f_p$  can be quite difficult. Until recently, an effective method for computing  $f_p$  was only known for elliptic curves [23, §IV.10] and for genus 2 curves if  $p \neq 2$  [10].

It was shown in [3] that  $f_p$  can effectively be computed from the *stable reduction* of  $Y$  at  $p$ . Moreover, for certain families of curves (the *superelliptic curves*) we gave a rather simple recipe for computing the stable reduction. The latter result needed the assumption that  $p$  does not divide the degree  $n$ . In [18] this restriction is removed for superelliptic curves of prime degree.

In the present paper we systematically study the case of *Picard curves*. These are superelliptic curves of genus 3 and degree 3, given by an equation of the form

$$Y : y^3 = f(x) = x^4 + a_3x^3 + a_2x^2 + a_1x + a_0,$$

with  $f \in \mathbb{Q}[x]$  separable. Picard curves form in some sense the next family of curves to study after hyperelliptic curves. They are interesting for many reasons and have been intensively studied, see e.g. [8, 9, 13, 15].

Our main results classify all possible configurations for the stable reduction of a Picard curve at a prime  $p$ , and use this to determine restrictions on the conductor exponents. For instance, we prove the following.

**Theorem 1.1** *Let  $Y$  be a Picard curve over  $\mathbb{Q}$ .*

- (a) *Then  $Y$  has bad reduction at  $p = 3$ , and  $f_3 \geq 4$ .*
- (b) *For  $p = 2$  we have  $f_2 \neq 1$ .*
- (c) *For  $p \geq 5$  we have  $f_p \in \{0, 2, 4, 6\}$ .*

Theorem 3.6 is a somewhat stronger version of the first statement. Theorem 4.4 contains the last two statements. We also give explicit examples, showing that at least part of our results are sharp. Our result can be seen as a complement, for Picard curves, to a result of Brumer–Kramer [4, Theorem 6.2], who prove an upper bound for  $f_p$  for abelian varieties of fixed dimension. Since the conductor of a curve coincides with that of its Jacobian, the result applies to our situation, as well. A more careful case-by-case analysis, combined with ideas from [4], could probably be used to obtain a more precise list of possible values for the conductor exponent at  $p = 2, 3$ , as well.

In the last section we discuss the problem of constructing Picard curves with small conductor. As a consequence of the Shafarevich conjecture (aka Faltings’

Theorem), there are at most a finite number of nonisomorphic curves of given genus and of bounded conductor. But except in very special cases, no effective proof of this theorem is known.

In his recent PhD thesis, the first named author has made an extensive search for Picard curves with good reduction outside a small set of small primes, and computed their conductor. The Picard curve with the smallest conductor that was found is the curve

$$Y : y^3 = x^4 - 1,$$

which has conductor

$$N_Y = 2^6 3^6 = 46656.$$

We propose as a subject for further research to either prove that the above example is the Picard curve over  $\mathbb{Q}$  with the smallest possible conductor, or to find (one or all) counterexamples. We believe that the methods presented in this paper may be very helpful to achieve this goal.

## 2 Semistable Reduction

We first introduce the general setup concerning the stable reduction and the conductor exponents of Picard curves. As explained in the introduction, the conductor exponent is a local invariant, encoding information about the ramification of the local Galois representation associated with the curve. Therefore, we may replace the number field  $K$  by its strict henselization. In other words, we may work from the start over a henselian field of mixed characteristic with algebraically closed residue field.

### 2.1 Setup and Notation

Throughout Sects. 2–4 the letter  $K$  will denote a field of characteristic zero that is henselian with respect to a discrete valuation. We denote the valuation ring by  $\mathcal{O}_K$ , the maximal ideal of  $\mathcal{O}_K$  by  $\mathfrak{p}$  and the residue field by  $k = \mathcal{O}_K/\mathfrak{p}$ . We assume that  $k$  is algebraically closed of characteristic  $p > 0$ . The most important example for us is when  $K = \mathbb{Q}_p^{\text{nr}}$  is the maximally unramified extension of the  $p$ -adic numbers. Then  $\mathfrak{p} = (p)$  and  $k = \overline{\mathbb{F}}_p$ .

Let  $Y/K$  be a Picard curve, given by the equation

$$Y : y^3 = f(x), \tag{1}$$

where  $f \in K[x]$  is a separable polynomial of degree 4. We set  $X := \mathbb{P}_K^1$  and interpret (1) as a finite cover  $\phi : Y \rightarrow X$ ,  $(x, y) \mapsto x$ , of degree 3.

By the Semistable Reduction Theorem (see [5]), there exists a finite extension  $L/K$  such that the curve  $Y_L := Y \otimes_K L$  has semistable reduction. Since  $g(Y) = 3 \geq 2$ , there even exists a (unique) distinguished semistable model  $\mathcal{Y} \rightarrow \text{Spec } \mathcal{O}_L$  of  $Y_L$ , the *stable model* [5, Corollary 2.7]. The special fiber  $\bar{Y} := \mathcal{Y}_s$  of  $\mathcal{Y}$  is called the *stable reduction* of  $Y$ . It is a stable curve over  $k$  [5, § 1], and it only depends on  $Y$ , up to unique isomorphism.

It is no restriction to assume that the extension  $L/K$  is Galois and contains a third root of unity  $\zeta_3 \in L$ . Then the cover  $\phi_L : Y_L \rightarrow X_L$  (the base change of  $\phi$  to  $L$ ) is a Galois cover. Its Galois group  $G$  is cyclic of order 3, generated by the element  $\sigma$  which is determined by

$$\sigma(y) = \zeta_3 y.$$

Let  $\Gamma := \text{Gal}(L/K)$  denote the Galois group of the extension  $L/K$ . The group  $\Gamma$  acts faithfully and in a natural way on the scheme  $Y_L = Y \otimes_K L$ . We denote by  $\tilde{G}$  the subgroup of  $\text{Aut}(Y_L)$  generated by  $G$  and the image of  $\Gamma$ . By definition,  $\tilde{G}$  is a semidirect product,

$$\tilde{G} = G \rtimes \Gamma.$$

The action of  $\Gamma$  on  $G$  via conjugation is determined by the following formula: for  $\tau$  in  $\Gamma$  we have

$$\tau \sigma \tau^{-1} = \begin{cases} \sigma & \text{if } \tau(\zeta_3) = \zeta_3, \\ \sigma^2 & \text{if } \tau(\zeta_3) = \zeta_3^2. \end{cases} \quad (2)$$

Because of the uniqueness properties of the stable model, the action of  $\tilde{G}$  on  $Y_L$  extends to an action on  $\mathcal{Y}$ . By restriction, we see that  $\tilde{G}$  has a natural,  $k$ -linear action<sup>1</sup> on  $\bar{Y}$ . This action will play a decisive role in our analysis of the stable reduction  $\bar{Y}$ . For the rest of this subsection we focus on the action of the subgroup  $G \subset \tilde{G}$ . The role of the subgroup  $\Gamma \subset \tilde{G}$  will become important later.

### Remark 2.1

- (a) The quotient scheme  $\mathcal{X} := \mathcal{Y}/G$  is a semistable model of  $X_L = \mathbb{P}_L^1$ , see e.g. [16, Cor. 1.3.3.i]. Since the map  $\mathcal{Y} \rightarrow \mathcal{X}$  is finite and  $\mathcal{Y}$  is normal,  $\mathcal{Y}$  is the normalization of  $\mathcal{X}$  in the function field of  $Y_L$ . This means that  $\mathcal{Y}$  is uniquely determined by the cover  $Y \rightarrow X$  and a suitable semistable model  $\mathcal{X}$  of  $X_L$ .

<sup>1</sup>By  $k$ -linear action we mean that the action is compatible with the structure of  $\bar{Y}$  as a  $k$ -scheme.

- (b) Let  $\bar{X} := \mathcal{X} \otimes k$  denote the special fiber of  $\mathcal{X}$  and  $\bar{\phi} : \bar{Y} \rightarrow \bar{X}$  the induced map. We note that  $\bar{\phi}$  is a finite  $G$ -invariant map. It is not true in general that  $\bar{Y}/G = \bar{X}$ . However, the natural map  $\bar{Y}/G \rightarrow \bar{X}$  is radical and in particular a homeomorphism (see e.g. [16, p. 101]).
- (c) Every irreducible component  $W \subset \bar{Y}$  is smooth. To see this note that the quotient of  $W$  by its stabilizer in  $G$  is homeomorphic to an irreducible component  $Z \subset \bar{X}$ , which is a smooth curve of genus 0. If  $W$  has a singular point, then  $\sigma$  acts on  $W$  and permutes the two branches of  $W$  passing through this point. But since  $\sigma$  has order 3, this is impossible.

Let  $\Delta_{\bar{Y}}$  denote the component graph of  $\bar{Y}$ : the vertices are the irreducible components of  $\bar{Y}$  and the edges correspond to the singular points. The stability condition for  $\bar{Y}$  means that an irreducible component of genus 0 corresponds to a vertex of  $\Delta_{\bar{Y}}$  of degree  $\geq 3$ . The number of loops of  $\Delta_{\bar{Y}}$  is given by the well known formula

$$\gamma(\bar{Y}) := \dim_{\mathbb{Q}} H^1(\Delta_{\bar{Y}}, \mathbb{Q}) = r - s + 1, \tag{3}$$

where  $r$  is the number of edges and  $s$  the number vertices of  $\Delta_{\bar{Y}}$ .

The curve  $\bar{X}$  is also semistable, but in general not stable. Since  $\bar{X}$  has arithmetic genus 0, the component graph  $\Delta_{\bar{X}}$  is a tree, and every vertex corresponds to a smooth curve of genus 0. It follows from Remark 2.1 that  $\Delta_{\bar{X}} = \Delta_{\bar{Y}}/G$ .

**Lemma 2.2** *If  $W \subset \bar{Y}$  is an irreducible component, then  $\sigma(W) = W$ .*

*Proof* To derive a contradiction, we assume that  $W_1, W_2, W_3 \subset \bar{Y}$  are three distinct components that form a single  $G$ -orbit. Then  $W_i \xrightarrow{\sim} Z := \bar{\phi}(W_i)$ . Since  $Z$  is a component of  $\bar{X}$ , we conclude that  $g(W_i) = 0$ , for  $i = 1, 2, 3$ . The stability condition on  $\bar{Y}$  implies that each  $W_i$  contains at least three singular points of  $\bar{Y}$ . Hence  $Z$  also contains at least three singular points of  $\bar{X}$ .

Let  $\bar{Y} \rightarrow \bar{Y}_0$  denote the unique morphism which contracts all components of  $\bar{Y}$  except the  $W_i$  and which is an isomorphism on the intersection of  $\cup_i W_i$  with the smooth locus of  $\bar{Y}$ . Similarly, let  $\bar{X} \rightarrow \bar{X}_0$  be the map contracting all components of  $\bar{X}$  except  $Z$ . These maps fit into a commutative diagram

$$\begin{array}{ccc} \bar{Y} & \longrightarrow & \bar{Y}_0 \\ \downarrow & & \downarrow \\ \bar{X} & \longrightarrow & \bar{X}_0, \end{array}$$

where the vertical arrows are quotient maps by the group  $G$  (at least for the underlying topological spaces). Also,  $\bar{X}_0 \cong Z$ .

Let  $\bar{x} \in Z$  be one of the singular points of  $\bar{X}$  lying on  $Z$ , and let  $T \subset \bar{X}$  be the closed subset which is contracted to  $\bar{x} \in Z = \bar{X}_0$ . Then  $T$  is a nonempty and connected union of irreducible components of  $\bar{X}$  and hence a semistable curve of genus 0. In particular, the component graph of  $T$  is a tree. Let  $Z' \subset T$  be a tail

component. As a component of  $\bar{X}$ ,  $Z'$  intersects the rest of  $\bar{X}$  in at most two points. Let  $W' \subset \bar{Y}$  be an irreducible component lying above  $Z'$ . The stability of  $\bar{Y}$  implies that  $\sigma(W') = W'$  and that the action of  $\sigma$  on  $W'$  is nontrivial. (Otherwise  $W'$  would be homeomorphic to  $Z'$ , and hence  $W'$  would be a component of genus 0 intersecting the rest of  $\bar{Y}$  in at most two points.) It follows that the inverse image  $S \subset \bar{Y}$  of  $T$  is connected. Note that  $S$  meets the component  $W_i$  in the unique point on  $W_i$  above  $\bar{x}$ . Since  $S$  is connected, it follows that the map  $\bar{Y} \rightarrow \bar{Y}_0$  contracts  $S$  to a single point.

We conclude that the curve  $\bar{Y}_0$  has at least three distinct singular points where all three components  $W_i$  meet. Equation (3) implies that  $\gamma(\bar{Y}_0)$  is at least 1. It follows that the arithmetic genus of  $\bar{Y}_0$  is  $\geq 4$ , and hence  $g(\bar{Y}) \geq 4$  as well. This is a contradiction, and the lemma follows.  $\square$

## 2.2 The Conductor Exponent

Let  $\mathfrak{c}_p$  be the conductor of the  $\text{Gal}(\bar{K}/K)$ -representation  $H_{\text{et}}^1(Y_{\bar{K}}, \mathbb{Q}_\ell)$ , see [21]. By definition, this is an ideal of  $\mathcal{O}_K$  of the form

$$\mathfrak{c}_p = \mathfrak{p}^{f_p},$$

with  $f_p \geq 0$ . The integer  $f_p$  is called the conductor exponent of  $Y/K$ .<sup>2</sup>

We recall from [3] an explicit formula for  $f_p$ , in terms of the action of  $\Gamma = \text{Gal}(L/K)$  on  $\bar{Y}$ . For this we let  $\Gamma^u \subset \Gamma$ , for  $u \geq 0$ , denote the  $u$ th higher ramification group (in the upper numbering). We set  $\bar{Y}^u := \bar{Y}/\Gamma^u$ . Note that  $\bar{Y}^u$  is a semistable curve for all  $u$ . Note also that  $\Gamma = \Gamma^0$  because the residue field  $k$  is assumed to be algebraically closed.

**Proposition 2.3** *The conductor exponent of the curve  $Y/K$  is given by*

$$f_p = \epsilon + \delta, \tag{4}$$

where

$$\epsilon := 6 - \dim H_{\text{et}}^1(\bar{Y}^0, \mathbb{Q}_\ell) \tag{5}$$

and

$$\delta := \int_0^\infty (6 - 2g(\bar{Y}^u)) du. \tag{6}$$

*Proof* See [3, Theorem 2.9] and [2, Corollary 2.14].  $\square$

---

<sup>2</sup>When working in a local context,  $f_p$  is often simply called the conductor of  $Y$ .

The étale cohomology group  $H_{\text{et}}^1(\bar{Y}^u, \mathbb{Q}_\ell)$  decomposes as

$$H_{\text{et}}^1(\bar{Y}^u, \mathbb{Q}_\ell) = \bigoplus_W H_{\text{et}}^1(W, \mathbb{Q}_\ell) \oplus H^1(\Delta_{\bar{Y}^u}, \mathbb{Q}_\ell),$$

where the first sum runs over the set of irreducible components  $W$  of the normalization of  $\bar{Y}^u$  and  $\Delta_{\bar{Y}^u}$  is the graph of components of  $\bar{Y}^u$ . (See [3, Lemma 2.7.(1)].) Therefore, the second term in (5) can be written as

$$\dim H_{\text{et}}^1(\bar{Y}^0, \mathbb{Q}_\ell) = \sum_W \dim H_{\text{et}}^1(W, \mathbb{Q}_\ell) + \dim H^1(\Delta_{\bar{Y}^0}). \tag{7}$$

The arithmetic genus of  $\bar{Y}^u$ , which occurs in (6), is given by the formula

$$g(\bar{Y}^u) = \sum_W g(W) + \dim H^1(\Delta_{\bar{Y}^u}). \tag{8}$$

For future reference we note that  $\dim H_{\text{et}}^1(W, \mathbb{Q}_\ell) = 2g(W)$ . The integer  $\gamma(\bar{Y}^0) := \dim H^1(\Delta_{\bar{Y}^0})$  can be interpreted as the number of loops of the graph  $\Delta_{\bar{Y}^0}$ . It is bounded by  $g(\bar{Y}^0)$ , and hence by  $g(Y) = 3$ .

**Lemma 2.4** *The following statements are equivalent.*

- (a)  $\delta = 0$ .
- (b)  $\Gamma^u$  acts trivially on  $\bar{Y}$ , for all  $u > 0$ .
- (c) The curve  $Y$  has semistable reduction over a tamely ramified extension of  $K$ .

*Proof* Assume that  $\delta = 0$ . By (6) this means that  $3 = g(\bar{Y}) = g(\bar{Y}^u)$  for all  $u > 0$ . Using (8) one easily shows that this means that  $\Gamma^u$  acts trivially on the component graph  $\Delta_{\bar{Y}}$  of  $\bar{Y}$ . Moreover, for every component  $W \subset \bar{Y}$  we have  $g(W) = g(W/\Gamma^u)$ . It follows that  $\Gamma^u$  acts trivially on  $\bar{Y}$ . We have proved the implication (a) $\Rightarrow$ (b). The implication (b) $\Rightarrow$ (c) follows from [11, Theorem 4.44]. The implication (c) $\Rightarrow$ (a) follows immediately from the definition of  $\delta$ .  $\square$

### 3 The Wild Case: $p = 3$

In this section we assume that  $p = 3$ . We first analyze the special fiber of the stable model of  $Y_L$ , and show that there are essentially five reduction types. From Sect. 3.2 we consider the case where  $K$  is absolutely unramified, and derive a lower bound for the conductor exponent  $f_3$ .

### 3.1 The Stable Model

We keep all the notation introduced in Sect. 2. In addition, we assume that  $p = 3$ . By Lemma 2.2 every irreducible component of  $\bar{Y}$  is fixed by the generator  $\sigma$  of  $G$ . Therefore, irreducible components of  $\bar{Y}$  fall into two different classes.

**Definition 3.1** An irreducible component  $W \subset \bar{Y}$  is called *étale* if the restriction  $\sigma|_W \in \text{Aut}_k(W)$  is nontrivial. If  $\sigma|_W$  is the identity, then  $W$  is called an *inseparable component*.

Let  $W \subset \bar{Y}$  be an irreducible component, and let  $Z := \bar{\phi}(W) \subset \bar{X}$  be its image. Then  $Z$  is an irreducible component of  $\bar{X}$  and hence a smooth curve of genus 0. Lemma 2.2 shows that  $\sigma(W) = W$ . It follows that  $W/G \rightarrow Z$  is a homeomorphism. If  $W$  is an inseparable component, then  $W \rightarrow Z$  is a purely inseparable homeomorphism (since  $W \rightarrow Z$  has degree 3, this can only happen when  $p = 3$ ). It follows that every inseparable component has genus zero.

If  $W$  is an étale component, then  $Z \cong W/G$ , and  $W \rightarrow Z$  is a  $G$ -Galois cover. For future reference we recall that the Riemann–Hurwitz formula for wildly ramified Galois covers of curves yields

$$2g(W) - 2 = -2 \cdot 3 + \sum_z 2(h_z + 1), \quad (9)$$

where the sum runs over the branch points of  $W \rightarrow Z$  and  $h_z$  is the (unique) jump in the filtration of the higher ramification groups in the lower numbering. We have that  $h_z \geq 1$  is prime to  $p$  [20, § IV.2, Cor. 2 to Prop. 9].

**Theorem 3.2** *We are in exactly one of the following five cases.*

- (a) *The curve  $\bar{Y}$  is smooth and irreducible.*
- (b) *The curve  $\bar{Y}$  has exactly two components  $W_1, W_2$ . Both of them are étale, they meet in a single point, and have genus  $g(W_1) = 2$ ,  $g(W_2) = 1$ .*
- (c) *There are three étale components  $W_1, W_2, W_3$ , all of genus one, and one inseparable component  $W_0$ , which has genus zero. For  $i = 1, 2, 3$ ,  $W_i$  intersects  $W_0$  in a unique point, and these intersection points are precisely the singular points of  $\bar{Y}$ .*
- (d) *There are two components  $W_1, W_2$ , all of which are étale. Their genus is  $g(W_1) = 1$  and  $g(W_2) = 0$ . There are exactly three singular points, which form an orbit under the action of  $G$ , and where  $W_1$  and  $W_2$  meet.*
- (e) *There are three components  $W_1, W_2, W_3$ , all of which are étale, Their genus is  $g(W_1) = g(W_2) = 0$  and  $g(W_3) = 1$ . Furthermore, there are four singular points. Three of the singular points are points of intersection of  $W_1$  and  $W_2$ , and form an orbit under the action of  $G$ . The fourth singular point is the point of intersection of  $W_2$  and  $W_3$ .*

*Proof* Let  $r_1$  (resp.  $s_1$ ) be the number of singular points (resp. irreducible components) of  $\bar{Y}$  which are fixed by  $\sigma$ , and let  $r_2$  (resp.  $s_2$ ) be the number of orbits of



singular point (resp. irreducible components) of  $\bar{Y}$  of length 3. Lemma 2.2 states that  $s_2 = 0$ . Therefore, (3) becomes

$$\gamma(\bar{Y}) = r - s + 1 = r_1 + 3r_2 - s + 1. \quad (10)$$

Because  $\Delta_{\bar{X}} = \Delta_{\bar{Y}}/G$  is a tree, we have

$$\gamma(\bar{X}) := \dim H^1(\Delta_{\bar{X}}) = r_1 + r_2 - s + 1 = 0. \quad (11)$$

Combining (10) and (11) we obtain

$$\gamma(\bar{Y}) = 2r_2. \quad (12)$$

Since  $0 \leq \gamma(\bar{Y}) \leq 3$ , we conclude that  $\gamma(\bar{Y}) \in \{0, 2\}$  and  $r_2 \in \{0, 1\}$ .

**Case 1**  $r_2 = 0$  and  $\gamma(\bar{Y}) = 0$ .

In this case  $\Delta_{\bar{Y}}$  is a tree, and the sum of the genera of all irreducible components is 3. In particular, there are at most three components of genus  $> 0$ . Moreover, the stability condition implies that every component of genus zero contains at least three singular points of  $\bar{Y}$ . It is an easy combinatorial exercise to see that this leaves us with exactly four possibilities for the tree  $\Delta_{\bar{Y}}$ . Going through these four cases we will see that one of them is excluded, while the remaining three correspond to Cases (a)–(c) of Theorem 3.2.

The first case is when  $\bar{Y}$  has a unique irreducible component. Then  $\bar{Y}$  is smooth. This is Case (a) of the lemma. Secondly, there may be two irreducible components, of genus 1 and 2, and a unique singular point. This corresponds to Case (b).

Thirdly, there may be three irreducible components, each of genus 1, and two singular points. We claim that this case cannot occur. Indeed, one of the three components would contain two singular points, and each of these two points must be a fixed point of  $\sigma$ . It follows that the  $G$ -cover  $W \rightarrow Z = W/G$  is ramified in at least two points. The Riemann–Hurwitz formula (9) implies that  $g(W) \geq 2$ . This yields a contradiction, and we conclude that this case does not occur.

Finally, in the last case, there are four singular points and four irreducible components. Three of them have genus 1 and one has genus zero. The component of genus zero necessarily contains all three singular points. A similar argument as in the previous case shows that the genus-0 component cannot be étale. This corresponds to Case (c).

**Case 2**  $\gamma(\bar{Y}) = 2$  and  $r_2 = 1$ .

In this case the sum of the genera of all components is equal to 1. Therefore, there must be a unique component of genus 1, and all other components have genus 0. Let  $W_1$  and  $W_2$  be two components which meet in a singular point  $\bar{y}$  such that  $\sigma(\bar{y}) \neq \bar{y}$ . Since  $\sigma(W_i) = W_i$  for  $i = 1, 2$  (Lemma 2.2),  $W_1$  and  $W_2$  are étale components and intersect each other in exactly three points (the  $G$ -orbit of  $\bar{y}$ ).

If there are no further components, we are in Case (d). Assume that there exists a third component  $W_3$ . Let  $T \subset \bar{Y}$  be the maximal connected union of components which contains  $W_3$  but neither  $W_1$  nor  $W_2$ . Then  $T$  contains a unique component  $W_0$  which meets either  $W_1$  or  $W_2$  in a singular point. The component graph of  $T$  is a tree, and we consider  $W_0$  as its root. By the stability condition, every tail component of  $T$  must have positive genus, so  $T$  has a unique tail. If  $W_0$  is not this tail, it has genus 0 and intersects the rest of  $\bar{Y}$  in exactly 2 points. This contradicts the stability condition. We conclude that  $\bar{Y}$  has exactly three components, of genus  $g(W_1) = g(W_2) = 0$  and  $g(W_3) = 1$ . This is Case (e) of the lemma. Now the proof is complete.  $\square$

### 3.2 A Lower Bound for $f_3$

We continue with the assumptions from the previous subsection. In addition, we assume that  $K$  is absolutely unramified. By this we mean that  $\mathfrak{p} = (3)$ . Under this assumption, we prove a lower bound for the conductor exponent  $f_3 := f_{\mathfrak{p}}$ . In fact, we will give a lower bound for  $\epsilon$ , where  $f_3 = \epsilon + \delta$  is the decomposition from Proposition 2.3. If  $L/K$  is at most tamely ramified, then  $\delta = 0$  (Lemma 2.4). In this case, our bounds are sharp.

Since  $K$  is absolutely unramified, the third root of unity  $\zeta_3 \in L$  is *not* contained in  $K$ . Therefore, there exists an element  $\tau \in \Gamma = \text{Gal}(L/K)$  such that  $\tau(\zeta_3) = \zeta_3^2$ . Let  $m$  be the order of  $\tau$ . After replacing  $\tau$  by a suitable odd power of itself we may assume that  $m$  is a power of 2. We keep this notation fixed for the rest of this paper. Recall that the semidirect product  $\tilde{G} = G \rtimes \Gamma$  acts on  $\bar{Y}$  in a natural way.

The following observation is crucial for our analysis of the conductor exponent.

**Lemma 3.3** *Let  $W \subset \bar{Y}$  be an étale component such that  $\tau(W) = W$ . Then inside the automorphism group of  $W$  we have*

$$\tau \circ \sigma \circ \tau^{-1} = \sigma^2 \neq \sigma. \quad (13)$$

*In particular,  $\tau|_W$  is nontrivial.*

*Proof* The statement follows immediately from Eq. (2) and Definition 3.1.  $\square$

Despite its simplicity, Lemma 3.3 has the following striking consequence. Note that we consider potentially good but not good reduction as bad reduction in this paper.

**Proposition 3.4** *Assume that  $\mathfrak{p} = (3)$ . Then every Picard curve  $Y$  over  $K$  has bad reduction.*

*Proof* Lemma 3.3 implies that  $Y$  acquires semistable reduction only after passing to a ramified extension  $L \ni \zeta_3$ . Therefore  $Y/K$  does not have good reduction. The

fact that  $f_3 \neq 0$  follows from Proposition 2.3, together with the fact that  $\tau$  acts nontrivially on each irreducible component of  $\bar{Y}$  (Lemma 3.3).  $\square$

In order to prove more precise lower bounds for  $f_3$ , we need to analyze the action of  $\sigma$  and  $\tau$  on  $\bar{Y}$  in more detail.

**Lemma 3.5** *Let  $W \subset \bar{Y}$  be an étale component. Then one of the following cases occurs:*

$g(W)$	$r$	$h$	$g(W/\Gamma^0)$
0	1	1	0
1	1	2	0
2	2	(1, 1)	1
3	1	4	0

Here  $r$  is the number of ramification points of the  $G$ -cover  $W \rightarrow Z := W/G$  and  $h$  lists the set of lower jumps. The fourth column gives an upper bound for the genus of  $W/\Gamma^0$ .

*Proof* Recall that we have assumed that the order  $m$  of  $\tau$  is a power of 2.

The Riemann–Hurwitz formula (9) immediately yields the cases for  $g(W)$ ,  $r$ , and  $h$  stated in the lemma, together with one additional possibility: the curve  $W$  has genus 3 and  $\phi : W \rightarrow Z \cong \mathbb{P}^1$  is branched at two points, with lower jump 1 and 2, respectively. We claim that this case does not occur.

Assume that  $W$  is an étale component of  $\bar{Y}$  such that  $\phi : W \rightarrow Z$  is branched at 2 points. Lemma 3.3 implies that  $\tau$  acts nontrivially on  $W$ . Since  $\tau$  normalizes  $\sigma$  and the two ramification points have different lower jumps, it follows that  $\tau$  fixes both ramification points  $w_i$  of  $\phi$ . We conclude that  $H := \langle \sigma, \tau \rangle$  acts on  $W$  as a nonabelian group of order 6 fixing the  $v_i$ .

We write  $h_i$  for lower jump of  $w_i$ . Lemma 2.6 of [14] implies that  $\gcd(h_i, m)$  is the order of the prime-to-3 part of the centralizer of  $H$ . Since  $\gcd(h_1, m) \neq \gcd(h_2, m)$  we obtain a contradiction, and conclude that this case does not occur.

We compute an upper bound for the genus of  $W/\langle \tau \rangle$  in each of the remaining cases. This is also an upper bound for  $g(W/\Gamma^0)$ .

In the case that  $g(W) = 0$  there is nothing to prove. In the case that  $g(W) = 1$ , the automorphism  $\tau$  fixes the unique ramification point of  $\phi$ , hence  $g(W/\Gamma^0) = 0$ .

Assume that  $g(W) = 2$ . The Riemann–Hurwitz formula immediately implies that  $g(W/\langle \tau \rangle) \leq 1$ .

Finally, we consider the case that  $g(W) = 3$ , i.e.  $Y$  has potentially good reduction. As before, we have that  $\tau$  fixes the unique fixed point of  $\sigma$ . Put  $H = \langle \sigma, \tau \rangle$ . Lemma 3.3 together with the assumption that the order  $m$  of  $\tau$  is a power of 2 implies that the order of the prime-to- $p$  centralizer of  $H$  is  $\gcd(h = 4, m) = m/2$ . It follows that  $m = 8$ . Since  $\tau$  has at least one fixed point on  $W$ , namely the unique fixed point of  $\sigma$ , the Riemann–Hurwitz formula implies that  $g(W/\langle \tau \rangle) = 0$ . This finishes the proof of the lemma.  $\square$

We have now all the necessary tools to prove our main theorem.

**Theorem 3.6** *Assume  $\mathfrak{p} = (3)$ , and let  $Y$  be a Picard curve over  $K$ . The conductor exponent  $f_3$  of  $Y/K$  satisfies*

$$f_3 \geq 4.$$

Moreover:

- (a) *If  $f_3 \leq 6$  then  $Y$  achieves semistable reduction over a tamely ramified extension  $L/K$ .*
- (b) *If  $f_3 = 4$  then we are in Case (b) or Case (c) from Theorem 3.2.*
- (c) *If  $f_3 = 5$  then we are in Case (d) or in Case (e) of Theorem 3.2.*

*Proof* We use the assumptions and notations from the beginning of Sect. 3.2. Recall that the inertia subgroup  $\Gamma^0 \subset \Gamma := \text{Gal}(L/K)$  acts on the geometric special fiber  $\bar{Y}$  of the stable model of  $Y_L$  and that the quotient  $\bar{Y}^0 = \bar{Y}/\Gamma^0$  is again a semistable curve.

*Claim* We have that

$$\dim H_{\text{et}}^1(\bar{Y}^0, \mathbb{Q}_\ell) \leq 2. \tag{14}$$

Note that (14), together with (4) and (5), immediately implies the first statement  $f_3 \geq 4$  of the theorem.

Recall from (8) and (3) that the contribution of a smooth component  $W$  of  $\bar{Y}^0$  to  $\dim H_{\text{et}}^1(\bar{Y}^0, \mathbb{Q}_\ell)$  is  $2g(W)$ . The contribution of  $H^1(\Delta_{\bar{Y}})$  to  $\dim H_{\text{et}}^1(\bar{Y}^0, \mathbb{Q}_\ell)$  is  $\gamma(\bar{Y}^0)$ , which is less than or equal to  $g(\bar{Y}^0)$ .

Let  $W \subset \bar{Y}$  be an irreducible component, and denote by  $W^0 \subset \bar{Y}^0$  its image in  $\bar{Y}^0$ . Clearly,  $g(W^0) \leq g(W)$ . Moreover, if  $\tau(W) = W$  then Lemma 3.5 shows that  $g(W^0) \leq 1$ .

Let us consider each case of Theorem 3.2 separately. In Case (a),  $\bar{Y}$  is smooth and irreducible of genus 3. Then  $\bar{Y}^0$  is also smooth and irreducible, and Lemma 3.5 shows that  $g(\bar{Y}^0) = 0$ . So in Case (a) we have proved  $\dim H_{\text{et}}^1(\bar{Y}^0, \mathbb{Q}_\ell) = 0$ , which is strictly stronger than (14). Similarly, in Case (b) Lemma 3.5 shows that  $\bar{Y}^0$  consists of two irreducible components which meet in a single point. One of these components has genus zero, the other one has genus  $\leq 1$ . Therefore, (14) holds in Case (b).

Assume that we are in Case (c). Let  $W_1, W_2, W_3$  denote the three components of genus 1, and  $W_i^0, i = 1, 2, 3$ , their images in  $\bar{Y}^0$ . Since the order of  $\tau$  is a power of two,  $\tau$  fixes exactly one of these components (say  $W_1$ ), or all three. In the first case,  $g(W_1^0) = 0$  by Lemma 3.5, and  $W_2^0 = W_3^0$ . Therefore,  $\dim H_{\text{et}}^1(\bar{Y}^0, \mathbb{Q}_\ell) = 1$ . In the second case,  $g(W_i^0) = 0$  for  $i = 1, 2, 3$ , and  $\dim H_{\text{et}}^1(\bar{Y}^0, \mathbb{Q}_\ell) = 0$ . In both cases, (14) holds.

Now assume that we are in Case (d). The action of  $\Gamma^0$  must fix both components  $W_1, W_2$ , since  $g(W_1) \neq g(W_2)$ . Lemma 3.5 shows that  $g(W_i/\Gamma^0) = 0$ , for  $i =$

1, 2. Also,  $\tau$  permutes the three singular points of  $\bar{Y}$ . But these points form one orbit under the action of  $G$ . Hence it follows from (13) that  $\tau$  fixes exactly one singular point and permutes the other two. We conclude that the curve  $\bar{Y}^0$  has two smooth components of genus 0 which meet in at most two points. We conclude that  $\dim H_{\text{et}}^1(\bar{Y}^0, \mathbb{Q}_\ell) \leq 1$ . A similar analysis shows that the same conclusion holds in Case (e). This proves the claim (14).

While proving the claim, we have shown the following stronger conclusion:

$$\dim H_{\text{et}}^1(\bar{Y}^0, \mathbb{Q}_\ell) \in \begin{cases} \{0\}, & \text{Case (a),} \\ \{0, 2\}, & \text{Case (b), (c),} \\ \{0, 1\}, & \text{Case (d), (e).} \end{cases} \tag{15}$$

It follows that  $\epsilon = 6$  in Case (a),  $\epsilon \in \{4, 6\}$  in the Cases (b) and (c), and  $\epsilon \in \{5, 6\}$  in the Cases (d) and (e).

The remaining statement that  $Y$  acquires stable reduction over a tamely ramified extension  $L$  of  $K$  in the case that  $f_3 \leq 6$  follows from Lemma 2.4.  $\square$

**Corollary 3.7** *If  $\mathfrak{p} = (3)$  and  $Y$  has potentially good reduction, then  $f_3 \geq 6$ .*

### 3.3 Examples

In this section we discuss two explicit examples of Picard curves over  $\mathbb{Q}_3^{\text{nr}}$  in some detail. These examples show, among other things, that the lower bounds for  $f_3$  given by Theorem 3.6 are sharp.

Let us fix some notation. We set  $K := \mathbb{Q}_3^{\text{nr}}$ . Given a suitable finite extension  $L/K$ , we denote by  $v_L$  the unique extension of the 3-adic valuation to  $L$  (which is normalized such that  $v_L(3) = 1$ ). We let  $F(X_L)$  denote the function field of  $X_L := \mathbb{P}_L^1$ , and identify  $F(X_L)$  with the rational function field  $L(x)$ . For a Picard curve  $Y$  over  $K$  given by  $y^3 = f(x)$  for a quartic polynomial  $f \in K[x]$  the function field  $F(Y_L)$  of  $Y_L$  is the degree-3 extension of  $F(X_L)$  obtained by adjoining the function  $y$ .

Let  $\mathcal{X}$  be a semistable model of  $X_L$ , and let  $Z_1, \dots, Z_n \subset \bar{X} := \mathcal{X} \otimes \mathbb{F}_L$  denote the irreducible components of the special fiber. Since each  $Z_i$  is a prime divisor on  $\mathcal{X}$ , it gives rise to a discrete valuation  $v_i$  on  $F(X_L)$ , extending  $v_L$ . It has the property that the residue field of  $v_i$  can be naturally identified with the function field of  $Z_i$ . Since  $X_L$  is simply a projective line and  $\mathcal{X}$  is a semistable model, the valuations  $v_i$  have a simple description, as follows. For all  $i$ , there exists a coordinate  $x_i \in F(X_L)$  such that  $v_i$  is the Gauss valuation on  $F(X_L) = L(x_i)$  with respect to  $x_i$ . The coordinate  $x_i$  is related to  $x$  by a fractional linear transformation

$$x = \frac{a_i x_i + b_i}{c_i x_i + d_i},$$

with  $a_i d_i - b_i c_i \neq 0$ . It can be shown that the model  $\mathcal{X}$  is uniquely determined by the set  $\{v_1, \dots, v_n\}$ , see [3] or [17].

Let  $\mathcal{Y}$  denote the normalization of  $\mathcal{X}$  inside the function field  $F(Y_L)$ . Then  $\mathcal{Y}$  is a normal integral model of  $Y_L$ . In general,  $\mathcal{Y}$  has no reason to be semistable, and it is not clear in general how to describe its special fiber  $\bar{Y} := \mathcal{Y} \otimes k$ . However, each irreducible component  $W \subset \bar{Y}$  corresponds again to a discrete valuation  $w$  on  $F(Y_L)$  extending  $v_L$ , such that the residue field of  $w$  is the function field of  $W$ . It can be shown that this gives a bijection between the irreducible components of  $\bar{Y}$  and the set of discrete valuations on  $F(Y_L)$  extending one of the valuations  $v_i$  (see e.g. [17, § 3]). In many situations, the knowledge of all extensions of the  $v_i$  to  $F(Y_L)$  will give enough information to decide whether the model  $\mathcal{Y}$  is semistable and to describe its special fiber.

We need one more piece of notation. For  $m > 1$  prime to 3 we set

$$L_m := K(\pi)/K$$

where  $\pi^m = -3$ . Then  $L_m/K$  is a tamely ramified Galois extension of degree  $m$ . The Galois group  $\Gamma := \text{Gal}(L_m/K)$  is cyclic and generated by the element  $\tau \in \Gamma_m := \text{Gal}(L_m/K)$  determined by

$$\tau(\pi) = \zeta_m \pi,$$

where  $\zeta_m \in K$  is a primitive  $m$ th root of unity (which exists because  $k$  is algebraically closed). Note also that  $L_m$  contains the third root of unity

$$\zeta_3 := \frac{-1 + \pi^{m/2}}{2}.$$

We remark that the choice of  $\tau$  and  $m$  agrees with the notation chosen in Sect. 3.2

*Example 3.8* Let  $Y$  be the Picard curve over  $K$  given by the equation

$$y^3 = x^4 + 1. \tag{16}$$

We claim that  $Y$  has potentially good reduction, which is attained over the tame extension  $L := L_8 = K(\pi)/K$ , with  $\pi^8 = -3$ .

To prove this, we apply the coordinate changes

$$x = \pi^3 x_1, \quad y = 1 + \pi^4 y_1$$

to (16). After a brief calculation, we obtain the new equation

$$y_1^3 - \pi^4 y_1^2 - y_1 = x_1^4. \tag{17}$$

Equation (17) is equivalent to (16) in the sense that it defines a curve over  $K$  which is isomorphic to  $Y$ . Also, (17) defines an integral model  $\mathcal{Y}$  of  $Y_L$ . Its special fiber is the curve over  $k = \bar{\mathbb{F}}_3$  given by the (affine) equation

$$\bar{Y} : y_1^3 - y_1 = x_1^4.$$

This is a smooth curve of genus 3. It follows that  $\mathcal{Y}$  has good reduction over  $L$ , as claimed.

Since  $Y$  acquires stable reduction over a tame extension  $L/K$ , Lemma 2.4 implies that  $f_3 = \epsilon$ . Equations (5) and (15) imply that  $f_3 = 6$ .

For completeness, we compute the action of  $\Gamma^0 = \langle \tau \rangle$  on  $\bar{Y}$  explicitly. We consider  $\tau$  as an automorphism of the structure sheaf of  $\mathcal{Y}$ . By definition, we have

$$\tau(\pi) = \zeta_8 \pi, \quad \tau(x) = x, \quad \tau(y) = y.$$

It follows that

$$\tau(x_1) = \zeta_8^5 x_1, \quad \tau(y_1) = -y_1.$$

This describes  $\tau|_{\bar{Y}}$  as an automorphism of  $\bar{Y}$  of order 8, as expected from the proof of Lemma 3.5.

*Example 3.9* Let  $Y/K$  be the Picard curve

$$Y : y^3 = f(x) := 3x^4 + x^3 - 54. \tag{18}$$

We claim that  $Y$  has semistable reduction over the tame extension  $L := L_4/K$ . Moreover, the stable reduction  $\bar{Y}$  is as in Case (b) of Theorem 3.2, and  $f_3 = 4$ .

First we define a semistable model  $\mathcal{X}$  of  $X_L := \mathbb{P}_L^1$  by specifying two discrete valuations  $v_1, v_2$  on  $F(X_L)$  which extend  $v_L$ . By construction, the special fiber  $\mathcal{X}_s$  of  $\mathcal{X}$  consists of two irreducible components  $\bar{X}_1, \bar{X}_2$  which correspond to  $v_1, v_2$ . We then show that the normalization  $\mathcal{Y}$  of  $\mathcal{X}$  in  $F(Y_L)$  is the stable model of  $Y_L$ , and determine its special fiber  $\bar{Y}$  and the action of the inertia group of  $L/K$  on  $\bar{Y}$ .

The valuation  $v_1$  is defined as the Gauss valuation on  $F(X_L) = F(x_1)$  with respect to the coordinate  $x_1$ , which is related to  $x$  by

$$x = \pi^2 x_1. \tag{19}$$

We claim that  $v_1$  has a unique extension  $w_1$  to  $F(Y_L)$  that is unramified. To show this, we need a so-called *p-approximation* of  $f$  with respect to  $v_1$ , see [18]. In fact, we can write

$$f = \pi^6(x_1^3 + \pi^6(2 - x_1^4)).$$

Here we have used the relation  $\pi^4 = -3$ . This suggests the coordinate change

$$y = \pi^2(x_1 + \pi^2 y_1). \quad (20)$$

After a short calculation we obtain a new equation for  $Y_L$ :

$$y_1^3 - \pi^2 x_1 y_1^2 - x_1^2 y_1 = 2 - x_1^4. \quad (21)$$

If we consider (21) as defining an affine curve over  $\mathcal{O}_L$ , its special fiber is the affine curve over  $k$  with equation

$$\bar{y}_1^3 - \bar{x}_1^2 \bar{y}_1 = -1 - \bar{x}_1^4. \quad (22)$$

In fact, (22) defines an irreducible affine curve with a cusp singularity in  $(\bar{x}_1, \bar{y}_1) = (0, -1)$ . It follows that the inverse image in  $\bar{Y}$  of  $\bar{X}_1$  is an irreducible component  $W_1$  of multiplicity one birationally equivalent to the curve given by (22). To compute the geometric genus of  $W_1$  we substitute  $\bar{y}_1 = -1 + \bar{x}_1 \bar{z}_1$  into (22) and obtain the Artin–Schreier equation

$$\bar{z}_1^3 - \bar{z}_1 = -\bar{x}_1^{-1} - \bar{x}_1. \quad (23)$$

Using the Riemann–Hurwitz formula, one sees that  $W_1$  has geometric genus 2.

We now consider the second valuation  $v_2$  of  $F(X_L)$ , defined as the Gauss valuation with respect to the coordinate  $x_2$  given by

$$x = 3(1 + \pi x_2). \quad (24)$$

After a short calculation we can write

$$f = 3^3((-1 + \pi x_2)^3 - 2\pi^6 x_2^2 + 3^2(\dots)). \quad (25)$$

This suggests to define a new function  $y_2 \in F(Y_L)$  via the change of coordinate

$$y = 3((-1 + \pi x_2) + \pi^2 y_2). \quad (26)$$

Plugging in (26) into (18) and using (25) we arrive at the equation

$$y_2^3 + \pi^2(-1 + \pi x_2)y_2^2 - (-1 + \pi x_2)^2 y_2 = -2x_2^2 + \pi^2(\dots). \quad (27)$$

Reducing (27) modulo  $\pi$  we obtain the irreducible equation

$$\bar{y}_2^3 - \bar{y}_2 = \bar{x}_2^2, \quad (28)$$

which defines a curve of genus 1. It follows that the inverse image of  $\bar{X}_2$  in  $\bar{Y}$  is an irreducible projective curve  $W_2$  of geometric genus 1.



So  $\bar{Y}$  consists of two irreducible components  $W_1$  and  $W_2$  of geometric genus 2 and 1. On the other hand,  $\mathcal{Y}_s$  is known to have arithmetic genus 3. By a standard argument (involving the computation of the arithmetic genus of a generically reduced, singular curve) we can conclude that  $W_1, W_2$  are smooth and meet transversely in a single point. This shows that  $Y$  has semistable reduction over the tame extension  $L_4/K$ , with a stable model of type (b).

Let us try to analyze the action of  $\Gamma = \Gamma^0 = \langle \tau \rangle$  on  $\bar{Y}$ . By definition,  $\tau(\pi) = \zeta_4\pi$ ,  $\tau(x) = x$  and  $\tau(y) = y$ . From (19) and (20) we deduce that  $\tau|_{W_1}$  is given by

$$\tau(\bar{x}_1) = -\bar{x}_1, \quad \tau(\bar{y}_1) = \bar{y}_1, \quad \tau(\bar{z}_1) = -\bar{z}_1.$$

From (24) and (26) we see that

$$\tau(\bar{x}_2) = \zeta_4^3\bar{x}_2, \quad \tau(\bar{y}_2) = -\bar{y}_2.$$

It follows that the curve  $\bar{Y}^0 := \bar{Y}/\Gamma^0$  has two irreducible smooth components,  $W_1^0 = W_1/\Gamma^0$  and  $W_2^0 = W_2/\Gamma^0$ , meeting in a single point. An easy calculation (compare with the proof of Lemma 3.5) shows that  $g(W_1^0) = 1$  and  $g(W_2^0) = 0$ . It follows that  $g(\bar{Y}^0) = 1$  and  $\dim H^1(\Delta_{\bar{Y}^0}) = 0$  and hence  $f_3 = 6 - 2 = 4$ .

*Remark 3.10* The two examples discussed above are quite special. Typically, the extension  $L/K$  needs to be wildly ramified, and have rather large degree. It is then hard (and often practically impossible) to do computations as above by hand. Most of the examples in [1] and this paper have been computed with the help of (earlier versions of) Julian R uth’s Sage packages `mac_lane` and `completion` (available at <https://github.com/saraedum>), and the algorithms from [3] and [18].

## 4 The Tame Case: $p \neq 3$

In this section we assume that the residue characteristic  $p$  of our ground field  $K$  is different from 3. In this case it is much easier to analyze the semistable reduction of Picard curves and to compute the conductor exponent  $f_p$  than for  $p = 3$ . The theoretical background for this are the *admissible covers*, see [7, 11, § 10.4.3], or [26]. In the case of superelliptic curves the computation of  $f_p$  has already been described in detail in [3], hence we can be much briefer than in the previous section.

### 4.1 The Stable Model

Let  $K$  be as in Sect. 2.1, with  $p \neq 3$ . Let  $Y/K$  be a Picard curve, given by an equation

$$Y : y^3 = f(x),$$

where  $f \in K[x]$  is a separable polynomial of degree 4. Let  $L_0/K$  denote the splitting field of  $f$ . Let  $L/L_0$  be a finite extension with ramification index 3 such that  $L/K$  is a Galois extension. Then [3], Corollary 4.6 implies that  $Y$  acquires semistable reduction over  $L$ .

We note in passing that  $L/K$  is tamely ramified unless  $p = 2$ . This follows from the definition of the Galois extension  $L_0/K$ , whose degree divides  $4! = 24$ .

A semistable model  $\mathcal{Y}$  of  $Y_L$  may be constructed as follows, see [3, § 4]. Let  $D \subset X = \mathbb{P}_K^1$  denote the branch divisor of the cover  $\phi : Y \rightarrow X$ , consisting of the set of zeros of  $f$  and  $\infty$ . Since  $L$  contains the splitting field of  $f$ , the pullback  $D_L \subset Y_L$  consists of 5 distinct  $L$ -rational points. Let  $(\mathcal{X}, \mathcal{D})$  denote the *stably marked model* of  $(X_L, D_L)$ . By this we mean that  $\mathcal{X}$  is the minimal semistable model of  $X_L$  with the property that the schematic closure  $\mathcal{D} \subset \mathcal{X}$  of  $D_L$  is étale over  $\text{Spec } \mathcal{O}_L$  and contained inside the smooth locus of  $\mathcal{X} \rightarrow \text{Spec } \mathcal{O}_L$ . Let  $\bar{X} := \mathcal{X} \otimes_{\mathbb{F}_L}$  denote the special fiber of  $\mathcal{X}$  and  $\bar{D} = \mathcal{D} \cap \bar{X}$  the specialization of  $D_L$ . Then  $(\bar{X}, \bar{D})$  is a stable 5-marked curve of genus zero. This means that  $\bar{X}$  is a tree of projective lines, where every irreducible component has at least three points which are either marked (i.e. lie in the support of  $\bar{D}$ ) or are singular points of  $\bar{X}$ .

Let  $\mathcal{Y}$  denote the normalization of  $\mathcal{X}$  with respect to the cover  $Y_L \rightarrow X_L$ . Theorem 3.4 from [3] shows that  $\mathcal{Y}$  is a quasi-stable model of  $Y_L$ . A priori, it is not clear whether  $\mathcal{Y}$  is the stable model of  $Y$ . The following case-by-case analysis will show that it is.

We will use the fact that the natural map  $\mathcal{Y} \rightarrow \mathcal{X}$  is an *admissible cover* with branch locus  $\mathcal{D}$ . In particular, the induced map

$$\bar{\phi} : \bar{Y} \rightarrow \bar{X}$$

between the special fiber of  $\mathcal{Y}$  and of  $\mathcal{X}$  is generically étale and identifies  $\bar{X}$  with the quotient scheme  $\bar{Y}/G$ .

We describe the restriction of the map  $\bar{\phi}$  to an irreducible component  $\bar{X}_i$  of  $\bar{X}$ . Without loss of generality we may assume that  $K$  (and hence  $L$ ) contains a primitive third root of unity  $\zeta_3$ , which we fix. For each branch point  $\xi$  of  $\bar{\phi}|_{\bar{X}_i}$  the *canonical generator of inertia*  $g \in G$  is characterized by  $g^*u \equiv \zeta_3 u \pmod{u^2}$ , where  $u$  is a local parameter at  $\bar{\phi}|_{\bar{X}_i}^{-1}(\xi_i)$ . A branch point of  $\bar{\phi}|_{\bar{X}_i}$  is either the specialization of a branch point of  $\phi$  or a singular point of  $\bar{X}$ .

Assume that  $\xi$  is the specialization of a branch point. An elementary calculation shows that the canonical generator of inertia is equal to  $\sigma$  if  $\xi$  is the specialization of  $\infty$  and is equal to  $\sigma^2$  otherwise. Now let  $\xi$  be a singularity of  $\bar{X}$ , and denote the irreducible components intersecting in  $\xi$  by  $\bar{X}_1$  and  $\bar{X}_2$ . Then the canonical generators  $g_i$  of the restrictions  $\bar{\phi}|_{\bar{X}_i}$  at  $\xi$  satisfy

$$g_1 = g_2^{-1}.$$

(This last condition says that  $\bar{\phi}$  is an admissible cover.)

The upshot is that the map  $\bar{\phi} : \bar{Y} \rightarrow \bar{X}$  is completely determined and easily described by the stably marked curve  $(\bar{X}, \bar{D})$ .

The following lemma lists the 5 possibilities for  $\bar{X}$ . Note that we need to distinguish between  $\infty$  and the other 4 branch points. The proof is elementary, and therefore omitted.

**Lemma 4.1** *With assumptions and notations as in the beginning of the section, we have the following five possibilities for  $\bar{X}$ .*

- (a) *The curve  $\bar{X}$  is irreducible.*
- (b) *The curve  $\bar{X}$  consists of two irreducible components  $\bar{X}_1$  and  $\bar{X}_2$ . Three of the branch points of  $\phi$  including  $\infty$  specialize to  $\bar{X}_1$ , the other two to  $\bar{X}_2$ .*
- (c) *The curve  $\bar{X}$  consists of three irreducible components  $\bar{X}_1$ ,  $\bar{X}_2$ , and  $\bar{X}_3$ , where  $\bar{X}_1$  and  $\bar{X}_3$  intersect  $\bar{X}_2$ . The branch point  $\infty$  specializes to  $\bar{X}_2$ , two other branch points specialize to  $\bar{X}_1$ , and two to  $\bar{X}_3$ .*
- (d) *The curve  $\bar{X}$  consists of two irreducible components  $\bar{X}_1$  and  $\bar{X}_2$ . Three of the branch points of  $\phi$  different from  $\infty$  specialize to  $\bar{X}_1$ , the other two to  $\bar{X}_2$ .*
- (e) *The curve  $\bar{X}$  consists of three irreducible components  $\bar{X}_1$ ,  $\bar{X}_2$ , and  $\bar{X}_3$ , where  $\bar{X}_1$  and  $\bar{X}_3$  intersect  $\bar{X}_2$ . Two branch points including  $\infty$  specialize to  $\bar{X}_1$ , two other branch points specialize to  $\bar{X}_3$ , and the last one to  $\bar{X}_2$ .*

The following result immediately follows from the possibilities for  $\bar{X}$ , together with the fact that  $\bar{\phi}$  is an admissible cover.

**Theorem 4.2** *Let  $K$  be as in Sect. 2.1, with  $p \neq 3$ . Let  $Y$  be a Picard curve over  $K$ ,  $L/K$  a finite Galois extension over which  $Y$  has semistable reduction. Let  $\mathcal{Y}$  denote the stable model of  $Y_L$  over  $\mathcal{O}_L$  and  $\bar{Y} := \mathcal{Y} \otimes k$  the special fiber. Then  $\bar{Y}$  is as in one of the following five cases.*

- (a) *The curve  $\bar{Y}$  is smooth.*
- (b) *The curve  $\bar{Y}$  consists of two irreducible components, of genus 2 and 1, which intersect in a unique singular point.*
- (c) *The curve  $\bar{Y}$  has three irreducible components  $W_1, W_2, W_3$  which are each smooth of genus 1. There are two singular points where  $W_1$  (resp.  $W_3$ ) intersects  $W_2$ .*
- (d) *There are two irreducible components  $W_1, W_2$  of genus 0 and 1, respectively, and three singular points where  $W_1$  and  $W_2$  intersect.*
- (e) *There are three irreducible components  $W_1, W_2, W_3$ , of genus 0, 0 and 1, respectively, and 4 singular points. The components  $W_1, W_2$  meet in three of these singular points, while  $W_2$  and  $W_3$  meet in the fourth.*

## 4.2 The Conductor Exponent in the Tame Case

In the tame case, there are no useful lower bounds for the conductor exponent. In particular,  $Y$  may have good reduction in which case we have  $f_p = 0$ . Also,

unlike for  $p = 3$ , nothing is gained by assuming that the ground field  $K$  is totally unramified. Still, some useful restrictions on  $f_{\mathfrak{p}}$  can be proved (see Theorem 4.4 below).

We start by recalling a well known criterion for good reduction, see e.g. [8, § 7]. Let

$$Y : y^3 = f(x) = a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0$$

be a Picard curve over  $K$ . Replacing  $(x, y)$  by  $(a_4^{-1}x, a_4^{-1}y)$  and multiplying both sides of the defining equation by  $a_4^3$ , we may assume that  $a_4 = 1$ . Let  $\Delta(f) \in K^\times$  denote the discriminant of  $f$ . (Since we assume that  $f$  is separable, we have  $\Delta(f) \neq 0$ .) After replacing  $(x, y)$  by  $(u^{-3}x, u^{-4}y)$  and multiplying by  $u^{12}$  on both sides, for a suitable  $u \in K^\times$ , we may further assume that all coefficients  $a_i \in \mathcal{O}_K$  are integral. In particular, it follows that  $\Delta(f) \in \mathcal{O}_K$ . Since

$$\Delta(u^{12}f(u^{-3}x)) = u^{36}\Delta(f),$$

by the right choice of  $u$ , we may assume that

$$0 \leq \text{ord}_{\mathfrak{p}}(\Delta(f)) < 36. \tag{29}$$

**Lemma 4.3** *Assume that the Picard curve  $Y$  is given by a minimal equation over  $\mathcal{O}_K$ , as above. Then  $Y$  has good reduction if and only if  $\Delta(f) \in \mathcal{O}_K^\times$ .*

*Proof* See [8, Lemma 7.13]. □

Note that the forwards direction of Lemma 4.3 also follows from Theorem 4.2. Here is what we can say in general about the conductor exponent.

**Theorem 4.4** *Let  $K$  be as before, with  $p \neq 3$ , and  $Y$  a Picard curve over  $K$ . Let  $f_{\mathfrak{p}}$  denote the conductor exponent for  $Y$ , relative to the prime ideal  $\mathfrak{p}$  of  $\mathcal{O}_K$ . Then the following holds.*

- (a) *If  $f_{\mathfrak{p}} = 0$  then the stable reduction of  $Y$  is as in Case (a), (b), or (c) of Theorem 4.2. Furthermore, the splitting field  $L_0/K$  of  $f$  is unramified at  $\mathfrak{p}$ .*
- (b) *If  $p = 2$  then  $f_{\mathfrak{p}} \neq 1$ .*
- (c) *If  $p \geq 5$  then  $f_{\mathfrak{p}} \in \{0, 2, 4, 6\}$ .*

*Proof* We start by proving Statement (a). Note that  $f_{\mathfrak{p}} = 0$  if and only if  $\delta = 0$  and  $\dim H_{\text{et}}^1(\bar{Y}^0, \mathbb{Q}_\ell) = 6$ . The second condition, together with the discussion after Proposition 2.3, implies that  $\gamma(\bar{Y}^0)$ . Statement (a) now follows immediately from Theorem 4.4.

*Claim* The integer  $\epsilon$ , defined in Proposition 2.3, is even. The discussion following Proposition 2.3 implies that  $f_{\mathfrak{p}}$  is odd if and only if  $\dim H^1(\Delta_{\bar{y}_0})$  is odd. The case distinction in Theorem 4.2 implies that  $\dim H^1(\Delta_{\bar{y}_0})$  is at most 2. Therefore to prove the claim, it suffices to show that  $\gamma(\bar{Y}^0) = \dim H^1(\Delta_{\bar{y}_0}) \neq 1$ . We prove this in the

case that  $\bar{Y}$  is as in (d) of Theorem 4.2. The argument in the case that  $\bar{Y}$  is as in (e) is very similar. In the other cases there is nothing to prove.

Assume that  $\bar{Y}$  is as in (d) of Theorem 4.2. Then  $\bar{X}$  is as (d) of Lemma 4.1 and  $\bar{\phi}$  maps  $W_i$  to  $\bar{X}_i$ . Since  $\infty$  is  $K$ -rational, the monodromy group  $\Gamma$  fixes it. It follows that  $\Gamma$  acts on the component  $\bar{X}_2$  to which  $\infty$  specializes. (This is similar to the argument in the proof of [3, Lemma 5.4].) Since there is exactly one other branch point specializing to  $\bar{X}_2$ , this point is fixed by  $\Gamma$ , as well. Similarly,  $\Gamma$  fixes the unique singularity. Since  $\Gamma$  fixes at least 3 points on the genus-0 curve  $\bar{X}_2$ , it acts trivially on  $\bar{X}_2$ . Equation (2) implies that the action of  $\Gamma$  on  $\bar{Y}$  descends to  $\bar{X}$ . It follows that  $\Gamma$  acts on  $W_2$  via a subgroup of  $G$ . We conclude that  $\Gamma$  either fixes the three singularities of  $\bar{Y}$  or cyclically permutes them. It follows that  $\gamma(\bar{Y}^0)$  is 2 or 0. This proves the claim.

Assume that  $p = 2$ . Using Eq. (6) one shows that if  $\delta \neq 0$  then  $\delta \geq 2$ . Therefore Statement (b) follows from the claim.

For Statement (c) recall that  $L/K$  is at most tamely ramified for  $p \geq 5$ . It follows that  $\delta = 0$ , and hence that  $f_p = \epsilon$  is bounded by  $2g(Y) = 6$ . Statement (c) now follows from the claim.  $\square$

*Remark 4.5*

- (a) The condition  $f_p = 0$  in Theorem 4.4(a) is equivalent to the condition that the Jacobian variety of  $Y$  has good reduction over  $K$ . This is the case if and only if  $Y$  has stable reduction already over  $K$ , and the graph of components  $\Delta_{\bar{Y}}$  is a tree. This observation is similar to the statement of Lemma 2.4.
- (b) For  $p = 2$  the conductor exponent  $f_2$  may be odd. An example can be found in Example 5.5.
- (c) The bound on  $f_p$  for  $p = 5, 7$  in Theorem 4.4.(c) is slightly sharper than the bound for  $f_p$  for general abelian varieties of dimension 3 from [4, Thm. 6.2]. The reason is that Brumer and Kramer obtain an upper bound for  $\delta$ . For Picard curves and  $p = 5, 7$  we have  $\delta = 0$ , whereas this is not necessarily the case for general curves of genus 3.

For  $p = 2$  the result of [4] yields the upper bound  $f_p \leq 28$ . Distinguishing the possibilities for the stable reduction and combining our arguments with those of [4] it might be possible to improve the bound in this case.

*Example 4.6* Consider the Picard curve

$$Y : y^3 = f(x) = x^4 + 14x^2 + 72x - 41$$

over  $K := \mathbb{Q}_5^{\text{nr}}$ . We claim that  $Y$  has semistable reduction over  $K$ , and that the reduction type is as in Case (b) of Theorem 4.2. Therefore,  $f_5 = 0$ .

We will argue in a similar way as in Sect. 3.3, see in particular Example 3.9, see also [3, § 6 and § 7]. The first observation is that

$$f = x^4 + 14x^2 + 72x - 41 \equiv (x + 3)^2(x^2 + 4x + 1) \pmod{5}. \tag{30}$$

By Hensel's Lemma,  $f$  has two distinct roots  $\alpha_1, \alpha_2 \in \mathcal{O}_K$  with  $\alpha_i^2 + 4\alpha_i + 1 \equiv 0 \pmod{5}$ . The other two roots of  $f$  are congruent to  $-3 \pmod{5}$ . Substituting  $x = -58 + 5^3x_1$  into  $f$ , we see that

$$f \equiv 5^6(3x_1^2 + 4x_1 + 2) \pmod{5^7}. \quad (31)$$

It follows that  $f$  has two more roots  $\alpha_3, \alpha_4 \in K$  of the form  $\alpha_i = -58 + 5^3\beta_i$ , with  $\beta_i \in \mathcal{O}_K$  and  $3\beta_i^2 + 4\beta_i + 2 \equiv 0 \pmod{5}$ . So  $f$  splits over  $K$ .

Let  $(\mathcal{X}, \mathcal{D})$  be the stably marked model of  $(X, D)$ , where  $X = \mathbb{P}_K^1$  and  $D = \{\infty, \alpha_1, \dots, \alpha_4\}$ . The calculation of the  $\alpha_i$  above shows that  $\mathcal{X}$  is the  $\mathcal{O}_K$ -model of  $X$  corresponding to the set of valuations  $\{v_0, v_1\}$ , where  $v_0$  (resp.  $v_1$ ) is the Gauss valuation on  $K(x)$  with respect to the parameter  $x$  (resp. to  $x_1$ ). Let  $\mathcal{Y}$  be the normalization of  $\mathcal{X}$  in the function field of  $Y$ . We claim that the special fiber  $\bar{Y}$  of  $\mathcal{Y}$  consists of two irreducible components  $W_0, W_1$  of geometric genus 2 and 1, respectively. By the same argument as in Example 3.9, this already implies that  $\mathcal{Y}$  is semistable and that the special fiber is as in Case (b) of Theorem 4.2.

To prove the claim it suffices to find generic equations for  $W_0$  and  $W_1$ . For  $W_0$  we just have to reduce the original equation for  $Y$  modulo 5. By (30) we obtain

$$W_0 : \bar{y}^3 = (\bar{x} + 3)^2(\bar{x}^2 + 4\bar{x} + 1),$$

which shows that  $g(W_0) = 2$ . For  $W_1$  we write  $f$  as a polynomial in  $x_1$ , substitute  $y = 5^2w$ , divide by  $5^6$  and reduce modulo 5. By (31) we obtain

$$W_1 : \bar{w}^3 = 3\bar{x}_1^2 + 4\bar{x}_1 + 2,$$

which shows that  $g(W_1) = 1$ . Now everything is proved.  $\square$

*Remark 4.7* The example above is again rather special, since  $f_5 = 0$  even though  $Y$  has bad reduction at  $p = 5$ . (See also Definition 5.4.)

## 5 Searching for Picard Curves over $\mathbb{Q}$ with Small Conductor

In this last section we briefly address the problem of constructing Picard curves with small conductor. We think this is an interesting problem which deserves further investigation. The main background result here is the *Shafarevich conjecture* (which is a theorem due to Faltings). We use this theorem via the following corollary.

**Theorem 5.1 (Faltings)** *Fix a number field  $K$  and an integer  $g \geq 2$ .*

- (a) *For any finite set  $S$  of finite places of  $K$  there exist at most a finite number of isomorphism classes of smooth projective curves of genus  $g$  over  $K$  with good reduction outside  $S$ .*

(b) For any constant  $N > 0$  there exists at most a finite number of isomorphism classes of curves of genus  $g$  over  $K$  with conductor  $\leq N$ .

*Proof* Satz 6 in [6] states that there are at most a finite number of  $d$ -polarized abelian varieties of dimension  $g$  over  $K$  with good reduction outside  $S$ , for fixed  $K, g, d$  and  $S$ . Statements (a) and (b) follow from this. For (a), one simply uses Torelli’s theorem (see [6, p. 365, Korollar 1]). To deduce (b) we use that the conductor of a curve  $Y$  is the same as the conductor of its Jacobian, and that an abelian variety over  $K$  has bad reduction at a finite place  $\mathfrak{p}$  of  $K$  if and only if  $f_{\mathfrak{p}} > 0$  (see e.g. [22, Theorem 1]).  $\square$

Unfortunately, no effective proof of Theorem 5.1 is known in general.<sup>3</sup> However, for some special classes of curves effective proofs are known, see e.g. [25].

The problem we wish to discuss here is whether the statement of Theorem 5.1 can be made computable in the case of Picard curves. More precisely: given a finite set  $S$  of rational primes (or a bound  $N > 0$ ), can we compute the finite set of curves with good reduction outside  $S$  (resp. with conductor  $\leq N$ )? Note that this is not equivalent to (and may be much easier than) having an effective proof of Theorem 5.1 for Picard curves. For the first problem, the answer is known to be affirmative:

**Proposition 5.2** *There exists an algorithm which, given as input a number field  $K$  and finite set  $S$  of finite places of  $K$ , computes the set of isomorphism classes of all Picard curves  $Y/K$  with good reduction outside  $S$ .*

*Proof* This is the main result of [12]. The algorithm is an adaption to Picard curves of the algorithm given by Smart for hyperelliptic curves, see [24]. The idea is that it suffices to determine the finite set of equivalence classes of binary forms of degree 4 over  $K$  whose discriminant is an  $S$ -unit (corresponding to the polynomial  $f(x)$ ). The latter problem can be reduced to solving an  $S$ -unit equation, for which effective algorithms are known.  $\square$

*Example 5.3* Let  $K = \mathbb{Q}$  and  $S = \{3\}$ . Then there are precisely 63 isomorphism classes of Picard curves over  $\mathbb{Q}$  with good reduction outside  $S$ . See [12].

For example, the curve

$$Y : y^3 = f(x) = x^4 - 3x^3 - 24x^2 - x$$

has good reduction outside  $S = \{3\}$  (the discriminant of  $f$  is  $\Delta(f) = 3^{10}$ ). The stable reduction  $\bar{Y}$  of  $Y$  at  $p = 3$  is as in Case (c) of Theorem 3.2, the exponent conductor is  $f_3 = 10$  (see [1, Appendix A1.1]). This is the lowest value for the conductor which occurs for the curves in the list of [12]. The conductor exponents of all 63 Picard curves from [12] have been computed in [1, Appendix A1.2]. From this calculation it follows that the conductor exponent  $f_3$  only takes the values  $f_3 = 10, 11, 12, 13, 15, 17, 19, 21$ .

---

<sup>3</sup>The precise meaning of an *effective proof* is that it provides an explicitly computable bound on the height of the curve or abelian variety in question.

The upper bound on the conductor exponent from abelian varieties of genus 3 from [4], Theorem 6.2 yields  $f_3 \leq 21$ . The result stated above therefore implies that this bound is also obtained for Picard curves.

Unfortunately we do not know any algorithm for solving (b), i.e. for finding all Picard curves with bounded conductor. The reason that the method for (a) does not solve (b) is the existence of *exceptional primes*.

**Definition 5.4** Let  $Y$  be a Picard curve over  $\mathbb{Q}$  and  $p$  a prime number. Then  $p$  is called *exceptional* with respect to  $Y$  if  $Y$  has bad reduction at  $p$  and  $f_p = 0$  (the latter means that the Jacobian of  $Y$  has good reduction at  $p$ ).

Exceptional primes are rather rare. It can easily be shown, using the arguments from this paper, that if  $p$  is an exceptional prime for  $Y$  then the splitting field of the polynomial  $f$  is unramified at  $p$ , and

$$\text{ord}_p(\Delta(f)) \in \{6, 12\}.$$

*Example 5.5* We consider the Picard curve over  $\mathbb{Q}$

$$Y : y^3 = f(x) = x^4 + 14x^2 + 72x - 41.$$

The discriminant of  $f$  is  $\Delta(f) = -2^{10}3^45^6$ . So  $Y$  has good reduction outside  $S = \{2, 3, 5\}$ . We have shown in Example 4.6 that  $f_5 = 0$ , i.e. that 5 is an exceptional prime. Using the methods of [3] and [18] one can prove that  $f_2 = 19$  and  $f_3 = 13$  (see e.g. this SageMathCloud worksheet: <http://tinyurl.com/hp3qzmo>, [19]). All in all, the conductor of  $Y$  is

$$N_Y = 2^{19}3^{13} = 835884417024.$$

Although  $S$  is small and  $p = 5$  is an exceptional prime,  $N_Y$  is relatively large. We have tried but were not able to find a similar example with exceptional primes and a significantly smaller conductor. Nevertheless, the fact that exceptional primes exist means that we cannot easily bound the size of the set  $S$  while searching for Picard curves with bounded conductor.

Here is an example of a Picard curve with a relatively small conductor.

*Example 5.6* Consider the Picard curve

$$Y/\mathbb{Q} : y^3 = f(x) = x^4 - 1.$$

The discriminant of  $f$  is  $\Delta(f) = -256 = -2^8$ . It follows that  $Y$  has good reduction outside  $S = \{2, 3\}$ . By [1, § 5.1.3], we have  $f_2 = 6$  and  $f_3 = 6$ . Therefore,

$$N_Y = 2^63^6 = 46656.$$



The first named author has made an extensive search for Picard curves over  $\mathbb{Q}$  with small conductor [1, § 5.3]. Among all computed examples, the curve  $Y$  was the one with the smallest conductor.

A remarkable property of the curve  $Y$  is that for every (rational) prime  $p$  it admits a map to  $\mathbb{P}^1$  of order prime to  $p$ , which becomes Galois over an extension: besides the degree-3 map  $\phi$  given by  $(x, y) \mapsto x$ , we have the map  $(x, y) \mapsto y$ , which has degree 4. In fact, the full automorphism group of  $Y$  has order 48, and is maximal in the sense that  $Y/\text{Aut}_{\mathbb{C}}(Y)$  is a projective line, and the natural cover is branched at three points.

It is instructive to compare the above example with the curve

$$Y' : y^3 = x^4 + 1.$$

This is a twist of  $Y$ . The curve  $Y$  and  $Y'$  become isomorphic over  $\mathbb{Q}[i]$ , yet have different conductors. In fact,

$$N_{Y'} = 2^{16}3^6,$$

see [1, § 5.1.2].

We propose to study the following problem.

**Problem 5.7** Prove that the curve from Example 5.6 is the only Picard curve (up to isomorphism) with conductor  $N_Y \leq 46,656$ , or find explicit counterexamples.

Proposition 5.2 and our main results (Theorems 3.6 and 4.4) suggest the following strategy for construction Picard curves with small conductor and thereby finding counterexamples. If we ignore the possibility of exceptional primes, a Picard curve with conductor  $\leq 2^6 3^6$  must have good reduction outside  $S$ , where  $S$  is one of the following sets:

- $\{2, 3, p\}$ ,  $p \leq 13$ ,
- $\{3, p\}$ ,  $p \leq 23$ .

To find all such curves looks challenging but within reach. It should also be very useful to take into account the local restrictions on the polynomial  $f$  imposed by our results on curves with a specific value for  $f_p$ . On the other hand, without an effective proof of Theorem 5.1(b) for Picard curves, it is not clear at the moment how one could actually prove that the curve from Example 5.6 (or any other curve we may find) has minimal conductor.

## References

1. M. Börner,  $L$ -functions of curves of genus  $\geq 3$ . Ph.D. thesis, Universität Ulm, 2016, <http://dx.doi.org/10.18725/OPARU-4137>
2. I.I. Bouw, S. Wewers, Semistable reduction of curves and computation of bad Euler factors of  $L$ -functions. Notes for a minicourse at ICERM (2015), [https://www.uni-ulm.de/fileadmin/website\\_uni\\_ulm/mawi.inst.100/mitarbeiter/wewers/course\\_notes.pdf](https://www.uni-ulm.de/fileadmin/website_uni_ulm/mawi.inst.100/mitarbeiter/wewers/course_notes.pdf)
3. I.I. Bouw, S. Wewers, Computing  $L$ -functions and semistable reduction of superelliptic curves. *Glasg. Math. J.* **59**, 77–108 (2017)
4. A. Brumer, K. Kramer, The conductor of an abelian variety. *Compos. Math.* **92**(2), 227–248 (1994)
5. P. Deligne, D. Mumford, The irreducibility of the space of curves of given genus. *Publ. Math. IHES* **36**, 75–109 (1969)
6. G. Faltings, Endlichkeitssätze für abelsche Varietäten über Zahlkörpern. *Invent. Math.* **73**, 349–366 (1983)
7. J. Harris, D. Mumford, On the Kodaira dimension of the moduli space of curves. *Invent. Math.* **67**, 23–86 (1982)
8. R.P. Holzapfel, *The Ball and Some Hilbert Problems* (Birkhäuser, Basel, 1995)
9. K. Koike, A. Weng, Construction of CM Picard curves. *Math. Comput.* **74**(249), 499–518 (2005)
10. Q. Liu, Conducteur et discriminant minimal de courbes de genre 2. *Compos. Math.* **94**(1), 51–79 (1994)
11. Q. Liu, *Algebraic Geometry and Arithmetic Curves* (Oxford University Press, Oxford, 2006)
12. B. Malmskog, C. Rasmussen, Picard curves over  $\mathbf{Q}$  with good reduction away from 3. *LMS Comput.* (2016). arXiv:1407.7892
13. E. Picard, Sur des fonctions de deux variables indépendantes analogues aux fonctions modulaires. *Acta Math.* **2**(1), 114–135 (1883)
14. R. Pries, Wildly ramified covers with large genus. *J. Number Theory* **119**(2), 194–209 (2006)
15. J.R. Quine, Jacobian of the Picard curve, in *Extremal Riemann Surfaces (San Francisco, CA, 1995)*. Contemporary Mathematics, vol. 201 (American Mathematical Society, Providence, RI, 1997), pp. 33–41
16. M. Raynaud, Spécialisation des revêtements en caractéristique  $p > 0$ . *Ann. Sci. Éc. Norm. Supér.* **32**(1), 87–126 (1999)
17. J. Rüth, Models of curves and valuations. Ph.D. thesis, Universität Ulm, 2014, <http://dx.doi.org/10.18725/OPARU-3275>
18. J. Rüth, S. Wewers, Semistable reduction of superelliptic curves of degree  $p$  (in preparation)
19. I. SageMath, SageMathCloud Online Computational Mathematics (2016), <https://cloud.sagemath.com/>
20. J.P. Serre, *Corps Locaux*, Troisième édition (Hermann, Paris, 1968). Publications de l'Université de Nancago, No. VIII
21. J.P. Serre, Facteurs locaux des fonctions zêta des variétés algébriques (définitions et conjectures). Séminaire Delange-Pisot-Poitou (Théorie des Nombres) **19**(2), 1–15 (1969)
22. J.P. Serre, J. Tate, Good reduction of abelian varieties. *Ann. Math.* **88**(3), 492–517 (1968)
23. J.H. Silverman, *Advanced Topics in the Arithmetic of Elliptic Curves*. Graduate Text in Mathematics, vol. 151 (Springer, New York, 1994)
24. N.P. Smart,  $S$ -unit equations, binary forms and curves of genus 2. *Proc. Lond. Math. Soc.* **75**(2), 271–307 (1997)
25. R. von Känel, An effective proof of the hyperelliptic Shafarevich conjecture. *J. Théor. Nombres Bordeaux* **26**(2), 507–530 (2014)
26. S. Wewers, Deformation of tame admissible covers of curves, in *Aspects of Galois Theory*, ed. by H. Völklein. LMS Lecture Note Series, vol. 256 (Cambridge University Press, Cambridge, 1999), pp. 239–282



Winfried Bruns, Richard Sieg, and Christof Söger

**Abstract** In this article we describe mathematically relevant extensions to Normaliz that were added to it during the support by the DFG SPP “Algorithmische und Experimentelle Methoden in Algebra, Geometrie und Zahlentheorie”: nonpointed cones, rational polyhedra, homogeneous systems of parameters, bottom decomposition, class groups and systems of module generators of integral closures.

**Keywords** Hilbert basis • Hilbert series • Rational cone • Rational polyhedron • Bottom decomposition • Class group • Triangulation • Linear diophantine system

**Subject Classifications** 52B20, 13F20, 14M25, 91B12

## 1 Introduction

The software package Normaliz [13] has been developed by the algebra and discrete mathematics group at Osnabrück since 1998. It is a tool for the computation of lattice points in rational polyhedra. Meanwhile it has been cited about 130 times in the literature (see [13]) with applications to algebraic geometry, commutative algebra, polytope theory, integer programming, combinatorial topology, group theory, theoretical physics and other areas. There exist interfaces to the major computer algebra systems CoCoA [1, 2], GAP [19], Macaulay2 [18] and Singular [17] and to polymake [20], a comprehensive tool for the computation of polytopes. Normaliz is used by other packages, notably by Regina [16], a tool for the exploration of 3-manifolds, and by SecDec [6] in the computation of multiscale integrals.

---

W. Bruns (✉) • R. Sieg  
Universität Osnabrück, Institut für Mathematik, 49069 Osnabrück, Germany  
e-mail: [wbruns@uos.de](mailto:wbruns@uos.de); [richard.sieg@uos.de](mailto:richard.sieg@uos.de)

C. Söger  
Alter Mühlenweg 1, 49504 Lotte, Germany  
e-mail: [csoeger@uos.de](mailto:csoeger@uos.de)

During the second half of the SPP 1489 “Algorithmische und Experimentelle Methoden in Algebra, Geometrie und Zahlentheorie” Normaliz was supported by the SPP. In this article we want to give an overview of those developments during the period of support that concern important mathematical aspects. For the mathematical background and unexplained terminology we refer the reader to Bruns and Gubeladze [7].

The main algorithms of Normaliz have been documented in the papers by Bruns with Koch [10], Ichim [9], Hemmecke, Ichim, Köppe and Söger [12], Söger [11] and Ichim and Söger [14]. See [14] for the performance of Normaliz on its main tasks.

Let  $A$  be a  $e \times d$  matrix with integer entries, and  $a \in \mathbb{Z}^e$ . Then the set

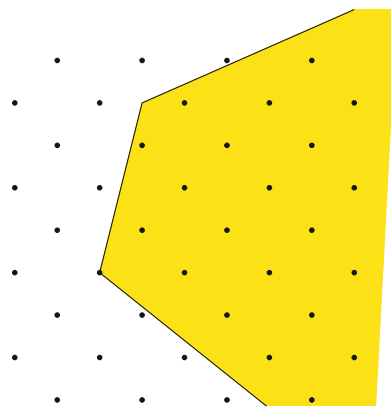
$$P = \{x \in \mathbb{R}^d; Ax \geq a\} \quad (1)$$

is called a *rational polyhedron*. Moreover, let  $B$  be a  $f \times d$  matrix of integers,  $b \in \mathbb{Z}^f$ ,  $C$  be a  $g \times d$  matrix of integers and  $c, m \in \mathbb{Z}^g$ . Then

$$L = \{x \in \mathbb{Z}^d : Bx = b, Cx \equiv c(m)\} \quad (2)$$

is an *affine sublattice* in  $\mathbb{R}^d$ , and it is the task of Normaliz to “compute” the set  $P \cap L$  (Fig. 1). So Normaliz can be considered as a tool for solving linear diophantine systems of inequalities, equations and congruences. Rational polyhedra and affine lattices can also be, and often are, described in terms of parametrizations or generators, and the conversion between the two descriptions for  $P$  and  $L$  separately is a basic task prior to the computation of  $P \cap L$ .

**Fig. 1** Lattice points in a polyhedron



The main computation goals of Normaliz are

*Generation*: find a (finite) system of generators of  $N = P \cap L$ ;

*Enumeration*: Compute the Hilbert series

$$H_N(t) = \sum_{x \in N} t^{\deg x}$$

with respect to a grading on  $\mathbb{Z}^d$ .

Of course, *Generation* must be explained, and *Enumeration* must even be modified somewhat to make sense in the general case.

The core case for Normaliz computations is the homogeneous one, in which the vectors  $a, b, c$  in (1) and (2) are 0, under the additional assumption that  $P$ , which in the homogeneous case is a cone  $C$ , is pointed, i.e. it does not contain a nontrivial linear subspace. The affine lattice  $L$  is then a subgroup of  $\mathbb{Z}^d$ , and for *Generation* Normaliz must compute a Hilbert basis of the monoid  $M = C \cap L$ , i.e., a minimal system of generators of the monoid  $M$ .

For a long time Normaliz could only handle homogeneous systems in the pointed case. These restrictions have been removed in two steps: version 2.11.0 (April 2014) introduced inhomogeneous systems and version 3.1.0 (February 2016) finally removed the condition that  $P$  has a vertex. These extensions will be discussed in Sects. 3 and 4, where also *Generation* and *Enumeration* will be made precise.

The Hilbert series is (the Laurent series expansion of) a rational function of type

$$H_N(t) = \frac{Q(t)}{(1 - t^{g_1}) \cdots (1 - t^{g_r})}$$

with a Laurent polynomial  $Q(t) \in \mathbb{Z}[t, t^{-1}]$ . In the general case there is no canonical choice for the exponents  $g_1, \dots, g_r$  in the denominator. One good possibility is to take them as the degrees of, in the language of commutative algebra, a homogeneous system of parameters (hsop). Such degrees can be found if one analyzes the face lattice of the recession cone of the system (1) and (2); the cone of solutions of the associated homogeneous system. This approach will be developed in Sect. 7. The option to use an hsop was introduced in version 3.1.2 (September 2016).

The primal algorithm of Normaliz is based on triangulations. A critical magnitude for the algorithm is the sum of the determinants of the simplicial cones in the triangulation. Since version 3.0.0 (September 2015) this determinant sum can be optimized by using a bottom decomposition. In Sect. 5 we explain how a bottom decomposition can be computed.

A normal affine monoid  $M$  has a well-defined class group. By a theorem of Chouinard (see [7, 4.F]) it coincides with the class group of the monoid algebra  $K[M]$  for an arbitrary field  $K$ . Since version 3.0.0 Normaliz computes the class group, as explained in Sect. 8.

The primal algorithm of Normaliz finds the Hilbert basis by first computing a system of generators of  $M$  as a module (in a natural way) over an input (or precomputed) monoid  $M_0$ . Therefore it can be used to find a minimal system of module generators of  $M$  over  $M_0$ . In more picturesque language these generators are

called “fundamental holes” of  $M_0$ . See Kohl et al. [22] for a package making use of this Normaliz feature.

There are several other extensions and options that have been introduced during the support of the Normaliz project by the SPP:

1. new input format (with backward compatibility),
2. standard sorting of vector lists in the output,
3. completely revised linear algebra with permanent overflow check,
4. automatic choice of integer type (64 bit or infinite precision),
5. computation of integer hulls as an option,
6. refinement of the triangulation to a disjoint decomposition,
7. subdivision of “large” simplicial cones by using SCIP [3] or approximation methods (see Bruns et al. [15]),
8. a normality test that avoids the computation of the full Hilbert basis,
9. improvement of the Fourier-Motzkin algorithm in connection with pyramid decomposition (see [14]),
10. revision of the dual algorithm,
11. various improvements in the algorithms that save memory and computation time,
12. improvements in NmzIntegrate (see Bruns and Söger [11]).

The file CHANGELOG in the Normaliz distribution gives an overview of the evolution.

The package HeLP [5] is an example for the application of Normaliz in another project of the SPP.

## 2 The Normaliz Primal Algorithm

The heart of Normaliz are two algorithms. The *primal algorithm* can be applied both to *Generation* and *Enumeration*. Among the two it is the considerably more complicated one. The *dual algorithm* can only be used for *Generation*. We refer the reader to Bruns and Ichim [9] for its description.

Since some details of the primal algorithm play a role in the following, we include a brief outline. The primal algorithm starts from a pointed rational cone  $C \subset \mathbb{R}^d$  given by a system of generators  $x_1, \dots, x_n$  and a sublattice  $L \subset \mathbb{Z}^d$  that contains  $x_1, \dots, x_n$ . (Other types of input data are first transformed into this format.) The algorithm is composed as follows:

1. Initial coordinate transformation to  $E = L \cap (\mathbb{R}x_1 + \dots + \mathbb{R}x_n)$ ;
2. Fourier-Motzkin elimination computing the support hyperplanes of  $C$ ;
3. pyramid decomposition and computation of the lexicographic triangulation  $\Delta$ ;
4. evaluation of the simplicial cones in the triangulation:
  - (a) enumeration of the set of lattice points  $E_\sigma$  in the fundamental domain of a simplicial subcone  $\sigma$ ,
  - (b) reduction of  $E_\sigma$  to the Hilbert basis  $\text{Hilb}(\sigma)$ ,

- (c) Stanley decomposition for the Hilbert series of  $\sigma' \cap L$  where  $\sigma'$  is a suitable translate of  $\sigma$ ;
- 5. Collection of the local data:
  - (a) reduction of  $\bigcup_{\sigma \in \Delta} \text{Hilb}(\sigma)$  to  $\text{Hilb}(C \cap L)$ ,
  - (b) accumulation of the Hilbert series of the intersections  $\sigma' \cap L$ ;
- 6. reverse coordinate transformation to  $\mathbb{Z}^d$ .

The algorithm does not strictly follow this chronological order, but interleaves steps 2–5 in an intricate way to ensure low memory usage and efficient parallelization. The steps 2 and 5 are treated in [9]. Steps 3 and 4 are described in [14]; the translates  $\sigma'$  in 4c are chosen in such a way that  $C \cap L$  is the disjoint union of their lattice points.

In view of the initial and final coordinate transformations 1 and 6 it is no essential restriction to assume that  $\dim C = d$  and  $L = \mathbb{Z}^d$ , as we will often do in the following.

### 3 Nonpointed Cones and Nonpositive Monoids

In this section we discuss only the homogeneous situation in which the polyhedron  $P \subset \mathbb{R}^d$  is a cone  $C$  and the affine lattice  $L$  is a subgroup of  $\mathbb{R}^d$ . Since [7] contains an extensive treatment of the mathematical background, we content ourselves with a brief sketch and references to [7].

The basic finiteness result in polyhedral convex geometry is the theorem of Minkowski-Weyl [7, 1.15]. It shows that one can equivalently describe cones by generators or by inequalities.

**Theorem 3.1** *The following conditions are equivalent for a subset  $C$  of  $\mathbb{R}^d$ :*

1. *there exist (integer) vectors  $x_1, \dots, x_n$  such that  $C = \mathbb{R}_+x_1 + \dots + \mathbb{R}_+x_n$ ;*
2. *there exist linear forms (with integer coefficients)  $\sigma_1, \dots, \sigma_s$  on  $\mathbb{R}^d$  such that  $C = \{x \in \mathbb{R}^d : \sigma_i(x) \geq 0, i = 1, \dots, s\}$ .*

With the additional requirement of integrality in the theorem,  $C$  is called a *rational cone*. If  $\dim \mathbb{R}C = d$  and the number of linear forms is chosen to be minimal, the  $\sigma_i$  in the theorem are uniquely determined up to positive scalars, and they are even unique if we additionally require that the coefficients are coprime integers. In this case we call  $\sigma_1, \dots, \sigma_s$  the *support forms* of  $C$ . The map  $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^s, \sigma(x) = (\sigma_1(x), \dots, \sigma_s(x))$ , is called the *standard map* of  $C$ . Clearly,  $\sigma$  maps  $\mathbb{Z}^d$  to  $\mathbb{Z}^s$  in the rational case.

The conversion from generators to inequalities in the description of cones is usually called *convex hull computation* and the converse transformation is *vertex enumeration*. These two transformations are two sides of the same coin and algorithmically completely identical since they amount to the dualization of a cone.

While this is not the main task of Normaliz, it often outperforms dedicated packages. See the recent benchmarks by Assarf et al. [4] and Köppe and Zhou [21].

In view of the remarks in Sect. 2 we can assume that  $\dim C = d$  and that the subgroup  $L \subset \mathbb{Z}^d$  is  $\mathbb{Z}^d$  itself. Thus the task is to compute the monoid  $M = C \cap \mathbb{Z}^d$ . The basic finiteness result for such monoids is *Gordan's lemma* [7, 2.9]:

**Theorem 3.2** *There exist  $x_1, \dots, x_n \in \mathbb{R}^d$  such that  $M = \mathbb{Z}_+x_1 + \dots + \mathbb{Z}_+x_n$ .*

At this point it is useful to borrow some terminology from number theory. We call  $U(M) = \{x \in M : -x \in M\}$  the *unit group* of  $M$ . Clearly,  $U(M) = \{x \in M : \sigma(x) = 0\}$ . The unit group is the maximal subgroup of  $\mathbb{Z}^d$  that is contained in  $M$ . One calls  $M$  *positive* if  $U(M) = 0$ . Similarly, the *maximal linear subspace* of  $C$  is  $U(C) = \text{Ker } \sigma$ . It is not hard to see that the positivity of  $M$  is equivalent to the *pointedness* of  $C$ : one has  $U(C) = \mathbb{R}U(M)$ , and therefore  $U(M) = 0$  if and only if  $U(C) = 0$ .

An element  $x \in M \setminus U(M)$  is called *irreducible* if a decomposition  $x = y + z$  with  $y, z \in M$  is only possible with  $y \in U(M)$  or  $z \in U(M)$ . The role of the irreducible elements in the generation of  $M$  is illuminated by the following theorem [7, 2.14 and 2.26].

**Theorem 3.3** *Let  $M = C \cap \mathbb{Z}^d$ . Then the following hold:*

1. *every element  $x$  of  $M$  can be written in the form  $x = u + y_1 + \dots + y_m$  where  $u$  is a unit and  $y_1, \dots, y_m$  are irreducible;*
2. *up to differences by units, there exist only finitely many irreducibles in  $M$ ;*
3. *let  $H \subset M$ ; then the following are equivalent:*
  - a.  *$M = U(M) + \mathbb{Z}_+H$  and  $H$  is minimal with this property;*
  - b.  *$H$  contains exactly one element of each residue class of irreducibles modulo  $U(M)$ .*
4.  *$M \cong U(M) \oplus \sigma(M)$ .*

If  $H$  satisfies the equivalent conditions in statement 3 we call it a *Hilbert basis* of  $M$ . Clearly, together with a basis of the free abelian group  $U(M)$  the Hilbert basis gives a minimal finite description of  $M$ . Statement 4 shows that  $U(M)$  and  $\sigma(M)$  are independent of each other. Moreover, the submonoid of  $M$  generated by  $H$  is isomorphic to  $\sigma(M)$ .

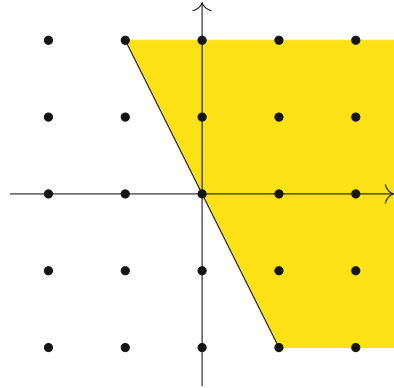
Note that  $H$  is uniquely determined if  $M$  is positive, so that we can denote it by  $\text{Hilb}(M)$ . In the general case  $H$  is a Hilbert basis of  $M$  if and only if  $\sigma(H) = \text{Hilb}(\sigma(M))$ . Therefore *Generation* can be split into two subtasks: (i) find  $U(M)$ , the kernel of the  $\mathbb{Z}$ -linear map  $\sigma|_{\mathbb{Z}^d}$  and (ii) find the Hilbert basis of the positive monoid  $\sigma(M)$ . The first task is a matter of solving a homogeneous diophantine system of linear equations, and the second is what Normaliz has done from its very beginnings.

The theory above can be developed for arbitrary affine monoids; see [7, Ch. 2]. However, the direct sum decomposition  $M \cong U(M) \oplus \sigma(M)$  is not always possible.

What we have described for Hilbert bases, applies similarly to extreme rays of cones. These are only defined modulo  $U(C)$  in the general case.



**Fig. 2** A nonpointed cone



Normaliz’ dual algorithm for the computation of Hilbert bases effectively does all its computations in the pointed cone  $\sigma(C)$ ; see [9]. Nevertheless, versions before 2.11.0 did not output the results if the cone was not pointed.

The primal algorithm could have been modified for Hilbert basis computations of nonpointed cones, but we do not see a way for the computation of Hilbert series in the nonpointed case. Moreover, the passage to the quotient modulo the maximal linear subspace reduces the dimension and therefore speeds up the computation. Let us look at a simple example (Fig. 2). The output shows:

```

1 Hilbert basis elements of degree 1:
0 1

0 further Hilbert basis elements of higher degree:

1 extreme rays:
0 1

1 basis elements of maximal subspace:
1 -2
    
```

Since in the vast majority of cases Normaliz is applied to positive monoids, Normaliz does not (always) try to compute  $U(M)$  beforehand—very likely it is 0. The computation of  $\sigma$  requires the computation of the support hyperplanes of  $C$ . Eventually these will be known, but their computation is inevitably intertwined with the computation of the triangulation, and would essentially have to be done twice. Therefore Normaliz takes the following “bold” approach in the primal algorithm:

1. Start the computation and proceed under the assumption that  $C$  is pointed.
2. As soon as the support hyperplanes have been computed, decide positivity.

3. If it should fail, throw an exception, perform the coordinate transformation to the pointed quotient, and restart the computation.

After *Generation* let us discuss *Enumeration*. A linear form  $\deg : \mathbb{Z}^d \rightarrow \mathbb{Z}$  is called a *grading* on  $M$  if  $\deg x \geq 0$  for all  $x \in M$  and  $\deg x > 0$  for  $x \in M \setminus U(M)$ . The *Hilbert series* of  $M$  with respect to  $\deg$  is the formal power series

$$H_M(t) = \sum_{x \in M} t^{\deg x}.$$

If  $M$  is positive there exist only finitely many elements in each degree, and the definition of  $H_M(t)$  makes sense. This is not the case if  $U(M) \neq 0$ —there exist already infinitely many elements of degree 0. Hence, if  $M$  is not positive, the only Hilbert series that we can associate to it, is that of  $\sigma(M)$ . In fact, since  $\deg(x) = 0$  for  $x \in U(M)$ ,  $\deg$  induces a grading on  $\sigma(M)$ : if  $\sigma(x) = \sigma(y)$ , then  $x - y \in U(M)$ , and so  $\deg x = \deg y$ . Therefore *Normaliz* (always) computes  $H_{\sigma(M)}(t)$ , and the invariants that depend on the Hilbert series are also computed for  $\sigma(M)$ .

## 4 Inhomogeneous Systems

In algebraic geometry one passes from an affine variety to a projective one by *homogenization*, and the same technique is used in discrete convex geometry to reduce algorithms for polyhedra to algorithms for cones. Let  $P \subset \mathbb{R}^d$  be an arbitrary polyhedron. Then the *cone over  $P$*  is the *closed set*

$$C(P) = \overline{\mathbb{R}_+(P \times \{1\})} \subset \mathbb{R}^{d+1}.$$

This amounts to passing from an inhomogeneous system to a homogeneous one by introducing a homogenizing variable, the  $(d + 1)$ th coordinate. Setting the homogenizing variable equal to 1, we get the inhomogeneous system back. In fact, it is not hard to see that one obtains a system of inequalities for  $C(P)$  by homogenizing such a system for  $P$  and adding the inequality  $x_{d+1} \geq 0$ .

If we set the homogenizing variable equal to 0 we obtain the *associated homogeneous system*, and its solution set is called the *recession cone* in our case:

$$\text{rec}(P) = \{x \in \mathbb{R}^d : (x, 0) \in C(P)\}.$$

It is useful to introduce the *level* of a point  $x \in \mathbb{R}^{d+1}$ ,

$$\text{lev}(x) = x_{d+1}.$$

By (de)homogenizing the Minkowski-Weyl theorem 3.1 one arrives at *Motzkin's theorem*; see [7, 1.27]:

**Theorem 4.1** *Let  $P$  be a nonempty subset of  $\mathbb{R}^d$ . Then the following are equivalent:*

1.  $P$  is a polyhedron;
2. there exist a nonempty polytope  $Q$  and a cone  $C$  such that  $P = Q + C$ .

A polytope is a bounded polyhedron; a special case of Theorem 4.1 is Minkowski’s theorem:  $P$  is a polytope if and only if  $P$  is the convex hull of finitely many points.

For the cone  $C$  in the theorem one has no choice:  $C = \text{rec}(P)$ . The polytope  $P$  is unique only if it is chosen minimal and  $\text{rec}(P)$  is pointed. In this case it must be the convex hull of the vertices of  $P$ . In the general case the vertices, like the extreme rays of cones, are only defined modulo the maximal linear subspace  $U(\text{rec}(P))$ .

One can interpret Theorem 4.1 as saying that polyhedra are finitely generated:  $Q$  is the convex hull of finitely many points, and the cone  $C$  is finitely generated. Finite generation holds also for lattice points, as we will see now.

In the same way as polyhedra, one homogenizes an affine lattice: from  $L \subset \mathbb{Z}^d$  one passes to the subgroup  $\bar{L}$  of  $\mathbb{Z}^{d+1}$  generated by  $L \times \{1\}$ . Normaliz goes this way, and then reduces the situation to the case  $\bar{L} = \mathbb{Z}^{d+1}$  by preliminary coordinate transformations. For simplicity we will therefore assume that  $\bar{L} = \mathbb{Z}^{d+1}$ .

We want to compute the set  $N = P \cap \mathbb{Z}^d$ . The homogenization of  $N$  is the monoid  $M = C(P) \cap \mathbb{Z}^{d+1}$ . By analogy with  $\text{rec}(P)$  we define the *recession monoid*

$$\text{rec}(N) = \{x \in \mathbb{Z}^d : (x, 0) \in M\}.$$

**Theorem 4.2** *Suppose that  $N \neq \emptyset$ .*

1. Then there exist finitely many lattice points  $y_1, \dots, y_m \in N$  such that

$$N = \bigcup_{i=1}^m x_i + \text{rec}(N).$$

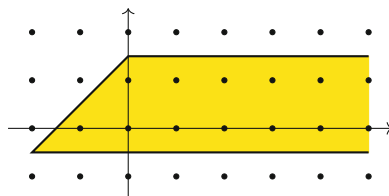
2. The number  $m$  is minimal if and only if there exists a Hilbert basis  $H$  of  $M$  such that

$$\{y_1, \dots, y_m\} = \{y \in \mathbb{Z}^d : (y, 1) \in H\}.$$

3. If  $H$  is a Hilbert basis of  $M$ , then  $\{x \in \mathbb{Z}^d : (x, 0) \in H\}$  is a Hilbert basis of  $\text{rec}(N)$ .

Part 1 is [7, 2.12], and the statements about Hilbert bases are easy to prove. The theorem entitles us to call  $N$  a *finitely generated module* over  $\text{rec}(N)$ . The computation goal *Generation* can now be made precise in the inhomogeneous case as well: compute a Hilbert basis of  $\text{rec}(N)$  and a *system of module generators*  $\{y_1, \dots, y_m\}$ . By the theorem it is enough to compute a Hilbert basis of  $M$ . However, it would be foolish to overlook the shortcut that is possible: all candidates  $x \in M$

Fig. 3 A polyhedron in  $\mathbb{R}^2$



with  $\text{lev}(x) > 1$  can be immediately discarded. This holds both for the primal and the dual algorithm of Normaliz. (The primal algorithm does only produce elements  $x$  with  $\text{lev}(x) \geq 0$ . For the dual algorithm that processes the inequalities defining  $C(P)$  one must start with the inequality  $\text{lev}(x) \geq 0$ .)

As a simple example we consider the polyhedron in Fig. 3.

Normaliz writes the results in homogenized coordinates:

```

2 module generators:
-1 0 1
0 1 1

1 Hilbert basis elements of recession monoid:
1 0 0
    
```

The result can be checked by inspection.

The set  $N$  has a *disjoint* decomposition into residue classes mod  $G = \text{gp}(\text{rec}(N))$  (where  $\text{gp}(M)$  is the group generated by  $M$ ):

$$N = \bigcup_{i=1}^r N_i, \quad N_i \neq \emptyset, \quad N_i \cap N_j = \emptyset \text{ if } i \neq j, \quad x \equiv y \pmod G \text{ for all } x, y \in N_i.$$

If  $y_1, \dots, y_m$  is a system of module generators of  $N$  as an  $\text{rec}(N)$ -module, then obviously  $r \leq m$ ; in particular,  $r$  is finite. It is justified to call  $r$  the *module rank* of  $N$  over  $\text{rec}(N)$  because of the following functorial process. Let  $K$  be a field and let  $R = K[\text{rec}(N)]$  be the monoid  $K$ -algebra defined by  $\text{rec}(N)$ . Let  $K[N]$  be the  $K$ -vector space with basis  $N$ . The “multiplication”  $\text{rec}(N) \times N \rightarrow N, (x, y) \mapsto x + y$  makes  $K[N]$  a module over  $R$  [7, p. 51]. Since  $R$  is an integral domain,  $K[N]$  has a well-defined rank, which is exactly  $r$ , as one sees by passage to the field of fractions of  $R$ . An intermediate step of this passage is the Laurent polynomial ring  $L = K[G]$ , and we can get  $K[N] \otimes_R L$  by introducing  $K$ -coefficients to  $N + G$ . This set decomposes into the subsets  $N_i + G$ , and one has  $N_i + G = x + G$  for every  $x \in N_i$ . Therefore  $K[N] \otimes_R L$  is the direct sum of  $r$  free  $L$ -modules of rank 1. In the example above, the module rank is 2.

If  $y_1, \dots, y_m$  have been computed, then it is very easy to find the module rank  $r$ : we simply count their pairwise different residue classes modulo  $G$ . But we can also compute  $r$  as the number of lattice points in a polytope, and Normaliz resorts

to this approach if a system of module generators is unknown. The polytope is a cross-section of  $P$  with a complement of  $\text{rec}(P)$ :

**Theorem 4.3** *Let  $z_1, \dots, z_s$  be a  $\mathbb{Z}$ -basis of  $G = \text{gp}(\text{rec}(N))$ . There exist  $z_{s+1}, \dots, z_d \in \mathbb{Z}^d$  such that  $z_1, \dots, z_d$  is a  $\mathbb{Z}$ -basis of  $\mathbb{Z}^d$ . Set  $H = \mathbb{Z}z_{s+1} + \dots + \mathbb{Z}z_d$ , and let  $\pi : \mathbb{R}^d \rightarrow \mathbb{R}H$  denote the projection defined by  $\pi|_G = 0$  and  $\pi|_H = \text{id}_H$ .*

*Then  $\pi(P)$  is a (rational) polytope, and the module rank  $r$  is the number of lattice points in  $\pi(P)$ .*

*Proof* The first statement amounts to the existence of a complement  $H$  of  $G$  in  $\mathbb{Z}^d$ , i.e., a subgroup  $H$  with  $\mathbb{Z}^d = G + H$  and  $G \cap H = 0$ . Such a complement exists if and only if  $\mathbb{Z}^d/G$  is torsionfree. Let  $z \in \mathbb{Z}^d$  such that  $kz \in G$  for some  $k \in \mathbb{Z}, k > 0$ . Since  $G = \mathbb{R} \text{rec}(P) \cap \mathbb{Z}^d$ , we must have  $x \in G$ .

The polyhedron  $P$  is the Minkowski sum  $Q + \text{rec}(P)$  with a polytope  $Q$ . Since  $\text{rec}(P) \subset \mathbb{R}G$ , we have  $\pi(P) = \pi(Q)$ , and therefore  $\pi(P)$  is a polytope. Clearly, the lattice points in  $P$  are mapped to lattice points in  $\pi(P)$ , and two such points have the same image if and only if they differ by an element in  $G$ .

The only critical question is whether every lattice point in  $\pi(P)$  is hit by a lattice point in  $P$  by the application of  $\pi$ . There is nothing to show if  $G = 0$  since  $\pi$  is the identity on  $\mathbb{R}^d$  then. So assume that  $G \neq 0$ . Let  $p \in \pi(P)$  be a lattice point,  $p = \pi(q)$  with  $q \in P$ . One has  $\pi(p - q) = 0$ , and therefore  $p - q \in \mathbb{R}G$ . Note that  $\mathbb{R}G = \mathbb{R} \text{rec}(P)$ . In other words,  $\text{rec}(P)$  is a fulldimensional cone in  $\mathbb{R}G$ . It contains a lattice point  $x$  in its (relative) interior. Thus  $(p - q) + kx \in \text{rec}(P)$  for  $k \in \mathbb{Z}, k \gg 0$ , and  $q + (p - q) + kx \in \mathbb{Z}^d$  is a preimage of  $p$  in  $P$  for  $k \gg 0$ .

Let us now discuss *Enumeration* in the inhomogeneous case. As in the homogeneous case, we can only compute the Hilbert series of  $N = P \cap \mathbb{Z}^{d+1}$  modulo  $U(\text{rec}(N))$ . Therefore it is enough to discuss the case in which  $\text{rec}(P)$  or, equivalently,  $C(P)$  is pointed.

Normaliz computes the Hilbert series via a *Stanley decomposition*. This is a disjoint decomposition of the set of lattice points  $P \cap \mathbb{Z}^d$  into subsets of the form

$$D = u + \sum_{i=1}^r \mathbb{Z}_+ v_i$$

where  $r$  varies between 0 and  $\dim P$  and  $v_1, \dots, v_r$  are linearly independent. Provided  $\deg v_i > 0$  for  $i = 1, \dots, r$ , the Hilbert series of  $D$  is given by

$$H_D(t) = \frac{t^{\deg u}}{(1 - t^{\deg v_1}) \dots (1 - t^{\deg v_r})}. \tag{3}$$

In order to get the Hilbert series of  $P \cap \mathbb{Z}^d$ , it only remains to sum the Hilbert series of the components of the Stanley decomposition.

In[14] the computation of the Stanley decomposition in the homogeneous case is described in detail. Therefore we only discuss how to derive the Stanley decomposition of  $P \cap \mathbb{Z}^d$  from a Stanley decomposition of  $C(P) \cap \mathbb{Z}^{d+1}$ . We must

intersect all components of the Stanley decomposition of  $C(P) \cap \mathbb{Z}^{d+1}$  with the hyperplane  $L_1$  of level 1 points. Since the levels of all participating vectors are integral and  $\geq 0$ , in a sum of level 1 exactly one summand must have level 1 and the others must have level 0.

**Proposition 4.4** *Suppose that  $C(P)$  is pointed, and that  $D$  is a component in the Stanley decomposition of  $C(P)$ . Let  $v_1, \dots, v_e$  be the vectors of level 1 among  $v_1, \dots, v_r$ , and  $v_{e+1}, \dots, v_f$  those of level 0. Then the following hold:*

1. *if  $\text{lev}(u) = 1$ , then  $D \cap L_1 = u + \sum_{i=e+1}^f \mathbb{Z}_+ v_i$ .*
2. *If  $\text{lev}(u) = 0$ , then  $D \cap L_1$  is the disjoint union of the sets  $u + v_j + \sum_{i=e+1}^f \mathbb{Z}_+ v_i$ ,  $j = 1, \dots, e$  (and thus empty if  $e = 0$ ).*
3. *if  $\text{lev}(u) > 1$ , then  $D \cap L_1 = \emptyset$ .*

Note that  $f - e \leq \dim P$  if  $D \cap L_1 \neq \emptyset$ . The proposition shows that the computation of a Stanley decomposition of  $P \cap \mathbb{Z}^d$  is as easy (or difficult) as the computation for  $C(P) \cap \mathbb{Z}^{d+1}$ .

In the homogeneous case all degrees are nonnegative. In the inhomogeneous case this requirement would be an unnecessary restriction. Normaliz takes care of this aspect by computing a *shift*. For our simple example above we obtain with  $\text{deg}(x_1, x_2) = x_1$

```
Hilbert series:
1 1
denominator with 1 factors:
1: 1

shift = -1
```

Thus the Hilbert series is

$$t^{-1} \frac{1+t}{1-t} = \frac{t^{-1}+1}{1-t}.$$

Normaliz lets the user specify a linear form  $\delta$  that plays the role of the dehomogenization. This is already useful for compatibility with the input formats of other packages: often the first coordinate is used for (de)homogenization.

*Remark 4.5* Inhomogeneous systems are often created by strict linear inequalities  $\lambda(x) > 0$  where  $\lambda$  is linear (in addition to non-strict ones). These can be treated as inhomogeneous systems, but Normaliz also offers a variant called “excluded faces”. Then homogenization (with its increase in dimension) is avoided at the expense of an inclusion-exclusion approach. This variant can also be used by NnmzIntegrate.

## 5 Bottom Decomposition

As mentioned above, Normaliz computes a triangulation of the cone  $C$  whose rays are given by the input (or precomputed) system of generators, a partial triangulation for Hilbert bases and a full one for Hilbert series.

The complexity of the Normaliz algorithm depends mainly on two parameters. The first is the size of the triangulation. The second is the determinant sum (or normalized volume) that determines the time needed for the evaluation of the simplicial cones in the triangulation. In the following  $\text{vol}$  denotes the  $\mathbb{Z}^d$ -normalized volume in  $\mathbb{R}^d$ . It is the Euclidean volume multiplied by  $d!$ .

Let  $\sigma$  be a simplicial cone generated by linearly independent vectors  $v_1, \dots, v_d$ . Then the normalized volume of the *basic simplex* spanned by 0 and  $v_1, \dots, v_d$  is the absolute value of the determinant of the  $d \times d$ -matrix with rows  $v_1, \dots, v_d$ . Therefore we call it the *determinant*  $\det \sigma$  of  $\sigma$ . It is also the number of lattice points in the semi-open parallelotope

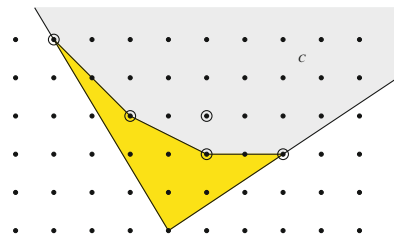
$$\text{par}(v_1, \dots, v_d) = \{a_1 v_1 + \dots + a_d v_d : 0 \leq a_i < 1, i = 1, \dots, d\},$$

which is also referred as the *fundamental domain* of  $\sigma$ . Normaliz must generate these points when evaluating  $\sigma$  for the Hilbert basis or Hilbert series. Therefore the determinant sum  $\text{detsum } \Sigma = \sum_{\sigma \in \Sigma} \det \sigma$  of  $\Sigma$  is a critical complexity parameter. In the following we explain how to optimize it.

**Definition 5.1** Let  $G \subset \mathbb{Z}^d$  be a finite set. We call the polyhedron  $\text{conv}^\wedge(G) = \{x \in \mathbb{R}^d : x = \sum_{g \in G} a_g g, a_g \geq 0, \sum_{g \in G} a_g \geq 1\}$  the *upper convex hull* of  $G$ . The *bottom*  $B(G)$  of  $G$  is the polyhedral complex of the compact facets of  $\text{conv}^\wedge(G)$  (or just their union).

Let  $C$  be the cone generated by  $G$ . Then  $\text{conv}^\wedge(G) = \text{conv}(G) + C$ , and  $B(G)$  is nonempty if and only if  $C$  is pointed, or, equivalently,  $\text{conv}^\wedge(G)$  has a vertex. In this case the bottom is indeed a set of polytopes of dimension  $\dim C - 1$  since their union is in bijective correspondence with a cross-section of  $C$ . Figure 4 illustrates the notion of bottom.

Fig. 4 The bottom



As usual, we assume from now on that  $C \subset \mathbb{R}^d$  is pointed and of dimension  $d$ , and that the monoid  $M = C \cap \mathbb{Z}^d$  is to be computed.

**Definition 5.2** The cones  $\mathbb{R}_+F$  where  $F$  runs through the facets in  $B(G)$  form the *bottom decomposition* of  $C$  with respect to  $G$ .

A triangulation  $\Sigma$  of  $C$  is a *bottom triangulation* with respect to  $G$  if every simplicial cone  $\sigma \in \Sigma$  is generated by elements of  $G \cap F$  where  $F$  is a facet of  $B(G)$ .

Bottom triangulations are optimal with respect to determinant sum:

**Proposition 5.3** *Let  $\Sigma$  be a bottom triangulation of  $C$  with respect to  $G$ . Then  $\text{detsum}(\Sigma) \leq \text{detsum}(\Delta)$  for all triangulations  $\Delta$  with rays in  $G$ .*

*Proof* The union of the basic simplices of  $\Sigma$  is the union of the polytopes  $\text{conv}(0, F)$  where  $F$  runs through the facets of  $B(G)$  (see Fig. 4). Therefore its determinant sum is the volume of the union  $D$  of these polytopes. But  $D$  is contained in the union of the basic simplices of the simplicial cones in  $\Delta$ , and therefore the volume of  $D$  bounds  $\text{detsum} \Delta$  from below.

Evidently, if the points of  $G$  lie in one hyperplane, all triangulations of  $C$  with rays through  $G$  have the same determinant sum, namely the normalized volume of the polytope  $\text{conv}(G, 0)$ . However, in general the determinant sums can differ widely. Therefore it makes sense to compute a bottom triangulation. First we determine the compact facets of  $\text{conv}^\wedge(G)$ . As usual, let us say that the facet  $F$  of the  $d$ -dimensional polyhedron  $Q \subset \mathbb{R}^d$  is *visible* from  $x \in \mathbb{R}^d$  if  $\lambda(x) < 0$  for the affine-linear form  $\lambda$  defining the hyperplane through  $F$  (normed such that  $\lambda(y) \geq 0$  for  $y \in Q$ .)

**Proposition 5.4** *Let  $F$  be a facet of  $\text{conv}^\wedge(G)$ . Then the following are equivalent:*

1.  $F$  belongs to  $B(G)$ ;
2.  $F$  is visible from 0.

*Proof* We choose  $\lambda$  as an affine-linear form defining  $F$  and a point  $x$  of  $F$ . Let  $H$  be the hyperplane spanned by  $F$ . Suppose first that  $\lambda(0) = 0$ . Then  $\lambda$  vanishes on the whole ray from 0 through  $x$ , and since this ray belongs to  $\text{conv}^\wedge(G)$  from  $x$  on, it is impossible that  $H \cap \text{conv}^\wedge(G)$  is compact. The assumption that  $\lambda(x) > 0$  implies that  $\lambda$  has negative values on this ray in points beyond  $x$ , and this is impossible as well. This proves  $1 \implies 2$ .

Conversely assume that  $F$  is visible from 0, but not compact. Then it is not contained in the compact polytope  $P = \text{conv}(G)$ . Let  $y$  be a point in  $F \setminus P$ ,  $y = \sum_{g \in G} a_g g$  with  $a = \sum a_g \geq 1$ , all  $a_g \geq 0$ . Then  $y/a \in P$ , and since  $\lambda(y) = 0$  and  $\lambda(y/a) \geq 0$ , it follows that  $\lambda(0) \geq 0$  since  $y/a$  lies between 0 and  $y$ . This is a contradiction.

Normaliz uses lexicographic triangulations (see [14]). These are uniquely determined by the order in which the elements are successively added in building the cone. Therefore we can triangulate  $\mathbb{R}_+F$  separately for all bottom facets  $F$  using



**Table 1** Effect of bottom decomposition

Input	Triangulation size	Determinant sum	Computation time
Inequalities	347,225,775,338	4,111,428,313,448	112:43:17 h
Inequalities, -b	288,509,390,884	1,509,605,641,358	84:26:19 h
Hilbert basis, -b	335,331,680,623	1,433,431,230,802	97:50:05 h

only points in  $G \cap F$ . These triangulations coincide on the intersections of the cones  $\mathbb{R}_+ F$  and can be patched to a triangulation of  $\mathbb{R}_+ C$ .

Normaliz does not blindly compute triangulations, taking the set  $G$  in the order in which it is given. In the presence of a grading it first orders the generating set by increasing degree, and this has already a strong effect on the determinant sum. Nevertheless, bottom decomposition can often improve the situation further.

If the Hilbert basis of  $C \cap \mathbb{Z}^d$  can be computed quickly by the dual algorithm, one can use it as input for a second run that computes the Hilbert series. (Since version 3.2.0, Normaliz tries to guess whether the primal or the dual algorithm is better for the given input, but the algorithm can also be chosen by the user.) It is clear that bottom decomposition with  $G$  being the Hilbert basis, produces the smallest determinant sum of any triangulation of  $C$  with rays through integer points. But the Hilbert basis has often many more elements than the set of extreme rays, and this can lead to a triangulation with a much larger number of simplicial cones. Despite of reducing the determinant sum, it may have a negative effect on computation time. The following example, a Hilbert series computation in social choice theory (input file `CondEffPlur.in` of the Normaliz distribution; see [9, 14] or Schürmann [23]), demonstrates the effect; see Table 1. With the input “inequalities”, Normaliz first computes the extreme rays and then applies the primal algorithm to compute the Hilbert series. The option `-b` forces bottom decomposition. The computation times were taken on a system equipped with 4 Xeon E5-2660 at 2.20 GHz, using 30 parallel threads.

At present Normaliz computes the bottom facets as suggested by Proposition 5.4. Since we must homogenize the polyhedron  $\text{conv}^\wedge(G)$ , this amounts to doubling the set  $G$  to  $G \times \{0\} \cup G \times \{1\} \in \mathbb{R}^{d+1}$ . The advantage of this approach is that one simultaneously computes the facets of  $C$  and the bottom facets. Nevertheless, the time spent on this computation can outweigh the saving by a smaller determinant sum. Therefore Normaliz only applies bottom decomposition if asked for by the user or if the bottom is very “rough”. Roughness is measured by the ratio of the largest degree of a generator and the smallest. At present bottom decomposition is activated if the roughness is  $\geq 10$ .

We will try to improve the efficiency of bottom decomposition by speeding up its computation. The following proposition suggests a potential approach:

**Proposition 5.5** *With the notation introduced above, let  $z \in C$ . Then the following are equivalent for a set  $F \subset \mathbb{R}^d$ :*

1.  $F$  is a facet of  $B(G)$ ;
2.  $F$  is a facet of  $\text{conv}(G) + \mathbb{R}_+ z$  that is visible from 0;

The easy proof is left to the reader. If one chooses  $z = 0$  in Proposition 5.5, then one must compute all facets of the polytope  $\text{conv}(G)$ , not only those in the bottom, but also those in the “roof”. Choosing  $z \neq 0$ , for example in the interior of  $C$ , “blows the roof off”, and it may be the better choice.

## 6 Integral Closure as a Module

Let  $M \subset \mathbb{Z}^d$  be a positive affine monoid,  $L \supset \text{gp}(M)$  a subgroup of  $\mathbb{Z}^d$ , and  $C$  the cone generated by  $M$ . Then  $\overline{M}_L = C \cap L$  is the integral closure of  $M$ . It is not only a finitely generated monoid itself, but also a finitely generated  $M$ -module: there exist  $y_1, \dots, y_m \in \overline{M}_L$  such that  $\overline{M}_L = \bigcup_{i=1}^m y_i + M$ . If  $M$  (and therefore  $\overline{M}_L$ ) is positive, then the set  $\{y_1, \dots, y_m\}$  is unique once it is chosen minimal. It contains 0 since  $M \subset \overline{M}_L = C \cap L$ .

Geometrically one can interpret the difference  $\overline{M}_L = C \cap L \setminus M$  as the set of “gaps” or “holes” of  $M$  in  $\overline{M}_L = C \cap L$ , and the nonzero elements of  $\{y_1, \dots, y_m\}$  are the “fundamental holes” in the terminology of [22]. Since version 3.0.0 Normaliz computes the set  $\{y_1, \dots, y_m\}$ , and therefore the fundamental holes.

In the following we assume  $L = \mathbb{Z}^d$ , and set  $\widetilde{M} = \overline{M}_{\mathbb{Z}^d}$ . (In [7]  $\widetilde{M}$  is reserved for the normalization  $\overline{M}_{\text{gp}(M)}$ .) Evidently the Hilbert basis elements of  $\widetilde{M}$  outside  $M$  belong to  $\{y_1, \dots, y_m\}$ , but in general this set is much larger than the Hilbert basis. Let  $M$  be the monoid generated by linearly independent vectors  $v_1, \dots, v_d$ . Then the lattice points in  $\text{par}(v_1, \dots, v_d)$  form a system of module generators of  $\widetilde{M}$ , but in general they do not all belong to the Hilbert basis; see Fig. 5 where  $C$  is generated by  $(2, 1)$  and  $(1, 3)$ . The Hilbert basis elements inside  $G$  are only  $(1, 1)$  and  $(1, 2)$ .

Since Normaliz computes the sets  $\text{par}(v_1, \dots, v_d)$  for the simplicial cones  $\mathbb{R}_+v_1 + \dots + \mathbb{R}_+v_d$  in a triangulation of  $C$  with rays in a given generating set of  $M$ , it is only a matter of restricting the “reducers” in the “global” reduction to elements of  $G$ .

**Proposition 6.1** *Let  $G \subset \mathbb{Z}^d$  generate the positive affine monoid  $M \subset \mathbb{Z}^d$ , and let  $\Sigma$  be a triangulation of  $C$  with rays in  $G$ . Then the union  $H$  of the sets  $\text{par}(\sigma) \cap \mathbb{Z}^d$ ,  $\sigma \in \Sigma$  generates the module  $\widetilde{M}$  over  $M$ .*

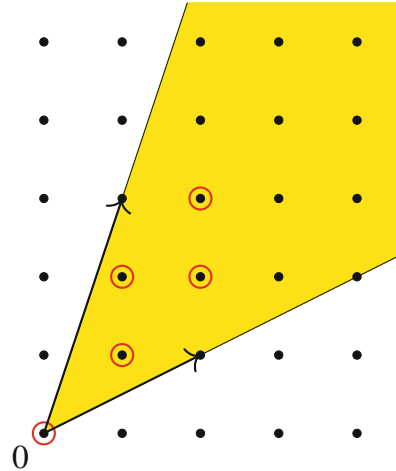
*An element  $y \in H$  belongs to the minimal generating set of  $\widetilde{M}$  if and only if  $y - x \notin C$ , for all  $x \in G$ ,  $x \neq 0$ .*

*Proof* Only the second statement may need a justification. We can of course assume that  $0 \notin G$ . Suppose first that  $z = y - x \in C$  for some  $x \in G$ . Then  $z \in \widetilde{M}$  and  $y + M \subset z + M$  so that  $y$  does not belong to the minimal generating set.

Conversely, if  $y - x \notin C$  for all  $x \in G$ , then there is no element  $z \in \widetilde{M}$ ,  $z \neq y$ , such that  $y \in z + M$ , and so  $y$  belongs to the minimal generating set.

Normaliz computes minimal sets of module generators not only in the discussed homogeneous case, but also in the inhomogeneous case in which the module is the set of lattice points in a polyhedron  $P$  and  $G$  generates  $\text{rec}(P)$  (since version 3.1.0).

**Fig. 5** Module generators of integral closure



## 7 Homogeneous Systems of Parameters

As above, we consider monoids  $M = C \cap L$  where  $C \subset \mathbb{R}^d$  is a rational pointed cone and  $L \subset \mathbb{Z}^d$  is a subgroup. We may right away assume that  $d = \dim C$  and  $L = \mathbb{Z}^d$ . Since we want to discuss Hilbert series, we need a grading  $\deg : \mathbb{Z}^d \rightarrow \mathbb{Z}$  such that  $\deg(x) > 0$  for  $x \in M, x \neq 0$ . Additionally we assume that  $\deg$  takes the value 1 on  $\text{gp}(M)$ , a standardization that Normaliz always performs. The following classical theorem shows that the Hilbert series can be expressed as a rational function.

**Theorem 7.1 (Ehrhart, Stanley, Hilbert-Serre)**

1. The Hilbert series  $H_M(t) = \sum_{x \in M} t^{\deg(x)}$  is (the power series expansion of) a rational function that can be written in the form

$$H_M(t) = \frac{Q(t)}{(1 - t^\ell)^d} \tag{4}$$

where  $Q(t) = 1 + h_1 t + \dots + h_s t^s$  is a polynomial of degree  $s < d\ell$  with nonnegative integer coefficients  $h_i$ , and  $\ell$  is the least common multiple of the degrees of the extreme integral generators of  $C$ .

2. There exists a (unique) quasipolynomial  $q_M(k)$  of degree  $d-1$  and period dividing  $\ell$  such that  $\#\{x \in M : \deg(x) = k\} = q_M(k)$  for all  $k > s - d\ell$ .

It is not difficult to derive the first claim from the existence of a Stanley decomposition so that  $H_M(t)$  is a sum of terms given by (3). This explains that all coefficients of the numerator polynomial are nonnegative. There is also an access via commutative algebra which we will explain below.

A quasipolynomial of period  $\pi > 0$  and degree  $g$  is a function  $q : \mathbb{Z} \rightarrow \mathbb{C}$  that can be represented in the form

$$q(k) = q_0^{(k)} + q_1^{(k)}k + \dots + q_g^{(k)}k^g$$

with  $q_i^{(k)} = q_i^{(j)}$  for all  $i$  whenever  $j \equiv k \pmod{\pi}$ ; moreover, one has  $q_g^{(k)} \neq 0$  for at least one  $k$  and  $\pi$  is chosen as small as possible. The quasipolynomial in Theorem 7.1 is called the *Hilbert quasipolynomial* of  $M$ .

We use the terms ‘‘Hilbert series’’ and ‘‘Hilbert quasipolynomial’’. One could equally well name these objects after Ehrhart. In fact, the Hilbert series of  $M$  is nothing but the Ehrhart series of the polytope that one obtains by intersecting  $C$  with the hyperplane of degree 1 elements in  $\mathbb{R}$ .

While Theorem 7.1 gives a representation of  $H_M(t)$  in which all parameters have a natural combinatorial description, it is not completely satisfactory since the denominator often has a very large degree and one can do better. It is our goal to find a representation of  $H_M(t)$  as a fraction whose

1. denominator is of the form  $(1 - t^{g_1}) \dots (1 - t^{g_d})$  and of small degree  $g_1 + \dots + g_d$  and such that
2. the coefficients of the numerator polynomial are nonnegative integers and have a combinatorial interpretation.

We will give an example showing that in general there is no canonical choice of the denominator. Nevertheless it makes sense to search for a good choice. Of course, if all extreme generators have degree 1, then the denominator of (4) is  $(1 - t)^d$ , and there is nothing to discuss.

By default Normaliz proceeds as follows: It reduces the fraction (4) to lowest terms and obtains a representation

$$H_M(t) = \frac{\widetilde{Q}(t)}{\zeta_{q_1}^{e_1} \dots \zeta_{q_u}^{e_u}}$$

with cyclotomic polynomials  $\zeta_k$ ,  $1 = q_1 < q_2 < \dots < q_u$ . Then it takes  $g_d$  as the lcm of all  $q_i$ , replaces their product by  $(1 - t^{g_d})$  and proceeds with then remaining cyclotomic factors etc. In this way the  $g_k$  express the periods of the coefficients in the Hilbert quasipolynomial:  $g_i$  is the lcm of the periods of the coefficients  $q_d, \dots, q_{d-i+1}$ . We will refer to the denominator of this representation as *standard denominator*. This choice is easy to compute and natural in its way, but not satisfactory if one wants a combinatorial interpretation of the coefficients in the numerator, as the following example shows.

Consider the cone  $C = \mathbb{R}_+(1, 2) + \mathbb{R}_+(2, 1)$  with the grading  $\text{deg}(x_1, x_2) = x_1 + x_2$  (known as the *total grading*). Then Hilbert series with standard denominator is:

$$H_M(t) = \frac{1 - t + t^2}{(1 - t)(1 - t^3)},$$

with coprime numerator and denominator, and the denominator even has the desired form  $(1 - t^{g_1})(1 - t^{g_2})$ . However, the numerator has a negative coefficient.

Commutative algebra suggests us to choose  $g_1, \dots, g_d$  as the degrees of the elements in a *homogeneous system of parameters* (hsop for short). Since version 3.1.2 Normaliz can compute such degrees. However, one must use this option with care since it requires the analysis of the face lattice of  $C$ , an impossible task if  $C$  has a large number of facets.

Let  $R = \bigoplus_{i=0}^{\infty} R_i$  be a finitely generated  $\mathbb{Z}$ -graded algebra over some infinite field  $K = R_0$  of Krull dimension  $\dim R = d$ . Its graded maximal ideal is given by  $\mathfrak{m} = \bigoplus_{i>0} R_i$ . In our case,  $R$  is the monoid algebra  $K[M]$  which is Cohen-Macaulay by a theorem of Hochster's, since  $M$  is normal, see [7, Theorem 6.10].

We call homogeneous elements  $\theta_1, \dots, \theta_d \in \mathfrak{m}$  a *homogeneous system of parameters* if  $\mathfrak{m} = \text{Rad}(\theta_1, \dots, \theta_d)$  or, equivalently,  $\dim R/\theta = 0$ , where  $\theta = (\theta_1, \dots, \theta_d)$ .

The existence of such a system is guaranteed in the  $\mathbb{Z}$ -graded case by the *prime avoidance lemma*, see [7, Lemma 6.2]:

**Lemma 7.2** *Let  $R$  be a  $\mathbb{Z}$ -graded ring and  $I \subset R$  an ideal generated in positive degree. Let  $\mathfrak{p}_1, \dots, \mathfrak{p}_r$  be prime ideals such that  $I \not\subset \mathfrak{p}_i$  for  $i = 1, \dots, r$ . Then there exists a homogeneous element  $x \in I$  with  $x \notin \mathfrak{p}_1 \cup \dots \cup \mathfrak{p}_r$ .*

For any ideal  $I$  in  $R$  generated in positive degree of height  $\text{ht}(I) = h$ , the lemma provides the existence of elements  $\theta_1, \dots, \theta_h$  such that  $\text{ht}(\theta_1, \dots, \theta_i) = i$  for all  $i = 1, \dots, h$ .

If  $\theta_1, \dots, \theta_d$  is an hsop for  $K[M]$ , the Hilbert series can be written in the form

$$H_M(t) = \frac{h_0 + h_1 t + \dots + h_m t^m}{(1 - t^{g_1}) \dots (1 - t^{g_d})},$$

where  $g_j = \text{deg } \theta_j$ . Furthermore  $h_i$  counts the number of elements of degree  $i$  in a homogeneous basis of  $K[M]$  over  $K[\theta_1, \dots, \theta_d]$  and in particular  $h_i$  is non-negative (see [7, Theorem 6.40]).

To reach our mentioned goal of finding a nice representation of the Hilbert series, we therefore compute (the degrees of) an hsop for the monoid algebra  $K[M]$ .

Our main idea for the construction of an hsop is generating elements  $\theta_i$  with  $\text{ht}(\theta_1, \dots, \theta_i) = i$  from the extreme integral generators of the cone  $C$ . We denote them by  $x_1, \dots, x_n \in \mathbb{Z}^n$  and note that  $\text{ht}(x_1, \dots, x_n) = d$ , where  $x_1, \dots, x_n$  are seen as monomials in  $K[M]$ . This claim will be justified below.

We successively insert the monomials  $x_j$  into a monomial ideal and compute its height. Note that in each step the height of this ideal can only increase by at most

one via Krull's principal ideal theorem, see [8, Theorem A.1]. If

$$\text{ht}(x_1, \dots, x_j) = i > i - 1 = \text{ht}(x_1, \dots, x_{j-1}),$$

we let

$$\theta_i := \lambda_1 x_1^{a_1} + \dots + \lambda_j x_j^{a_j},$$

where  $\lambda_k \in K$  are generic coefficients and the exponents  $a_k$  are chosen in such a way that  $\theta_i$  is homogeneous of degree  $\text{lcm}(\deg(x_1), \dots, \deg(x_j))$ . We point out that the height does not change if we replace the  $x_i$  by powers of them. Furthermore, all current monomials  $x_1, \dots, x_j$  are needed in general to ensure that  $\text{ht}(\theta_1, \dots, \theta_i) = i$ .

We are left with the task to compute  $\text{ht}(x_1, \dots, x_j)$ . The minimal prime ideals of a monomial ideal  $I$  in the monoid algebra  $K[M]$  are of the form  $\mathfrak{p}_F = K\{M \setminus F\}$ , where  $F$  runs through all faces of  $C$  which are maximal with respect to disjointness to  $I$ . Furthermore the height of a prime ideal is given by the codimension of its respective face, i.e.  $\text{ht}(\mathfrak{p}_F) = d - \dim(F)$  (see for instance [7, Corollary 4.35 and Proposition 4.36]). (In particular, the ideal generated by the monomials  $x_1, \dots, x_n$  has height  $d$ : the only face disjoint to them is  $\{0\}$ .) In conclusion

$$\text{ht}(x_1, \dots, x_j) = \min_{F \text{ face}} \{\text{codim}(F); F \cap (x_1, \dots, x_j) = \emptyset\}.$$

These considerations lead to a step-by-step algorithm to compute the *heights vector*  $h \in \mathbb{Z}_+^n$  with  $h_j = \text{ht}(x_1, \dots, x_j)$ , see Algorithm 1.

---

### Algorithm 1 Heights

---

```

1:  $h_0 \leftarrow 1$ 
2:  $\mathcal{G} \leftarrow$  facets of  $C$ 
3:  $m \leftarrow d$ 
4: for  $j = 1, \dots, n$  do
5:    $\mathcal{G}_1 \leftarrow \{G_k \in \mathcal{G}; x_j \notin G_k\}$ 
6:    $\mathcal{G}_2 \leftarrow \{G_k \in \mathcal{G}; x_j \in G_k\}$ 
7:   if  $\mathcal{G}_1 \neq \emptyset$  then
8:     if  $\max_{G_k \in \mathcal{G}_1} \{\dim(G_k)\} < m$  then  $m \leftarrow m - 1$ ;  $h_j = h_{j-1} + 1$ 
9:     else  $h_j = h_{j-1}$ 
10:  else  $h_j = h_{j-1} + 1$ 
11:  for all facets  $F_\ell$  with  $x_j \notin F_\ell$  do
12:    for all  $G_k \in \mathcal{G}_2$  do
13:       $G_{k,\ell} \leftarrow G_k \cap F_\ell$ 
14:   $\mathcal{G} \leftarrow \mathcal{G}_1 \cup \{\text{maximal faces from } G_{k,\ell}\}$ 

```

---

Some of the facets can be neglected in the process of taking intersections with the faces in step  $j$  due to the following criteria:

1. The facet contains the current generator  $x_j$ ;
2. The facet only involves generators appearing in faces in  $\mathcal{G}_1$  or  $x_1, \dots, x_{j-1}$ ;
3. Facets only involving the generators  $x_1, \dots, x_j$  can be ignored for all following iterations.

Once the heights vector  $h$  is computed, the degrees of the corresponding hsop can be determined as mentioned before, although not all initial generators need to appear in the lcm to compute the homogeneous degree. More precisely, let  $\ell$  denote the smallest index such that  $h_\ell = h_{\ell+1}$ . Since  $\text{ht}(x_1, \dots, x_j, x_{j+1}) = h_{j+1} = h_j + 1 = \text{ht}(x_1, \dots, x_j) + 1$  for  $j = 1, \dots, \ell - 1$  we have

$$\text{deg}(\theta_i) = \begin{cases} \text{deg}(x_i), & \text{if } i \leq \ell, \\ \text{lcm}(\text{deg}(x_{\ell+1}), \dots, \text{deg}(x_i)), & \text{if } i > \ell. \end{cases}$$

We finally calculate the numerator of the new representation of the Hilbert series, by multiplying the form with cyclotomic polynomials in the denominator with the product  $(1 - t^{g_1}) \dots (1 - t^{g_d})$ , where  $g_j = \text{deg}(\theta_j)$ .

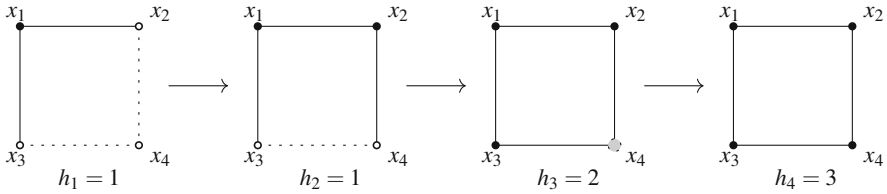
We note that for the simplicial case the extreme integral generators  $x_1, \dots, x_d$  already form an hsop. Therefore the choice of their degrees in the denominator of the Hilbert series can be considered a canonical. In the above simplicial example  $C = \mathbb{R}_{+x_1} + \mathbb{R}_{+x_2}$  with  $x_1 = (1, 2)$  and  $x_2 = (2, 1)$  the series can be expressed as:

$$H_M(t) = \frac{1 + t^2 + t^4}{(1 - t^3)^2},$$

where the degrees appearing in the denominator come from the extreme integral generators of  $C$ . The numerator has non-negative coefficients and counts the number of homogeneous basis elements of  $K[M]$  as a  $K[x_1, x_2]$ -module per degree, in this case  $(0, 0)$ ,  $(1, 1)$  and  $(2, 2)$  of degree 0, 2 and 4 respectively. This example also shows that using the Hilbert basis instead of the extreme integral generators as a generating system for  $M$  sometimes yield smaller exponents in the denominator, namely  $(1 - t^2)(1 - t^3)$ . However, using the Hilbert basis for the algorithm increases the complexity of taking intersections remarkably, which is the most expensive step.

As an example, let  $C = Q \times \{1\}$  be the cone over a square  $Q$ , see Fig. 6. The degree is given by  $\text{deg}(x_i) = i$  for  $i = 1, \dots, 4$ . (This choice is eligible since the only condition for this configuration is that the two sums of the degrees of antipodal points agree.) We get the following sequence of heights, which is also illustrated in Fig. 6 where dotted lines indicate the maximal disjoint faces:

$$\begin{aligned} h_1 = \text{ht}(x_1) = 1, \quad h_2 = \text{ht}(x_1, x_2) = 1, \quad h_3 = \text{ht}(x_1, x_2, x_3) = 2, \\ h_4 = \text{ht}(x_1, x_2, x_3, x_4) = 3. \end{aligned}$$



**Fig. 6** Sequence of heights for a cone over a square

The degrees for the corresponding hsop are given by  $\deg(\theta_1) = 1$ ,  $\deg(\theta_2) = 6$  and  $\deg(\theta_3) = 12$  and the Hilbert series has the form

$$H_M(t) = \frac{1 + t^2 + t^3 + 2t^4 + 2t^6 + t^7 + 2t^8 + 2t^{10} + t^{11} + t^{12} + t^{14}}{(1 - t)(1 - t^6)(1 - t^{12})}.$$

The heights vector and the degrees of the corresponding hsop can also be seen on the terminal if Normaliz is run with the verbosity option:

Heights vector: 1 1 2 3

Degrees of HSOP: 1 6 12

The Hilbert series with standard denominator for this cone is

$$H_M(t) = \frac{1 + t^3 + t^4 - t^5 + t^6 + t^7 + t^{10}}{(1 - t)(1 - t^2)(1 - t^{12})},$$

which again has a negative coefficients in the numerator.

If the order of the generators would be  $x_2, x_3, x_1, x_4$  the degrees and hence the exponents in the denominator of the Hilbert series are smaller, namely  $\deg(\theta_1) = 2$ ,  $\deg(\theta_2) = 3$ ,  $\deg(\theta_3) = 4$  and

$$H_M(t) = \frac{1 + t + t^2 + t^3 + t^4}{(1 - t^2)(1 - t^3)(1 - t^4)}.$$

However, considerations about the best possible order of generators would involve knowledge about the algebraic structure and defining equations (in this case  $x_1x_4 = x_2x_3$ ) of the input data, which are not accessible in Normaliz. Moreover, there is no clear answer to the question what an optimal choice for the exponents in the denominator should look like. Nevertheless, a possibility to improve the current representation would be a dynamic choice of the generators, where the next generator is chosen to lie in as many faces as possible, e.g.  $x_1, x_4, x_2, x_3$  in the above example. Future versions of Normaliz may contain this choice.



## 8 Class Group

The monoids  $M = C \cap L$  where  $C \subset \mathbb{R}^d$  is a rational cone and  $L$  a subgroup of  $\mathbb{Z}^d$  are exactly the normal affine monoids. For such a monoid  $M$  and a field  $K$  the monoid algebra  $K[M]$  is a normal Noetherian domain, which has a divisor class group  $\text{Cl}(K[M])$ , the group of isomorphism classes of divisorial ideals. It is not hard to prove that every isomorphism class is represented by a monomial divisorial ideal, and if one analyzes which monomial ideals are divisorial and when two such ideals are isomorphic modules, then one obtains *Chouinard's theorem*, see [7, Corollary 4.56]:

**Theorem 8.1** *Let  $\sigma : \text{gp}(M) \rightarrow \mathbb{Z}^s$  be the standard map. Then the divisor class group  $\text{Cl}(K[M])$  (identical to the divisor class group  $\text{Cl}(M)$  of  $M$ ) is given by  $\mathbb{Z}^s / \sigma(\text{gp}(M))$ .*

If  $\dim C = d$  and  $L = \mathbb{Z}^d$ , one has  $\text{gp}(M) = \mathbb{Z}^d$ . Therefore  $\text{Cl}(M) = \mathbb{Z}^s / \sigma(\mathbb{Z}^d)$ . Since  $\sigma$  is known, the computation of the divisor class group is a cheap by-product. Let  $A$  be the matrix whose columns are the support forms with coordinates in the dual basis to the unit vectors in  $\mathbb{Z}^d$ . Then the rows generate  $\sigma(\mathbb{Z}^d) \subset \mathbb{Z}^s$ , and it is only a matter of computing the Smith normal form of  $A$ . It immediately yields a decomposition  $\text{Cl}(M) = \mathbb{Z}^r \oplus (\mathbb{Z}/c_1\mathbb{Z})^{e_1} \oplus \dots \oplus (\mathbb{Z}/c_u\mathbb{Z})^{e_u}$  such that  $c_1 \mid \dots \mid c_u$ .

**Acknowledgements** The second author was partially supported by the German Research Council DFG-GRK 1916 and a doctoral fellowship of the German Academic Exchange Service.

## References

1. J. Abbott, A.M. Bigatti, C. Söger, Integration of libnormaliz in CoCoALib and CoCoA 5, in *Mathematical Software – ICMS 2014. 4th International Congress, Seoul, August 5–9, 2014. Proceedings* (Springer, Berlin, 2014), pp. 647–653
2. J. Abbott, A.M. Bigatti, G. Lagorio, CoCoA-5: a system for doing Computations in Commutative Algebra. Available at <http://cocoa.dima.unige.it>
3. T. Achterberg, SCIP: solving constraint integer programs. *Math. Program. Comput.* **1**, 1–41 (2009). Available from <http://mpc.zib.de/index.php/MPC/article/view/4>
4. B. Assarf et al., Computing convex hulls and counting integer points with polymake. Preprint (2015). arXiv:1408.4653
5. A. Bächle, L. Margolis, HeLP – a GAP-package for torsion units in integral group rings. Preprint (2016). arXiv:1507.08174.
6. S. Borowka et al., SecDec – a program to evaluate dimensionally regulated parameter integrals numerically. Available from <https://secdec.hepforge.org/>
7. W. Bruns, J. Gubeladze, *Polytopes, Rings and K-theory* (Springer, Berlin, 2009)
8. W. Bruns, J. Herzog, *Cohen-Macaulay Rings* (Cambridge University Press, Cambridge, 1998)
9. W. Bruns, B. Ichim, Normaliz: algorithms for affine monoids and rational cones. *J. Algebra* **324**, 1098–1113 (2010)
10. W. Bruns, R. Koch, Computing the integral closure of an affine semigroup. *Univ. J. Math.* **39**, 59–70 (2001)

11. W. Bruns, C. Söger, Generalized Ehrhart series and integration in Normaliz. *J. Symb. Comput.* **68**, 75–86 (2015)
12. W. Bruns, R. Hemmecke, B. Ichim, M. Köppe, C. Söger. Challenging computations of Hilbert bases of cones associated with algebraic statistics. *Exp. Math.* **20**, 25–33 (2011)
13. W. Bruns, B. Ichim, T. Römer, R. Sieg, C. Söger, Normaliz. Algorithms for rational cones and affine monoids. Available at <http://normaliz.uos.de>
14. W. Bruns, B. Ichim, C. Söger, The power of pyramid decomposition in Normaliz. *J. Symb. Comput.* **74**, 513–536 (2016)
15. W. Bruns, R. Sieg, C. Söger, The subdivision of large simplicial cones in Normaliz, in *MathematicalSoftware – ICMS 2016. 5th International Conference Berlin, July 11–14, 2016. Proceedings* (Springer, Berlin, 2016), p. 1026
16. B.A. Burton, Regina: software for 3-manifold theory and normal surfaces. Available from <http://regina.sourceforge.net/>
17. W. Decker, G.-M. Greuel, G. Pfister, H. Schönemann, Singular 4-0-2 — a computer algebra system for polynomial computations. Available at <http://www.singular.uni-kl.de>
18. D.R. Grayson, M.E. Stillman, Macaulay2, a software system for research in algebraic geometry. Available at <http://www.math.uiuc.edu/Macaulay2/>
19. S. Gutsche, M. Horn, C. Söger, NormalizInterface for GAP. Available at <https://github.com/gap-packages/NormalizInterface>
20. M. Joswig, B. Müller, A. Paffenholz, Polymake and lattice polytopes, in *DMTCS proc. AK*, ed. by C. Krattenthaler et al., Proceedings of FPSAC 2009, pp. 491–502
21. M. Köppe, Y. Zhou, New computer-based search strategies for extreme functions of the Gomory–Johnson infinite group problem. Preprint (2016). arXiv:1506.00017v3
22. F. Kohl, Y. Li, J. Rauh, R. Yoshida, Semigroups – a computational approach. Preprint (2017). arXiv:1608.03297
23. A. Schürmann, Exploiting polyhedral symmetries in social choice. *Soc. Choice Welf.* **40**, 1097–1110 (2013)

# Integral Frobenius for Abelian Varieties with Real Multiplication



Tommaso Giorgio Centeleghe and Christian Theisen

**Abstract** In this paper we introduce the concept of *integral Frobenius* to formulate an integral analogue of the classical compatibility condition linking the collection of rational Tate modules  $V_\lambda(A)$  arising from abelian varieties over number fields with real multiplication. Our main result gives a recipe for constructing an integral Frobenius when the real multiplication field has class number one. By exploiting algorithms already existing in the literature, we investigate this construction for three modular abelian surfaces over  $\mathbf{Q}$ .

**Keywords** Integral Tate module • Abelian variety • Real multiplication

**Subject Classifications** 11G10 Abelian varieties of dimension  $> 1$

## 1 Introduction

Let  $K$  be a number field, and  $A$  an abelian variety over  $K$  with real multiplication. By this we shall mean throughout that it is given a totally real number field  $E$  of degree  $[E : \mathbf{Q}] = \dim(A)$  together with an embedding

$$\iota : O_E \hookrightarrow \text{End}_K(A) \tag{1}$$

of its ring of integers  $O_E$  in the ring of  $K$ -endomorphisms of  $A$ . To simplify our notation we will omit the reference to  $\iota$ , and regard  $O_E$  as a given subring of the endomorphism ring of  $A$ .

---

T.G. Centeleghe (✉) • C. Theisen  
IWR Universität Heidelberg, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany  
e-mail: [tommaso.centeleghe@gmail.com](mailto:tommaso.centeleghe@gmail.com); [christian.theisen90@web.de](mailto:christian.theisen90@web.de)

© Springer International Publishing AG, part of Springer Nature 2017  
G. Böckle et al. (eds.), *Algorithmic and Experimental Methods  
in Algebra, Geometry, and Number Theory*,  
[https://doi.org/10.1007/978-3-319-70566-8\\_6](https://doi.org/10.1007/978-3-319-70566-8_6)

Let  $\lambda$  be any finite prime of  $E$ , denote by  $E_\lambda$  the completion of  $E$  at  $\lambda$ , and by  $O_\lambda \subset E_\lambda$  the corresponding valuation ring. Consider the  $\lambda$ -adic Tate module

$$T_\lambda(A) = \varprojlim_n A[\lambda^n],$$

and its rational version

$$V_\lambda(A) = T_\lambda(A) \otimes \mathbf{Q}.$$

From the embedding  $O_E \subseteq \text{End}_K(A)$  one deduces a structure of  $O_\lambda$ -module on  $T_\lambda(A)$  and a structure of  $E_\lambda$ -vector space on  $V_\lambda(A)$ . The module  $T_\lambda(A)$  is free of rank two over  $O_\lambda$ , and  $V_\lambda(A)$  is two-dimensional over  $E_\lambda$  (see [9, Prop. 2.2.1]). These structures are compatible with the action of the absolute Galois group  $G_K = \text{Gal}(\bar{K}/K)$ , where  $\bar{K}$  is a fixed algebraic closure of  $K$ . Thus we have two Galois representations:

$$\begin{aligned} \rho_\lambda : G_K &\longrightarrow \text{GL}_{O_\lambda}(T_\lambda(A)) \simeq \text{GL}_2(O_\lambda), \\ \rho_\lambda^0 : G_K &\longrightarrow \text{GL}_{E_\lambda}(V_\lambda(A)) \simeq \text{GL}_2(E_\lambda). \end{aligned}$$

Let now  $\mathfrak{p}$  be a finite prime of  $K$  where  $A$  has good reduction  $A_\mathfrak{p}$ . Denote by  $k_\mathfrak{p}$  be the residue field of  $\mathfrak{p}$ , by  $q$  its cardinality and by  $p$  its characteristic. Let  $\text{Frob}_\mathfrak{p} \in G_K$  be an arithmetic Frobenius element at  $\mathfrak{p}$ . As it is well known, the representations

$$\{\rho_\lambda^0\}_{\lambda \nmid p}$$

are all unramified at  $\mathfrak{p}$  and satisfy a compatibility condition that can be formulated as follows.

*There exists a semi-simple conjugacy class  $\Sigma_\mathfrak{p}^0 \subset \text{GL}_2(E)$  such that for every  $\lambda \nmid p$  the image of  $\Sigma_\mathfrak{p}^0$  in  $\text{GL}_2(E_\lambda)$  defines the conjugacy class of  $\rho_\lambda^0(\text{Frob}_\mathfrak{p})$ .*

Moreover, the characteristic polynomial of  $\Sigma_\mathfrak{p}^0$  has coefficients in  $O_E$ , and can be computed from the Frobenius isogeny  $\pi_\mathfrak{p}$  of  $A_\mathfrak{p}$  (see Sect. 3).

We find it natural to investigate an integral analogue of the property above. To this purpose we raise the following questions.

1. Is there a conjugacy class

$$\Sigma_\mathfrak{p} \subset \text{GL}_2(O_E[1/p])$$

such that for any  $\lambda \nmid p$  the action of  $\text{Frob}_\mathfrak{p}$  on  $T_\lambda(A)$  is described by the conjugacy class of  $\text{GL}_2(O_\lambda)$  containing the image of  $\Sigma_\mathfrak{p}$ ?

2. If such a  $\Sigma_\mathfrak{p}$  exists, how can we describe it?

**Definition 1.1** A conjugacy class  $\Sigma_\mathfrak{p}$  satisfying the requirement of question 1 will be called an *integral Frobenius* of  $A$  at  $\mathfrak{p}$ .

In this paper, in the case where  $E$  has class number one, we construct an explicit matrix  $\sigma_p \in \text{GL}_2(\mathcal{O}_E[1/p])$  with entries in  $\mathcal{O}_E$  and show that its conjugacy class is an integral Frobenius of  $A$  at  $p$ . Our main result generalizes a theorem of Duke and T oth (see [4, Theorem 2.1]) who treated the case of elliptic curves using a different technique. We remark that the integral questions raised above can be considered for more general compatible systems of Galois representations arising from geometry.

## 2 The Main Result

In this section we keep the notation of the introduction and further assume that  $E$  has class number one.

The injectivity of the reduction map (see [3, § 1.4.4])

$$r_p : \text{End}_K(A) \rightarrow \text{End}_{k_p}(A_p)$$

will be used throughout to identify the ring  $\mathcal{O}_E \subseteq \text{End}_K(A)$  with a subring of  $\text{End}_{k_p}(A_p)$ . In particular, if  $\lambda$  is a prime of  $E$  not dividing  $p$ , we can make sense of the Tate modules  $T_\lambda(A_p)$  and  $V_\lambda(A_p)$ , defined as in the characteristic zero case.

The ring  $\text{End}_{k_p}(A_p)$  has two distinguished elements (not necessarily distinct) given by the Frobenius isogeny  $\pi_p : A_p \rightarrow A_p$  relative to  $k_p$  and the corresponding Verschiebung  $q/\pi_p : A_p \rightarrow A_p$ . The existence of the embedding  $\mathcal{O}_E \subseteq \text{End}_{k_p}(A_p)$  implies that  $A_p$  is  $k_p$ -isogenous to a power of a  $k_p$ -simple abelian variety over  $k_p$  (see Proposition 3.1). Hence the  $\mathbf{Q}$ -subalgebra

$$\mathbf{Q}(\pi_p) \subseteq \text{End}_{k_p}(A_p) \otimes \mathbf{Q}$$

generated by  $\pi_p$  is a number field, and  $\pi_p$  a Weil  $q$ -number of it.

The element  $\pi_p$  plays a central role in the problem formulated in the introduction, in that, by means of the natural identification

$$T_\lambda(A_p) = T_\lambda(A),$$

the action induced by  $\pi_p$  on  $T_\lambda(A_p)$  corresponds to the Galois action of the arithmetic Frobenius  $\text{Frob}_p$  on  $T_\lambda(A)$ . The semi-simplicity of  $\pi_p$  acting on  $V_\lambda(A_p)$  (see [13, p. 138]) can then be used to deduce that of  $\text{Frob}_p$  acting on  $V_\lambda(A)$ . The characteristic polynomial of these  $E_\lambda$ -linear actions, denoted by

$$h_p(x) = x^2 - a_p x + s_p,$$

is independent of  $\lambda$  and has coefficients in  $\mathcal{O}_E$  (see Proposition 3.2).

After having recalled these basic facts, we give a recipe to construct an integral Frobenius  $\sigma_p \in \text{GL}_2(\mathcal{O}_E[1/p])$ . The construction is divided in two cases, according to whether the discriminant of  $h_p(x)$  is zero or not.

- $a_p^2 - 4s_p = 0$ . This condition is equivalent to  $\pi_p \in O_E$  (see Proposition 3.2), and in this case the problem is trivial: since  $\pi_p$  acts on  $T_\lambda(A)$  via scalar multiplication, the matrix

$$\sigma_p = \begin{pmatrix} \pi_p & 0 \\ 0 & \pi_p \end{pmatrix} \quad (2)$$

gives an integral Frobenius of  $A$  at  $\mathfrak{p}$ . Since  $E$  is totally real, the Weil number  $\pi_p$  is square root of  $q$  and

$$h_p(x) = (x - \pi_p)^2 = x^2 - 2\pi_p x + q.$$

- $a_p^2 - 4s_p \neq 0$ . This is the interesting case of the problem, and the definition of  $\sigma_p$  is more involved. For more details we refer to Sect. 4. The  $E$ -algebra  $L = E[\pi_p] \subseteq \text{End}_{k_p}(A_p) \otimes \mathbf{Q}$  is semi-simple and has dimension two over  $E$ . Inside  $L$  there is a chain of  $O_E$ -orders given by

$$O_E[\pi_p] \subseteq S_p \subseteq O_L,$$

where  $O_L$  denotes the integral closure of  $O_E[\pi_p]$  in  $L$ , and  $S_p$  is defined as

$$S_p = E[\pi_p] \cap \text{End}_{k_p}(A_p),$$

the intersection being taken in  $\text{End}_{k_p}(A_p) \otimes \mathbf{Q}$ . The  $O_E$ -discriminant of  $O_E[\pi_p]$  is the principal ideal  $(a_p^2 - 4s_p)$ , which can be written as

$$(a_p^2 - 4s_p) = \delta_{O_L} \cdot \mathfrak{b}_{O_L}^2,$$

where  $\delta_{O_L}$  is the  $O_E$ -discriminant of  $O_L$ , and  $\mathfrak{b}_{O_L}$  is the  $O_E$ -conductor of  $O_E[\pi_p]$  in  $O_L$ . Let  $\mathfrak{b}_p \subseteq O_E$  be the divisor of  $\mathfrak{b}_{O_L}$  corresponding to the intermediate order  $S_p$  (see Proposition 4.2), and choose a generator  $b_p \in \mathfrak{b}_p$ . Let  $u_p \in O_E$  be any element such that the ratio  $(\pi_p - u_p)/b_p$  belongs to  $S_p$  (see Proposition 4.3). The matrix  $\sigma_p$  is defined by the formula

$$\sigma_p = \begin{pmatrix} u_p - \frac{u_p^2 - a_p u_p + s_p}{b_p} & \\ b_p & a_p - u_p \end{pmatrix}, \quad (3)$$

from which it is easily checked that its characteristic polynomial is  $h_p(x)$ .

Our main result says:

**Theorem 2.1** *The matrix  $\sigma_p$  has coefficients in  $O_E$  and defines an integral Frobenius  $\Sigma_p$  of  $A$  at  $\mathfrak{p}$ .*

In more concrete terms, the theorem says that for any finite prime  $\lambda$  of  $E$  not dividing  $p$ , the Tate module  $T_\lambda(A)$  admits an  $O_\lambda$ -basis such that the action of  $\text{Frob}_p$  on  $T_\lambda(A)$  in the coordinates of this basis is given by  $\sigma_p$ . In particular we deduce:

**Corollary 2.2** *For any ideal  $\mathfrak{n} \subseteq O_E$  relatively prime to  $\mathfrak{p}$ , the matrix  $\sigma_p$  describes the action of  $\text{Frob}_p$  on the  $\mathfrak{n}$ -torsion points  $A[\mathfrak{n}]$  of  $A$ , in the coordinates of a suitable  $O_E/\mathfrak{n}$ -basis.*

The corollary, which is essentially a reformulation of the main result, emphasises a connection of our work to [4].

In the non-trivial case  $a_p^2 - 4s_p \neq 0$ , write

$$\sigma_p = \begin{pmatrix} u_p & 0 \\ 0 & u_p \end{pmatrix} + b_p \begin{pmatrix} 0 & -\frac{u_p^2 - a_p u_p + s_p}{b_p^2} \\ 1 & -\frac{2u_p - a_p}{b_p} \end{pmatrix}. \tag{4}$$

From the construction of  $\sigma_p$  it follows that the matrix  $(\sigma_p - u_p)/b_p$  appearing in the right hand side of (4) has coefficients in  $O_E$  (see Proposition 4.3). Thus from Theorem 2.1 we deduce the following interesting property of the ideal  $b_p$ . For any prime-to- $p$  ideal  $\mathfrak{n} \subseteq O_E$  we have:

$$\begin{aligned} &\text{Frob}_p \text{ acts on } A[\mathfrak{n}] \text{ as scalar} \\ &\text{multiplication by an element in } O_E/\mathfrak{n} \iff \mathfrak{n} \text{ divides } b_p. \end{aligned}$$

This equivalence can be linked to prime splitting phenomena in Galois extensions of number fields. Extend the definition of  $b_p$  by setting it equal to zero when  $a_p^2 - 4s_p = 0$ . Let  $\mathfrak{n}$  be a nonzero ideal of  $O_E$ , and consider the projective Galois representation

$$\mathbb{P}(\bar{\rho}_\mathfrak{n}) : G_K \longrightarrow \mathbb{P}\text{Aut}_{O_E/\mathfrak{n}}(A[\mathfrak{n}]) \simeq \text{PGL}_2(O_E/\mathfrak{n})$$

obtained from the  $\mathfrak{n}$ -torsion of  $A$ . Let  $K(\mathbb{P}A[\mathfrak{n}])/K$  be the Galois extension of  $K$  corresponding to the kernel of  $\mathbb{P}(\bar{\rho}_\mathfrak{n})$ .

**Corollary 2.3** *Let  $\mathfrak{p}$  a prime of  $K$  where  $A$  has good reduction, and let  $\mathfrak{n} \subseteq O_E$  an ideal relatively prime to  $p$ . Then  $\mathfrak{p}$  splits completely in  $K(\mathbb{P}A[\mathfrak{n}])/K$  if and only if  $\mathfrak{n}$  divides  $b_p$ .*

In the case  $a_p^2 - 4s_p \neq 0$ , the matrix  $\sigma_p$ , whose definition might appear quite mysterious, represents the multiplication action of  $\pi_p$  on the ring  $S_p$  in the coordinates of a suitable  $O_E$ -basis (see Sect. 4, Remark 4.4). The main result relies on the key observation that the  $\ell$ -adic Tate module  $T_\ell(A)$  is free of rank one over  $S_p \otimes \mathbf{Z}_\ell$ . In Sect. 3 we prove some basic facts on reduction in positive characteristic of abelian varieties with real multiplication. In Sect. 4 we discuss orders in quadratic extensions of number fields useful to understand the details of the construction of  $\sigma_p$ . In Sect. 5 we give the proof of Theorem 2.1. Lastly, in

Sects. 6 and 7 we exploit algorithms already existing in the literature (see [1] and [5]) to make computational investigations with three modular abelian surfaces over  $\mathbf{Q}$ . This article is a development of the Master Thesis of the second author at the University of Heidelberg.

### 3 Reduction of Abelian Varieties with Real Multiplication

We keep the notation and assumptions of the first two sections, so that  $A$  is an abelian variety over a number field  $K$  with real multiplication by  $E$ , and  $\mathfrak{p}$  is a place of  $K$  where  $A$  has good reduction  $A_{\mathfrak{p}}$ . We denote by  $k_{\mathfrak{p}}$  the residue field of  $K$  at  $\mathfrak{p}$ , by  $p$  its characteristic, and by  $q = p^a$  its cardinality, where  $a = [k_{\mathfrak{p}} : \mathbf{F}_p]$ . As before,  $\pi_{\mathfrak{p}} : A_{\mathfrak{p}} \rightarrow A_{\mathfrak{p}}$  denotes the Frobenius isogeny relative to  $k_{\mathfrak{p}}$ . The reduction of the real multiplication on  $A$  gives inclusion

$$O_E \subseteq \text{End}_{k_{\mathfrak{p}}}(A_{\mathfrak{p}}). \tag{5}$$

The existence of this subring has the following consequence.<sup>1</sup>

**Proposition 3.1** *The abelian variety  $A_{\mathfrak{p}}$  is isotypical, i.e., it is  $k_{\mathfrak{p}}$ -isogenous to  $B^n$ , where  $B$  is a  $k_{\mathfrak{p}}$ -simple abelian variety and  $n$  is an integer  $> 0$ .*

*Proof* Consider a  $k_{\mathfrak{p}}$ -isogeny

$$f : A_{\mathfrak{p}} \longrightarrow \prod_{1 \leq i \leq h} B_i^{n_i}$$

from  $A_{\mathfrak{p}}$  into the product of powers of  $k_{\mathfrak{p}}$ -simple, pairwise non- $k_{\mathfrak{p}}$ -isogenous abelian varieties  $B_i$ , with  $n_i > 0$ . We clearly have  $[E : \mathbf{Q}] = \dim(A_{\mathfrak{p}}) = \sum_i n_i \dim(B_i)$ . To prove the lemma we have to show that  $h = 1$ .

The isogeny  $f$  induces an identification

$$\text{End}_{k_{\mathfrak{p}}}(A) \otimes \mathbf{Q} \simeq \prod_{1 \leq i \leq h} M_{n_i}(D_i), \tag{6}$$

where  $M_{n_i}(D_i)$  is the ring of  $n_i$ -by- $n_i$  matrices with coefficients in the division ring  $D_i = \text{End}_{k_{\mathfrak{p}}}(B_i) \otimes \mathbf{Q}$ .

Let  $\pi_i \in D_i$  be the Frobenius isogeny of  $B_i$  relative to  $k_{\mathfrak{p}}$ . The subfield  $\mathbf{Q}(\pi_i) \subseteq D_i$  is the center of  $D_i$  (see [13, Theorem 2 (a)]), and a standard formula from Honda-Tate theory says that

$$2 \dim(B_i) = s_i [\mathbf{Q}(\pi_i) : \mathbf{Q}], \tag{7}$$

where  $s_i$  is the index of  $D_i$ , i.e., the square root of the degree  $[D_i : \mathbf{Q}(\pi_i)]$ .

---

<sup>1</sup>The fact that the variety  $A_{\mathfrak{p}}$  arises as reduction from characteristic zero plays no role in Propositions 3.1 and 3.2.



The inclusion  $E \subseteq \text{End}_{k_p}(A) \otimes \mathbf{Q}$  projects into each factor of the decomposition (6) and gives, for any  $i$ , an embedding

$$\mu_i : E \longrightarrow M_{n_i}(D_i). \tag{8}$$

We first complete the proof of the proposition assuming that there exist an index  $i_0$  such that  $\pi_{i_0}$  is not a real Weil  $q$ -number.

Under this assumption we see that the compositum  $L_{i_0} = \mu_{i_0}(E)\mathbf{Q}(\pi_{i_0})$  inside  $M_{n_{i_0}}(D_{i_0})$  is a semi-simple commutative subalgebra of  $M_{n_{i_0}}(D_{i_0})$  containing the center  $\mathbf{Q}(\pi_{i_0})$  and strictly containing the field  $\mu_{i_0}(E) \simeq E$ . Since the degree over  $\mathbf{Q}(\pi_{i_0})$  of any commutative semi-simple subalgebra  $L \subseteq M_{n_{i_0}}(D_{i_0})$  is bounded by  $n_{i_0}s_{i_0}$ , we conclude from (7) that

$$2[E : \mathbf{Q}] \leq [L_{i_0} : \mathbf{Q}] \leq n_{i_0}2 \dim(B_{i_0}),$$

which readily implies that  $h = 1$ , given that  $[E : \mathbf{Q}] = \sum_i n_i \dim(B_i)$ .

We are left with proving the proposition in the case where all Frobenius isogenies  $\pi_i$  define real Weil  $q$ -numbers. If  $a$  is odd the proposition holds simply because there is *only one* real Weil  $q$ -number, up to conjugation, namely that given by  $\sqrt{q}$ , a real quadratic algebraic integer.

If  $a$  is even there are precisely two distinct conjugacy classes of real Weil  $q$ -numbers, given by the integers  $q^{a/2}$  and  $-q^{a/2}$ , and the isogeny  $f$  above has the form

$$f : A \longrightarrow B_1^{n_1} \times B_2^{n_2},$$

for some  $n_1, n_2 \geq 0$ , where the Frobenius isogenies  $\pi_1$  and  $\pi_2$  are given by multiplication by  $q^{a/2}$  and  $-q^{a/2}$ , respectively. As it turns out, both  $B_1$  and  $B_2$  are supersingular elliptic curves (which are not  $k_p$ -isogenous to each other) with all geometric endomorphisms defined over  $k_p$ . Their endomorphism algebra  $D_1$  and  $D_2$  are both isomorphic to the definite  $\mathbf{Q}$ -quaternion  $D$  ramified at  $p$ , and we have  $[E : \mathbf{Q}] = n_1 + n_2$ .

Arguing by contradiction, assume that both  $n_1$  and  $n_2$  are  $> 0$ , and consider as above the two embeddings

$$\mu_i : E \rightarrow M_{n_i}(D),$$

for  $i = 1, 2$ . Since  $2n_i$  is the degree over  $\mathbf{Q}$  of any commutative semi-simple subalgebra  $L_i \subseteq M_{n_i}(D)$ , we easily see that  $n_1 = n_2$  and that  $\mu_i(E)$  is maximal commutative subfield of  $M_{n_i}(D)$ . It follows that  $\mu_1(E)$  is a splitting field for  $M_{n_1}(D)$  and hence it is also a splitting field for  $D$ . This is a contradiction since  $\mu_1(E)$  is totally real, whereas every splitting field of the definite quaternion  $D$  cannot have a real place. This completes the proof of the proposition.  $\square$

Proposition 3.1 is equivalent to the statement that  $\mathbf{Q}(\pi_p)$  is a field (and not just a product of fields). In this way we see that  $\pi_p$  defines a Weil  $q$ -number of the

number field  $\mathbf{Q}(\pi_p)$ . The complex conjugate of  $\pi_p$ , with respect to any embedding  $\mathbf{Q}(\pi_p) \subset \mathbf{C}$ , is the Verschiebung isogeny  $q/\pi_p$  of  $A_p$ .

Consider now the commutative subalgebra  $E[\pi_p] \subseteq \text{End}_{k_p}(A_p) \otimes \mathbf{Q}$ , and let  $g_p(x)$  be the minimal polynomial of  $\pi_p$  over  $E$ . Since  $\mathbf{Q}(\pi_p)$  is semi-simple, so is  $E[\pi_p]$  and  $g_p(x)$  has non-zero discriminant. Moreover, the degree  $[E[\pi_p] : E]$  is either 1 or 2, according to whether  $\pi_p$  belongs to  $O_E$  or not, respectively. This can be seen using (7) and reasoning as in the proof of Proposition 3.1. Set now

$$h_p(x) = \begin{cases} g_p^2(x), & \text{if } \pi_p \in O_E \\ g_p(x), & \text{if } \pi_p \notin O_E \end{cases}.$$

Since  $\pi_p$  is an algebraic integer, the polynomials  $h_p(x)$  and  $g_p(x)$  have coefficients in  $O_E$ . Moreover notice that  $\pi_p \in O_E$  if and only if  $\pi_p \in E$ .

**Proposition 3.2** *Let  $\lambda$  be a prime of  $E$  not dividing  $p$ . The polynomial  $h_p(x)$  is the characteristic polynomial of the  $E_\lambda$ -linear action induced by  $\pi_p$  on  $V_\lambda(A)$ . Its discriminant is zero if and only if  $\pi_p \in O_E$ .*

*Proof* If  $\pi_p \in O_E$ , then  $V_\lambda(\pi_p)$  is given by scalar multiplication by  $\pi_p$  itself. We have  $g_p(x) = (x - \pi_p)$  and  $h_p(x) = g_p^2(x)$ . If  $\pi_p \notin O_E$ , then  $g_p(x)$  has degree two and thus we must have  $h_p(x) = g_p(x)$ . The last statement of the proposition follows from the fact that  $g_p(x)$  has distinct roots.  $\square$

*Remark 3.3* Let  $a_p$  and  $s_p$  respectively denote the trace and the determinant of the  $E_\lambda$ -linear action induced by  $\pi_p$  on  $V_\lambda(A_p)$ , so that we have

$$h_p(x) = x^2 - a_p x + s_p.$$

The following can be said about the coefficients of  $h_p(x)$ . If  $\pi_p \in O_E$ , then  $\pi_p$  is a real Weil  $q$ -number, hence its square is equal to  $q$ . In this case we have  $a_p = 2\pi_p$  and  $s_p = q$ . If  $\pi_p \notin O_E$  and  $\pi_p$  is not real, then  $h_p(x)$  is irreducible in  $E[x]$ , and we have  $a_p = \pi_p + q/\pi_p$  and  $s_p = q$ . Finally, if  $\pi_p \notin O_E$  and  $\pi_p$  is real, then  $a_p = 0$  and  $s_p = -q$ . This last case can only occur if  $a$  is odd, and  $h_p(x)$  is reducible if and only if  $O_E$  contains a square root of  $q$ .

We conclude the section with the following observation.

**Proposition 3.4** *Assume that there is a place  $\mathfrak{p}$  of  $K$  of good reduction for  $A$  such that  $\text{End}_{k_p}(A_p)$  is commutative. Then  $E$  is the unique subfield of  $\text{End}_K(A) \otimes \mathbf{Q}$  which is totally real and has degree  $\dim(A)$ .*

*Proof* Let  $E' \subseteq \text{End}_K(A) \otimes \mathbf{Q}$  be a totally real number field with  $[E' : \mathbf{Q}] = \dim(A)$ . We shall show that the image of  $E'$  in  $\text{End}_{k_p}(A_p) \otimes \mathbf{Q}$  under the reduction map (also denoted by  $E'$ ) is equal to the number field  $\mathbf{Q}(\pi_p + q/\pi_p)$ , which depends only on the reduction of  $A$  modulo  $\mathfrak{p}$ , and not on the choice of  $E'$  inside  $\text{End}_K(A) \otimes \mathbf{Q}$ .

The assumption on the place  $\mathfrak{p}$  is equivalent to ask that  $A_p$  be  $k_p$ -simple, and that its endomorphism ring tensored with  $\mathbf{Q}$  be given by  $\mathbf{Q}(\pi_p)$ , for some non-real Weil  $q$ -number  $\pi_p$ . Formula (7) from Honda-Tate theory applied to the  $k_p$ -simple variety

$A_p$  implies that

$$\dim(A_p) = [\mathbf{Q}(\pi_p) : \mathbf{Q}]/2 = [\mathbf{Q}(t_p) : \mathbf{Q}],$$

where  $t_p = \pi_p + q/\pi_p$ .

Arguing once again as in the proof of Proposition 3.1, one can show that  $E' \subseteq \text{End}_{k_p}(A_p) \otimes \mathbf{Q}$  must contain  $t_p$ . Since  $E'$  and  $\mathbf{Q}(t_p)$  have the same degree over  $\mathbf{Q}$  they coincide, and the proposition follows.  $\square$

## 4 Quadratic Orders

In this section we clarify some aspects of the recipe given in Sect. 2 for the construction of the integral Frobenius  $\sigma_p$  in the non-trivial case where  $\pi_p \notin O_E$ .

Denote by  $L$  the subalgebra  $E[\pi_p] \subseteq \text{End}_{k_p}(A_p) \otimes \mathbf{Q}$  generated by  $E$  and  $\pi_p$ . In our notation for  $L$ , for simplicity, we dropped any reference to the prime  $p$ . Hopefully, this will not lead to any confusion.

Thanks to the assumption  $\pi_p \notin O_E$ , the polynomial  $h_p(x)$  has distinct roots (see Proposition 3.2), and there is an isomorphism of  $E$ -algebras

$$L \simeq E[x]/(h_p(x)).$$

Thus  $L$  is either a quadratic field extension of  $E$  or it is isomorphic to  $E^2$ , respectively according to whether  $h_p(x)$  is irreducible or not in  $E[x]$ .

In what follows by an  $O_E$ -order  $S$  of  $L$ , or simply an order of  $L$ , we shall mean a subring  $S \subset L$  containing  $O_E$  and defining an  $O_E$ -lattice of  $L$ . Any such order  $S$  is locally free of rank two over the localizations  $(O_E)_\lambda$  of  $O_E$  at each nonzero prime ideal  $\lambda$ . There is a notion of  $O_E$ -discriminant  $\delta_S$  of an order  $S \subset L$  (see [10, III §2]). Without entering in the details here, we recall that  $\delta_S$  is an ideal of  $O_E$  which, locally at any nonzero prime  $\lambda$ , is computed as the determinant of the usual bilinear pairing given by the  $E$ -linear trace map

$$(x, y) \mapsto \text{Tr}(xy).$$

The  $O_E$ -discriminant of  $O_E[\pi_p]$  is generated by the discriminant  $a_p^2 - 4s_p$  of  $h_p(x)$ .

If  $O_L$  denotes the integral closure of  $O_E$  in  $L$ , we have a chain of inclusions of orders

$$O_E[\pi_p] \subseteq S_p \subseteq O_L, \tag{9}$$

where

$$S_p = L \cap \text{End}_{k_p}(A_p)$$

is the order appearing in Sect. 2 in the definition of  $\sigma_p$ . We observe that  $O_L$  is the ring of integers of  $L$  when  $h_p(x)$  is irreducible, and it is isomorphic to  $O_E^2$  otherwise.

Let now  $S \subset L$  be an order containing  $\pi_p$ , and  $\mathfrak{b}_S \subseteq O_E$  the nonzero ideal given by the annihilator of the torsion module  $S/O_E[\pi_p]$ .

**Proposition 4.1** *For any  $O_E$ -order  $S \subset L$  containing  $\pi_p$  we have*

$$O_E[\pi_p] = O_E + \mathfrak{b}_S S \quad \text{and} \quad \delta_{O_E[\pi_p]} = \delta_S \cdot \mathfrak{b}_S^2$$

*The ideal  $\mathfrak{b}_S$  will be called the  $O_E$ -conductor of  $O_E[\pi_p]$  in  $S$ .*

*Proof* Both equalities of the proposition can be proved after localization at each nonzero prime ideal  $\lambda \subset O_E$ , where the statements becomes easy to verify since the localizations  $(O_E[\pi_p])_\lambda$  and  $S_\lambda$  are free of rank two over the discrete valuation ring  $(O_E)_\lambda$ . □

The  $O_E$ -conductor of the order  $S_p$  entered in the recipe of the integral Frobenius from Sect. 2, where it was denoted by  $\mathfrak{b}_p$ . If  $S, S' \subset L$  are orders containing  $\pi_p$  then from Proposition 4.1 we deduce that  $S \subset S'$  if and only if  $\mathfrak{b}_S | \mathfrak{b}_{S'}$ . In particular, we have that  $\mathfrak{b}_S | \mathfrak{b}_{O_L}$  for any  $S$ . The next proposition shows that the invariant  $\mathfrak{b}_S$  suffices to determine the order  $S$ .

**Proposition 4.2** *The map  $\psi$  sending an  $O_E$ -order  $S \subset L$  containing  $\pi_p$  to the conductor  $\mathfrak{b}_S$  gives a bijection*

$$\psi : \left\{ \begin{array}{l} O_E\text{-orders } S \subset L \\ \text{containing } \pi_p \end{array} \right\} \xrightarrow{\sim} \left\{ \begin{array}{l} \text{ideals } \mathfrak{b} \subseteq O_E \\ \text{dividing } \mathfrak{b}_{O_L} \end{array} \right\}$$

*Proof* Let  $S \subset L$  be an  $O_E$ -order containing  $\pi_p$ . Consider the short exact sequence of  $O_E$ -modules

$$0 \longrightarrow O_E[\pi_p] \longrightarrow L \xrightarrow{r} E/O_E \oplus (E/O_E) \cdot \bar{\pi}_p \longrightarrow 0,$$

where  $\bar{\pi}_p$  denotes the image of  $\pi_p$  in  $L/E$ . The quotient  $S/O_E[\pi_p]$  is a submodule of the right term of the sequence which intersects the first summand trivially. Since there are isomorphisms of  $O_E$ -modules

$$(E/O_E) \cdot \bar{\pi}_p \simeq E/O_E \simeq \varinjlim_{0 \neq \mathfrak{n}} O_E/\mathfrak{n},$$

where the direct limit is taken over all nonzero ideals of  $O_E$ , we see that for any nonzero ideal  $\mathfrak{b} \subset O_E$  there is a unique submodule  $M_{\mathfrak{b}} \subset (E/O_E) \cdot \bar{\pi}_p$  whose annihilator is  $\mathfrak{b}$ . We conclude that

$$S = r^{-1}(0 \oplus M_{\mathfrak{b}_S}),$$

and hence  $S$  is uniquely determined by  $\mathfrak{b}_S$  and  $\psi$  is injective.

If  $\mathfrak{b} \subseteq O_E$  is an ideal dividing  $\mathfrak{b}_{O_L}$ , then

$$O_E + \frac{\mathfrak{b}_{O_L}}{\mathfrak{b}} O_L$$

is an  $O_E$ -order of  $L$  in which  $O_E[\pi_p]$  sits with conductor  $\mathfrak{b}$ . This shows that  $\psi$  is surjective and completes the proof of the proposition.  $\square$

We assume for the rest of the section that  $E$  has class number one. This assumption, besides the principality of any ideal of  $O_E$ , ensures that any  $O_E$ -order  $S \subset L$  is free of rank two as an  $O_E$ -module (see [8, Theorem 1.32]).

**Proposition 4.3** *Assume that  $E$  has class number one, and let  $\mathfrak{b} \subseteq O_E$  be a nonzero ideal. Then  $\mathfrak{b}$  divides  $\mathfrak{b}_{O_L}$  if and only if there exists  $u \in O_E$  such that the following conditions are satisfied:*

1.  $h'_p(u) = 2u - a_p \in \mathfrak{b}$ ;
2.  $h_p(u) = u^2 - a_p u + s_p \in \mathfrak{b}^2$ .

*Under these conditions, the reduction of  $u$  modulo  $\mathfrak{b}$  is uniquely determined, and if  $b$  is a generator of  $\mathfrak{b}$  the pair*

$$(1, (\pi_p - u)/b) \tag{10}$$

*is an  $O_E$ -basis of the order  $S \subset L$  with  $\mathfrak{b}_S = \mathfrak{b}$ .*

*Proof* Reasoning as in the proof of Proposition 4.2, we see that the ideal  $\mathfrak{b}$  divides  $\mathfrak{b}_{O_L}$  if and only if there exists  $u \in O_E$  such that the ratio  $(\pi_p - u)/b$  belongs to  $O_L$ , where  $b$  is a generator of  $\mathfrak{b}$ . This is to say that  $\mathfrak{b}$  divides  $\mathfrak{b}_{O_L}$  if and only if the minimal, monic polynomial of  $(\pi_p - u)/b$  over  $E$  has coefficients in  $O_E$ . Since this polynomial is given by

$$x^2 + \frac{2u - a_p}{b}x + \frac{u^2 - a_p u + s_p}{b^2},$$

the first part of the proposition follows. This also shows that the pair (10) is a basis of the order corresponding to  $\mathfrak{b}_S$  under the bijection  $\psi$  from Proposition 4.2. From this it is easy to see that  $u$  is uniquely determined modulo  $\mathfrak{b}$ . The proposition follows.  $\square$

*Remark 4.4* The matrix  $\sigma_p$  constructed in Sect. 2 represents the multiplication action of  $\pi_p$  on  $S_p$  on the coordinates induced by an  $O_E$ -basis of the form  $(1, (\pi_p - u_p)/b_p)$ , where  $b_p$  is a generator of  $\mathfrak{b}_p$  and  $u_p$  is an element of  $O_E$  chosen to satisfy the two congruences of Proposition 4.3.

We point out the following corollary of Proposition 4.3.

**Corollary 4.5** *Let  $\mathfrak{b} \subseteq O_E$  an ideal which is relatively prime to (2). Then  $\mathfrak{b}$  divides  $\mathfrak{b}_{O_L}$  if and only if  $\mathfrak{b}^2$  divides the discriminant  $a_p^2 - 4s_p$ .*

*Proof* The “only if” part is clear from the relationship between discriminant and conductor. To see the if part, assume that  $\mathfrak{b}^2$  divides  $(a_p^2 - 4s_p)$  and let  $u \in O_E$  be an element such that the first condition of the proposition is satisfied, i.e.,

$$2u \equiv a_p \pmod{\mathfrak{b}}.$$

Such a  $u$  exists since  $\mathfrak{b}$  and (2) are relatively prime. Then

$$4h_p(u) = (2u - a_p)^2 - (a_p^2 - 4s_p)$$

is divisible by  $\mathfrak{b}^2$ , since so are both summand. Since  $\mathfrak{b}$  and (2) are relatively prime, we conclude that  $\mathfrak{b}^2$  divides  $h_p(u)$  and the second condition of the proposition is also satisfied. Thus  $\mathfrak{b}$  divides  $\mathfrak{b}_{O_L}$ .  $\square$

We conclude the section with an observation that will be useful in our computations. Choose a generator  $b_{O_L}$  of  $\mathfrak{b}_{O_L}$  and an element  $u_p$  such that the pair

$$\left(1, \frac{\pi_p - u_p}{b_{O_L}}\right)$$

is an  $O_E$ -basis of  $O_L$ , and set  $\mathbf{e}_2 = (\pi_p - u_p)/b_{O_L}$ . From Propositions 4.2 and 4.3 we deduce the following corollary.

**Corollary 4.6** *Let  $S \subset L$  an  $O_E$ -order containing  $\pi_p$ , let  $\mathfrak{b}_S$  the  $O_E$ -conductor of  $O_E[\pi_p]$  in  $S$ , and let  $b_S$  a generator of  $\mathfrak{b}_S$ . The pair*

$$\left(1, \frac{b_{O_L}}{b_S} \cdot \mathbf{e}_2\right) = \left(1, \frac{\pi_p - u_p}{b_S}\right) \quad (11)$$

*is an  $O_E$ -basis of  $S$ .*

As  $b_S$  varies through the divisors of  $b_{O_L}$ , formula (11) parametrizes all  $O_E$ -orders  $S \subset L$  containing  $\pi_p$ , by exhibiting  $O_E$ -basis of them.

## 5 Proof of the Main Result

We first prove an abstract lemma that will be the key to our proof of Theorem 2.1. Let  $R$  be a ring isomorphic to a finite product  $\prod R_i$  of discrete valuation rings  $R_i$ , with total ring of fractions  $M$ . Consider the free module  $R^2$  of rank two, and assume that we are given an  $R$ -linear map  $F : R^2 \rightarrow R^2$  such that the  $R$ -subring  $R[F] \subset \text{End}_R(R^2)$  generated by  $F$  is free of rank two as an  $R$ -module.

The map  $F$  is given by a collection of  $R_i$ -linear maps  $F_i : R_i^2 \rightarrow R_i^2$ , and the above requirement is equivalent to ask that  $F_i$  is not given by multiplication by an element of  $R_i$ , for every index  $i$ . The ring

$$S = \text{End}_{R[F]}(R^2)$$

of  $R$ -linear endomorphisms of  $R^2$  commuting with  $F$  is an order of  $R[F] \otimes_R M$  containing  $R[F]$  and which acts on  $R^2$  in the obvious way.

**Lemma 5.1**  $R^2$  is a free  $S$ -module of rank one.

*Proof* The ring  $S$  decomposes as the product  $\prod S_i$ , where  $S_i = \text{End}_{R_i[F_i]}(R_i^2)$ . Therefore the general form of the lemma follows from the special case where  $R$  is a discrete valuation ring, which we treat next. Denote by  $\mathfrak{m}$  the maximal ideal of  $R$ , choose a uniformizer  $\omega$ , and let  $k$  be the residue field.

The  $R$ -order  $S$  of the  $M$ -algebra  $R[F] \otimes_R M$  is free of rank two over  $R$ , and therefore

$$S = R \oplus R \cdot F_0,$$

for some  $F_0 \in S$  which does not belong to the subring  $R \subset S$ . We claim that the morphism

$$F_0 \bmod \mathfrak{m} : R^2/\mathfrak{m}R^2 \longrightarrow R^2/\mathfrak{m}R^2$$

is not given by multiplication by any element of  $k$ . For otherwise there exists  $\lambda \in R$  such that  $F_0 - \lambda$  sends  $R^2$  to  $\mathfrak{m}R^2$ . This implies that  $(F_0 - \lambda)/\omega \in S$ , which contradicts the fact that  $(1, F_0)$  is an  $R$ -basis of  $S$ .

The claim says precisely that there exists  $r \in R^2 \setminus \mathfrak{m}R^2$  such that

$$F_0(r) \notin R \cdot r + \mathfrak{m}R^2.$$

From Nakayama's Lemma we deduce that  $(r, F_0(r))$  is an  $R$ -basis of  $R^2$ , since the reductions of its elements generate  $R^2/\mathfrak{m}R^2$ . From this it readily follows that the map

$$S \ni s \mapsto s(r) \in R^2$$

is an isomorphism of  $S$ -modules. This completes the proof of the lemma. □

We now give the proof of Theorem 2.1, the main result of the paper.

*Proof* The result is trivial if  $\pi_p \in E$ , therefore we continue assuming  $\pi_p \notin E$ . Let  $\ell$  be a prime different from  $p$ , the residual characteristic of  $\mathfrak{p}$ . By a well known result of Tate (see [13]), there is a natural isomorphism

$$r_\ell^0 : \text{End}_{k_p}^0(A_p) \otimes \mathbf{Q}_\ell \xrightarrow{\sim} \text{End}_{\mathbf{Q}_\ell[\pi_p]}(V_\ell(A)).$$

Since  $\pi_p \notin E$ , the subalgebra  $L = E[\pi_p] \subseteq \text{End}_{k_p}^0(A_p)$  is a maximal commutative semi-simple subring, and hence it coincides with its own commutator. This implies that the restriction of  $r_\ell^0$  to  $L \otimes \mathbf{Q}_\ell$  induces an isomorphism

$$s_\ell^0 : L \otimes \mathbf{Q}_\ell \xrightarrow{\sim} \text{End}_{L \otimes \mathbf{Q}_\ell}(V_\ell(A_p)). \tag{12}$$

Now, the integral version of  $r_\ell^0$ , which is given by

$$r_\ell : \text{End}_{k_p}(A_p) \otimes \mathbf{Z}_\ell \xrightarrow{\sim} \text{End}_{\mathbf{Z}_\ell[\pi_p]}(T_\ell(A)), \tag{13}$$

is also an isomorphism. From (12) and (13) we conclude that the map

$$s_\ell : S_p \otimes \mathbf{Z}_\ell \longrightarrow \text{End}_{S_p \otimes \mathbf{Z}_\ell}(T_\ell(A_p))$$

arising as the restriction of  $r_\ell$  to  $S_p \otimes \mathbf{Z}_\ell$  is also an isomorphism. Since  $T_\ell(A_p)$  is free of rank two over  $O_E \otimes \mathbf{Z}_\ell$  and  $\pi_p \notin O_E$ , Lemma 5.1 gives that  $T_\ell(A_p)$  is a free  $S_p \otimes \mathbf{Z}_\ell$ -module of rank one,<sup>2</sup> and hence

$$T_\lambda(A_p) \text{ is a free } S_p \otimes_{O_E} O_\lambda\text{-module of rank one.} \tag{14}$$

Theorem 2.1 now follows from the fact that the matrix  $\sigma_p$  describes, by construction, the multiplication action of  $\pi_p$  on  $S_p$  in a suitable basis.  $\square$

## 6 Computations

Our aim in the remaining part of the paper is to explain how two algorithms already present in the literature (see [5] and [1]) can be used to compute the integral Frobenius at several primes of good reduction for certain modular abelian *surfaces* over  $\mathbf{Q}$ . We are grateful to the authors of these algorithms for providing us with the nice opportunity to make experimental tests. All our auxiliary computations, like those in [5] and [1], have been performed using Magma (see [2]).

### 6.1 The Main Algorithms

The first algorithm is the result of joint work of González-Jiménez et al. (see [5]). The input from which they start is a cuspidal, normalized eigenform  $f = \sum a_n q^n \in S_2(\Gamma_0(N))$  of weight 2, trivial nebentype and conductor  $N$  such that its Fourier coefficient field  $E_f$  is a (real) quadratic extension of  $\mathbf{Q}$ . The modular abelian surface  $A_f$  attached to  $f$  via the classical Shimura construction (see [12]) is a  $\mathbf{Q}$ -subvariety of the Jacobian  $\text{Jac}X_0(N)$  of the modular curve  $X_0(N)$ , and has good reduction away from  $N$ . The Hecke action induces an inclusion

$$O_f \subseteq \text{End}_{\mathbf{Q}}(A_f), \tag{15}$$

where  $O_f = \mathbf{Z}[(a_p)_{p \nmid N}]$  is the order of  $E_f$  generated by the Fourier coefficients of  $f$  indexed by primes not dividing  $N$ .

---

<sup>2</sup>More generally, this freeness holds if  $S_p \otimes \mathbf{Z}_\ell$  is a Gorenstein ring (see [11, Remark, p. 502]).



Assuming that the canonical polarization on  $A_f$  coming from that of the Jacobian of the modular curve  $X_0(N)$  is a power of a principal one, the three authors compute a hyperelliptic, genus two equation

$$y^2 = F(x),$$

where  $F(x) \in \mathbf{Z}[x]$  has degree 5 or 6, whose desingularization defines a curve  $C_f$  over  $\mathbf{Q}$  such that there is an isomorphism

$$\text{Jac}(C_f) \simeq A_f \tag{16}$$

of principally polarized abelian varieties over  $\mathbf{Q}$ . In table at the end of their paper they list the hyperelliptic equations that they obtained for the 75 modular abelian surfaces of conductor  $\leq 500$  whose canonical polarization satisfies the required condition. We remark that their output, and hence also ours, is correct only up to numerical approximation. However, several tests in favour of its correctness are performed by the authors.

Notice that if (15) extends to the whole ring of integers  $O_{E_f} \subset E_f$ , then  $A_f$  is an abelian surface with real multiplication by  $E_f$ , according to the definition we gave in Sect. 1. Furthermore, if  $E_f$  has class number one then it makes sense to try and compute the integral Frobenius of  $A_f$  at primes  $p \nmid N$ .

The second algorithm on which our computations depend is due to Bisson (see [1]). The input is a smooth genus two curve  $C$  over a finite field  $\mathbf{F}$  with  $q$  elements such that its Jacobian  $\text{Jac}(C)$  is an absolutely simple, ordinary abelian surface over  $\mathbf{F}$ . The curve is assumed to be represented by a hyperelliptic equation  $y^2 = \bar{F}(x)$ , for a suitable polynomial  $\bar{F}(x) \in \mathbf{F}[x]$  of degree 5 or 6. Under these assumptions, the algorithm returns the endomorphism ring of the principally polarized abelian surface given by  $\text{Jac}(C)$ , which is an order of the quartic number field  $\mathbf{Q}(\pi)$  generated by the Frobenius isogeny  $\pi$  of  $\text{Jac}(C)$  relative to  $\mathbf{F}$ .

## 6.2 *Synthesis of the Algorithms*

The strategy we suggest for computing the integral Frobenius at primes of good reduction for a modular abelian surface  $A_f$  over  $\mathbf{Q}$  consists of the following steps.

1. Start from a cuspidal, normalized eigenform  $f \in S_2(\Gamma_0(N))$  whose coefficient field  $E_f$  is quadratic. The first goal is to use [5] to find a hyperelliptic equation of a genus two curve  $C_f$  over  $\mathbf{Q}$  such that the isomorphism (16) holds.

There are 465 modular surfaces of conductor  $\leq 500$ . In the 75 cases where the canonical polarization is a power of a principal one, [5] provides the hyperelliptic equations of the corresponding curves  $C_f$ . In the remaining cases, one can still try to use the same algorithm to solve (16) in  $C_f$  by constructing a principal polarization

on  $A_f$ . In [6] this problem is carefully analyzed and sufficient conditions for the existence of  $C_f$  are given.

We continue assuming that step 1 was successful, and perform now two checks.

2. Check that the inclusion (15) extends to the ring of integers  $O_{E_f}$ .

This maximality condition is often satisfied in practice. With the help of Magma we verified that for 428 modular surfaces of conductor  $\leq 500$  the order  $O_f$  is already the maximal order of  $E_f$ . Moreover, using [1], we verified that for only two of the surfaces  $A_f$  considered in [5] the ring  $\text{End}_{\mathbf{Q}}(A_f)$  fails to be the maximal order. These surfaces are those with conductor 224, where  $\text{End}_{\mathbf{Q}}(A_f)$  sits in  $O_{E_f}$  with index two.

3. Check that the class number of  $E_f$  is one.

This condition is required by our method for constructing integral Frobenia. Among surfaces of conductor  $\leq 500$  the condition fails only once in conductor 276.

Assuming that the three steps above are successfully completed, we enter now the second part of the strategy. Let  $p$  be a prime  $\nmid N$ , denote by  $A_{f,p}$  the reduction of  $A_f$  at  $p$ , and by  $\pi_p$  the Frobenius isogeny of  $A_{f,p}$  relative to its base field  $\mathbf{F}_p$ . By the Eichler-Shimura relation, we have

$$\pi_p + p/\pi_p = a_p \in O_{E_f},$$

where  $a_p$  is the  $p$ th Hecke eigenvalue of  $f$ , and hence the characteristic polynomial of  $\sigma_p$  is given by

$$h_p(x) = x^2 - a_p x + p. \tag{17}$$

If  $a_p^2 - 4p = 0$ , then

$$\pi_p \in O_{E_f} \subseteq \text{End}_{\mathbf{F}_p}(A_{f,p}),$$

and the integral Frobenius  $\sigma_p$  is the scalar matrix given by multiplication by  $\pi_p$ . We remark that in the computation we performed we never run in such an example.

We therefore continue assuming  $a_p^2 - 4p \neq 0$ , which also implies that  $\pi_p$  is not a real Weil  $p$ -number, for otherwise we would have  $h_p(x) = x^2 - p$  (see Sect. 3, Remark 3.3), a contradiction to (17).

4. Consider the quadratic  $E_f$ -algebra

$$L = E_f[\pi_p] \subseteq \text{End}_{\mathbf{F}_p}(A_{f,p}) \otimes \mathbf{Q},$$

and compute the ideal  $\mathfrak{b}_{O_L}$  given by the  $O_{E_f}$ -conductor of  $O_{E_f}[\pi_p]$  in its integral closure  $O_L \subset L$ . Compute further a generator  $b_{O_L}$  of  $\mathfrak{b}_{O_L}$  and an element  $u_p \in O_{E_f}$  such that the element

$$\mathbf{e}_2 = \frac{\pi_p - u_p}{b_{O_L}}$$

completes  $1 \in O_{E_f}$  to an  $O_{E_f}$ -basis of  $O_L$ .

Using Propositions 4.1, 4.3 and Corollary 4.5, the required computation can be carried out using basic Magma functions on the arithmetic of real quadratic fields. Notice that the  $O_{E_f}$ -basis  $(1, \mathbf{e}_2)$  of  $O_L$  satisfies the useful property of Corollary 4.6. The crucial information that we need to compute for the recipe of the integral Frobenius is the ideal  $\mathfrak{b}_p$  given by the  $O_{E_f}$ -conductor of  $O_{E_f}[\pi_p]$  inside  $S_p$ , where  $S_p$  is the order  $L \cap \text{End}_{\mathbf{F}_p}(A_{f,p})$ . If the conductor  $\mathfrak{b}_{O_L}$  is the trivial ideal  $O_{E_f}$ , then the chain (9) becomes

$$O_{E_f}[\pi_p] = S_p = O_L,$$

and hence the ideal  $\mathfrak{b}_p$  is trivial, and the integral Frobenius is simply given by the companion matrix

$$\sigma_p = \begin{pmatrix} 0 & -p \\ 1 & a_p \end{pmatrix}. \tag{18}$$

We then continue assuming that the ideal  $\mathfrak{b}_{O_L}$  is a proper ideal of  $O_E$ . In this case there is more than one possibility for the order  $S_p$ , and to decide which one occurs we want to use Bisson’s algorithm to compute the ring  $\text{End}_{\mathbf{F}_p}(A_{f,p})$ . In order to do so we first have to make sure that the assumptions of his algorithm are satisfied. We discuss these in the next three steps. If one of these assumptions fails, then our strategy will not lead to the computations of the integral Frobenius of  $A_f$  at  $p$ .

5. In the case where  $a_p^2 - 4p \neq 0$  and the ideal  $\mathfrak{b}_{O_L}$  is proper, check whether the affine model  $\mathbf{Z}[x, y]/(y^2 - F(x))$  of  $C_f$  has good reduction at the prime  $p$ .

It can happen that the model of  $C_f$  coming from the algorithm in [5] has singular reduction at a prime  $p \nmid N$ . In our computations this never occurred in a case where  $a_p^2 - 4p \neq 0$  and  $\mathfrak{b}_{O_L} \subsetneq O_{E_f}$ .

6. In the case where  $a_p^2 - 4p \neq 0$  and the ideal  $\mathfrak{b}_{O_L}$  is proper, check whether the abelian surface  $A_{f,p}$  is ordinary.

Recall that a Weil  $p$ -number  $\pi$  is ordinary if and only if the algebraic integer  $\pi + p/\pi$  is a  $p$ -adic unit. In our case this amounts to check if  $a_p$  is relatively prime to  $p$  in  $O_{E_f}$ , which can easily be done in Magma.

7. In the case where  $a_p^2 - 4p \neq 0$  and the ideal  $\mathfrak{b}_{O_L}$  is proper, check whether the abelian surface  $A_{f,p}$  is absolutely irreducible.

The abelian surface  $A_{f,p}$  is either  $\mathbf{F}_p$ -isogenous to the square of an elliptic curve or it is  $\mathbf{F}_p$ -simple (see Proposition 3.1). Since our base field is  $\mathbf{F}_p$  and the Weil number  $\pi_p$  is not real, we know from Honda-Tate theory (see [14]) that

$$A_{f,p} \text{ is } \mathbf{F}_p\text{-simple} \iff \pi_p + p/\pi_p = a_p \notin \mathbf{Z} \iff [\mathbf{Q}(\pi) : \mathbf{Q}] = 4.$$

In order to proceed we then must require  $a_p \notin \mathbf{Z}$  and still have to check whether  $A_{f,p}$  is absolutely simple or not. This amounts to verify the equality of number fields

$$\mathbf{Q}(\pi_p^N) = \mathbf{Q}(\pi) \tag{19}$$

for any integer  $N \geq 2$ . Since  $\mathbf{Q}(\pi)$  is a CM quartic extension of  $\mathbf{Q}$  it suffices to check (19) for all integers  $N \geq 2$  such that  $\varphi(N) \leq 4$ , where  $\varphi$  denotes the Euler  $\varphi$ -function. These values are 2, 3, 4, 5, 6, 10 and 12, and (19) can be verified using Magma.

8. In the case where  $a_p^2 - 4p \neq 0$  and the ideal  $\mathfrak{b}_{O_L}$  is proper, assuming that the affine model of  $C_f$  has good reduction at  $p$  and the abelian surface  $A_{f,p}$  is absolutely irreducible and ordinary, use [1] to compute  $\text{End}_{\mathbf{F}_p}(A_{f,p})$ . Then extract from it the information of the ideal  $\mathfrak{b}_p$ .

Since  $A_{f,p}$  is  $\mathbf{F}_p$ -simple and  $\pi_p$  is not real, we have that the CM quartic field  $\mathbf{Q}(\pi_p)$  coincides with the algebra  $\text{End}_{\mathbf{F}_p}(A_{f,p}) \otimes \mathbf{Q}$ . So that the ring of  $\mathbf{F}_p$ -endomorphisms of  $A_{f,p}$  is an order of  $\mathbf{Q}(\pi_p)$ . Notice that we also have

$$L = \mathbf{Q}(\pi) \text{ and } S_p = \text{End}_{\mathbf{F}_p}(A_{f,p}).$$

The output of Bisson’s algorithm is a  $\mathbf{Z}$ -basis of the  $\mathbf{Z}$ -order  $\text{End}_{\mathbf{F}_p}(A_{f,p}) \subset \mathbf{Q}(\pi_p)$ , expressed in terms of the basis  $(1, \pi_p, \pi_p^2, \pi_p^3)$  of  $\mathbf{Q}(\pi_p)$ . We are left with converting this output in an “ $O_{E_f}$ -linear” format, suitable for our purposes. To do this we use the equality  $\pi_p + p/\pi_p = a_p$  to embed the totally real field  $E_f$  in the number field  $\mathbf{Q}(\pi_p)$ . In this way, using step 4 and Corollary 4.6, we can control all  $O_{E_f}$ -orders  $S$  containing  $O_{E_f}[\pi_p]$  by exhibiting for each of them an  $O_{E_f}$ -basis of the form

$$\left(1, \frac{b_{O_L}}{b_S} \cdot \mathbf{e}_2\right) \tag{20}$$

inside the number field  $\mathbf{Q}(\pi)$  where Bisson’s output lives. Letting  $b_S$  vary through a set of generators of all divisors of  $\mathfrak{b}_{O_L}$ , we can then easily determine the unique element for which the  $O_{E_f}$ -span of the pair (20) gives the lattice from Bisson’s algorithm. This element is the generator  $b_p$  of the ideal  $\mathfrak{b}_p$  we were after.

## 7 Tables of Results

We applied the strategy explained in the previous section to three modular abelian surfaces  $A_f$  over  $\mathbf{Q}$ . The goal is to compute as many integral Frobenia as possible at primes  $p$  of good reduction in the range  $2, \dots, 1997$ . The three Hecke cuspidal newforms  $f$  we chose have conductor  $N = 23, 125$  and  $133$ . They all lie in the first Galois orbit of the corresponding space  $S_2(\Gamma_0(N))$ , according to Magma

enumeration. The hyperelliptic equations we used for the genus two curves  $C_f$  are those appearing in [5].

The heading of the six columns of each table follows the notations of the paper. The first column consists of primes  $p$ . The second, third and fourth columns ( $a_p$ ,  $u_p$  and  $b_p$ ) contain the elements of  $O_{E_f}$  needed to construct the integral Frobenius  $\sigma_p$ , they are expressed with respect to the  $\mathbf{Q}$ -basis  $(1, a)$  of  $E_f$  used by Magma to parametrize the coefficient field  $E_f$ . The fifth and sixth columns respectively give the prime factorizations of the ideals  $\mathfrak{b}_p$  and  $\mathfrak{b}_{O_L}$  in  $O_{E_f}$ .<sup>3</sup> If  $\ell$  is a rational prime which does not split in  $E_f$ , then the corresponding prime of  $O_{E_f}$  is denoted by  $(\ell)$  or  $\lambda_\ell$ , according to whether  $\ell$  is inert or ramifies, respectively. If  $\ell$  is split, then the corresponding primes are denoted by  $\lambda_{\ell,1}$  and  $\lambda_{\ell,2}$ .

In every table we listed all primes  $p \leq 1997$  where the given surface has good reduction and such that the order  $O_{E_f}[\pi_p]$  is not the maximal order of  $L = E[\pi_p]$ . When we were not able to apply Bisson’s algorithm (or when the algorithm did not terminate), a dash (–) appears in place of the entries  $u_p$  and  $b_p$ . In certain cases we did obtain an output from Bisson’s algorithm even though its basic assumptions on the input surface  $A_{f,p}$  were not satisfied. These primes appear marked in the tables: the symbol (\*) indicates that  $A_{f,p}$  is not ordinary, and the symbol (\*\*) denotes that it is not absolutely simple, but just  $\mathbf{F}_p$ -simple.

Finally, in every example considered, the coefficient field  $E_f$  is the real quadratic field  $\mathbf{Q}(\sqrt{5})$  of discriminant 5, and the order  $O_f$  is the maximal order. In the last two examples, the Galois representation on the 2-torsion  $A_f[2]$  defines two extensions of  $\mathbf{Q}$  with Galois group isomorphic to  $A_5$ , the alternating group in five letters. The computation of the integral Frobenius, when successful, reveal the primes that are completely split in these extensions.

### 7.1 First Example

Let  $f \in S_2(\Gamma_0(23))$  be the unique normalized cusp form of weight 2 and level 23. The element  $a \in E_f$  used by Magma to parametrize  $E_f$  has minimal polynomial  $x^2 + x - 1$ . The first few coefficients of the Fourier expansion of  $f$  are

$$f = q + aq^2 - (2a + 1)q^3 - (a + 1)q^4 + 2aq^5 + \dots$$

In Table 1 we can experimentally observe a reducibility phenomenon predicted by a famous result of Mazur (see [7]): since the prime 11 divides  $N - 1$ , Mazur predicts the existence of a prime  $\lambda$  of  $E_f$  lying above 11 such that

$$\bar{\rho}_\lambda \simeq 1 \oplus \chi_{11},$$

---

<sup>3</sup>These ideals are always defined in our computations as we never found a prime  $p$  for which  $a_p^2 - 4p = 0$ .

**Table 1** Integral Frobenius for  $A_f$ , where  $f \in S_2(\Gamma_0(23))$

$p$	$a_p$	$u_p$	$b_p$	Fac( $\mathfrak{b}_p$ )	Fac( $\mathfrak{b}_{O_L}$ )
19	-2	-	-	-	(3)
43	0	-	-	-	(2)
53	$-2 + 4a$	0	1	(1)	(2)
59	$4 + 4a$	1	2	(2)	(2)
61	$-2 - 8a$	0	1	(1)	(2)
67	$-4 + 2a$	$9 + a$	$2 + 3a$	$\lambda_{11,1}$	$\lambda_{11,1}$
89	$-8 - 4a$	$7 + 9a$	$2 + 3a$	$\lambda_{11,1}$	$\lambda_{11,1}$
101(**)	$2 + 4a$	1	2	(2)	(2)
149	$14 + 16a$	0	1	(1)	(2)
167	$4 - 4a$	1	2	(2)	(2)
173	$18 + 8a$	1	2	(2)	(2)
199	$-16 + 6a$	$3 + 3a$	$2 + 3a$	$\lambda_{11,1}$	$\lambda_{11,1}$
211	$-16 - 12a$	1	2	(2)	(2)
223	4	-	-	-	(2)
233	$-9 + 4a$	0	1	(1)	$\lambda_{31,2}$
271	8	-	-	-	(2)
307	$12 - 4a$	1	2	(2)	(2)
311	$7 + 10a$	0	1	(1)	$\lambda_5$
317	$18 + 12a$	1	2	(2)	(2)
331	$-11 - 14a$	$4a$	$2 + 3a$	$\lambda_{11,1}$	$\lambda_{11,1}$
347	$-16a$	1	2	(2)	(2)
353	$-3 + 20a$	$4 + 10a$	$2 + 3a$	$\lambda_{11,1}$	$\lambda_{11,1}$
379	$12 + 20a$	0	1	(1)	(2)
383	$12 - 8a$	0	1	(1)	(2)
397	$-17 - 12a$	$8 + 5a$	$2 + 3a$	$\lambda_{11,1}$	$\lambda_{11,1}$
401	$-8 - 10a$	0	1	(1)	$\lambda_5$
409	$9 + 20a$	0	1	(1)	$\lambda_5$
419	$-12 + 12a$	$10 + 8a$	$2 + 3a$	$\lambda_{11,1}$	$(2)\lambda_{11,1}$
431	$-20 + 4a$	0	1	(1)	(2)
449	$-10 - 8a$	1	2	(2)	(2)
463	-20	-	-	-	$(2)\lambda_{11,2}\lambda_{11,1}$
563	$-28 - 8a$	0	1	(1)	(2)
569	$-16 - 10a$	0	1	(1)	$\lambda_5$
593	$2 - 8a$	1	2	(2)	(2)
599	$24 + 16a$	1	2	(2)	(2)
607	$24 + 4a$	1	2	(2)	(2)
617	$-10 + 4a$	$8 + 5a$	$2 + 3a$	$\lambda_{11,1}$	$(2)\lambda_{11,1}$
619	$12 + 12a$	0	1	(1)	(2)
631	$20a$	0	1	(1)	(2)
661	$-18 - 8a$	$5 + 6a$	$2 + 3a$	$\lambda_{11,1}$	$(2)\lambda_{11,1}$
677	18	-	-	-	(2)
683	$13 + 22a$	1	$2 + 3a$	$\lambda_{11,1}$	$\lambda_{11,1}$

(continued)

Table 1 (continued)

$p$	$a_p$	$u_p$	$b_p$	$\text{Fac}(b_p)$	$\text{Fac}(b_{O_L})$
691	$12 - 8a$	1	2	(2)	(2)
719	$-8 + 8a$	1	2	(2)	$(2)\lambda_{11,2}$
727	$-24 - 6a$	$10 + 8a$	$2 + 3a$	$\lambda_{11,1}$	$\lambda_{11,1}$
751	$-12 + 20a$	0	1	(1)	$(2)\lambda_5$
787	$32 - 12a$	0	1	(1)	(2)
797	$-22 - 20a$	0	1	(1)	(2)
809(**)	$22 - 16a$	1	2	(2)	(2)
821	$-34 - 8a$	1	2	(2)	$(2)^2$
827	$4 - 4a$	0	1	(1)	(2)
829	$18 + 36a$	1	2	(2)	(2)
853	$-18 + 12a$	1	2	(2)	(2)
859	$-13 - 6a$	$10 + 8a$	$2 + 3a$	$\lambda_{11,1}$	$\lambda_{11,1}$
877	$-34 - 4a$	1	2	(2)	(2)
881	$38 + 10a$	$8 + 5a$	$2 + 3a$	$\lambda_{11,1}$	$\lambda_5\lambda_{11,1}$
883	4	—	—	—	(2)
911(**)	$14 + 28a$	$1 + 2a$	3	(3)	(3)
941	$-2 + 14a$	0	1	(1)	(3)
947	$-17 + 10a$	$8 + 5a$	$2 + 3a$	$\lambda_{11,1}$	$\lambda_{11,1}$
953	$18 + 4a$	0	1	(1)	(2)
991	24	—	—	—	$(2)\lambda_{11,2}\lambda_{11,1}$
997	$2 + 24a$	1	2	(2)	(2)
1009(**)	$6 + 12a$	0	1	(1)	(2)
1013	$-29 - 8a$	$2 + 7a$	$2 + 3a$	$\lambda_{11,1}$	$\lambda_{11,1}$
1069	$26 + 18a$	0	1	(1)	(3)
1091	$4 + 40a$	0	1	(1)	(2)
1097	$-18 - 24a$	1	2	(2)	(2)
1117	$14 - 28a$	1	2	(2)	(2)
1123	$-34 + 12a$	$5 + 6a$	$2 + 3a$	$\lambda_{11,1}$	$\lambda_{11,1}$
1151(**)	$-24 - 48a$	1	2	(2)	(2)
1163	$-8 - 20a$	1	2	(2)	(2)
1171	$16 - 18a$	0	1	(1)	(3)
1181	$-2 - 16a$	1	2	(2)	(2)
1213	$28 + 36a$	0	1	(1)	(3)
1217	$4 - 28a$	0	1	(1)	(3)
1231	$-16 - 24a$	1	2	(2)	(2)
1259	$-24 - 12a$	0	1	(1)	(2)
1277	$-7 - 8a$	$2 + 7a$	$2 + 3a$	$\lambda_{11,1}$	$\lambda_{11,1}$
1279	$-24 - 42a$	$10 + a$	$1 + 3a$	$\lambda_{11,2}$	$\lambda_{11,2}$
1301	$47 + 4a$	0	1	(1)	(3)
1303	$12 + 20a$	0	1	(1)	(2)
1319	$4 - 16a$	1	2	(2)	(2)

(continued)

**Table 1** (continued)

$p$	$a_p$	$u_p$	$b_p$	$\text{Fac}(b_p)$	$\text{Fac}(b_{O_L})$
1321	$8 - 24a$	$4 + 10a$	$2 + 3a$	$\lambda_{11,1}$	$\lambda_{11,1}$
1409	$-31 - 44a$	1	$2 + 3a$	$\lambda_{11,1}$	$\lambda_{11,1}$
1451	$-8 + 32a$	1	2	(2)	(2)
1453	2	—	—	—	$(2)\lambda_{11,2}\lambda_{11,1}$
1459	$66 + 10a$	0	1	(1)	$\lambda_5$
1481	$6 + 8a$	1	2	(2)	(2)
1483	$-36 + 8a$	0	1	(1)	(2)
1489	$-4 + 36a$	0	1	(1)	(3)
1499	$-13 + 2a$	0	1	(1)	$\lambda_{11,2}$
1523	$-24 - 56a$	0	1	(1)	(2)
1543	$-41 - 18a$	2	3	(3)	(3)
1549	43	—	—	—	(3)
1553	$-6 - 8a$	1	2	(2)	(2)
1559	$39 - 10a$	0	1	(1)	$\lambda_5$
1607	$-46 - 28a$	$10 + 8a$	$2 + 3a$	$\lambda_{11,1}$	$\lambda_{11,1}$
1613	$18 + 48a$	1	2	(2)	(2)
1663	$-8 - 44a$	0	1	(1)	(2)
1667	$-36 - 48a$	0	1	(1)	(2)
1669	$-38 - 32a$	3	4	$(2)^2$	(2)
1697	$-38 - 8a$	1	2	(2)	(2)
1721(**)	$4 + 8a$	0	1	(1)	$(3)^2$
1733	$-47 - 4a$	0	1	(1)	(3)
1783	$-57 - 6a$	$10 + 8a$	$2 + 3a$	$\lambda_{11,1}$	$\lambda_{11,1}$
1787	$40 - 4a$	1	2	(2)	(2)
1789	$-18 + 16a$	1	2	(2)	(2)
1811	$28 + 52a$	0	1	(1)	(2)
1831	$-52 - 10a$	0	1	(1)	$\lambda_5$
1861	$-30 - 44a$	0	1	(1)	(2)
1867	$20 + 44a$	1	2	(2)	(2)
1871	$-12 + 12a$	$21 + 8a$	$4 + 6a$	$(2)\lambda_{11,1}$	$(2)\lambda_{11,1}$
1873	$-38 - 8a$	0	1	(1)	(2)
1877	$22 + 32a$	0	1	(1)	$(2)^2$
1879	$-20 - 12a$	1	2	(2)	(2)
1889	$-2 - 44a$	1	2	(2)	(2)
1901	$-14 + 8a$	$2 + a$	3	(3)	$(2)(3)$
1913	$-62 - 16a$	1	2	(2)	(2)
1931	$2 + 20a$	0	1	(1)	$\lambda_5$
1949	$-58 - 20a$	0	1	(1)	(2)
1997	$-46 - 8a$	0	1	(1)	(2)



where  $\bar{\rho}_\lambda$  is the residual Galois representation of  $\rho_\lambda$ ,  $\chi_{11}$  is the mod 11 cyclotomic character, and 1 is the trivial character. The consequence of this result relevant for our computation is that for every prime  $p \neq 23$  with  $p \equiv 1 \pmod{23}$  the ideal  $\mathfrak{b}_p$  appearing in the definition of the integral Frobenius is divisible by  $\lambda$ . Such ideal  $\lambda$  is denoted by  $\lambda_{11,1}$  in the table.

## 7.2 Second Example

Let now  $f \in S_2(\Gamma_0(125))$  be the normalized cusp form of weight 2 and level 125 lying in the first Galois orbit of eigenforms. The element  $a \in E_f$  has also in this case minimal polynomial  $x^2 + x - 1$ . The first few coefficients of the Fourier expansion of  $f$  are

$$f = q + aq^2 - (a + 2)q^3 - (a + 1)q^4 - (a + 1)q^6 + \dots$$

Consider the Galois representation

$$\bar{\rho}_{(2)} : G_{\mathbf{Q}} \longrightarrow \text{Aut}_{O_{E_f/(2)}}(A_f[2]) \simeq \text{GL}_2(O_{E_f/(2)}). \tag{21}$$

defined by the 2-torsion  $A_f[2]$  of  $A_f$ . The rational prime 2 is inert in  $E_f \simeq \mathbf{Q}(\sqrt{5})$ , denote by  $\mathbf{F}_4$  its residue field. Since  $\bar{\rho}_{(2)}$  has trivial determinant we see that  $\bar{\rho}_{(2)}$  is valued in the special linear group  $\text{SL}_2(\mathbf{F})$ , which is isomorphic to  $A_5$ , the alternating group in five letters.

After analyzing the reduction modulo 2 of the first few Hecke eigenvalues of  $f$ , and using elementary group theory, one can deduce that

$$\text{Im}(\bar{\rho}_{(2)}) \simeq \text{SL}_2(\mathbf{F}), \tag{22}$$

i.e.,  $\bar{\rho}_{(2)}$  defines an  $A_5$ -extension  $K/\mathbf{Q}$ . According to Corollary 2.3, a rational prime  $p \nmid 2 \cdot 5$  splits completely in  $K$  if and only if (2) divides  $\mathfrak{b}_p$ , which, by Chebotarev, happens for a set of primes of density  $1/60 \sim 0.017$ . In Table 2 we observe this splitting phenomenon for  $p = 887, 1657$  and  $1699$ .

Lastly, notice that for every prime  $p \equiv 1 \pmod{5}$  for which we were able to compute  $\sigma_p$ , we have that the unique prime of  $E_f$  lying above 5 divides  $\mathfrak{b}_p$ . Reasoning as in the first example, this suggests that there is a decomposition

$$\bar{\rho}_{\lambda_5} \simeq 1 \oplus \chi_5,$$

where  $\chi_5$  denotes the mod 5 cyclotomic character. However, with our methods we are not able to prove this.

**Table 2** Integral Frobenius for  $A_f$ , where  $f$  lies in the first Galois orbit of  $S_2(\Gamma_0(125))$

$p$	$a_p$	$u_p$	$b_p$	$\text{Fac}(b_p)$	$\text{Fac}(b_{O_L})$
11	-3	-	-	-	$\lambda_5$
31(*)	$-3 - 5a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
41	-3	-	-	-	$\lambda_5$
61	$2 + 5a$	-	-	-	$\lambda_5\lambda_{19,1}$
71	-3	-	-	-	$\lambda_5^2$
89(**)	$6 + 12a$	0	1	(1)	(2)
101	-3	-	-	-	$\lambda_5$
131	$12 + 15a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
137	$4 - a$	$2 + a$	3	(3)	(3)
151	$-13 + 5a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
173	$-13 - 8a$	$1 + 2a$	3	(3)	(3)
181	$2 - 10a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
191	12	-	-	-	$(2)\lambda_5$
211	$12 + 10a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
229	$-3 - a$	0	1	(1)	$\lambda_{11,1}$
233	$-1 + 16a$	$1 + 2a$	3	(3)	(3)
241	$-3 + 10a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
251	$-18 - 15a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
271	$12 - 5a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
281	12	-	-	-	$\lambda_5(7)$
311	$-3 - 15a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
313	$-12a$	0	1	(1)	$\lambda_{11,1}$
317	$-14 + 8a$	$2 + a$	3	(3)	$(2)(3)$
331	$-13 + 5a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
353	$-22 - 16a$	0	1	(1)	(2)
379	$7 + 9a$	2	3	(3)	(3)
401	12	-	-	-	$\lambda_5$
421	$17 + 5a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
431	$12 + 15a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
439	$1 - 18a$	2	3	(3)	(3)
457	-18	-	-	-	$(2)^2$
461	$12 - 15a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
491	$12 - 15a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
503	$8 - 11a$	$1 + 2a$	3	(3)	(3)
509(**)	$-6 - 12a$	0	1	(1)	(2)
521	$-18 - 15a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
541	$-18 + 10a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
547	$-27 - 3a$	-	-	-	$\lambda_{59,1}$
557	$-8 + 20a$	$2 + a$	3	(3)	(3)
563	$20 + 8a$	0	1	(1)	(2)
571	-13	-	-	-	$\lambda_5(3)$

(continued)

**Table 2** (continued)

$p$	$a_p$	$u_p$	$b_p$	$\text{Fac}(b_p)$	$\text{Fac}(b_{O_L})$
587	$4 + 4a$	0	1	(1)	(2)
601	$-33 - 20a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
631	2	—	—	—	$\lambda_5(3)$
641	$-33 - 15a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
647	$-17 + 2a$	$2 + a$	3	(3)	(3)
661	$-18 - 20a$	$2 + 2a$	$1 + 2a$	$\lambda_5$	$(2)\lambda_5$
677	$30 + 16a$	0	1	(1)	(2)
691	$42 + 10a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
701	$27 + 15a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
727	$-24a$	0	1	(1)	(2)
743	$-34 - 5a$	$1 + 2a$	3	(3)	(3)
751	$17 + 20a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
757	27	—	—	—	$\lambda_{11,2}\lambda_{11,1}$
761	-18	—	—	—	$(2)^3\lambda_5$
811	$-28 - 10a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
821	$-3 - 15a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
859	$4 + 18a$	2	3	(3)	(3)
863	$-10 - 2a$	$1 + 2a$	3	(3)	(3)
881	$-3 - 30a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
887	$-36 + 4a$	1	2	(2)	(2)
911	$12 + 30a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
941	$-3 + 15a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
971	$-3 + 30a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
991	$-43 - 10a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1021	$17 + 20a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1031	$-3 - 30a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1051	$-28 - 45a$	1	$3 + 6a$	$\lambda_5(3)$	$\lambda_5(3)$
1061	$27 + 30a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1091	$-3 - 30a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1097	$-17 + 2a$	$2 + a$	3	(3)	(3)
1151	$12 + 45a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1171	$-3 + 25a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5^2$
1181	$-18 - 15a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1193	$-24 + 20a$	0	1	(1)	$\lambda_{11,1}$
1201	$-3 - 5a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1231	$-18 - 5a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1291	$2 + 20a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1301	$-18 + 30a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1321	$2 - 40a$	$2 + 2a$	$1 + 2a$	$\lambda_5$	$(2)\lambda_5$
1361	42	—	—	—	$(2)\lambda_5$
1367	$1 - 7a$	$2 + a$	3	(3)	(3)
1381	$27 + 25a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$

(continued)

**Table 2** (continued)

$p$	$a_p$	$u_p$	$b_p$	$\text{Fac}(b_p)$	$\text{Fac}(b_{O_L})$
1399	$-50$	—	—	—	(3)
1433	$14 - 4a$	0	1	(1)	(2)
1451	12	—	—	—	$(2)\lambda_5$
1471	$-18 - 20a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1481	$-48 - 15a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1511	$-3 + 30a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1531	2	—	—	—	$\lambda_5(3)$
1549	$-20 - 45a$	2	3	(3)	(3)
1571	$-18 + 30a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1583	$-1 + 16a$	$1 + 2a$	3	(3)	(3)
1601	$-3 + 15a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1607	$-48 - 20a$	0	1	(1)	(2)
1621	$47 + 45a$	1	$3 + 6a$	$\lambda_5(3)$	$\lambda_5(3)$
1657	$42 + 60a$	1	2	(2)	(2)
1663	$-60 - 12a$	0	1	(1)	(2)
1669	$-32 - 9a$	2	3	(3)	(3)
1699	$40 + 40a$	1	2	(2)	(2)
1721	$-3 - 30a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1741	$-13 + 20a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1777	$-30 - 24a$	0	1	(1)	(2)
1789	$1 - 18a$	2	3	(3)	(3)
1801	$2 - 10a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1811	$-63$	—	—	—	$\lambda_5^2$
1823	$-43 - 23a$	—	—	—	(3)
1831	$12 - 35a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1861	$17 + 45a$	1	$3 + 6a$	$\lambda_5(3)$	$\lambda_5(3)$
1871	27	—	—	—	$\lambda_5$
1901	$27 + 30a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1931	$27 + 45a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$
1951	$-33 - 20a$	1	$1 + 2a$	$\lambda_5$	$\lambda_5$

### 7.3 Third Example

In our last example we consider a normalized cuspidal  $f \in S_2(\Gamma_0(133))$  of weight 2 and conductor 133 lying in the first Galois orbit of eigenforms. The first few coefficients of the Fourier expansion of  $f$  are

$$f = q + aq^2 + aq^3 - 3(a + 1)q^4 - (2a + 3)q^5 - (3a + 1)q^6 + \dots$$

where  $a \in E_f$  has minimal polynomial  $x^2 + 3x + 1$ . The same argument used in the second example shows that  $\bar{\rho}_{(2)}$  defines an  $A_5$ -extension  $K/\mathbf{Q}$ . Looking at Table 3, we observe that the primes 839, 941, 1663, 1783 and 1789 are completely split in  $K$ .

**Table 3** Integral Frobenius for  $A_f$ , where  $f$  lies in the first Galois orbit of  $S_2(\Gamma_0(133))$

$p$	$a_p$	$u_p$	$b_p$	$\text{Fac}(b_p)$	$\text{Fac}(b_{O_L})$
5	$-3 - 2a$	—	—	—	$\lambda_5$
11	$-3 + a$	0	1	(1)	(3)
29	$-3 + a$	0	1	(1)	(3)
47(**)	$-15 - 10a$	$a$	3	(3)	(3)
59	$-15 - 6a$	0	1	(1)	$\lambda_{11,1}$
79	$-10$	—	—	—	(3)
131	$-3 - 5a$	0	1	(1)	$\lambda_5(3)$
137	$6 - 4a$	0	1	(1)	(2)
173	$6 + 10a$	0	1	(1)	(3)
181	$-25 - 9a$	0	1	(1)	(3)
193	$-5 - 9a$	2	3	(3)	(3)
229	$-14 - 12a$	0	1	(1)	(2)
239	$15 + 4a$	$2a$	3	(3)	(3)
251	$-3 - 11a$	0	1	(1)	(3)
311	$3 - 5a$	4	$3 + 2a$	$\lambda_5$	$\lambda_5$
317	$30 + 12a$	0	1	(1)	(2)
389	$15 + 11a$	0	1	(1)	$\lambda_{19,2}$
431(**)	$-30 - 20a$	0	1	(1)	(3)
439	8	—	—	—	(2)(3)
443	$-12 + a$	$2a$	3	(3)	(3)
449	$-12 + 7a$	$2a$	3	(3)	(3)
457	$28 + 9a$	0	1	(1)	(3)
479	$51 + 25a$	0	1	(1)	$\lambda_5$
491	$-12 - 4a$	0	1	(1)	(2)
503(**)	$24 + 16a$	0	1	(1)	(2)(3)
509	30	—	—	—	(2)
541(**)	$18 + 12a$	0	1	(1)	(2)
571	$-23 - 18a$	0	1	(1)	(3)
599	$-6 - 5a$	2	$3 + 2a$	$\lambda_5$	$\lambda_5\lambda_{11,1}$
619	$10 - 9a$	0	1	(1)	(3)
631	$1 + 6a$	0	1	(1)	$\lambda_{11,1}$
661	$-26 - 24a$	0	1	(1)	(2)
677	$-42 - 19a$	0	1	(1)	(3)
719	$12 + 16a$	0	1	(1)	(2)
757	$10 + 12a$	0	1	(1)	(2)
787	$20 + 12a$	0	1	(1)	(2)
839	$24 + 20a$	1	2	(2)	$(2)\lambda_5$
857	$-69 - 37a$	$a$	3	(3)	(3)
911	$-6 + 8a$	$a$	3	(3)	(3)
941	$6 - 20a$	1	2	(2)	(2)
971	$-33 - 10a$	0	1	(1)	$\lambda_5$
977	$10a$	—	—	—	$\lambda_{89,1}$

(continued)

**Table 3** (continued)

$p$	$a_p$	$u_p$	$b_p$	$\text{Fac}(b_p)$	$\text{Fac}(b_{O_L})$
983	$-57 - 26a$	$2a$	3	(3)	(3)
1051	$38 + 15a$	0	1	(1)	$\lambda_5$
1061(**)	$3 + 2a$	$a$	3	(3)	(3)
1087	$37 + 27a$	2	3	(3)	(3)
1109	$6 - 4a$	0	1	(1)	(2)
1117	$-7 + 18a$	0	1	(1)	(3)
1217	$33 - 8a$	$2a$	3	(3)	(3)
1231	$-46 - 27a$	0	1	(1)	(3)
1249	$-44 - 18a$	2	3	(3)	(3)
1259	$21 + 25a$	3	$3 + 2a$	$\lambda_5$	$\lambda_5$
1303	$-54 - 30a$	0	1	(1)	$\lambda_{11,1}$
1361	$6 - 20a$	0	1	(1)	(2)
1367	$-36 - 8a$	0	1	(1)	(2)
1409	$66 + 32a$	0	1	(1)	(2)(3)
1447	$-56 - 48a$	0	1	(1)	(2)
1451	$-33 - 37a$	—	—	—	(3)
1483	$62 + 45a$	1	3	(3)	(3)
1487	$-84 - 53a$	—	—	—	(3)
1493	$-54 - 28a$	0	1	(1)	(2)
1531	43	—	—	—	$\lambda_5^2(3)$
1553	$-75 - a$	0	1	(1)	(7)
1567	$-38 - 18a$	0	1	(1)	(3) $\lambda_{11,2}$
1609	$-17 + 15a$	—	—	—	$\lambda_{109,1}$
1663	$44 + 24a$	1	2	(2)	(2)
1669	$49 + 18a$	2	3	(3)	(3)
1723	$56 + 24a$	0	1	(1)	(2)
1733	$-24 - 13a$	0	1	(1)	(3)
1741	$7 + 15a$	0	1	(1)	$\lambda_5$
1753	$-10 + 27a$	1	3	(3)	(3)
1759	$14 - 9a$	0	1	(1)	(3)
1783	$-32 + 12a$	1	2	(2)	(2)
1789	$-6 - 12a$	1	2	(2)	(2)
1823	$-84 - 23a$	—	—	—	(3)
1847(**)	$-48 - 32a$	1	2	(2)	(2)
1871	24	—	—	—	(2)
1873	$-43 - 18a$	0	1	(1)	(3)
1879	-35	—	—	—	(3)
1889	$-105 - 52a$	—	—	—	(3)
1907	$-60 - 7a$	0	1	(1)	(3)
1933	$10 - 9a$	0	1	(1)	(3)
1973	$-6 - 16a$	0	1	(1)	(2)
1987	$-10 - 27a$	1	3	(3)	(3)

**Acknowledgements** We wish to thank Gebhard Böckle for all the useful discussions we had on this project. We thank Gaetan Bisson for an intense email correspondence where he helped us understanding better his work. The first author thanks Jordi Guàrdia and Josep González for the warm hospitality he received during his visit to Universitat Politècnica de Catalunya in July 2015. This work was supported by the DFG Priority Program SPP 1489 and the Luxembourg FNR.

## References

1. G. Bisson, Computing endomorphism rings of abelian varieties of dimension two. *Math. Comput.* **84**, 1977–1989 (2015)
2. W. Bosma, J. Cannon, C. Playoust, The Magma algebra system. I. The user language. *J. Symb. Comput.* **24**, 235–265 (1997)
3. C.-L. Chai, B. Conrad, F. Oort, *Complex Multiplication and Lifting Problems*. Mathematical Surveys and Monographs, vol. 195 (American Mathematical Society, Providence, RI, 2014)
4. W. Duke, Á. Tóth, The splitting of primes in division fields of elliptic curves. *Exp. Math.* **11**(4), 555–565 (2002)
5. E. González-Jiménez, J. González, J. Guàrdia, Computations on modular Jacobian surfaces, in *Algorithmic Number Theory (Sidney 2002)*. Lecture Notes in Computer Science, vol. 2369 (Springer, Berlin, 2002), pp. 189–197
6. J. González, J. Guàrdia, V. Rotger, Abelian surfaces of  $GL_2$ -type as Jacobians of curves. *Acta Arith.* **116**(3), 263–287 (2005)
7. B. Mazur, Modular curves and the Eisenstein ideal. *Inst. Hautes Études Sci. Publ. Math.* **47**, 33–186 (1977)
8. W. Narkiewicz, *Elementary and Analytic Theory of Algebraic Numbers*, 3rd edn. (Springer, Berlin, 2004)
9. K. Ribet, Galois action on division points of abelian varieties with real multiplications. *Am. J. Math.* **98**(3), 751–804 (1976)
10. J.-P. Serre, *Corps Locaux*, Quatrième édition, corrigée (Hermann, Paris, 2004)
11. J.-P. Serre, J. Tate, Good reduction of abelian varieties, *Ann. Math.* **88**(3), 492–517 (1968)
12. G. Shimura, Introduction to the arithmetic theory of automorphic forms, *Publications of the Mathematical Society of Japan*, vol. 11. Kanô Memorial Lectures, 1 (Princeton University Press, Princeton, NJ, 1994)
13. J. Tate, Endomorphisms of abelian varieties over finite fields. *Invent. math.* **2**, 134–144 (1966)
14. J. Tate, *Classes d'isogénie des variétés abéliennes sur un corps fini*, Sém. Bourbaki 21e année, no. 352 (1968/69)

# Monodromy of the Multiplicative and the Additive Convolution



Michael Dettweiler and Mirjam Jöllenbeck

**Abstract** We give an algorithmic approach for the computation of the monodromy of the additive and the multiplicative convolution in terms of singular cohomology.

**Keywords** Convolution • Monodromy

**Subject Classifications** 14D05, 32S40

## 1 Introduction

Convolution integrals of the form

$$f * g(y) := \int f(x)g(y-x)dx \quad (\text{additive convolution})$$

or

$$f \star g(y) := \int f(x)g\left(\frac{y}{x}\right) dx \quad (\text{multiplicative convolution})$$

play an important role in many areas of mathematics and physics, see e.g. [15] and [18].

---

M. Dettweiler (✉)  
University of Bayreuth, Bayreuth, Germany  
e-mail: [michael.dettweiler@uni-bayreuth.de](mailto:michael.dettweiler@uni-bayreuth.de); [michael.dettweiler@unibayreuth.de](mailto:michael.dettweiler@unibayreuth.de)

M. Jöllenbeck  
e-mail: [joellenbeck.mirjam@gmail.com](mailto:joellenbeck.mirjam@gmail.com)



It is often important to know the analytic continuation of the functions  $f * g$  and  $f \star g$  (i.e., their monodromy) if  $f$  and  $g$  are complex-valued analytic functions with regular singularities. It is the aim of this article to provide a general method to compute the monodromy of additive and multiplicative convolutions in a systematic and algorithmic way. In this way, we e.g. obtain explicit monodromy tuples of the motivic local systems with  $G_2$ -monodromy, considered in [5, Cor. 2.4.2]:

$$\begin{pmatrix} 1 & 0 & 0 & 2 & 2 & 0 & 0 \\ 0 & 1 & 0 & -2 & 0 & 2 & 0 \\ 0 & 0 & 1 & 2 & 2 & 2 & 2 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 & 4 & 4 & 1 \end{pmatrix}, \begin{pmatrix} 5 & 4 & 0 & 2 & 2 & 0 & 0 \\ -4 & 1 & 4 & -2 & 0 & 2 & 0 \\ 4 & 4 & 5 & 6 & 10 & 10 & 2 \\ -2 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & -2 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -2 & -4 & -4 & -1 \end{pmatrix}.$$

The knowledge of such matrices is especially important for refined considerations of monodromy groups in terms of finite, arithmetic or thin subgroups of linear algebraic groups, cf. [12].

Using a shearing transformation  $(x, y) \mapsto (x, y + x)$  (resp.  $(x, y) \mapsto (x, xy)$ ) the convolution integrals  $f * g$  or  $f \star g$  can be seen as local sections of a sheaf theoretic higher direct image

$$R^1 \pi_1(\mathcal{V}_1 \boxtimes \mathcal{V}_2)$$

where  $\pi$  is either the addition map on  $\mathbb{A}^1$  (if  $f * g$  is considered) or  $\pi$  denotes the multiplication map on  $\mathbb{G}_m$  (if  $f \star g$  is considered). Because we are mainly interested in irreducible objects, we study middle (or intermediate) convolutions  $\mathcal{V}_1 * \mathcal{V}_2$  or  $\mathcal{V}_1 \star \mathcal{V}_2$  of the form

$$\text{im}(R^1 \pi_1(\mathcal{V}_1 \boxtimes \mathcal{V}_2) \rightarrow R^1 \pi_*(\mathcal{V}_1 \boxtimes \mathcal{V}_2)),$$

which, by shearing back, can be interpreted as suitable variations of parabolic cohomology groups, cf. Eq. (10) and Sects. 3.1 and 4.1.

The monodromy representation of  $\mathcal{V}_1 * \mathcal{V}_2$  or  $\mathcal{V}_1 \star \mathcal{V}_2$  in dependence of the monodromy representations of the initial local systems  $\mathcal{V}_1$  and  $\mathcal{V}_2$  is then determined using the unifying theory of braid groups and the parabolic cohomology of variations of local systems, developed in [10].

This article complements a series of articles on the convolutions worked out within the realm of the SPP 1489:

- Classification of orthogonally rigid local systems with  $G_2$ -monodromy using middle convolution [6].
- Motives for rigid  $G_2$ -local systems [9].
- Classification of irregular rigid  $D$ -modules whose differential Galois group is equal to  $G_2$  [13].

- Hodge theoretical description of the additive middle convolution with Kummer sheaves [8].
- The global and local Hodge data for general additive and multiplicative convolutions are worked out in [7] (building up on [8]).

An associated program [14] for the computation of variations of parabolic cohomology and especially both types of convolutions will appear on the homepage of one of the authors.

## 2 Preliminary Results

### 2.1 Group Theoretical Definitions and Tensor Product of Representations

Throughout the article,  $R$  denotes an integral domain with fraction field  $K$ . If  $V$  is a free  $R$ -module, then  $\text{GL}(V)$  denotes the  $R$ -linear isomorphisms of  $V$ . As usual, the group of  $R$ -linear isomorphisms of  $R^n$  is denoted by  $\text{GL}_n(R)$ . Linear automorphisms act from the right, i.e., if  $A \in \text{GL}(V)$  and  $v \in V$ , then  $vA$  denotes the image of  $v$  under  $A$ .

A Jordan block of eigenvalue  $\alpha \in R$  and of length  $l$  is denoted by  $J(\alpha, l)$ . We write

$$J(\alpha_1, n_1) \oplus \cdots \oplus J(\alpha_k, n_k)$$

for a block matrix in  $\text{GL}_{n_1+\dots+n_k}(R)$  which is in Jordan normal form and whose Jordan blocks are  $J(\alpha_1, n_1), \dots, J(\alpha_k, n_k)$ . Let  $V_1, \dots, V_t$  be free  $R$ -modules having rank  $n_1, \dots, n_t$  (respectively). Set  $n := \prod_{i=1}^t n_i$  and  $V := V_1 \otimes \cdots \otimes V_t$ . For  $g_i \in \text{GL}(V_i)$ , define the elements  $g_1 \otimes \cdots \otimes g_t \in \text{GL}(V)$  by setting

$$(v_1 \otimes \cdots \otimes v_t)(g_1 \otimes \cdots \otimes g_t) := v_1 g_1 \otimes \cdots \otimes v_t g_t.$$

This tensor product of matrices is also called the *Kronecker product*. Let  $\rho_1 : H_1 \rightarrow \text{GL}(V_1)$  and  $\rho_2 : H_2 \rightarrow \text{GL}(V_2)$  be two representations, then the tensor product defines a representation

$$\rho_1 \otimes \rho_2 : H_1 \times H_2 \longrightarrow \text{GL}(V_1 \otimes V_2), (h_1, h_2) \longmapsto h_1 \otimes h_2.$$

It is often important to compute the Jordan normal form of the tensor product  $A \otimes B$  of two matrices  $A \in \text{GL}_n(K)$  and  $B \in \text{GL}_m(K)$ . In characteristic 0 this can be done using the following well known lemma (cf. [17, Table 5, Case  $A_1$  on p. 300]):

**Lemma 2.1.1** *Let  $K$  be an algebraically closed field of characteristic zero, let  $\alpha, \beta \in K$ , and let  $n_1 \leq n_2$ . Let  $J(\alpha, n_1) \in \text{GL}_{n_1}(K)$  and  $J(\beta, n_2) \in \text{GL}_{n_2}(K)$  be*

two Jordan blocks. Then the Jordan normal form of  $J(\alpha, n_1) \otimes J(\beta, n_2)$  is given by

$$\bigoplus_{i=0}^{n_1-1} J(\alpha\beta, n_1 + n_2 - 1 - 2i).$$

## 2.2 Braid Groups and Affine Fibrations

We will write  $\mathbb{A}^1, \mathbb{P}^1, \dots$  instead of  $\mathbb{A}^1(\mathbb{C}), \mathbb{P}^1(\mathbb{C}), \dots$  and view these objects equipped with their associated topological and complex analytic structures. Let  $X$  be a connected topological manifold and let  $\pi_1(X, x_0)$  denote the fundamental group of  $X$  with base point  $x_0$ . The product of two elements  $\alpha, \beta \in \pi_1(X, x_0)$  is given by (the homotopy class of)  $\alpha\beta$ , where the path product  $\alpha\beta$  is given by first walking through  $\alpha$  and then through  $\beta$ . Let  $r \in \mathbb{N}$  with  $r \geq 2$  and let  $U_0 = \mathbb{A}^1 \setminus \mathbf{u}$ , where  $\mathbf{u} := \{u_1, \dots, u_r\}$  is a finite subset of  $\mathbb{A}^1$ . We will identify  $U_0$  with  $\mathbb{P}^1 \setminus (\mathbf{u} \cup \{\infty\})$  in the obvious way. Using a suitable homeomorphism  $\kappa : \mathbb{P}^1 \rightarrow \mathbb{P}^1$ , (called a *marking* in [10, Section 1.2]) and the conventions of loc. cit., Section 2.3, one obtains generators  $\alpha_1, \dots, \alpha_{r+1}$  of  $\pi_1(U_0, u_0)$  which satisfy the product relation  $\alpha_1 \cdots \alpha_{r+1} = 1$ . If  $\mathbf{u} = \{u_1, \dots, u_r, u_{r+1} = \infty\}$  is elementwise real then  $\alpha_i$  approaches  $u_i$  in the upper half plane ( $i = 1, \dots, r$ ) and then moves counterclockwise around  $u_i$  once and goes back to  $u_0$  in the upper half plane. Let

$$\mathcal{O}_r := \{\mathbf{u} \subseteq \mathbb{A}^1 \mid |\mathbf{u}| = r\} \quad \text{and} \quad \mathcal{O}_{r,1} := \{(\mathbf{u}, x) \in \mathcal{O}_r \times \mathbb{P}^1 \mid x \notin \mathbf{u}\}.$$

The sets  $\mathcal{O}_r$  and  $\mathcal{O}_{r,1}$  are connected topological manifolds in a natural way and we set  $\mathcal{A}_r := \pi_1(\mathcal{O}_r, \mathbf{u})$  and  $\mathcal{A}_{r,1} := \pi_1(\mathcal{O}_{r,1}, (\mathbf{u}, u_0))$ . Then, the marking on  $U_0$  also defines standard generators  $\beta_1, \dots, \beta_{r-1}$  of  $\mathcal{A}_r$  which satisfy the usual relations of the standard generators of the Artin braid groups. If  $\mathbf{u}$  is elementwise real with  $u_1 < \dots < u_r$ , then  $\beta_i$  fixes all  $u_j$  with  $j \neq i, i + 1$  and interchanges  $u_i$  and  $u_{i+1}$  via a braid by moving  $u_i$  to  $u_{i+1}$  along the real axis and moving  $u_{i+1}$  to  $u_i$  in the upper half plane (except for the initial and the end point). One obtains a split exact sequence

$$1 \longrightarrow \pi_1(U_0, u_0) \longrightarrow \mathcal{A}_{r,1} \longrightarrow \mathcal{A}_r \longrightarrow 1 \tag{1}$$

such that the following equations holds (cf. [1, L. 1.8.2 and Cor. 1.8.3]):

$$\beta_i^{-1} \alpha_j \beta_i = \begin{cases} \alpha_i \alpha_{i+1} \alpha_i^{-1}, & \text{for } j = i, \\ \alpha_i, & \text{for } j = i + 1, \\ \alpha_j, & \text{otherwise.} \end{cases} \tag{2}$$

As usual, one sees that  $\mathcal{A}_r$  acts (product preserving) as follows on  $G^r$ , where  $G$  is any group:

$$(g_1, \dots, g_r)^{\beta_i} = (g_1, \dots, g_{i-1}, g_{i+1}, g_{i+1}^{-1} g_i g_{i+1}, g_{i+2}, \dots, g_r), \quad \forall (g_1, \dots, g_r) \in G^r. \quad (3)$$

Let

$$\mathcal{O}^r := \{(v_1, \dots, v_r) \in \mathbb{C}^r \mid v_i \neq v_j \text{ for } i \neq j\}.$$

Let  $\mathcal{A}^r := \pi_1(\mathcal{O}^r, (u_1, \dots, u_r))$  be the pure braid group. The map

$$\mathcal{O}^r \longrightarrow \mathcal{O}_r, (v_1, \dots, v_r) \longmapsto \{v_1, \dots, v_r\}$$

is a unramified covering map. Thus (via the lifting of paths)  $\mathcal{A}^r$  can be seen as a subgroup of  $\mathcal{A}_r$ . It is well known, that  $\mathcal{A}^r$  as such is generated by the following braids (cf. [16]):

$$\beta_{ij} := (\beta_i^2)^{\beta_{i+1}^{-1} \dots \beta_{j-1}^{-1}} = (\beta_{j-1}^2)^{\beta_{j-2} \dots \beta_i}, \quad (4)$$

where  $1 \leq i < j \leq r$ .

Let  $S$  be a smooth connected complex manifold, let  $X := \mathbb{P}_S^1 = \mathbb{P}^1 \times S$ , and let  $d \subseteq X$  be a smooth relative divisor of degree  $r + 1$  over  $S$  which contains the section  $\{\infty\} \times S$ . Let  $U := X \setminus d$ , let  $j : U \rightarrow X$  the natural inclusion, let  $\bar{\pi} : X \rightarrow S$  be the projection onto  $S$ , and let  $\pi : U \rightarrow S$  be the restriction of  $\bar{\pi}$  to  $U$ . Let further  $s_0 \in S$  and let  $U_0 := \pi^{-1}(s_0)$ . One has a continuous map

$$S \longrightarrow \mathcal{O}_r, s \longmapsto \pi'(\bar{\pi}^{-1}(s) \cap d) \setminus \infty,$$

where  $\pi'$  denotes the projection of  $X$  onto  $\mathbb{P}^1$ . This map induces a homomorphism of fundamental groups  $\phi : \pi_1(S, s_0) \rightarrow \mathcal{A}_r$ . Similarly, the map

$$U \longrightarrow \mathcal{O}_{r,1}, (u, s) \longmapsto (\pi'(\bar{\pi}^{-1}(s) \cap d) \setminus \infty, \pi'(u)),$$

gives rise to a homomorphism  $\tilde{\phi} : \pi_1(U, (u_0, s_0)) \rightarrow \mathcal{A}_{r,1}$ .

In [10] it is shown how to obtain a commuting diagram whose rows are split exact sequences:

$$\begin{array}{ccccccc} 1 & \longrightarrow & \pi_1(U_0, u_0) & \longrightarrow & \pi_1(U, (u_0, s_0)) & \longrightarrow & \pi_1(S, s_0) \longrightarrow 1 \\ & & \downarrow & & \tilde{\phi} \downarrow & & \phi \downarrow \\ 1 & \longrightarrow & \pi_1(U_0, u_0) & \longrightarrow & \mathcal{A}_{r,1} & \longrightarrow & \mathcal{A}_r \longrightarrow 1. \end{array} \quad (5)$$

Let  $\iota_1 : \pi_1(S, s_0) \rightarrow \pi_1(U, (u_0, s_0))$  denote the splitting of the upper row coming from the  $\infty$ -section, and let  $\iota_2 : \mathcal{A}_r \rightarrow \mathcal{A}_{r,1}$  be the splitting of the lower row induced from the section  $D \in \mathcal{O}_{r,1} \mapsto (D, \infty)$ . Then

$$\tilde{\phi} \circ \iota_1 = \iota_2 \circ \phi \tag{6}$$

(see loc. cit., Section 2.3 and Rem. 2.6).

### 2.3 Local Systems and Representations of Fundamental Groups

Let  $X$  be a connected topological manifold. A *local system of  $R$ -modules* is a sheaf  $\mathcal{V} \in \text{Sh}_R(X)$  for which there exists an  $n \in \mathbb{N}$  such that  $\mathcal{V}$  is locally isomorphic to  $R^n$ . The number  $n$  is called the *rank* of  $\mathcal{V}$  and is denoted by  $\text{rk}(\mathcal{V})$ . Let  $\text{LS}_R(X)$  denote the category of local systems of  $R$ -modules on  $X$ . Any local system  $\mathcal{V} \in \text{LS}_R(X)$  gives rise to its *monodromy representation*  $\rho_{\mathcal{V}} : \pi_1(X, x_0) \rightarrow \text{GL}(V)$ . (We always let  $\pi_1(X, x_0)$  act from the right on  $V$ .) Let  $\text{Rep}_R(\pi_1(X, x_0))$  denote the category of representations  $\pi_1(X, x_0) \rightarrow \text{GL}(V)$ , where  $V \simeq R^n$  for some  $n \in \mathbb{N}$ . One has an equivalence of categories:  $\text{LS}_R(X) \cong \text{Rep}_R(\pi_1(X, x_0))$  with  $\mathcal{V}$  corresponding to  $\rho_{\mathcal{V}}$ .

Let  $U_0 := \mathbb{A}^1 \setminus \mathbf{u}$ ,  $\mathbf{u} \in \mathcal{O}_r$ , and fix generators  $\alpha_1, \dots, \alpha_{r+1}$  of  $\pi_1(U_0, u_0)$  as in Sect. 2.2. If  $\mathcal{V}$  is a given local system on  $U_0$ , then

$$\begin{aligned} \mathcal{V} \in \text{LS}_R(U_0) &\longleftrightarrow \rho_{\mathcal{V}} \in \text{Rep}_R(\pi_1(U_0, u_0)) \\ &\longleftrightarrow T_{\mathcal{V}} := (T_1 := \rho_{\mathcal{V}}(\alpha_1), \dots, T_{r+1} := \rho_{\mathcal{V}}(\alpha_{r+1})) \\ &\qquad \in \text{GL}(V)^{r+1}, T_1 \cdots T_{r+1} = 1. \end{aligned}$$

We call  $T_{\mathcal{V}}$  the *monodromy tuple* of  $\mathcal{V}$ . The equivalence class of  $T_i$  under GL-conjugation is called *local monodromy at  $x_i$* .

### 2.4 Cohomology of Local Systems on $U_0$

The results of this section can be found in [10]. Let  $U_0 = \mathbb{A}^1 \setminus \mathbf{u}$  be as in Sect. 2.2. Let  $\mathcal{V}_0 \in \text{LS}_R(U_0)$  and let

$$T := T_{\mathcal{V}_0} = (T_1, \dots, T_{r+1}) \in \text{GL}(V)^{r+1}$$

denote its monodromy tuple. It is shown in [10] that the group  $H^1(U_0, \mathcal{V}_0)$  is isomorphic to  $H_T/E_T$ , where

$$H_T := \{(v_1, \dots, v_{r+1}) \in V^{r+1} \mid v_1(T_2 \cdots T_{r+1}) + v_2(T_3 \cdots T_{r+1}) + \cdots + v_{r+1} = 0\} \tag{7}$$

and

$$E_T := \{(v(T_1 - 1), \dots, v(T_{r+1} - 1)) \mid v \in V\}. \tag{8}$$

The isomorphism is given as the composition of the natural isomorphism

$$H^1(U_0, \mathcal{V}_0) \rightarrow H^1(\pi_1(U_0, x_0), V)$$

with the evaluation map, which associates to the equivalence class of a crossed homomorphism  $[\delta] \in H^1(\pi_1(U_0, x_0), V)$  the corresponding equivalence class of  $[(\delta(\alpha_1), \dots, \delta(\alpha_{r+1}))]$  in  $V^{r+1}/E_T$ .

Let  $j : \mathcal{U}_0 \rightarrow \mathbb{P}^1$  be the natural inclusion. It is shown in loc. cit. that the parabolic cohomology group  $H_p^1(U_0, \mathcal{V}_0) := H^1(\mathbb{P}^1, j_*(\mathcal{V}_0))$  is isomorphic to  $U_T/E_T$ , where

$$U_T := \{(v_1, \dots, v_{r+1}) \in H_T \mid v_i \in \text{im}(T_i - 1), i = 1, \dots, r + 1\}. \tag{9}$$

Here, the additional relations arise from the natural isomorphism

$$H_p^1(U_0, \mathcal{V}_0) \simeq \text{im}(H_c^1(U_0, \mathcal{V}_0) \rightarrow H^1(U_0, \mathcal{V}_0)). \tag{10}$$

### 2.5 Variation of Parabolic Cohomology

Recall the basic setting of [10]: Let  $S$  be a smooth connected complex manifold, let  $X := \mathbb{P}_S^1 = \mathbb{P}^1 \times S$ , and let  $d \subseteq X$  be a smooth relative divisor of degree  $r + 1$  over  $S$  which contains the section  $\{\infty\} \times S$ . Let  $U := X \setminus d$ , let  $j : U \rightarrow X$  be the natural inclusion, let  $\bar{\pi} : X \rightarrow S$  be the projection onto  $S$ , and let  $\pi : U \rightarrow S$  be the restriction of  $\bar{\pi}$  to  $U$ . Let further  $s_0 \in S$  and let  $U_0 := \pi^{-1}(s_0)$ .

A local system  $\mathcal{V} \in \text{LS}_R(U)$  is called a *variation of  $\mathcal{V}_0 \in \text{LS}_R(U_0)$*  over  $S$ , if  $\mathcal{V}_0 = \mathcal{V}|_{U_0}$ . Under the isomorphism  $\pi_1(U) \simeq \pi_1(U_0) \rtimes \pi_1(S)$  (cf. (5)),  $\mathcal{V}$  corresponds to

$$\rho : \pi_1(U) \rightarrow \text{GL}(V), \gamma\delta \mapsto \rho_0(\gamma) \cdot \chi(\delta), \tag{11}$$

where  $\rho_0$  is the restriction of  $\rho$  to  $\pi_1(U_0)$  and  $\chi$  is the restriction of  $\rho$  to  $\pi_1(S) \leq \pi_1(U)$ . The *parabolic cohomology* of this variation is by definition the first higher direct image  $\mathcal{W} := R^1\bar{\pi}_*(j_*\mathcal{V})$ . It is a local system on  $S$  whose stalk  $\mathcal{W}_{s_0}$  is canonically isomorphic to the parabolic cohomology group  $H_p^1(U_0, \mathcal{V}_0)$  (see loc. cit). Thus  $\mathcal{W}$  corresponds to its monodromy representation

$$\rho_{\mathcal{W}} : \pi_1(S, s_0) \longrightarrow \text{GL}(H_p^1(U_0, \mathcal{V}_0)) \cong \text{GL}(U_T/E_T),$$

where  $T := T_{\mathcal{V}_0}$  is the associated tuple of  $\mathcal{V}_0$  and  $U_T$  and  $E_T$  are as in the last section.

We want to determine the representation  $\rho_{\mathcal{W}}$ . For this, let  $\beta_1, \dots, \beta_{r-1}$  denote the generators of  $\mathcal{A}_r$ . Consider linear automorphisms  $\Phi(T, \beta_i)$  of  $V^{r+1}$  which are

defined as follows:

$$\begin{aligned} & (v_1, \dots, v_{r+1})^{\Phi(T, \beta_i)} \\ &= (v_1, \dots, v_{i-1}, v_{i+1}, \underbrace{v_{i+1}(1 - T_{i+1}^{-1}T_iT_{i+1}) + v_iT_{i+1}}_{(i+1)\text{th entry}}, v_{i+2}, \dots, v_{r+1}). \end{aligned} \tag{12}$$

These automorphisms multiply by the following rule:

$$\Phi(T, \beta) \cdot \Phi(T^\beta, \beta') = \Phi(T, \beta\beta'). \tag{13}$$

It is easy to see, that the spaces  $U_T$  and  $E_T$  under

$$\bar{\Phi}(T, \phi(\gamma)), \quad \gamma \in \pi_1(S, s_0),$$

are mapped isomorphically to the spaces  $U_{T^{\phi(\gamma)}}$ , resp.  $E_{T^{\phi(\gamma)}}$  (where  $\phi(\gamma)$  is as in Sect. 2.2 and acts as in (3) on  $\text{GL}(V^{r+1})^r$ ). Let

$$\bar{\Phi}(T, \phi(\gamma)) : U_T/E_T \longrightarrow U_{T^{\phi(\gamma)}}/E_{T^{\phi(\gamma)}}$$

be the isomorphism induced by  $\Phi(T, \phi(\gamma))$ .

For  $T \in \text{GL}(V)^{r+1}$  as above and  $h \in \text{GL}(V)$  we similarly obtain a linear map

$$\Psi(T, h) : H_T \longrightarrow H_{T^h}, \quad (v_1, \dots, v_{r+1}) \longmapsto (v_1h, \dots, v_{r+1}h),$$

descending to an isomorphism  $\bar{\Psi}(T, h) : H_T/E_T \longrightarrow H_{T^h}/E_{T^h}$ , where  $T^h$  arises from  $T$  by elementwise conjugation by  $h$ .

**Proposition 2.5.1**

(i)

$$\rho_{\mathcal{W}}(\gamma) = \bar{\Phi}(T, \phi(\gamma)) \circ \bar{\Psi}(T, \chi(\gamma)), \quad \forall \gamma \in \pi_1(S, s_0).$$

(ii) (Ogg-Shafarevich) *Suppose that  $R = K$  is a field and that the stabilizer  $V^{\pi_1(U_0)}$  is trivial. Then*

$$\text{rk}(\mathcal{W}) = \dim_K H_p^1(U_0, \mathcal{V}_0) = (r - 1) \dim_K V - \sum_{i=1}^{r+1} \dim_K \text{Ker}(T_i - 1),$$

where  $T = (T_1, \dots, T_{r+1})$  is the monodromy tuple of  $\mathcal{V}_0$ .

(iii) (Poincaré Duality) *Let  $\mathcal{V} \otimes \mathcal{V} \rightarrow R$  be a non-degenerate symmetric (resp. alternating) bilinear pairing of sheaves. Then the cup product defines a non-degenerate alternating (resp. symmetric) bilinear pairing of sheaves  $\mathcal{W} \otimes \mathcal{W} \rightarrow R$ .*

*Proof* Claim (i) follows immediately from [10, Thm. 2.5] (using the above Diagram (5) and (6)), (ii) is [10, Rem. 1.3]. See [11] for (iii).  $\square$

### 3 Multiplicative Convolution

#### 3.1 Definition of the Multiplicative Convolution

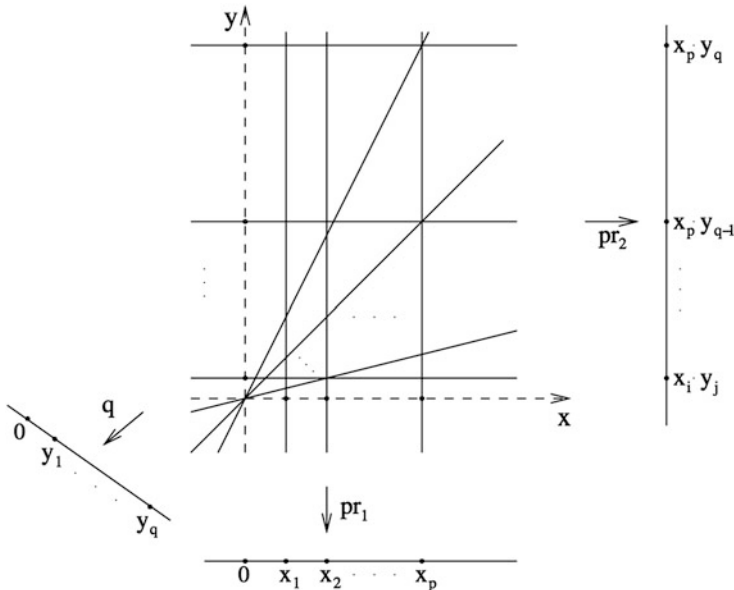
For  $\mathbf{u} := \{0, x_1, \dots, x_p\} \in \mathcal{O}_{p+1}$  and  $\mathbf{v} := \{0, y_1, \dots, y_q\} \in \mathcal{O}_{q+1}$  define

$$\mathbf{u} \cdot \mathbf{v} := \{x_i y_j \mid i = 1, \dots, p, j = 1, \dots, q\} \cup \{0\}.$$

We call  $\mathbf{u} \cdot \mathbf{v}$  *generic*, if the cardinality of  $\mathbf{u} \cdot \mathbf{v}$  is equal to  $pq + 1$ . Let  $U_1 := \mathbb{A}^1 \setminus \mathbf{u}$ ,  $U_2 := \mathbb{A}^1 \setminus \mathbf{v}$  and  $S := \mathbb{A}^1 \setminus \mathbf{u} \cdot \mathbf{v}$ . Set

$$\tilde{f}(x, y) := xy \prod_{i=1}^p (x - x_i) \prod_{j=1}^q (y - y_j x) \prod_{ij} (y - x_i y_j),$$

let  $f(x, y)$  denote the associated reduced polynomial, let  $\tilde{\mathbf{w}} := \{(x, y) \in \mathbb{A}^2 \mid f(x, y) = 0\}$ , and let  $U := \mathbb{A}^2 \setminus \tilde{\mathbf{w}}$ .



We have the quotient map  $q : U \rightarrow U_2, (x, y) \mapsto y/x$  and the partial completion  $j : U \rightarrow \mathbb{P}^1 \times S =: X, (x, y) \mapsto ([x, 1], y)$ . Define  $\mathbf{w} := X \setminus U$ . Then we are in the



situation of Sect. 2.5 with  $r = p + q + 1$  and  $\pi := \text{pr}_2 : U \rightarrow S, (x, y) \mapsto y$ . Let  $\overline{\text{pr}}_2 : X \rightarrow S$  be the second projection and for  $s_0 \in S$  let  $U_0 := \overline{\text{pr}}^{-1}(s_0)$  (note that  $U_0$  can be identified with  $\mathbb{A}^1 \setminus (\mathbf{u} \cup \{s_0/y_1, \dots, s_0/y_q\})$  via the projection  $\text{pr}_1$  to the  $x$ -coordinate).

Let  $\mathcal{V}_i \in \text{LS}_R(U_i)$  ( $i = 1, 2$ ) be irreducible and nonconstant. We further assume that  $\mathcal{V}_1$  has nontrivial local monodromy at at least two different points  $x_{i_1}, x_{i_2} \neq \infty$ . The *multiplicative middle convolution* (or *middle Hadamard product*) is then defined as a variation of parabolic cohomology groups

$$\mathcal{V}_1 \star \mathcal{V}_2 := R^1 \overline{\text{pr}}_{2*} (j_* (\text{pr}_1^* \mathcal{V}_1 \otimes q^* \mathcal{V}_2)) \in \text{LS}(S).$$

In the following we assume that the coefficient domain  $R$  is a field  $K$ .

### 3.2 Monodromy of the Multiplicative Middle Convolution

Choose homotopy generators  $\gamma_0, \dots, \gamma_{p+q}$  of  $\pi_1(U_0, x_0)$  as in Sect. 2.2, where we identify  $U_0$  with  $\mathbb{A}_x^1 \setminus (\mathbf{u} \cup \{s_0/y_q, \dots, s_0/y_1\})$  via  $\text{pr}_1$  (note that  $s_0/y_q < \dots < s_0/y_1$ ). Let  $(A_0, A_1, \dots, A_p, A_\infty) \in \text{GL}(V_1)^{p+2}$  be the monodromy tuple of  $\mathcal{V}_1$  (w.r. to the  $\text{pr}_{1*}(\gamma_i)$  ( $i = 0, \dots, p$ ) and let

$$B_\infty := \rho_{\mathcal{V}_2}(q_*(\gamma_0)), B_q := \rho_{\mathcal{V}_2}(q_*(\gamma_{p+1})), \dots, B_1 := \rho_{\mathcal{V}_2}(q_*(\gamma_{p+q}))$$

and  $B_0 := (B_0 \cdots B_q)^{-1}$ . Since the map  $q|_{U_0}$  is given by  $x \mapsto s_0/x$ , it interchanges 0 and  $\infty$  and maps  $\gamma_{p+1}, \dots, \gamma_{p+q}$  to simple closed loops approaching  $y_q = q(s_0/y_q), \dots, y_1$  (in this order) using a path in the lower half plane. Let  $\alpha_0, \dots, \alpha_q, \alpha_\infty$  be a standard generating system of  $\pi_1(\mathbb{A}^1 \setminus \{y_1, \dots, y_q\}, y_0 = s_0/x_0)$  as in Sect. 2.2, where  $\alpha_i$  approaches  $y_i$  inside the upper half plane, before encircling it. Then the following holds:

$$q_*(\gamma_{p+1}) = \alpha_q, \quad q_*(\gamma_{p+2}) = \alpha_{q-1}^{\alpha_q}, \dots, \quad q_*(\gamma_{p+q}) = \alpha_1^{\alpha_2 \cdots \alpha_q},$$

$$q_*(\gamma_0) = \alpha_\infty \quad \text{and} \quad q_*(\gamma_\infty) = \alpha_0^{\alpha_1 \cdots \alpha_q}.$$

Let  $(C_0, \dots, C_{q+1} = C_\infty) \in \text{GL}(V_2)$  is the monodromy tuple w.r. to  $\alpha_0, \dots, \alpha_\infty$ . Then we can express the  $B_i$ 's in terms of the  $C_i$ 's as follows:

$$B_\infty = C_\infty, B_1 = C_1^{C_2 \cdots C_q}, B_2 = C_2^{C_3 \cdots C_q}, \dots, B_q = C_q, B_0 = C_0^{C_1 \cdots C_q}. \tag{14}$$

Then the monodromy tuple of  $\text{pr}_1^* \mathcal{V}_2 \otimes \mathcal{Q}^* \mathcal{V}_2|_{U_0}$  with respect to  $(\gamma_0, \dots, \gamma_{p+q}, \gamma_\infty)$  is as follows:

$$(A_0 \otimes B_\infty, A_1 \otimes 1_{V_2}, \dots, A_p \otimes 1_{V_2}, 1_{V_1} \otimes B_q, \dots, 1_{V_1} \otimes B_1, A_\infty \otimes B_0). \tag{15}$$

Therefore Proposition 2.5.1 (ii) implies the following (cf. Proposition 2.1.1 for the computation of expressions like  $\dim(\ker(A_0 \otimes C_\infty - 1_{V_1 \otimes V_2}))$ ):

**Proposition 3.2.1**

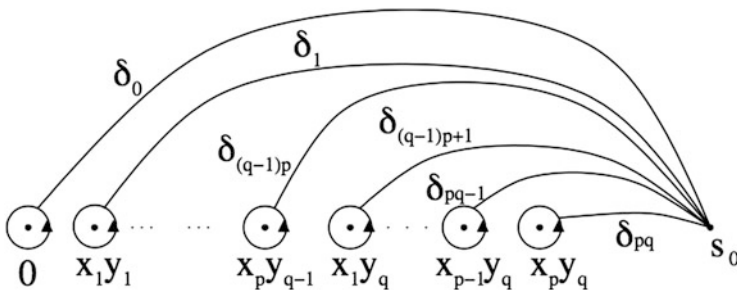
$$\begin{aligned} \text{rk}(\mathcal{V}_1 \star \mathcal{V}_2) &= (p + q)\text{rk}(\mathcal{V}_1)\text{rk}(\mathcal{V}_2) - \text{rk}(\mathcal{V}_2) \sum_{i=1}^p \dim \ker(A_i - 1_{V_1}) \\ &\quad - \text{rk}(\mathcal{V}_1) \sum_{j=1}^q \dim \ker(C_j - 1_{V_2}) \\ &\quad - \dim \ker(A_0 \otimes C_\infty - 1_{V_1 \otimes V_2}) - \dim \ker(A_\infty \otimes C_0 - 1_{V_1 \otimes V_2}). \end{aligned}$$

□

In the following we assume that  $\mathbf{u} \cdot \mathbf{v}$  is generic. Using a suitable deformation argument involving a marking (cf. Schoenflies’ theorem) we can assume that the elements in  $\mathbf{u} \cdot \mathbf{v}$  are real-valued and that

$$0 < x_1 y_1 < \dots < x_p y_1 < \dots < x_1 y_q < \dots < x_p y_q < s_0.$$

Choose a homotopy base of  $\pi_1(S, s_0)$  as follows:



For  $\mathbf{u} = \{0, x_1, \dots, x_p\}$  and  $\mathbf{v} = \{0, y_1, \dots, y_q\}$  as above define  $x_0 := 0$  and  $x_{p+1} := s_0/y_q, \dots, x_{p+q} := s_0/y_1$ , so that  $U_0$  is identified with  $\mathbb{A}^1 \setminus \{x_0, \dots, x_{p+q}\}$  via the first projection. Consider the braid group

$$\mathcal{A}_{p+q+1} := \pi_1(\mathcal{O}_{p+q+1}, \{x_0, \dots, x_{p+q}\}) = \langle \beta_0, \dots, \beta_{p+q-1} \rangle,$$

where the  $\beta_i$ 's are similar as in Sect. 2.2, i.e., the braid  $\beta_i$  fixes all elements  $\{x_0, \dots, x_{p+q}\} \setminus \{x_i, x_{i+1}\}$  and interchanges  $x_i, x_{i+1}$  using a counterclockwise rotation.

**Proposition 3.2.2** *Let  $\phi : \pi_1(S, s_0) \rightarrow \mathcal{A}_{p+q+1}$  be as in Sect. 2.2. Then the following holds:*

(i)

$$\phi(\delta_{(q-1)p+i}) = \beta_{i,p+1}, \quad i = 1, \dots, p,$$

and

$$\phi(\delta_{(j-1)p+i}) = \beta_{i,p+1}^{\beta_{p+1} \cdots \beta_{p+q-j}}, \quad i = 1, \dots, p, j = 1, \dots, q - 1,$$

where  $\beta_{k,l} := (\beta_k^2)^{\beta_{k+1}^{-1} \cdots \beta_{l-1}^{-1}}$ , for  $1 \leq k < l \leq p + q + 1$ .

(ii) For  $\delta_0$  as above

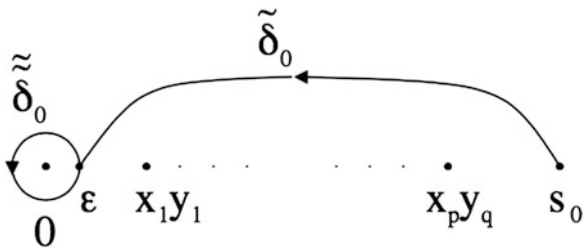
$$\phi(\delta_0) = (\beta_0^2 \cdot (\beta_1 \beta_0^2 \beta_1)) \cdot (\beta_2 \beta_1 \beta_0^2 \beta_1 \beta_2) \cdots (\beta_{q-1} \cdots \beta_1 \beta_0^2 \beta_1 \cdots \beta_{q-1})^{\beta^{-1}},$$

where

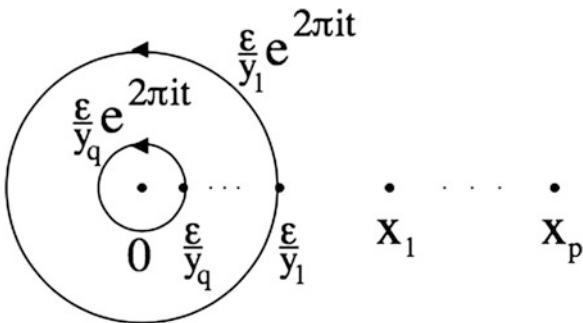
$$\beta := (\beta_p \cdots \beta_1)(\beta_{p+1} \cdots \beta_2) \cdots (\beta_{p+q} \cdots \beta_{q+1}).$$

*Proof* As explained in [3, Section 1], the expression of the various  $\phi(\delta_i)$ 's depends on the intersection-data of  $\mathbf{w}$ , cf. [2]. Here, crossing an exceptional value which involves just one simple crossing amounts to conjugation by the inverse of a  $\beta_k$  (if the  $k$ th and the  $k + 1$ th line meet when ordered locally by their real values). Turning around an intersection point which involves just one simple crossing amounts to a  $\beta_k^2$ . This leads to the expression of  $\delta_i$  for  $i \neq 0$  in a straightforward manner. (Let us indicate a direct method to obtain the expression of  $\phi(\delta_{(j-1)p+i})$  in terms of the  $\beta_i$ 's. It is immediate from the structure of  $\mathbf{w}$  (cf. to the image in Sect. 3) that  $\phi(\delta_{(j-1)p+i})$  is of the following form: The points  $x_0, \dots, x_p$  are fixed whereas  $x_{p+1}, \dots, x_{p+q}$  move in the upper half plane to the real axis with  $x_j$  moving counterclockwise around  $x_i$  and the other points of the set  $\{x_{p+1}, \dots, x_{p+q}\}$  moving around a closed disc, i.e., none of the points  $x_0, \dots, x_p$ . This braid is then homotopic to a braid which fixes all points  $\{x_0, \dots, x_{p+q}\} \setminus x_j$  with  $x_j$  crossing the points  $x_{p+1}, \dots, x_{j-1}$  in the lower half plane, then crossing the real axis, then crossing the points  $x_{i+1}, \dots, x_p$  in the upper half plane before encircling  $x_i$  and moving the same way back. The expression of the braid is then visibly  $\beta_{i,p+1}^{\beta_{p+1} \cdots \beta_{p+q-j}}$ , cf. [16, Section III.1.2].)

For  $\delta_0$  we argue as follows: First we write  $\delta_0$  as a product  $\tilde{\delta}_0 \tilde{\delta}_0^{-1}$  as follows:



Then, viewing  $\phi$  as a homomorphism of fundamental groupoids, the braid  $\phi(\tilde{\delta}_0)$  is as follows:



Since the paths

$$\frac{\epsilon}{y_q} e^{2\pi i t}, \quad \frac{\epsilon}{y_{q-1}} e^{2\pi i t}, \quad \dots, \quad \frac{\epsilon}{y_1} e^{2\pi i t}$$

occurring in the above expression of  $\phi(\tilde{\delta}_0)$  are pairwise disjoint, they commute. They contribute a factor of  $\phi(\tilde{\delta}_0)$  as follows (in the same ordering):

$$\beta_0^2, \quad \beta_1 \beta_0^2 \beta_1, \quad \dots, \quad \beta_{q-1} \cdots \beta_1 \beta_0^2 \beta_1 \cdots \beta_{q-1},$$

cf. [16, Figure 1.4 in Section III.1.2]. Here we have identified the initial base point  $P_1 := \{x_0, \dots, x_{p+q}\}$  of  $\mathcal{O}_{p+q+1}$  with

$$P_0 := \{0, \frac{\epsilon}{y_q}, \dots, \frac{\epsilon}{y_1}, x_1, \dots, x_p\}$$

by inserting a null-homotopic path  $P_t^{-1} \cdot P_t$  in between  $\phi(\tilde{\delta}_0)$  and  $\phi(\tilde{\delta}_0)$  with

$$P_t := \{0, (1-t)\frac{\epsilon}{y_q} + tx_1, \dots, (1-t)x_p + tx_{p+q}\}.$$

Finally, by a similar argument as above, involving the intersection data of  $\mathbf{w}$ , one sees that the formal conjugation by  $\phi(\delta_0)^{-1}$  amounts to conjugation with  $\beta^{-1}$ .  $\square$

**Proposition 3.2.3** *Let  $\chi : \pi_1(S, s_0) \rightarrow \mathrm{GL}(V_1 \otimes V_2)$  denote the homomorphism from Eq. (11) in the situation of the multiplicative convolution. Then  $\chi(\delta_0) = 1_{V_1} \otimes B_0$  and  $\chi(\delta_k) = 1$  for  $k = 1, \dots, pq$ .*

*Proof* The local system  $\mathrm{pr}_1^* \mathcal{V}_1$  clearly has no monodromy in  $\mathrm{pr}_2$ -direction as well as the local system  $q^* \mathcal{V}_2$ , away from 0. By construction of  $q^* \mathcal{V}_2$  one has  $\chi(\delta_0) = 1_{V_1} \otimes B_0$ .  $\square$

Summarizing, we obtain:

**Theorem 3.2.4** *Let  $\mathcal{V}_1, \mathcal{V}_2$  and*

$$g := (A_0 \otimes B_\infty, A_1 \otimes 1_{V_2}, \dots, A_p \otimes 1_{V_2}, 1_{V_1} \otimes B_q, \dots, 1_{V_1} \otimes B_1, A_\infty \otimes B_0) \in \mathrm{GL}(V_1 \otimes V_2)$$

*be as above. Let*

$$\rho_{\mathcal{V}_1 \star \mathcal{V}_2} : \pi_1(S, s_0) = \langle \delta_0, \dots, \delta_{pq} \rangle \rightarrow \mathrm{GL}(H_g/E_g)$$

*be the monodromy representation of the multiplicative middle convolution  $\mathcal{V}_1 \star \mathcal{V}_2$ . Then*

$$\rho_{\mathcal{V}_1 \star \mathcal{V}_2}(\delta_i) = \bar{\Phi}(g, \phi(\delta_i)) \cdot \bar{\Psi}(g, \chi(\delta_i)) \quad i = 0, \dots, pq,$$

*where  $\bar{\Phi}, \bar{\Psi}$  are as in Sect. 2.5 and where  $\phi(\delta_i), \chi(\delta_i)$  are as in Propositions 3.2.2 and 3.2.3.*

*Remark 3.2.5* In the non-generic case, we can assume (again using a suitable deformation argument involving Schoenflies' theorem) that consecutive real-valued exceptional values  $s_k < s_{k+1} < \dots < s_{k+d} \in \tilde{\mathbf{u}} \cdot \tilde{\mathbf{v}}$  of a small deformation  $\tilde{\mathcal{V}}_1 \star \tilde{\mathcal{V}}_2$  of  $\mathcal{V}_1 \star \mathcal{V}_2$  collapse to a single exceptional value  $s_k \in \mathbf{u} \cdot \mathbf{v}$ . Then

$$\rho_{\tilde{\mathcal{V}}_1 \star \tilde{\mathcal{V}}_2}(\delta_k) \cdots \rho_{\tilde{\mathcal{V}}_1 \star \tilde{\mathcal{V}}_2}(\delta_{k+d}) = \rho_{\mathcal{V}_1 \star \mathcal{V}_2}(\delta_k)$$

## 4 Additive Convolution

### 4.1 The Definition of the Additive Middle Convolution

For  $\mathbf{u} := \{x_1, \dots, x_p\} \in \mathcal{O}_p$  and  $\mathbf{v} := \{y_1, \dots, y_q\} \in \mathcal{O}_q$  set

$$\mathbf{u} * \mathbf{v} := \{x_i + y_j \mid i = 1, \dots, p, j = 1, \dots, q\}.$$

Let  $U_1 := \mathbb{A}^1 \setminus \mathbf{u}$ ,  $U_2 := \mathbb{A}^1 \setminus \mathbf{v}$  and  $S := \mathbb{A}^1 \setminus \mathbf{u} * \mathbf{v}$ . Set

$$\tilde{f}(x, y) := \prod_{i=1}^p (x - x_i) \prod_{j=1}^q (y - x - y_j) \prod_{i,j} (y - (x_i + y_j))$$

and let  $f \in \mathbb{C}[x, y]$  be the associated reduced polynomial. One has  $\tilde{f} = f$  if and only if  $|\mathbf{u} * \mathbf{v}| = i \cdot j$ , in which case we call  $\mathbf{u} * \mathbf{v}$  *generic*. Let

$$\tilde{\mathbf{w}} := \{(x, y) \in \mathbb{A}^2 \mid f(x, y) = 0\}$$

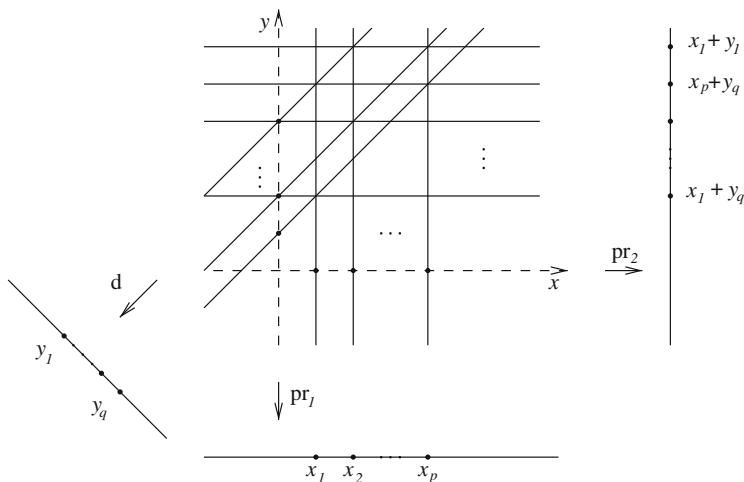
and let  $U := \mathbb{A}^2 \setminus \tilde{\mathbf{w}}$ . The set  $U$  is equipped with three maps:

$$\text{pr}_1 : U \longrightarrow U_1, \quad (x, y) \longmapsto x,$$

$$\text{pr}_2 : U \longrightarrow S, \quad (x, y) \longmapsto y,$$

and the *subtraction map*

$$d : U \longrightarrow U_2, \quad (x, y) \longmapsto y - x.$$



Let

$$j : U \longrightarrow X := \mathbb{P}_S^1, \quad (x, y) \longmapsto ([x, 1], y)$$

and let  $\mathbf{w} := X \setminus U$ . Since  $\mathbf{w}$  is a smooth relative divisor over  $S$ , we are in the situation of Sect. 2.5 with  $r \leq p + q$  and  $\pi = \text{pr}_2$ . The second projection  $\mathbb{P}_S^1 \rightarrow S$  is denoted by  $\overline{\text{pr}}_2$ . The fibre  $\text{pr}_2^{-1}(y_0)$  is denoted by  $U_0$ . The first projection yields an

identification of  $U_0$  with  $\mathbb{A}^1 \setminus (\mathbf{u} \cup (y_0 - \mathbf{v}))$ , where

$$\mathbf{u} \cup (y_0 - \mathbf{v}) := \mathbf{u} \cup \{y_0 - y_1, \dots, y_0 - y_q\} \in \mathcal{O}_{p+q}.$$

Let  $\mathcal{V}_i \in \text{LS}_R(U_i)$  ( $i = 1, 2$ ) be irreducible and nonconstant. We further assume that  $V_1$  has nontrivial local monodromy at at least two different points  $x_{i_1}, x_{i_2} \neq \infty$ . The local system  $\mathcal{V}_1 \boxtimes \mathcal{V}_2 := \text{pr}_1^* \mathcal{V}_1 \otimes \text{d}^* \mathcal{V}_2$  is a local system on  $U$  which is a variation of  $\mathcal{V}_1 \boxtimes \mathcal{V}_2|_{U_0}$  over  $S$ . The *additive middle convolution* of  $\mathcal{V}_1 \in \text{LS}_R(U_1)$  and  $\mathcal{V}_2 \in \text{LS}_R(U_2)$  is the local system

$$\mathcal{V}_1 * \mathcal{V}_2 := R^1(\overline{\text{pr}}_2)_*(j_*(\mathcal{V}_1 \boxtimes \mathcal{V}_2)) \in \text{LS}_R(S).$$

*Remark 4.1.1*

- (i) In [15], Katz gives a similar construction in a more general category of complexes of sheaves, which is (under [15, Prop. 2.8.4], and our assumptions on  $\mathcal{V}_1$  and  $\mathcal{V}_2$ ) equivalent to our construction. Explicit matrices for the monodromy in this case are given in [4].
- (ii) An important case of the middle convolution is Katz’ middle convolution functor  $\text{MC}_\chi$ , see [15]: Let  $\chi$  be a character of  $\pi_1(\mathbb{G}_m)$ ,  $\mathbb{G}_m = \mathbb{A}^1 \setminus \{0\}$ , and let  $\mathcal{V}_\chi \in \text{LS}_R(\mathbb{G}_m)$  be the associated local system. We call  $\mathcal{V}_\chi$  the *Kummer sheaf* associated to  $\chi$ . Then one obtains a functor

$$\text{LS}_R(U_1) \longrightarrow \text{LS}_R(U_1), \mathcal{V} \longmapsto \text{MC}_\chi(\mathcal{V}) := \mathcal{V} * \mathcal{V}_\chi.$$

### 4.2 Monodromy of the Additive Middle Convolution

Let in this section  $R = K$  be a field. Throughout this section we assume that  $\mathcal{V}_i \in \text{LS}_R(U_i)$  ( $i = 1, 2$ ) is irreducible and nonconstant, where  $U_1 = \mathbb{A}^1 \setminus \mathbf{u}$  and  $U_2 = \mathbb{A}^1 \setminus \mathbf{v}$  such that  $\mathbf{u} * \mathbf{v}$  is generic. We further assume that  $V_1$  has nontrivial local monodromy at at least two different points  $x_{i_1}, x_{i_2} \neq \infty$ . Let us fix a basepoint  $(x_0, y_0)$  in  $U$ . This induces basepoints  $x_0 = \text{pr}_1(x_0, y_0)$ ,  $y_0 - x_0 = \text{d}(x_0, y_0)$ ,  $y_0 = \text{pr}_2(x_0, y_0)$  of  $U_1$ ,  $U_2$  and  $S = \mathbb{A}^1 \setminus \mathbf{u} * \mathbf{v}$ . Let  $V_1$  denote the stalk of  $\mathcal{V}_1$  at  $x_0$  and let  $V_2$  denote the stalk of  $\mathcal{V}_2$  at  $y_0 - x_0$ . Let also

$$U_0 = \text{pr}_2^{-1}(y_0) = \mathbb{A}^1 \setminus \mathbf{u} \cup (y_0 - \mathbf{v})$$

be as in the last section.

The representation  $\rho_{\mathcal{V}_1 \boxtimes \mathcal{V}_2|_{U_0}} : \pi_1(U_0, (x_0, y_0)) \rightarrow \text{GL}(V_1 \otimes V_2)$  factors as

$$\rho_{\mathcal{V}_1 \boxtimes \mathcal{V}_2|_{U_0}} = (\rho_{\mathcal{V}_1} \otimes \rho_{\mathcal{V}_2}) \circ (\text{pr}_1 \times \text{d})_* , \tag{16}$$

where

$$(\text{pr}_1 \times \text{d})_* : \pi_1(U_0, (x_0, y_0)) \longrightarrow \pi_1(U_1, x_0) \times \pi_1(U_2, y_0 - x_0)$$

is the map which is induced by  $\text{pr}_1|_{U_0} \times \text{d}|_{U_0}$ . Let  $\alpha_1, \dots, \alpha_{p+q+1}$  be generators of  $\pi_1(U_0, (x_0, y_0))$  which are chosen as in the figure below. Let

$$(\gamma_1 := \text{pr}_{1*}(\alpha_1), \dots, \gamma_p := \text{pr}_{1*}(\alpha_p))$$

be the induced generators of  $\pi_1(U_1, x_0)$  and let

$$(\eta_1 := \text{d}_*(\alpha_{p+1}), \dots, \eta_q := \text{d}_*(\alpha_{p+q}))$$

be those of  $\pi_1(U_2, y_0 - x_0)$ . With respect to these generators, let

$$T_{\mathcal{V}_1} = (A_1, \dots, A_{p+1}) \in \text{GL}(V_1)^{p+1} \quad \text{and} \quad T_{\mathcal{V}_2} = (B_1, \dots, B_{q+1}) \in \text{GL}(V_2)^{q+1}$$

be the associated tuples. It follows from our choice of homotopy generators and (16) that

$$\begin{aligned} T_{\mathcal{V}_1 \boxtimes \mathcal{V}_2|_{U_0}} &= (C_1 := A_1 \otimes 1_{V_2}, \dots, C_p := A_p \otimes 1_{V_2}, \\ &C_{p+1} := 1_{V_1} \otimes B_1, \dots, C_{p+q} := 1_{V_1} \otimes B_q, A_{p+1} \otimes B_{q+1}). \end{aligned} \quad (17)$$

**Proposition 4.2.1** *Let  $\dim_K V_i = n_i$ . Then*

$$\begin{aligned} \text{rk}(\mathcal{V}_1 * \mathcal{V}_2) &= (p + q - 1)n_1n_2 - \sum_{i=1}^p n_2 \dim_K \ker(A_i - 1_{V_1}) \\ &- \sum_{j=1}^q n_1 \dim_K \ker(B_j - 1_{V_2}) - \dim_K \ker(A_{p+1} \otimes B_{q+1} - 1_{V_1 \otimes V_2}). \end{aligned} \quad (18)$$

*Proof* It follows from (17) and the properties of the tensor product that

$$\dim_K \ker(C_i - 1_{V_1 \otimes V_2}) = n_2 \dim_K \ker(A_i - 1_{V_1}), \quad i = 1, \dots, p$$

and

$$\dim_K \ker(C_i - 1_{V_1 \otimes V_2}) = n_1 \dim_K \ker(B_i - 1_{V_2}), \quad i = p + 1, \dots, p + q.$$

The claim follows now from Proposition 2.5.1 (ii). □

*Remark 4.2.2* The dimension  $\dim_K \ker(A_{p+1} \otimes B_{q+1} - 1_{V_1 \otimes V_2})$  can be easily computed using Lemma 2.1.1.



We want to describe the monodromy of  $\mathcal{V}_1 * \mathcal{V}_2$ . We can assume (using a suitable marking as in [10]) that we are in the following situation: The sets  $\mathbf{u} = \{x_1, \dots, x_p\}$ ,  $\mathbf{v} = \{y_1, \dots, y_q\}$ ,  $\{y_0\}$  are elementwise real and one has

$$x_1 < x_2 < \dots < x_p < y_0 - y_1 < y_0 - y_2 < \dots < y_0 - y_q.$$

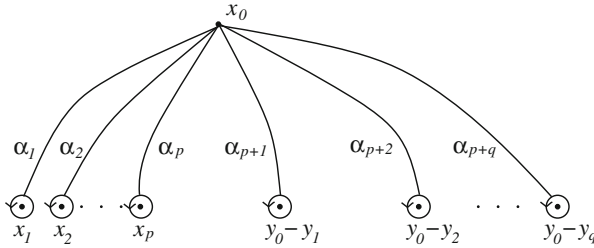
Moreover, we can assume that

$$|x_p - x_1| < |y_{i+1} - y_i| \quad \text{for } i = 1, \dots, q - 1. \tag{19}$$

Let us fix a basepoint  $(x_0, y_0)$  of  $U_0$  and of  $U$ . We assume that the imaginary part of  $x_0$  to is large enough, i.e., larger than the maximal imaginary part of  $\delta_{i,j}(t)$ , where  $\delta_{i,j}$  is as shown below. One obtains basepoints

$$x_0 = \text{pr}_1(x_0, y_0), \quad y_0 - x_0 = \text{d}(y_0, x_0), \quad y_0 = \text{pr}_2(x_0, y_0)$$

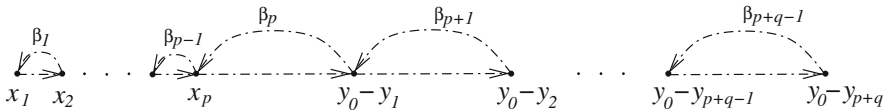
on  $U_1, U_2$  and  $S$  (respectively). We choose generators  $\alpha_1, \dots, \alpha_{p+q}$  of  $\pi_1(U_0, (x_0, y_0))$  as follows:



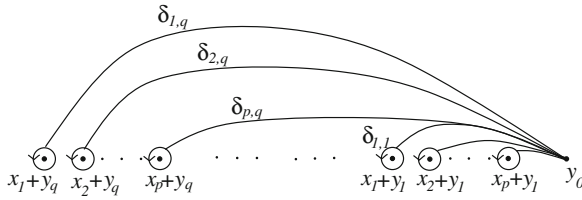
Next we choose generators  $\beta_1, \dots, \beta_{p+q-1}$  of

$$\mathcal{A}_{p+q} = \pi_1(\mathcal{O}_{p+q}, \mathbf{u} \cup \{y_0 - y_1, \dots, y_0 - y_q\})$$

as follows:



Then we choose generators  $\delta_{i,j}$ ,  $i = 1, \dots, p, j = 1, \dots, q$  of  $\pi_1(S, y_0)$  as follows:



**Proposition 4.2.3** *Let*

$$\phi : \pi_1(S, y_0) \rightarrow \mathcal{A}_{p+q} = \pi_1(\mathcal{O}_{p+q}, \mathbf{u} \cup (y_0 - \mathbf{v}))$$

be as in Sect. 2.2. Then

$$\phi(\delta_{i,1}) = \beta_{i,p+1}, \quad i = 1, \dots, p, \tag{20}$$

and

$$\phi(\delta_{i,j}) = \beta_{i,p+1}^{\beta_{p+1} \dots \beta_{p+j-1}}, \quad i = 1, \dots, p, j = 2, \dots, q, \tag{21}$$

where

$$\beta_{i,j} = (\beta_i^2)^{\beta_{i+1}^{-1} \dots \beta_{j-1}^{-1}}.$$

*Proof* Using (19) and the methods of [3] involving the intersection behaviour of  $\mathbf{w}$ , it is easy to see that

$$\phi(\delta_{i,1}) = \beta_{i,p+1}$$

and

$$\phi(\delta_{i,j}) = (\beta_{i+j-1,p+j})^{(\beta_{j-1}^{-1} \dots \beta_{p+j-2}^{-1}) \dots (\beta_2^{-1} \dots \beta_{p+1}^{-1}) (\beta_1^{-1} \dots \beta_p^{-1})}, \quad j = 2, \dots, q.$$

Using a suitable homotopy argument in  $\mathcal{O}_{p+q}$  (deform the paths with initial points  $y_0 - y_1, \dots, y_0 - y_{p+j-1}$  to paths with constant real part and large enough imaginary part), one can see that for  $j \geq 2$  these braids coincide with  $\beta_{i,p+1}^{\beta_{p+1} \dots \beta_{p+j-1}}$ .  $\square$

Using the choice of our setup, one obtains a diagram

$$\begin{array}{ccccccc}
 1 & \longrightarrow & \pi_1(U_0, (x_0, y_0)) & \longrightarrow & \pi_1(U, (x_0, y_0)) & \longrightarrow & \pi_1(S, y_0) \longrightarrow 1 \\
 & & \downarrow & & \downarrow & & \phi \downarrow \\
 1 & \longrightarrow & \pi_1(U_0, (x_0, y_0)) & \longrightarrow & \mathcal{A}_{p+q,1} & \longrightarrow & \mathcal{A}_{p+q} \longrightarrow 1, \\
 & & & & & & (22)
 \end{array}$$

such that the rows are split exact sequences and such that the vertical arrows are compatible with the splittings of the rows (see (6)).

**Proposition 4.2.4** *The monodromy of  $\mathcal{V}_1 * \mathcal{V}_2$  is given by*

$$\rho_{\mathcal{V}_1 * \mathcal{V}_2}(\gamma) = \bar{\Phi}(T_{\mathcal{V}_1 \boxtimes \mathcal{V}_2|_{U_0}}, \phi(\gamma)) \quad \forall \gamma \in \pi_1(S, y_0),$$

where  $T_{\mathcal{V}_1 \boxtimes \mathcal{V}_2|_{U_0}} = (C_1, \dots, C_{p+q+1})$  is as in (17) and  $\bar{\Phi}$  is as in Sect. 2.5.

*Proof* By the properties of the tensor product, the commutators

$$[C_i = A_i \otimes 1_{\mathcal{V}_2}, C_{p+j} = 1_{\mathcal{V}_1} \otimes B_j], \quad i = 1, \dots, p, j = 1, \dots, q,$$

vanish. It follows from the construction of  $\mathcal{V}_1 \boxtimes \mathcal{V}_2 = \text{pr}_1^* \otimes d^* \mathcal{V}_2$  that

$$\rho_{\mathcal{V}_1 \boxtimes \mathcal{V}_2}|_{\pi_1(S, y_0)} = \chi = 1.$$

By the above discussion, Eq. (6) can be assumed to hold for (22). Thus, Proposition 2.5.1 gives the claim. □

*Remark 4.2.5* In the non-generic case, a deformation argument shows that one obtains the associated tuple  $T_{\mathcal{V}_1 * \mathcal{V}_2}$  from the generic case, by multiplying the monodromy generators of the generic case suitably: We can assume (again using a suitable deformation argument involving Schoenflies’ theorem) that consecutive real-valued exceptional values  $s_k < s_{k+1} < \dots < s_{k+d} \in \tilde{\mathbf{u}} * \tilde{\mathbf{v}}$  of a small deformation  $\tilde{\mathcal{V}}_1 * \tilde{\mathcal{V}}_2$  of  $\mathcal{V}_1 * \mathcal{V}_2$  collapse to a single exceptional value  $s_k \in \mathbf{u} * \mathbf{v}$ . Then

$$\rho_{\tilde{\mathcal{V}}_1 * \tilde{\mathcal{V}}_2}(\delta_k) \cdots \rho_{\tilde{\mathcal{V}}_1 * \tilde{\mathcal{V}}_2}(\delta_{k+d}) = \rho_{\mathcal{V}_1 * \mathcal{V}_2}(\delta_k)$$

**Acknowledgements** The authors thank the referee for valuable comments. Michael Dettweiler gratefully acknowledges financial support within the DFG-Schwerpunkt SPP 1489.

## References

1. J.S. Birman, *Braids, Links and Mapping Class Groups*. Annals of Mathematics Studies, vol. 82 (Princeton University Press, Princeton, 1974)
2. D. Cohen, A. Suciuc, The braid monodromy of plane algebraic curves and hyperplane arrangements. *Comment. Math. Helvetici* **72**, 285–315 (1997)
3. M. Dettweiler, Plane curve complements and curves on Hurwitz spaces. *J. Reine Angew. Math.* **573**, 19–43 (2004)
4. M. Dettweiler, S. Reiter, Middle convolution of Fuchsian systems and the construction of rigid differential systems. *J. Algebra* **318**, 1–24 (2007)
5. M. Dettweiler, S. Reiter, Rigid local systems and motives of type  $G_2$ . *Compos. Math.* **146**, 929–963 (2010)
6. M. Dettweiler, S. Reiter, The classification of orthogonally rigid  $G_2$ -local systems and related differential operators. *Trans. Am. Math. Soc.* **366**, 5821–5851 (2014)
7. M. Dettweiler, S. Reiter, On the Hodge theory of the additive and the multiplicative convolution (in preparation, 2017)
8. M. Dettweiler, C. Sabbah, Hodge theory of the middle convolution. *Publ. Math. RIMS* **49**, 761–800 (2013)
9. M. Dettweiler, J. Schmidt, Rigid  $G_2$ -representations and motives of type  $G_2$ . *Isr. J. Math.* **212**, 81–106 (2016)
10. M. Dettweiler, S. Wewers, Variation of local systems and parabolic cohomology. *Isr. J. Math.* **156**, 157–185 (2006)
11. M. Dettweiler, S. Wewers, Variation of parabolic cohomology and Poincaré duality, in *Groupes de Galois Arithmétiques et Différentiels*, ed. by D. Bertrand, P. Debes, *Seminaires et Congres*, vol. 13 (Société Mathématique de France, Paris, 2006), pp. 145–164
12. E. Fuchs, C. Meiri, P. Sarnak, Hyperbolic monodromy groups for the hypergeometric equation and Cartan involutions. *J. Eur. Math. Soc.* **16**, 1617–1671 (2014)
13. K. Jakob, The classification of rigid irregular  $G_2$ -connections (2016). arXiv:1609.03292
14. M. Jöllenbeck, VarParCohom.m. Available at <http://zahlentheorie.uni-bayreuth.de>
15. N.M. Katz, *Rigid Local Systems*. Annals of Mathematics Studies, vol. 139 (Princeton University Press, Princeton, 1996)
16. G. Malle, B.H. Matzat, *Inverse Galois Theory*. Monographs in Mathematics (Springer, Berlin, 1999)
17. A.L. Onishchik, E.B. Vinberg, *Lie Groups and Algebraic Groups* (Springer, Berlin, 1990)
18. L. Schwartz, *Mathematische Methoden der Physik* (Bibliographisches Institut, Mannheim, 1974)

# Constructing Groups of ‘Small’ Order: Recent Results and Open Problems



Bettina Eick, Max Horn, and Alexander Hulpke

**Abstract** We investigate the state of the art in the computational determination and enumeration of the groups of small order. This includes a survey of the available algorithms and a discussion of their recent improvements. We then show how these algorithms can be used to determine or enumerate the groups of order at most 20,000 with few exceptions and we discuss the orders in this range which remain as challenging open problems.

**Keywords** Enumeration • Determination • Small groups • Algorithms

**Subject Classifications** 20D45, 20E22, 20-04

## 1 Introduction

The determination of the groups of a given order  $n$  up to isomorphism is one of the central problems in finite group theory. The aim is to determine a list  $\mathcal{L}_n$  of groups of order  $n$  so that every group of order  $n$  is isomorphic to exactly one group in the list  $\mathcal{L}_n$ . A slightly weaker but also interesting goal is to enumerate the isomorphism types of groups of a given order. The aim is to determine the cardinality  $|\mathcal{L}_n|$ , possibly without explicitly listing all groups in  $\mathcal{L}_n$ . There are asymptotic estimates

---

B. Eick (✉)

TU Braunschweig, Pockelsstraße 14, 38106 Braunschweig, Germany  
e-mail: [beick@tu-bs.de](mailto:beick@tu-bs.de); [b.eick@tu-braunschweig.de](mailto:b.eick@tu-braunschweig.de)

M. Horn

Justus-Liebig-Universität Gießen, Arndtstraße 2, 35392 Gießen, Germany  
e-mail: [max.horn@math.uni-giessen.de](mailto:max.horn@math.uni-giessen.de)

A. Hulpke

Colorado State University, Fort Collins, CO 80523-1874, USA  
e-mail: [hulpke@colostate.edu](mailto:hulpke@colostate.edu)

known for  $|\mathcal{L}_n|$ , see Pyber [23] for a survey, but no closed formula for  $|\mathcal{L}_n|$  is known. See also [11] for a discussion of properties of  $|\mathcal{L}_n|$  as a function in  $n$ .

The history of this group construction or enumeration problem goes back to the beginnings of abstract group theory: Cayley [10] introduced the abstract definition of groups and determined the groups of order at most 6. Many other group constructions and enumerations followed the work of Cayley. We refer to Besche et al. [6] for a history of group constructions and to Blackburn et al. [8] for details on enumerations of groups.

Senior and Lunn [25, 26] determined all groups of order at most 200 except 128 and 192. It is quite remarkable that they did this by hand and got it right! A natural question is: why did they omit 128 and 192? Nowadays it is known that these two orders yield by far the most groups in the range of orders at most 200: using the SmallGroups library [7] one observes that there are 2328 groups of order 128 and 1543 of order 192, while the maximum is 267 for every other order at most 200. There are 6065 groups of order at most 200 in total.

Why are there many groups for some orders and very few for others? For example, the SmallGroups library [7] asserts that there are 49,487,365,422 groups of order 1024 and only 4 groups of order 1025. The asymptotic results on  $|\mathcal{L}_n|$  as reported in [23] as well as the known values for  $|\mathcal{L}_n|$  in the SmallGroups library suggest that the largest multiplicity of a prime dividing an order  $n$  plays a major role; that is, if  $n = p_1^{e_1} \cdots p_r^{e_r}$  for different primes  $p_1, \dots, p_r$ , then  $e = \max\{e_1, \dots, e_r\}$  has a major impact on the number of groups of order  $n$ . Note that  $1024 = 2^{10}$ , while  $1025 = 5^2 \cdot 41$ .

Besche et al. [5, 6] determined the groups of order at most 2000 except 1024 and Eick and O'Brien [13] enumerated the groups of order 1024. The results are available in the SmallGroups library [7]. This group determination and enumeration was obtained with the massive help of computers and methods from computational group theory. The use of computers is essential due to the large numbers of groups. For example, there are eight orders in the range of orders at most 2000 with more than 100,000 groups.

Since then, computer technology and also the methods from computational group theory have improved significantly. For example, a new isomorphism test and automorphism group algorithm has been developed by Cannon and Holt [9] and a new method to construct finite solvable groups has been introduced by Eick and Horn [12]. Further, many of the methods in the computer algebra system GAP [29] have been improved; in particular, the machinery to construct subdirect products has been significantly updated by the third author. The combination of these advances permits us to extend significantly the range of orders  $n$  for which  $\mathcal{L}_n$  or at least  $|\mathcal{L}_n|$  is computable.

It is the aim of this paper to report on the available group construction and enumeration methods in GAP [29] and its packages and their application to the determination or enumeration of groups of order at most 20,000. There are currently 39 orders in the range at most 20,000 for which the number of groups of these orders are unknown. We list these orders and discuss the difficulty that they impose

on the group construction and enumeration methods. Thus we highlight the most challenging problems in the enumeration of finite groups of small orders.

The known numbers of groups  $|\mathcal{L}_n|$  for  $1 \leq n \leq 20,000$  can be obtained at a web-page prepared by the second author:

<http://groups.quendi.de>.

Among these orders with known numbers, there are 56 orders with more than one million groups. For example, each of the orders of the form  $2^9 \cdot p$  with  $p$  a prime yields more than 400 million groups. And the order  $2^{10}$  yields more than one billion groups.

## 2 Algorithms to Construct Finite Groups

In this section we give a brief overview of the available methods to construct finite groups. These methods fall into three different categories: methods to construct nilpotent groups, methods to construct solvable (non-nilpotent) groups and methods to construct non-solvable groups.

### 2.1 Construction of Nilpotent Groups

A finite nilpotent group is a direct product of its Sylow subgroups. Hence the construction of nilpotent groups directly translates to the construction of  $p$ -groups. For this purpose there is a well-established method available: the  $p$ -group generation method of O’Brien [21]. The basic approach of this method is to use induction along the lower exponent- $p$  central series. An implementation of this method is available in the GAP package [20].

### 2.2 Construction of Solvable (Non-nilpotent) Groups

#### 2.2.1 The Frattini Extension Method

The Frattini extension method by Besche and Eick [2] is a widely used method for the construction of finite solvable groups. Recall that the Frattini subgroup  $\Phi(G)$  of a finite group  $G$  is the intersection of all maximal subgroups of  $G$ . The basic approach of this method is to determine up to isomorphism a list of candidates for the Frattini factors of the groups of order  $n$  and then, for each candidate  $F$ , determine up to isomorphism all groups  $G$  of order  $n$  with  $G/\Phi(G) \cong F$ . The first step of this approach is usually comparatively fast and yields a comparatively short list of groups. It relies heavily on an effective determination of subdirect products. The second step is often more involved and requires the reduction of a given list of

groups to isomorphism type representatives. A first reduction can often be achieved by using the highly effective random isomorphism test of [2]. A final reduction is then obtained by general isomorphism testing methods such as those described in [9, 27]. The Frattini extension method allows us readily to restrict to the construction of non-nilpotent groups or those with certain normal or non-normal Sylow subgroups. An implementation of this method is available in the GAP package [3].

### 2.2.2 Solvable Group Construction

Eick and Horn [12] present an alternative method to construct the solvable groups of a given order. This is a direct generalisation of the  $p$ -group generation method. It is often slower than the Frattini extension method, but in some cases it was able to determine the groups of a given order where the Frattini extension method failed. Further, it is very useful in verifying the results of the Frattini extension method. Also this method restricts readily to construct non-nilpotent groups only. An implementation of this method in GAP exists and will be made available as a GAP package [17].

## 2.3 Construction of Non-solvable Groups

### 2.3.1 The Cyclic Extensions Method

Besche and Eick [2] outlined a rather crude approach towards constructing non-solvable groups. It starts from the library of perfect groups [16]; we refer to the work by Holt and Plesken [16] for this. It then iteratively constructs cyclic extensions of groups. The extensions obtained then must be reduced to isomorphism types. This requires an effective isomorphism test for non-solvable groups. Nowadays we use the method by Cannon and Holt [9] for this purpose whose implementation in GAP will be made available as part of GAP 4.9.

### 2.3.2 Archer's Methods: Supplements and $Z\phi$

Archer [1] described two effective methods to construct the non-solvable groups of a given order  $n$ . Both approaches require that the perfect groups of all orders  $m$  dividing  $n$  are determined up to isomorphism; see [16].

The supplement method additionally requires that the solvable groups of order  $n/m$  are classified. For a given perfect group  $H$  of order  $m$  and a given solvable group  $G$  of order  $n/m$ , it determines up to isomorphism all groups  $E$  having a normal subgroup  $M \trianglelefteq E$  with  $M \cong H$  and  $E/M \cong G$ .

The  $Z\phi$  method additionally requires that all solvable groups of the orders  $kn/m$  are known, where  $k$  ranges over the sizes of the centers of the perfect groups of



order  $m$ . For a given perfect group  $H$  of order  $m$  with center  $Z = Z(H)$  of order  $k$ , it considers all solvable groups  $G$  of order  $kn/m$  extending  $Z$  and it determines up to isomorphism all groups  $E$  having a normal subgroup  $M \trianglelefteq E$  with  $M \cong H$  and  $E/M \cong G/Z$ . Note that the  $Z\phi$  method does not apply in all cases on  $H$ ; we refer to [1, p. 75] for details. Note also that [1, Section 5.1] contains a limited version of the  $Z\phi$  method. This applies only if  $|Z(H)| \leq 2$  and, if  $|Z(H)| = 2$ , then if  $\gcd(|Out(H)|, |G|) \leq 2$ . While this is a significant restriction, it still applies to many of the cases that we need to consider.

Archer neither published the results of his enumeration, nor his implementations of the algorithms. We have implemented in GAP the limited version of the  $Z\phi$  method. This does not apply to all cases, but it allows readily to be combined with the cyclic extension method. We are using this combination as an alternative approach to construct non-solvable groups. If the limited  $Z\phi$  method applies, then it is usually more effective than the cyclic extension method. Our implementation of the limited  $Z\phi$  method will be made available as a package for GAP.

As part of the applications of our implementation, we recomputed and extended the table on p. 64 in [1]. We noted that the rows for  $|S| \in \{128, 256\}$  in this table were incorrect. The correct values, as well as the additional value for  $|S| = 384$ , are as follows:

$ S $	grps	$\mathcal{O}_K$	$\mathcal{O}_Z$	$\mathcal{O}_{Z,K}$
128	2328	16,996	8308	72,010
256	56,092	1,027,380	337,956	6,856,498
384	20,169	206,463	82,035	938,587

### 3 A Symbolic Enumeration Algorithm

Suppose that  $m \in \mathbb{N}$  is given and that the groups of order  $m$  are available; that is,  $\mathcal{L}_m$  is known. In this section we describe an effective algorithm to enumerate the groups of order  $m \cdot p$  for *all* primes  $p$  coprime to  $m$ . Our approach is based on a theorem by Taunt [28] and it extends the cyclic split extension method described in [2] and the ideas in [4].

For a group  $G$  of order  $m$  and  $l \mid m$  let  $\mathcal{O}_l$  denote a set of representatives of the  $\text{Aut}(G)$ -classes of normal subgroups  $K$  in  $G$  with  $G/K$  cyclic of order  $l$ . For  $K \in \mathcal{O}_l$  let  $\text{Aut}_K(G)$  denote the stabilizer of  $K$  in  $\text{Aut}(G)$ , let  $\overline{\text{Aut}}_K(G)$  be the subgroup of  $\text{Aut}(G/K)$  induced by the natural action of  $\text{Aut}_K(G)$  on  $G/K$  and let  $\text{ind}_K := [\text{Aut}(G/K) : \overline{\text{Aut}}_K(G)]$ .

Let  $(d_1, \dots, d_k)$  be the list of all divisors of  $m$ , with  $d_1 = 1$ . We set

$$w(G) := (w_{d_1}(G), \dots, w_{d_k}(G)), \quad \text{where} \quad w_{d_i}(G) := \sum_{K \in \mathcal{O}_{d_i}} \text{ind}_K$$

and we denote

$$w(m) := (w_{d_1}(m), \dots, w_{d_k}(m)), \quad \text{where} \quad w_{d_i}(m) := \sum_{G \in \mathcal{L}_m} w_{d_i}(G).$$

**Theorem 3.1** *Let  $m \in \mathbb{N}$  and  $\pi$  the set of those prime divisors of  $(d_2 - 1) \cdots (d_k - 1)$  that do not divide  $m$ .*

- a) *Let  $p$  be a prime with  $p \nmid m$ . If there exists a group of order  $mp$  without normal Sylow  $p$ -subgroup, then  $p \in \pi$ .*
- b) *Let  $p$  be a prime with  $p \nmid m$ . The number of isomorphism types of groups of order  $m \cdot p$  having a normal Sylow  $p$ -subgroup is*

$$\sum_{\substack{i \in \{1, \dots, k\} \\ \text{with } d_i | (p-1)}} w_{d_i}(m).$$

*Proof*

- a) By Sylow’s theorems the number of Sylow  $p$ -subgroups in a group of order  $mp$  is congruent to 1 modulo  $p$  and it divides  $m$ . Thus if there exists a group of order  $mp$  without normal Sylow  $p$ -subgroup, then  $p \mid (d_i - 1)$  for some  $i \in \{2, \dots, k\}$ .
- b) Suppose that  $H$  is group of order  $mp$  with normal Sylow  $p$ -subgroup. By the Schur-Zassenhaus theorem,  $H \cong C_p \rtimes_{\varphi} G$  for a group  $G$  of order  $m$  and some homomorphism  $\varphi : G \rightarrow \text{Aut}(C_p) \cong C_{p-1}$ . Taunt [28] proved that two split extensions  $C_p \rtimes_{\varphi_1} G$  and  $C_p \rtimes_{\varphi_2} G$  are isomorphic if and only if there exist  $\alpha \in \text{Aut}(G)$  and  $\beta \in \text{Aut}(C_p)$  so that  $\varphi_1(\alpha(g)) = \beta^{-1} \varphi_2(g) \beta$  in  $\text{Aut}(C_p)$  for each  $g \in G$ . As  $\text{Aut}(C_p)$  is abelian, this reduces to  $\varphi_1(\alpha(g)) = \varphi_2(g)$  for all  $g \in G$  and thus is independent of  $\beta$ . Based on this, one can readily observe that the different isomorphism types of split extensions  $C_p \rtimes_{\varphi} G$  with  $K = \ker(\varphi)$  correspond one-to-one to the elements of a transversal of  $\overline{\text{Aut}}_K(G)$  in  $\text{Aut}(G/K)$  and this yields the desired result.

Theorem 3.1 translates to an effective method to enumerate the groups of order  $mp$  for fixed  $m$  and arbitrary prime  $p \nmid m$ :

- (1) Let  $D = (d_1, \dots, d_k)$  be the list of divisors of  $m$ , with  $d_1 = 1$ .
- (2) For all groups  $G$  in  $\mathcal{L}_m$  determine  $w(G)$  with respect to  $D$ .
- (3) Using the values in (2), determine  $w_{d_1}(m), \dots, w_{d_k}(m)$ .
- (4) Determine the (finite) set  $\pi$  of those prime divisors of  $(d_2 - 1) \cdots (d_k - 1)$  that do not divide  $m$ .
- (5) For each  $p \in \pi$  determine the number  $a_p$  of groups of order  $mp$  without normal Sylow  $p$ -subgroups (for example, using the Frattini extension method and the construction of non-solvable groups).
- (6) Define  $a_p = 0$  if  $p \notin \pi$ .

(7) Given an arbitrary prime  $p$  with  $p \nmid m$ , it now follows that

$$|\mathcal{L}_{mp}| = a_p + \sum_{\substack{i \in \{1, \dots, k\} \\ \text{with } d_i | (p-1)}} w_{d_i}(m).$$

Note that this method can be adapted readily to count solvable and non-solvable groups separately and we use this frequently in applications.

## 4 Recent Improvements to Implementations in GAP

Many of the above algorithms rely on effective methods to determine automorphism groups and to decide isomorphism. Here we exhibit various improvements to the existing methods for these purposes. We discuss automorphisms and isomorphisms in the following two subsections and we note that all exhibited improvements will be made public with GAP 4.9.

### 4.1 Automorphism Groups

There are various methods known to determine automorphism groups. For finite  $p$ -groups we use the method by Eick et al. [15] as implemented in the GAP package [14], for finite solvable groups we use the method by Smith [27], and for finite non-solvable groups we use the method by Cannon and Holt [9]. Smith’s method is implemented in the GAP library. This implementation has recently been improved by the third author and it has been combined with an implementation of the method by Cannon and Holt [9].

In the remainder of this subsection, we discuss the recent improvements to the GAP implementation of Smith’s method. Let  $G$  be a finite solvable group. Smith’s method uses induction along a characteristic series of  $G$  with elementary abelian factors. Let  $M$  be a characteristic elementary abelian subgroup of  $G$  of order  $p^d$ , say. Then there is a natural homomorphism

$$\varphi : \text{Aut}(G) \rightarrow \text{Aut}(G/M) \times \text{Aut}(M).$$

By induction, we assume that  $\text{Aut}(G/M)$  is given. Note that  $\text{Aut}(M) \cong \text{GL}(d, p)$ . The principal idea of Smith’s method is to determine  $\text{Aut}(G)$  via determining the kernel and image of  $\varphi$ . The kernel of  $\varphi$  is naturally isomorphic to  $Z^1(G/M, M)$  and can be determined readily. The image of  $\varphi$  can be described by certain stabilizer calculations; these stabilizer calculations are the main bottlenecks of the method.

One idea towards reducing the bottlenecks is the following. Instead of starting a stabilizer computation with the full direct product  $\text{Aut}(G/M) \times \text{Aut}(M)$ , we

determine a priori a subgroup  $D \leq \text{Aut}(G/M) \times \text{Aut}(M)$  with  $\text{im}(\varphi) \leq D$  and then use  $D$  instead of  $\text{Aut}(G/M) \times \text{Aut}(M)$ . For example, a subgroup  $D$  can be determined as the stabilizer of each group in a collection of characteristic subgroups of  $G$ . This often breaks a single stabilizer calculation into a sequence of smaller calculations and thus reduces the bottleneck of the overall method.

Using characteristic subgroups is particularly helpful to reduce  $\text{Aut}(M)$ . In this case the stabilizer of each group in a collection of characteristic subgroups of  $M$  in  $\text{Aut}(M)$  translates to the stabilizer of a collection of invariant subspaces of  $\mathbb{F}_p^d$  in  $\text{GL}(d, p)$ . This can be determined readily via the method described by Schwingel [24]. We implemented this in GAP and use it in combination with Smith's method.

We exhibit a second idea towards reducing bottlenecks. Let  $D \leq \text{Aut}(G/M) \times \text{Aut}(M)$  with  $\text{im}(\varphi) \leq D$ . Then  $D$  acts naturally on the set of homomorphism  $G/M \rightarrow M$ . Let  $\sigma$  denote the homomorphism arising from the conjugation action of  $G/M$  on  $M$ . One step in Smith's method is to determine the stabilizer in  $D$  of  $\sigma$ . We first determine a permutation representation of  $D$  related to the action on homomorphisms and then use the highly effective permutation group machinery of GAP to determine the desired stabilizer.

## 4.2 Isomorphisms

In this section we discuss the GAP implementation of the method of Cannon and Holt [9] to decide if two finite groups  $G$  and  $H$  are isomorphic. We first determine various invariants of  $G$  and  $H$  to have a fast initial check for non-isomorphism.

The method of Cannon and Holt uses induction along a fully invariant series through  $G$  and  $H$ . In each induction step it decides isomorphism and computes the automorphism group of the considered quotient.

Two groups  $G$  and  $H$  are isomorphic if and only if there exists  $\alpha \in \text{Aut}(G \times H)$  with  $G^\alpha = H$ . This translates an isomorphism test to an automorphism group calculation. Note that it is not necessary for this approach to determine the full automorphism group of  $G \times H$ : if  $G$  and  $H$  are isomorphic, then  $\text{Aut}(G \times H)$  contains a subgroup  $W \cong \text{Aut}(G) \wr C_2$  and  $G$  and  $H$  are conjugate in  $W$ .

Further, with this method it is frequently useful to determine a collection of fully invariant subgroups of  $G$  and  $H$  a priori and to use these subgroups to reduce the calculation, since a fully invariant subgroup of  $G$  (such as, for example,  $G'$ ,  $\text{Fit}(G)$  or  $Z(G)$ ) has to map onto the corresponding subgroup of  $H$  and thus our aim is to determine  $\alpha \in W$  that maps these pairs of subgroups onto each other.

## 5 The Groups of Order at most 20,000

In this section we describe how we enumerated or constructed the groups of order at most 20,000 with few exceptions.

**Nilpotent Groups** We have constructed these as direct products of  $p$ -groups. The groups of order dividing  $p^7$  have been determined by Newman et al. [18, 22]. The groups of order dividing  $2^9$  have been constructed by Eick and O'Brien [13, 19] who also enumerated the groups of order  $2^{10}$ . The groups of order  $3^8$  have been determined by Vaughan-Lee [30]. Hence the nilpotent groups of order  $n$  are available for all  $n \in \{1, \dots, 20,000\}$  except for those  $n$  divisible by  $2^{10}$  or  $3^9$ ; and the nilpotent groups of order divisible by  $2^{10}$ , but not divisible by  $2^{11}$ , can be enumerated.

**Solvable, Non-nilpotent Groups** We have used the Frattini extension method or the solvable group construction to determine these groups. The Frattini extension method in combination with an improved isomorphism test for solvable groups has been used for the vast majority of orders in the range up to 20,000. The only exception are the groups of order  $2^8 \cdot 3^2 = 2304$  which were constructed with the solvable group construction method. Further, we used the method of Sect. 3 to enumerate groups of certain orders. Among the orders  $n$  in the range at most 20,000 there are 19,733 orders of the form  $m \cdot p$  with  $p$  a prime that does not divide  $m$ . We applied the method of Sect. 3 to a significant range of these orders. In particular, we enumerated the groups of order  $2^9 \cdot p$  for  $p$  an odd prime with this approach.

**Non-solvable Groups** We have used the combination of the cyclic extension method with the limited version of Archer's  $Z\phi$  method to construct these groups. We note that there are 448 orders in the range of orders at most 20,000 for which non-solvable groups exist. For example, we determined 99,926 non-solvable groups of order  $7680 = 2^9 \cdot 3 \cdot 5$ , and counted that there are more than 8,279,000 non-solvable groups of order  $15,360 = 2^{10} \cdot 3 \cdot 5$ .

## 6 Open Cases and Challenges

We first discuss enumerations of groups before we consider explicit constructions.

### 6.1 Enumeration

There are 39 orders in the range at most 20,000 for which we have not (yet?) enumerated the groups of these orders. Note that in all but one case, order  $15,360 = 2^{11} \cdot 3 \cdot 5$ , the non-solvable groups have been enumerated successfully. Thus, the following discussion is primarily concerned with solvable groups.

**First Case** Let  $E_1 = \{n \in \{1, \dots, 20,000\} \mid 2^{10} \mid n \text{ or } 3^9 \mid n\}$ . Then  $E_1$  contains 20 orders. For these 20, the nilpotent groups of each order are not explicitly constructed, let alone the non-nilpotent groups. Using the methods in [13], one can determine that there are 4,896,600,938 groups of order  $3^9$  and exponent-3

class 2. Further, it is known that there are 49,487,365,422 groups of order  $2^{10}$ . These numbers of groups are so large, that the enumeration of groups of orders in  $E_1$  appears to be infeasible. Just to give an idea of the problems that will arise in trying to address this case, we note that the methods of [13] can be used to determine that there are 1,774,274,116,992,170 groups of order  $2^{11}$  and exponent-2 class 2. These methods are available as part of the GAP package [14].

**Second Case** Let  $E_2$  the set of those  $n \in \{1, \dots, 20,000\}$  satisfying that  $2^9 \cdot p$  is a proper divisor of  $n$  for some  $p \in \{3, 5, 7\}$ . There are 18 orders in  $E_2$ . There are over 400,000,000 groups for each of the orders  $2^9 \cdot p$  with  $p \in \{3, 5, 7\}$ . Hence, again, these numbers are so large, that the enumeration of groups of orders in  $E_2$  appears to be infeasible.

**Third Case** Let  $E_3 = \{n \in \{1, \dots, 20,000\} \mid (2^8 \cdot p^2) \mid n \text{ for some } p \in \{3, 5, 7\}\}$ . There are 12 orders in  $E_3$ . These orders are difficult cases for the construction of solvable groups via the Frattini extension method. We have determined the 15,756,130 groups of order  $2^8 \cdot 3^2$  using the solvable group construction. This order is an exception in the set  $E_3$ .

**Exceptional Cases** Six orders remain which are not in  $E_1 \cup E_2 \cup E_3$  and the construction of the groups of these orders is an open problem. We list these orders in the following table. The nilpotent groups of each of these orders are determined. Where known, we exhibit in the table the numbers of nilpotent, solvable and non-solvable groups.

$n$	# nilpotent	# solvable	# non-solvable
$8748 = 2^2 \cdot 3^7$	18,620	Not known	0
$10,368 = 2^7 \cdot 3^4$	34,920	Not known	0
$13,122 = 2 \cdot 3^8$	1,396,077	Not known	0
$16,000 = 2^7 \cdot 5^3$	11,640	Not known	0
$17,496 = 2^3 \cdot 3^7$	46,550	Not known	0
$18,816 = 2^7 \cdot 3 \cdot 7^2$	4656	Not known	387

## 6.2 Construction

As observed in the previous paragraph, for all but 39 orders at most 20,000 we have enumerated the numbers of groups of these orders. For the vast majority of these orders we have also determined isomorphism type representatives explicitly: more precisely, there are 34 orders for which we have enumerated the corresponding groups only, but did not construct them. This includes the order  $2^{10}$  for which the groups have been enumerated using the methods in [13].

For another 32 of these orders, we have used the approach exhibited in Sect. 3 to enumerate the groups of these orders. These orders are of the form  $2^9p$  for  $p$  an odd prime, of the form  $2^8pq$  for  $p \in \{3, 5, 7\}$  and  $q$  a prime different from 2 and  $p$ , of the form  $2^73^2p$  for  $p$  a prime different from 2 and 3, and the orders

$$\begin{aligned} 8640 &= 2^6 \cdot 3^3 \cdot 5, & 9600 &= 2^7 \cdot 3 \cdot 5^2, & 13,440 &= 2^7 \cdot 3 \cdot 5 \cdot 7, \\ 16,320 &= 2^6 \cdot 3 \cdot 5 \cdot 17, & 17,280 &= 2^7 \cdot 3^3 \cdot 5, & 19,440 &= 2^4 \cdot 3^5 \cdot 5. \end{aligned}$$

Finally, we counted the groups of order  $12,500 = 2^2 \cdot 5^5$  using a modified version of the method described in Sect. 3, by exploiting that these groups always admit a normal Sylow 5-subgroup.

For all other 19,927 orders in the range of orders at most 20,000, we have explicitly determined the groups of the corresponding orders, unless they are already available in the SmallGroups library [7]. The resulting groups will be made available as a package for GAP.

## 7 Reliability of the Data

It is important to cross-check the computed data for group constructions and enumerations. One very useful way for doing this is to determine or enumerate the groups of a certain order in two different ways. We have done this in many cases. In all of them the results of the different methods agree with each other.

We computed non-solvable groups using both the cyclic extension method and the limited version of Archer’s  $Z\phi$  method whenever Archer’s method applies. Additionally, we used the method of Sect. 3 to obtain an independent enumeration of the groups whenever the order is of the type  $mp$  with  $p$  prime and  $m$  coprime to  $p$  and we used different types of such factorisations of the order when possible. Out of the 447 orders admitting non-solvable groups, we enumerated 441 with at least two different methods. This leaves only six cases which were not cross-checked.

For the solvable groups, we employed two methods: the Frattini extension method, as well as the enumeration approach from Sect. 3. In the range of order at most 20,000, the SmallGroups library [7] already covers 17,903 orders. Of the remaining 2097 orders, we enumerated 1875 orders with both methods, 183 orders with only one method, and 39 orders remain open, see Sect. 6.

**Acknowledgements** We thank Eamonn O’Brien for comments on drafts of this work. The second author was supported by the DFG Schwerpunkt SPP 1489. The third author was supported by Simons Foundation Collaboration Grant 244502.

## References

1. C. Archer, The extension problem and classification of nonsolvable groups. PhD Thesis, Université Libre de Bruxelles, 1998
2. H.U. Besche, B. Eick, Construction of finite groups. *J. Symb. Comput.* **27**, 387–404 (1999)
3. H.U. Besche, B. Eick, GrpConst - Construction of finite groups (1999). A refereed GAP 4 package, see [29]
4. H.U. Besche, B. Eick, The groups of order  $q^n \cdot p$ . *Commun. Algebra* **29**(4), 1759–1772 (2001)
5. H.U. Besche, B. Eick, E.A. O'Brien, The groups of order at most 2000. *Electron. Res. Announc. Am. Math. Soc.* **7**, 1–4 (2001)
6. H.U. Besche, B. Eick, E.A. O'Brien, A millennium project: constructing small groups. *Int. J. Algebra Comput.* **12**, 623–644 (2002)
7. H.U. Besche, B. Eick, E. O'Brien, SmallGroups - a library of groups of small order (2005). A GAP 4 package; Webpage available at [www.icm.tu-bs.de/ag\\_algebra/software/small/small.html](http://www.icm.tu-bs.de/ag_algebra/software/small/small.html)
8. S. Blackburn, P. Neumann, G. Venkataraman, *Enumeration of Finite Groups* (Cambridge University Press, Cambridge, 2007)
9. J.J. Cannon, D.F. Holt, Automorphism group computation and isomorphism testing in finite groups. *J. Symb. Comput.* **35**, 241–267 (2003)
10. A. Cayley, On the theory of groups, as depending on the symbolic equation  $\theta^n = 1$ . *Philos. Mag.* **4**(7), 40–47 (1854)
11. J. Conway, H. Dietrich, E. O'Brien, Counting groups: Gnus, Moas and other exotica. *Math. Intell.* **30**, 6–15 (2008)
12. B. Eick, M. Horn, The construction of finite solvable groups revisited. *J. Algebra* **408**, 166–182 (2014)
13. B. Eick, E.A. O'Brien, Enumerating  $p$ -groups. *J. Aust. Math. Soc.* **67**, 191–205 (1999)
14. B. Eick, E. O'Brien, AutPGrp - computing the automorphism group of a  $p$ -group, Version 1.8 (2016). A refereed GAP 4 package, see [29]
15. B. Eick, C.R. Leedham-Green, E.A. O'Brien, Constructing automorphism groups of  $p$ -groups. *Commun. Algebra* **30**, 2271–2295 (2002)
16. D. Holt, W. Plesken, *Perfect Groups* (Clarendon Press, Oxford, 1989)
17. M. Horn, B. Eick, GroupExt - Constructing finite groups (2013). A GAP 4 package, see [29]
18. M.F. Newman, E.A. O'Brien, M.R. Vaughan-Lee, Groups and nilpotent Lie rings whose order is the sixth power of a prime. *J. Algebra* **278**, 383–401 (2003)
19. E.A. O'Brien, The groups of order dividing 256. PhD thesis, Australian National University, Canberra, 1988
20. E. O'Brien, ANUPQ - the ANU  $p$ -Quotient algorithm (1990). Also available in MAGMA and as GAP package
21. E.A. O'Brien, The  $p$ -group generation algorithm. *J. Symb. Comput.* **9**, 677–698 (1990)
22. E.A. O'Brien, M.R. Vaughan-Lee, The groups with order  $p^7$  for odd prime  $p$ . *J. Algebra* **292**(1), 243–258 (2005)
23. L. Pyber, Group enumeration and where it leads us, in *European Congress of Mathematics, Volume II (Budapest, 1996)*, Progress in Mathematics, vol. 169 (Birkhäuser, Basel, 1998), pp. 187–199
24. R. Schwingel, Two matrix group algorithms with applications to computing the automorphism group of a finite  $p$ -group. PhD Thesis, QMW, University of London, 2000
25. J.K. Senior, A.C. Lunn, Determination of the groups of orders 101–161, omitting order 128. *Am. J. Math.* **56**(1–4), 328–338 (1934)
26. J.K. Senior, A.C. Lunn, Determination of the groups of orders 162–215 omitting order 192. *Am. J. Math.* **57**(2), 254–260 (1935)
27. M.J. Smith, Computing automorphisms of finite soluble groups. PhD thesis, Australian National University, Canberra, 1995



28. D. Taunt, Remarks on the isomorphism problem in theories of construction of finite groups. Proc. Camb. Philos. Soc. **51**, 16–24 (1955)
29. The GAP Group, GAP – groups, algorithms and programming, Version 4.4. Available from <http://www.gap-system.org> (2005)
30. M. Vaughan-Lee, B. Eick, SglPPow – Database of certain p-groups (2016). A GAP 4 package, see [29]

# Classifying Nilpotent Associative Algebras: Small Coclass and Finite Fields



Bettina Eick and Tobias Moede

**Abstract** We survey the state of the art in the classification of nilpotent associative  $\mathbb{F}$ -algebras by coclass using their associated coclass graphs  $\mathcal{G}_{\mathbb{F}}(r)$ . For arbitrary fields  $\mathbb{F}$ , we determine up to isomorphism the nilpotent associative  $\mathbb{F}$ -algebras of coclass 1 and their coclass graphs  $\mathcal{G}_{\mathbb{F}}(1)$ . For finite fields  $\mathbb{F}$  and arbitrary  $r$ , we propose a conjecture on the structure of the coclass graph  $\mathcal{G}_{\mathbb{F}}(r)$ ; this conjecture is based on computational investigations. We further show how computational methods apply in an enumeration of the isomorphism types of nilpotent associative  $\mathbb{F}$ -algebras of small dimensions over small finite fields  $\mathbb{F}$ .

**Keywords** Coclass theory • Nilpotent associative algebras •  $p$ -groups

**Subject Classifications** 16N40, 16W99, 16Z05, 20D15

## 1 Introduction

Let  $\mathbb{F}$  be an arbitrary field. An associative  $\mathbb{F}$ -algebra  $A$  is called **nilpotent** of class  $\text{cl}(A)$  if every product of  $\text{cl}(A) + 1$  elements of  $A$  is zero and there exist  $\text{cl}(A)$  elements in  $A$  whose product is non-zero. Note that a nilpotent associative  $\mathbb{F}$ -algebra does not contain an identity element. The **coclass** of a nilpotent associative  $\mathbb{F}$ -algebra  $A$  is defined as

$$\text{cc}(A) = \dim(A) - \text{cl}(A).$$

---

B. Eick  
Institut Computational Mathematics, TU Braunschweig, Pockelsstraße 14,  
38106 Braunschweig, Germany  
e-mail: [beick@tu-bs.de](mailto:beick@tu-bs.de)

T. Moede (✉)  
School of Mathematical Sciences, Monash University, Clayton, VIC 3800, Australia  
e-mail: [tobias.moede@monash.edu](mailto:tobias.moede@monash.edu); [t.moede@tu-braunschweig.de](mailto:t.moede@tu-braunschweig.de)

For each field  $\mathbb{F}$  and each  $r \in \mathbb{N}_0$  one can visualize the nilpotent associative  $\mathbb{F}$ -algebras of coclass  $r$  in a graph  $\mathcal{G}_{\mathbb{F}}(r)$ : The vertices of this graph correspond one-to-one to the isomorphism types of nilpotent associative  $\mathbb{F}$ -algebras of coclass  $r$  and there is an edge  $A \rightarrow B$  if  $\text{cl}(B) = \text{cl}(A) + 1$  and  $B/B^{\text{cl}(B)} \cong A$  holds, where  $B^{\text{cl}(B)}$  is the ideal of  $B$  spanned by all products of  $\text{cl}(B)$  elements in  $B$ .

The coclass of a nilpotent associative  $\mathbb{F}$ -algebra is a non-negative integer and thus each nilpotent associative  $\mathbb{F}$ -algebra is contained in one of the graphs  $\mathcal{G}_{\mathbb{F}}(r)$ . Hence the classification up to isomorphism of nilpotent associative  $\mathbb{F}$ -algebras translates to an investigation of the coclass graphs  $\mathcal{G}_{\mathbb{F}}(r)$  for each  $r \in \mathbb{N}_0$ . This provides a new approach towards a classification of nilpotent associative  $\mathbb{F}$ -algebras.

We consider the graphs  $\mathcal{G}_{\mathbb{F}}(r)$  in more detail. By construction, each connected component of the graph  $\mathcal{G}_{\mathbb{F}}(r)$  is a tree which we call a **maximal descendant tree**. The roots of these trees are the **roots** of the graph  $\mathcal{G}_{\mathbb{F}}(r)$ . An infinite path in  $\mathcal{G}_{\mathbb{F}}(r)$  is called **maximal** if it is not properly contained in another infinite path of  $\mathcal{G}_{\mathbb{F}}(r)$ .

**Theorem 1.1** ([5, Theorem 3] and [6, Theorem 1]) *Let  $\mathbb{F}$  be an arbitrary field and  $r \in \mathbb{N}_0$ .*

- a) *The roots of  $\mathcal{G}_{\mathbb{F}}(r)$  have dimension at most  $2r$ .*
- b) *The graph  $\mathcal{G}_{\mathbb{F}}(r)$  has finitely many maximal infinite paths if and only if  $r \leq 1$  or  $\mathbb{F}$  is a finite field.*

Theorem 1.1 b) suggests that the cases of algebras of coclass at most 1 and algebras over finite fields provide two promising areas for further investigations. We consider these in more detail in the following.

### 1.1 Coclass at most 1

It is not difficult to show that  $\mathcal{G}_{\mathbb{F}}(0)$  consists of a single infinite path. In [5] it is shown that  $\mathcal{G}_{\mathbb{F}}(1)$  consists of a single infinite tree having one infinite path starting at its root and there is an experimental investigation of  $\mathcal{G}_{\mathbb{F}}(1)$  exhibited for some finite fields  $\mathbb{F}$ .

Our first aim here is a complete classification up to isomorphism of the nilpotent associative  $\mathbb{F}$ -algebras of coclass 1 for all fields  $\mathbb{F}$ . We include a brief summary of this result here and refer to Sect. 3 for details.

For an arbitrary field  $\mathbb{F}$  and  $i \geq 2$  let  $U_i = (\mathbb{F}^*)^{i-1}$  denote the group of  $(i - 1)$ -th powers and  $S = (\mathbb{F}^*)^2$  the group of squares in the multiplicative group  $\mathbb{F}^*$ . Further, if there is an edge  $A \rightarrow B$  in a coclass graph, then we say that  $B$  is an **immediate descendant** of  $A$ .

**Theorem 1.2 (See Sect. 3 for a Proof)** *Let  $\mathbb{F}$  be an arbitrary field and let  $A_1 \rightarrow A_2 \rightarrow \dots$  denote the infinite path in  $\mathcal{G}_{\mathbb{F}}(1)$  starting at its root.*

- a) *Each vertex in  $\mathcal{G}_{\mathbb{F}}(1)$  has distance at most 1 from the infinite path.*
- b) *The algebra  $A_1$  has  $[\mathbb{F}^* : S] + |\mathbb{F}^*| + 3$  immediate descendants and, for  $i \geq 2$ , the algebra  $A_i$  has  $[\mathbb{F}^* : SU_i] + [\mathbb{F}^* : U_i] + 2$  immediate descendants.*

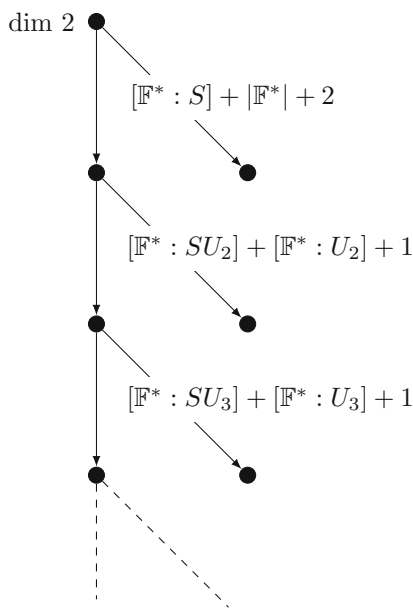
Theorem 1.2 fully describes  $\mathcal{G}_{\mathbb{F}}(1)$ . We visualize this graph in Fig. 1 in compact notation: a number  $n$  on an edge means that this edge exists  $n$  times.

We note the following immediate consequence of Theorem 1.2.

**Corollary 1.3** *Using the notation of Theorem 1.2, let  $a_i$  denote the number of immediate descendants of  $A_i$  in  $\mathcal{G}_{\mathbb{F}}(1)$ .*

- *If  $\mathbb{F}$  is a finite field of size  $q$ , then  $a_i = a_{i+(q-1)}$  for each  $i \geq 2$ .*
- *If  $\mathbb{F}$  is algebraically closed, then  $a_i = 4$  for each  $i \geq 2$ .*
- *If  $\mathbb{F} = \mathbb{R}$  and  $i \geq 2$ , then  $a_i = 6$  if  $i$  is even, and  $a_i = 4$  if  $i$  is odd.*
- *If  $\mathbb{F} = \mathbb{Q}$ , then  $a_i = \infty$  for all  $i \geq 1$ .*
- *Let  $P$  be a set of primes, let  $\mathbb{F}$  be the closure of  $\mathbb{Q}$  under taking  $p$ -th roots for all  $p \in P$  and let  $i \geq 2$ . Then  $a_i = 4$  if and only if all prime factors of  $i - 1$  lie in  $P$ .*

**Fig. 1** The coclass graph  $\mathcal{G}_{\mathbb{F}}(1)$



## 1.2 Finite Fields

We call a subtree of  $\mathcal{G}_{\mathbb{F}}(r)$  a **coclass tree** if it contains exactly one infinite path starting at its root. A coclass tree is called **maximal** if it is not properly contained in another coclass tree. Coclass trees play a crucial role in the study of  $\mathcal{G}_{\mathbb{F}}(r)$  for finite fields. We recall the following main result on coclass graphs for finite fields.

**Theorem 1.4** ([6, Corollary 2] and [6, Theorem 3]) *Let  $\mathbb{F}$  be a finite field and  $r \in \mathbb{N}_0$ .*

- a)  $\mathcal{G}_{\mathbb{F}}(r)$  consists of finitely many maximal descendant trees.
- b)  $\mathcal{G}_{\mathbb{F}}(r)$  consists of finitely many maximal coclass trees and finitely many other vertices.

Theorem 1.4 reduces the investigation of coclass graphs over finite fields to an investigation of their maximal coclass trees. We introduce some further notation to discuss this in more detail. Given a vertex  $A$  in  $\mathcal{G}_{\mathbb{F}}(r)$ , we say that  $B$  is a **descendant** of  $A$  if there is a path from  $A$  to  $B$  in  $\mathcal{G}_{\mathbb{F}}(r)$ . We denote with  $\mathcal{T}_A$  the full subtree of  $\mathcal{G}_{\mathbb{F}}(r)$  consisting of all descendants of  $A$ .

Let  $\mathcal{T}$  be a maximal coclass tree in  $\mathcal{G}_{\mathbb{F}}(r)$  with root  $A$  and denote its maximal infinite path by  $A = A_1 \rightarrow A_2 \rightarrow \dots$ . Then the **depth**  $\text{dep}(\mathcal{T})$  of  $\mathcal{T}$  is the maximal distance of a vertex in  $\mathcal{T}$  to its infinite path. The **rank**  $\text{rk}(\mathcal{T})$  is the dimension of  $A/A^2$ . Further, we say that  $\mathcal{T}$  is **virtually periodic** with **period**  $d$  if  $\mathcal{T}_{A_i}$  and  $\mathcal{T}_{A_{i+d}}$  are isomorphic as directed trees.

In [6] we proposed a conjecture on the periodic patterns in the maximal coclass trees of  $\mathcal{G}_{\mathbb{F}}(r)$  for each finite field  $\mathbb{F}$  and each  $r \in \mathbb{N}_0$ . Here we propose the following stronger and more detailed version of this conjecture.

*Conjecture (Stronger Version of [6, Conjecture 6])* Let  $\mathbb{F}$  be a finite field of size  $q$  and characteristic  $p$  and let  $r \in \mathbb{N}_0$ . Let  $\mathcal{T}$  be a maximal coclass tree of  $\mathcal{G}_{\mathbb{F}}(r)$  of depth  $d$  and rank  $e$ . Then:

- a) The depth of  $\mathcal{T}$  is bounded by  $d \leq r - e + 2$ .
- b) The tree  $\mathcal{T}$  is virtually periodic with period dividing  $p^{d-1}(q - 1)$ .

This conjecture holds trivially for  $r = 0$ . Theorem 1.2 proves that it holds for  $r = 1$ . It thus remains to investigate the conjecture for  $r \geq 2$ . In [6] we exhibited various experimental data for graphs  $\mathcal{G}_{\mathbb{F}}(2)$  and we note that these experiments support the conjecture. Our aim here is to exhibit further experimental support for the conjecture, see Sect. 4. We use the algorithm of [6] for this purpose.

	class 2				class $\geq 3$			
	dim 3	dim 4	dim 5	dim 6	dim 3	dim 4	dim 5	dim 6
$\mathbb{F}_2$	5	21	354	42319	1	5	49	729
$\mathbb{F}_3$	7	29	1703	3328650	1	5	54	
$\mathbb{F}_4$	7	31	6684	105547591	1	5	59	
$\mathbb{F}_5$	9	39	22052	1685636086	1	5	64	
$\mathbb{F}_7$	11	49	144894	118539109666	1	5	72	
$\mathbb{F}_8$	11	51	311520	651101343361	1	5		
$\mathbb{F}_9$	13	59	616331	2940062651968	1	5		

**Fig. 2** Numbers of nilpotent associative algebras over small finite fields. Empty entries indicate that the precise number is not known

### 1.3 Enumeration of Algebras

The classification by coclass is a new approach towards a detailed investigation of nilpotent associative  $\mathbb{F}$ -algebras. More classical is to use the dimension as primary invariant. We consider this here briefly.

In [7] there is an effective method introduced to count the number of isomorphism types of the finite  $p$ -groups of given order and exponent- $p$  class 2. A variation of this method allows to count the isomorphism types of nilpotent associative  $\mathbb{F}$ -algebras of a given dimension and class 2 for a finite field  $\mathbb{F}$ .

Further, using the algorithm of [6] and combining it with the Burnside-Lemma to count numbers of orbits it is possible to count the isomorphism types of nilpotent associative  $\mathbb{F}$ -algebras of a given dimension. We exhibit our results on enumerations of algebras in Fig. 2.

These numbers coincide with the classifications in [2, 8] for dimension at most 3 and with [1] for dimension at most 5 over  $\mathbb{F}_2$ . The fast growing numbers even for class 2 algebras indicate the difficulty of a classification by dimension only.

## 2 The Construction of Coclass Graphs

In this section we recall the main ideas used in the investigation of coclass graphs as far as we need them later.

### 2.1 Constructing Immediate Descendants

Let  $A$  be a finite-dimensional nilpotent  $\mathbb{F}$ -algebra for an arbitrary field  $\mathbb{F}$  and write  $e = \dim_{\mathbb{F}}(A/A^2)$ . Note that  $e$  is the minimal generator number of  $A$ . Let  $F$  be the non-unital free associative  $\mathbb{F}$ -algebra on  $e$  generators and let  $R \trianglelefteq F$  with  $A \cong F/R$ .

Then the **covering algebra**  $A^*$  of  $A$  is given by

$$A^* \cong F/(FR \cup RF),$$

where  $(FR \cup RF)$  is the ideal generated by  $FR \cup RF$  in  $F$ . The covering algebra was introduced in [3]. It was shown that  $A^*$  is finite-dimensional and nilpotent of class  $\text{cl}(A)$  or  $\text{cl}(A) + 1$ . Let

$$\varphi : A^* \rightarrow A$$

be the natural epimorphism. Then the kernel of  $\varphi$  is called the **multiplicator** of  $A$  and is denoted with  $M(A)$ . By construction,  $M(A)A^* = A^*M(A) = \{0\}$  holds. We call  $N(A) = (A^*)^{\text{cl}(A)+1}$  the **nucleus** of  $A^*$ .

**Theorem 2.1 ([3, Theorem 7])** *Let  $A$  be a finite-dimensional nilpotent  $\mathbb{F}$ -algebra over an arbitrary field  $\mathbb{F}$ .*

- a) *For each immediate descendant  $B$  of  $A$  there exists one (or several) proper subspaces  $U < M(A)$  with  $U + N(A) = M(A)$  so that  $A^*/U \cong B$ .*
- b) *If  $U$  is a proper subspace of  $M(A)$  with codimension 1 and  $U + N(A) = M(A)$ , then  $A^*/U$  is an immediate descendant of  $A$ .*

This theorem shows that immediate descendants can be associated with supplements of  $N(A)$  of codimension 1 in  $M(A)$ . It remains to solve the isomorphism problem for immediate descendants. For this purpose we note that each automorphism of  $A$  extends to an automorphism of  $A^*$ . This extended automorphism of  $A^*$  is not necessarily unique, but its action on  $M(A)$  is. Hence  $\text{Aut}(A)$  acts on  $M(A)$  and also on the set of proper supplements of  $N(A)$  in  $M(A)$ .

**Theorem 2.2 ([3, Theorem 10])** *Let  $A$  be a finite-dimensional nilpotent  $\mathbb{F}$ -algebra over an arbitrary field  $\mathbb{F}$ . Let  $B_1, B_2$  be two immediate descendants of  $A$  and suppose that  $B_i = A^*/U_i$  for  $i = 1, 2$  and two supplements  $U_1, U_2$  to  $N(A)$  in  $M(A)$ . Then  $B_1 \cong B_2$  if and only if there exists an automorphism  $\alpha \in \text{Aut}(A)$  with  $U_1^\alpha = U_2$ .*

Theorems 2.1 and 2.2 translate readily to a method to determine isomorphism type representatives of immediate descendants. This method requires the determination of the  $\text{Aut}(A)$ -orbits of proper supplements to  $N(A)$  in  $M(A)$ . If  $\mathbb{F}$  is a finite field, then this is always a finite calculation and translates to an implementable algorithm, see [6, Algorithm 13] for details. A GAP [9] implementation of this algorithm is available in the package `ccalgs`, see [4].

## 2.2 Exploring Coclass Graphs

The algorithm of Sect. 2.1 can be used to explore coclass graphs. Note that if  $A \rightarrow B$  is an edge in a coclass graph, then  $A$  and  $B$  have the same coclass and their class

differs by one. Thus  $\dim(B) = \dim(A) + 1$ , and  $B$  corresponds to a supplement  $U$  to  $N(A)$  in  $M(A)$  of codimension 1.

If  $\mathbb{F}$  is a finite field and  $r \in \mathbb{N}_0$ , then all algebras in a coclass graph  $\mathcal{G}_{\mathbb{F}}(r)$  up to some fixed dimension  $d$  can be computed. For this purpose we first determine the roots of the graph with the method exhibited in [6] and then we iteratively determine immediate descendants with the approach of Sect. 2.1.

### 3 Coclass 1

Let  $\mathbb{F}$  be an arbitrary field. The aim of this section is to determine up to isomorphism the nilpotent associative  $\mathbb{F}$ -algebras of coclass 1 and thus to prove Theorem 1.2. It follows from [5, Corollary 11] that the algebras  $A_i$  on the infinite path  $A_1 \rightarrow A_2 \rightarrow \dots$  starting at the root of  $\mathcal{G}_{\mathbb{F}}(1)$  can be described as  $A_i = \langle t, a \mid a^2, at, ta, t^{i+1} \rangle$ . Using the notation of Theorem 1.2, we proceed in the following steps:

- **Step 1:** Determine the immediate descendants of  $A_1$ .
- **Step 2:** Determine the immediate descendants of  $A_i$  for  $i \geq 2$ .
- **Step 3:** Summarize the resulting classifications of immediate descendants.
- **Step 4:** Show that each of the determined immediate descendants except the  $A_i$ 's does not have any further immediate descendants.

#### 3.1 Step 1

We use the general approach exhibited in Sect. 2.1 to prove Step 1. Recall that  $A_1 = \langle t, a \mid a^2, at, ta, t^2 \rangle$ . Then a straightforward calculation shows that

- $A_1^* = \langle t, a \mid \text{all products of length 3} \rangle$ ,
- $M(A_1) = \langle a^2, at, ta, t^2 \rangle$ , and
- $N(A_1) = M(A_1)$ .

By construction,  $\text{Aut}(A_1) \cong \text{GL}(2, \mathbb{F})$ . Let  $V = \mathbb{F}^2$  denote the natural module for  $\text{GL}(2, \mathbb{F})$ . Then  $\text{Aut}(A_1)$  acts on  $M(A_1)$  with respect to the basis  $\{a^2, at, ta, t^2\}$ , as  $\text{GL}(2, \mathbb{F})$  acts on  $T := V \otimes_{\mathbb{F}} V$ . Theorems 2.1 and 2.2 now assert that the isomorphism types of immediate descendants of  $A_1$  in  $\mathcal{G}_{\mathbb{F}}(1)$  correspond one-to-one to the  $\text{GL}(2, \mathbb{F})$ -orbits of subspaces of codimension 1 in  $M(A_1)$ .

Let  $\rho$  denote the map that maps each subspace  $C$  of  $M(A_1)$  to its orthogonal complement  $C^\dagger$  with respect to the standard scalar product. Then  $\rho$  is compatible with the action of  $\text{GL}(2, \mathbb{F})$ . It follows that the  $\text{GL}(2, \mathbb{F})$ -orbits of subspaces of codimension 1 in  $M(A_1)$  correspond one-to-one to the  $\text{GL}(2, \mathbb{F})$ -orbits of subspaces of dimension 1 in  $M(A_1)$ .



Let  $\varphi$  be a transversal function for  $S = \{x^2 \mid x \in \mathbb{F}^*\}$  in the multiplicative group  $\mathbb{F}^*$  and let  $U$  denote the subgroup of upper triangular matrices in  $GL(2, \mathbb{F})$ ; that is,

$$U = \left\{ \begin{pmatrix} x & y \\ 0 & z \end{pmatrix} \mid x, z \in \mathbb{F}^*, y \in \mathbb{F} \right\}.$$

As a first step, we describe the  $U$ -orbits of 1-dimensional subspaces of  $T$  in the following Lemma.

**Lemma 3.1** *Let  $w_2, w_3, w_4 \in \mathbb{F}$ .*

(a) *The  $U$ -orbit of  $\langle(1, w_2, w_3, w_4)\rangle$  has the representative*

- $\langle(1, 0, 0, 0)\rangle$  if  $w_2 = w_3$  and  $w_4 = w_3^2$ ,
- $\langle(1, 0, 0, \varphi(w_4 - w_3^2))\rangle$  if  $w_2 = w_3$  and  $w_4 \neq w_3^2$ ,
- $\langle(1, 0, 1, (w_4 - w_2w_3)/(w_3 - w_2)^2)\rangle$  if  $w_2 \neq w_3$ .

(b) *The  $U$ -orbit of  $\langle(0, 1, w_3, w_4)\rangle$  has the representative*

- $\langle(0, 1, -1, 1)\rangle$  if  $w_3 = -1$  and  $w_4 \neq 0$ ,
- $\langle(0, 1, w_3, 0)\rangle$  otherwise.

(c) *The  $U$ -orbit of  $\langle(0, 0, 1, w_4)\rangle$  has the representative  $\langle(0, 0, 1, 0)\rangle$ .*

(d) *The  $U$ -orbit of  $\langle(0, 0, 0, 1)\rangle$  consists of this element only.*

*Proof* Note that each 1-dimensional subspace contains a unique normed vector; that is, a vector whose first non-zero entry equals 1. We use normed vectors to represent 1-dimensional subspaces throughout the proof.

For  $x, z \in \mathbb{F}^*$  and  $y \in \mathbb{F}$  write

$$a = \begin{pmatrix} x & y \\ 0 & z \end{pmatrix}.$$

Then  $a$  describes a generic element in  $U$  and it acts on the tensor product  $T$  via

$$a \otimes a = \begin{pmatrix} x^2 & xy & xy & y^2 \\ 0 & xz & 0 & yz \\ 0 & 0 & xz & yz \\ 0 & 0 & 0 & z^2 \end{pmatrix}.$$

Write  $s = zx^{-1}$ .

(a) The normed vector associated with  $(1, w_2, w_3, w_4)(a \otimes a)$  is the vector  $(1, (zw_2 + y)x^{-1}, (zw_3 + y)x^{-1}, (yzw_2 + yzw_3 + z^2w_4 + y^2)x^{-2})$ . Choosing  $y = -zw_2$  translates this to  $(1, 0, (w_3 - w_2)zx^{-1}, (w_4 - w_2w_3)(zx^{-1})^2) = (1, 0, (w_3 - w_2)s, (w_4 - w_2w_3)s^2)$ . If  $w_3 = w_2$ , then this yields the stated result. If  $w_3 \neq w_2$ , then choosing  $s = (w_3 - w_2)^{-1}$  yields the stated result.

- (b) The normed vector associated with  $(0, 1, w_3, w_4)(a \otimes a)$  is  $(0, 1, w_3, (y(w_3 + 1) + zw_4)x^{-1})$ . If  $w_3 = -1$  and  $w_4 \neq 0$ , then choosing  $s = w_4^{-1}$  yields the stated result. Otherwise we choose  $y = -zw_4/(w_3 + 1)$  if  $w_3 \neq -1$  or  $y = 0$  if  $w_4 = 0$  to obtain the stated result.
- (c) Determining  $(0, 0, 1, w_4)(a \otimes a)$  and norming the result yields  $(0, 0, 1, (zw_4 + y)x^{-1})$ . Choosing  $y = -zw_4$  yields the stated result.
- (d) Determining  $(0, 0, 0, 1)(a \otimes a)$  and norming the resulting vector yields  $(0, 0, 0, 1)$  as claimed.

Based on Lemma 3.1 we determine the orbits of  $GL(2, \mathbb{F})$  on the 1-dimensional subspaces in  $T$  as follows.

**Theorem 3.2** *Orbit representatives for the  $GL(2, \mathbb{F})$ -orbits on the 1-dimensional subspaces of  $T$  are*

- $\langle(0, 0, 0, 1)\rangle$ ,
- $\langle(0, 0, 1, 0)\rangle$ ,
- $\langle(0, 1, -1, 0)\rangle$ ,
- $\langle(1, 0, 0, u)\rangle$ , with  $u$  in a transversal of  $S$  in  $\mathbb{F}^*$ , and
- $\langle(1, 0, 1, v)\rangle$ , with  $v \in \mathbb{F}^*$ .

*Proof* Again, let  $U$  denote the group of upper triangular matrices and note that  $GL(2, \mathbb{F}) = \langle t, U \rangle$  with

$$t = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The element  $t$  acts on  $T$  as

$$t \otimes t = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

As in the proof of Lemma 3.1, we use normed vectors to represent 1-dimensional subspaces. We obtain the  $GL(2, \mathbb{F})$ -orbits of normed vectors in  $T$  by closing the  $U$ -orbits under the action of  $(t \otimes t)$ . We now proceed in two steps.

Step 1: We show that the list in Theorem 3.2 is complete. For this purpose we observe that each of the orbit representatives of Lemma 3.1 can be mapped to one of the representatives in Theorem 3.2.

- Consider the representatives of  $U$ -orbits for the normed vectors of the form  $(1, w_2, w_3, w_4)$  with  $w_2, w_3, w_4 \in \mathbb{F}$ . Let  $u$  be in a transversal of  $S$  in  $\mathbb{F}^*$  and  $v \in \mathbb{F}^*$ .

$$(1, 0, 0, 0) \xrightarrow{t \otimes t} (0, 0, 0, 1),$$

$(1, 0, 0, u)$  with  $u$  in a transversal of  $S$  in  $\mathbb{F}^*$  remains fixed,

$(1, 0, 1, v)$  with  $v$  in  $\mathbb{F}^*$  remains fixed,

$$(1, 0, 1, 0) \xrightarrow{t \otimes t} (0, 1, 0, 1) \xrightarrow{U} (0, 1, 0, 0) \xrightarrow{t \otimes t} (0, 0, 1, 0).$$

- Consider the representatives of  $U$ -orbits for the normed vectors  $(0, 1, w_3, w_4)$  with  $w_3, w_4 \in \mathbb{F}$ . If  $w_4 \neq 0$ , then  $(0, 1, w_3, w_4)(t \otimes t) = (1, w_3w_4^{-1}, w_4^{-1}, 0)$  and such normed vectors have been considered above. If  $w_4 = 0$  and  $w_3 = -1$  then the result is trivial. If  $w_4 = 0$  and  $w_3 \neq -1$ , then  $U$  maps this vector to the normed vector  $(0, 1, w_3, (w_3 + 1)yx^{-1})$  and thus for  $y \neq 0$  to an element with last entry non-zero; these have been considered above. Finally,  $(0, 1, 0, 0)$  maps under  $(t \otimes t)$  to  $(0, 0, 1, 0)$ .
- Consider the representatives of  $U$ -orbits for the normed vectors of the form  $(0, 0, 1, w_4)$  with  $w_4 \in \mathbb{F}$ . These map under  $U$  to  $(0, 0, 1, 0)$  and thus are included.
- Finally, the vector  $(0, 0, 0, 1)$  is included.

Step 2: We show that the list in Theorem 3.2 is irredundant. As a first example, consider the vector  $(0, 0, 0, 1)$  of the list. Suppose that there exists

$$g = \begin{pmatrix} x & y \\ w & z \end{pmatrix} \in \text{GL}(2, \mathbb{F})$$

so that  $g$  maps this vector onto another representative in the list. Note that

$$(0, 0, 0, 1)(g \otimes g) = (w^2, wz, wz, z^2).$$

If  $w = 0$ , then this yields the normed vector  $(0, 0, 0, 1)$  and thus the vector itself. If  $w \neq 0$ , then this yields the normed vector  $(1, zw^{-1}, zw^{-1}, (zw^{-1})^2)$ . It is easy to see that this is not a representative in the list of Theorem 3.2.

As a second example, consider the vector  $(1, 0, 1, v)$  with  $v \in \mathbb{F}^*$  of the list. Note that

$$(1, 0, 1, v)(g \otimes g) = (w^2v + x^2 + xw, zwv + xy + yw, zwv + xy + xz, z^2v + y^2 + yz).$$

We denote  $(w_1, w_2, w_3, w_4) = (w^2v + x^2 + xw, zwv + xy + yw, zwv + xy + xz, z^2v + y^2 + yz)$ . Then the normed vector obtained from  $(w_1, w_2, w_3, w_4)$  is not among the first three representatives in the above list. If it would be of the form  $(1, 0, 0, u)$  with  $u$  in a transversal of  $S$  in  $\mathbb{F}^*$ , then it follows that

$$zwv + xy + yw = 0 \quad \text{and} \quad zwv + xy + xz = 0.$$

This implies  $\det(g) = xz - yw = 0$  which contradicts  $g \in \text{GL}(2, \mathbb{F})$ . Finally, suppose that the normed vector obtained from  $(w_1, w_2, w_3, w_4)$  is of the form  $(1, 0, 1, v')$  with  $v' \in \mathbb{F}^*$ . Then the following equations hold:

- (1)  $w_1 \neq 0$ ,
- (2)  $w_2 = 0$ ,
- (3)  $w_3 - w_1 = 0$ ,
- (4)  $w_4 - w_1 v' = 0$ , and
- (5)  $xz - wy \neq 0$ .

Using equations (2) and (3), it follows that  $w_1 = xz - wy$ . Write  $d = xz - wy$  and replace  $w_1$  by  $d$  in the equations. Then the ideal generated by equations (2), (3), (4) and  $d = xz - wy$  has a reduced Groebner basis containing the equation  $vd^2 - v'd^2$ . As  $d \neq 0$ , it follows that  $v = v'$  holds.

Similar calculations prove the claim for the other entries in the list of Theorem 3.2.

### 3.2 Step 2

Again we use the general approach of Sect. 2.1 to prove Step 2. Recall that  $A_i = \langle t, a \mid a^2, at, ta, t^{i+1} \rangle$  for  $i \geq 2$ . Then a straightforward calculation shows that

- $A_i^* = \langle t, a \mid a^3, a^2t, at^2, ata, tat, t^2a, ta^2, t^{i+2} \rangle$ , and
- $M(A_i) = \langle a^2, at, ta, t^{i+1} \rangle$ , and
- $N(A_i) = \langle t^{i+1} \rangle$ .

First, we determine  $\text{Aut}(A_i)$ . Note that  $A_i$  has two  $\text{Aut}(A_i)$ -invariant series: the upper annihilator series and the series of power ideals. The upper annihilator series has the ideals  $\text{Ann}_0(A_i) = \{0\}$  and  $\text{Ann}_j(A_i) = \langle t^{i-j+1}, t^{i-j+2}, \dots, t^i, a \rangle$  for  $1 \leq j \leq i$  with  $\text{Ann}_i(A_i) = A_i$ . The series of power ideals has the ideals  $A_i^1 = A_i$  and  $A_i^j = \langle t^j, t^{j+1}, \dots, t^i \rangle$  for  $2 \leq j \leq i$  with  $A_i^{i+1} = \{0\}$ .

**Lemma 3.3** *Each automorphism  $\alpha \in \text{Aut}(A_i)$  has the form*

$$\begin{aligned} \alpha(t) &= xt + ua + b \quad \text{with } x \in \mathbb{F}^*, u \in \mathbb{F}, b \in A_i^2, \\ \alpha(a) &= ya + vt^i \quad \text{with } y \in \mathbb{F}^*, v \in \mathbb{F}. \end{aligned}$$

*Proof* Each element of  $A_i$  can be written as  $xt + ua + b$  with  $x, u \in \mathbb{F}^*$  and  $b \in A_i^2$ . As  $\alpha$  is surjective, it follows that the image of  $\alpha$  has to cover  $A/\text{Ann}_{i-1}(A_i)$  and thus  $x \in \mathbb{F}^*$  follows.

The image of  $a$  under  $\alpha$  has to be an element of  $\text{Ann}_1(A_i) = \langle a, t^i \rangle$ . Thus  $\alpha(a) = ya + vt^i$  for  $y, v \in \mathbb{F}$ . Again, as  $\alpha$  is surjective, and the subgroup generated by  $\alpha(t)$  avoids  $\langle a \rangle$ , it follows that  $y \in \mathbb{F}^*$ .

Next, we determine the action of  $\text{Aut}(A_i)$  on  $M(A_i)$ .

**Lemma 3.4** *Let  $\alpha \in \text{Aut}(A_i)$  as in Lemma 3.3 depending on the parameters  $x, y \in \mathbb{F}^*$  and  $u, v \in \mathbb{F}$ . Then with respect to the basis  $\{a^2, at, ta, t^{i+1}\}$  of  $M(A_i)$ , it follows that  $\alpha$  acts as*

$$\begin{pmatrix} y^2 & 0 & 0 & 0 \\ yu & yx & 0 & vx \\ yu & 0 & yx & vx \\ 0 & 0 & 0 & x^{i+1} \end{pmatrix}.$$

*Proof* This follows by a direct calculation.

$$\begin{aligned} \alpha(a^2) &= \alpha(a)^2 = (ya + vt^i)(ya + vt^i) \\ &= y^2a^2 + vyt^ia + yvat^i + v^2t^{2i} \\ &= y^2a^2, \end{aligned}$$

since  $i \geq 2$  and thus  $t^ia = at^i = t^{2i} = 0$  in  $A_i^*$ .

$$\begin{aligned} \alpha(at) &= \alpha(a)\alpha(t) = (ya + vt^i)(xt + ua + b) \\ &= yxat + yua^2 + yab + vxt^{i+1} + vut^ia + vt^ib \\ &= yxat + yua^2 + vxt^{i+1}, \text{ and similarly} \\ \alpha(ta) &= yxta + yua^2 + vxt^{i+1}, \end{aligned}$$

since  $b \in A_i^2$  and thus  $ab = t^ib = 0$  and also  $t^ia = 0$  as above. With the same arguments it follows that

$$\begin{aligned} \alpha(t^{i+1}) &= \alpha(t)\alpha(t^i) = (xt + ua + b)(x^i t^i) \\ &= x^{i+1} t^{i+1}. \end{aligned}$$

Using Theorem 2.2, it follows that the central aim of this section translates to a determination of the  $\text{Aut}(A_i)$ -orbits of complements to  $N(A_i)$  in  $M(A_i)$ . Note that the action of  $\text{Aut}(A_i)$  on the vector space  $M(A_i)$  is compatible with the standard scalar product of the vector space. Hence the  $\text{Aut}(A_i)$ -orbits of complements to  $N(A_i)$  in  $M(A_i)$  correspond one-to-one to the  $\text{Aut}(A_i)$ -orbits of the spaces orthogonal to these complements.

**Lemma 3.5** *Identify  $M(A_i) \cong \mathbb{F}^4$  with respect to the basis  $\{a^2, at, ta, t^{i+1}\}$ . Then for each complement  $C$  to  $N(A_i)$  in  $M(A_i)$  there exist  $w_1, w_2, w_3 \in \mathbb{F}$  with  $C = \langle c_1, c_2, c_3 \rangle$ , where  $c_1 = (1, 0, 0, -w_1)$ ,  $c_2 = (0, 1, 0, -w_2)$ ,  $c_3 = (0, 0, 1, -w_3)$ . This implies that the orthogonal space  $C^\dagger$  then has the form  $C^\dagger = \langle (w_1, w_2, w_3, 1)^T \rangle$ .*

*Proof* This follows, since

$$\begin{pmatrix} 1 & 0 & 0 & -w_1 \\ 0 & 1 & 0 & -w_2 \\ 0 & 0 & 1 & -w_3 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ 1 \end{pmatrix} = 0.$$

Thus it now remains to determine the  $\text{Aut}(A_i)$ -orbits of vectors  $(w_1, w_2, w_3, 1)^T$  where the vectors have to remain normed from the right. The following is straightforward.

**Lemma 3.6** *Let  $\alpha \in \text{Aut}(A_i)$  as in Lemma 3.3 depending on the parameters  $x, y \in \mathbb{F}^*$  and  $u, v \in \mathbb{F}$ . Write  $s = yx^{-i}$  and  $t = (yuw_1 + vx)x^{-(i+1)}$ . Then  $\alpha$  maps the normed vector  $(w_1, w_2, w_3, 1)^T$  to*

$$\begin{aligned} (w_1, w_2, w_3, 1)^T &\xrightarrow{\alpha} (y^2w_1, yuw_1 + yxw_2 + vx, yuw_1 + yxw_3 + vx, x^{i+1})^T \\ &\xrightarrow{\text{norm}} (s^2x^{2i-(i+1)}w_1, sw_2 + t, sw_3 + t, 1)^T \end{aligned}$$

If  $v$  ranges over  $\mathbb{F}$ , then  $t$  also ranges over  $\mathbb{F}$ . Further if  $y$  ranges over  $\mathbb{F}^*$ , then  $s$  ranges over  $\mathbb{F}^*$ . This allows to determine the following orbit representatives for the action of  $\text{Aut}(A_i)$  on right-normed vectors.

**Theorem 3.7** *Orbit representatives for the  $\text{Aut}(A_i)$ -action on the spaces orthogonal to a complement to  $N(A_i)$  in  $M(A_i)$  are*

$$\begin{aligned} &\langle (0, 1, 0, 1)^T \rangle, \\ &\langle (0, 0, 0, 1)^T \rangle, \\ &\langle (v, 1, 0, 1)^T \rangle, \text{ where } v \text{ lies in a transversal of } U_i \text{ in } \mathbb{F}^*, \\ &\langle (w, 0, 0, 1)^T \rangle, \text{ where } w \text{ lies in a transversal of } SU_i \text{ in } \mathbb{F}^*. \end{aligned}$$

*Proof* Consider the action on right-normed vectors  $(w_1, w_2, w_3, 1)^T$  as in Lemma 3.6. As  $t \in \mathbb{F}$  is arbitrary, we can choose  $w_3 = 0$ . As  $s \in \mathbb{F}^*$  is arbitrary, we can choose  $w_2 \in \{0, 1\}$ . Suppose that  $w_2 = 0$ . Then  $s$  remains arbitrary and we can choose  $w_1$  in a transversal of  $SU_i$  in  $\mathbb{F}^*$  or  $w_1 = 0$ . Suppose that  $w_2 = 1$ . Then  $s$  is fixed and we can choose  $w_1$  in a transversal of  $U_i$  in  $\mathbb{F}^*$  or  $w_1 = 0$ .

### 3.3 Step 3

Steps 1 and 2 can now be used to determine the following classification of isomorphism types algebras of distance 1 from the infinite path in  $\mathcal{G}_{\mathbb{F}}(1)$ .

**Theorem 3.8** *Let  $\mathbb{F}$  be an arbitrary field. Then the nilpotent associative  $\mathbb{F}$ -algebras of distance 1 to the infinite path in  $\mathcal{G}_{\mathbb{F}}(1)$ , i.e. the immediate descendants of  $A_i$  that are not isomorphic to  $A_{i+1}$ , can be described by the following set of parametrized presentations.*

a) For the algebra  $A_1$ :

$$\langle t, a \mid a^2, at, t^2 \rangle,$$

$$\langle t, a \mid a^2, at + ta, t^2 \rangle,$$

$$\langle t, a \mid ua^2 - t^2, at, ta \rangle, \quad \text{with } u \text{ in a transversal of } S \text{ in } \mathbb{F}^*,$$

$$\langle t, a \mid a^2 - ta, at, vta - t^2 \rangle, \quad \text{with } v \in \mathbb{F}^*.$$

b) For the algebras  $A_i$  with  $i \geq 2$ :

$$\langle t, a \mid a^2, at - t^{i+1}, ta, t^{i+2} \rangle,$$

$$\langle t, a \mid a^2 - vt^{i+1}, at - t^{i+1}, ta, t^{i+2} \rangle, \quad \text{with } v \text{ in a transversal of } U_i \text{ in } \mathbb{F}^*,$$

$$\langle t, a \mid a^2 - wt^{i+1}, at, ta, t^{i+2} \rangle, \quad \text{with } w \text{ in a transversal of } SU_i \text{ in } \mathbb{F}^*.$$

*Proof*

a) The orbit representatives determined in Theorem 3.2 translate to subspaces of codimension 1 in  $M(A_1)$ . Taking quotients of  $A_1^*$  by these subspaces (and possibly removing redundant relators) gives algebras with the presentations claimed. Note that the quotient by the subspace corresponding to the orbit representative  $\langle (0, 0, 0, 1) \rangle$  is isomorphic to the mainline algebra  $A_2$  and thus is not listed. (Up to isomorphism this agrees with the algebras one can obtain by restricting to the cclass 1 algebras in the general classification of three-dimensional nilpotent-associative  $\mathbb{F}$ -algebras; see [2] or [8].)

b) The orbit representatives determined in Theorem 3.7 yield complements of  $N(A_i)$  in  $M(A_i)$  as indicated in Lemma 3.5. Taking quotients of the covering algebra  $A_i^*$  by these complements yields algebras with the claimed presentations.

### 3.4 Step 4

It remains to show that the algebras determined in Theorem 3.8 do not have immediate descendants. We consider the case  $i \geq 2$  as a first step. Note that for each of the algebras in Theorem 3.8 b) there exists  $x, y, z \in \mathbb{F}$  so that the algebra has a presentation of the form

$$B_{i+1}(x, y, z) := \langle t, a \mid a^2 - xt^{i+1}, at - yt^{i+1}, ta - zt^{i+1}, t^{i+2} \rangle.$$

First note that  $A_{i+1} \cong B_{i+1}(0, w, w)$  for each  $w \in \mathbb{F}$  via the isomorphism defined by  $t \mapsto t, a \mapsto a - wt^i$ . We show that  $B_{i+1}(x, y, z)$  does not have immediate descendants in all other cases, that is, if  $y \neq z$  or  $x \neq 0$ . To shorten notation we write  $B = B_{i+1}(x, y, z)$ .

First suppose that  $y \neq z$ . Write  $B = F/R$  and thus  $B^* = F/(FR \cup RF)$  for  $F$  free non-unital on 2 generators. Then  $N(B) \leq \langle t^{i+2} \rangle$  and  $yt^{i+2} = tat = zt^{i+2}$ . As  $y \neq z$  it follows that  $t^{i+2} = 0$  in  $B^*$ . Thus  $N(B) = \{0\}$  and  $B$  does not have immediate descendants by Theorem 2.1.

Now suppose that  $x \neq 0$ . Again write  $B = F/R$  and  $B^* = F/(FR \cup RF)$  as above. Then  $xt^{i+2} = ta^2 = zt^{i+1}a = zt^i ta = z^2 t^{2i+1} = 0$ , since  $2i + 1 \geq i + 2$  for  $i \geq 2$  and  $t^{i+2} = 0$ . As  $x \neq 0$  it follows that  $t^{i+2} = 0$  in  $B^*$  and hence  $N(B)$  is zero. Thus  $B$  cannot have any immediate descendants.

The case  $i = 1$  can be verified with similar calculations in the respective covering algebras, again showing that the nucleus is trivial in all cases.

## 4 Experiments for Larger Coclasses

In this section we discuss our computer experiments for the coclass graphs  $\mathcal{G}_{\mathbb{F}}(r)$  with  $r \geq 2$  and small finite fields  $\mathbb{F}$ .

### Coclass 2

In [5, Lemma 13] we showed that the number of maximal coclass trees in  $\mathcal{G}_{\mathbb{F}}(2)$  is  $|\mathbb{F}| + 4$ . We extended this to a very detailed conjecture containing the depth and (minimal) periods of the maximal coclass trees; see [6, Conjecture 5]. Note that coclass 2 is the first coclass for which finite connected components arise.

### Coclass 3

A full conjectural description of  $\mathcal{G}_{\mathbb{F}_2}(3)$  can be found at the website [4]. In Fig. 3 we give a brief overview of the conjectured features of the graphs  $\mathcal{G}_{\mathbb{F}_2}(2)$  (39 maximal coclass trees),  $\mathcal{G}_{\mathbb{F}_3}(2)$  (49 maximal coclass trees) and  $\mathcal{G}_{\mathbb{F}_4}(2)$  (55 maximal coclass trees). The table contains ranks, depths and periods of the maximal coclass trees contained in these graphs.



rank	depth	period	#trees
2	1	1	10
2	2	1	1
2	2	2	2
2	3	1	1
2	3	2	3
3	1	1	19
3	2	1	1
3	2	2	1
4	1	1	1

rank	depth	period	#trees
2	1	1	5
2	1	2	6
2	2	2	2
2	2	3	1
2	2	6	2
2	3	2	1
2	3	3	1
2	3	6	2
3	1	2	25
3	2	2	1
3	2	6	2
4	1	2	1

rank	depth	period	#trees
2	1	1	3
2	1	3	11
2	2	3	1
2	2	6	4
2	3	2	2
2	3	3	1
2	3	12	1
3	1	3	29
3	2	3	1
3	2	6	1
4	1	3	1

Fig. 3 Conjectured features of the maximal coclass trees in  $\mathcal{G}_{\mathbb{F}_2}(3)$ ,  $\mathcal{G}_{\mathbb{F}_3}(3)$  and  $\mathcal{G}_{\mathbb{F}_4}(3)$

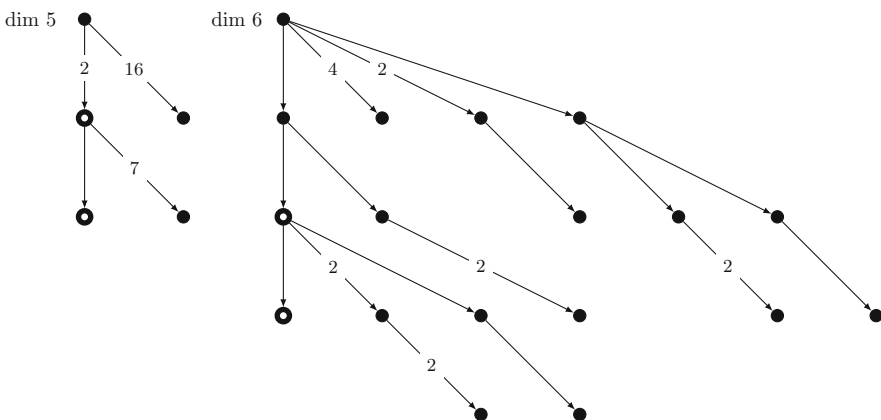


Fig. 4 Left: A maximal descendant tree in  $\mathcal{G}_{\mathbb{F}_2}(3)$  containing two maximal coclass trees. Right: A maximal coclass tree in  $\mathcal{G}_{\mathbb{F}_2}(3)$ , where the depth in the conjectured periodic part is different from the depth in the non-periodic part. The two circles indicate the conjectured periodicity, i.e. the segment between the two circles is repeated infinitely often

Experimentally, as illustrated in Fig. 4, two new phenomena arise in coclass 3:

- There are maximal descendant trees that contain more than one maximal coclass tree. For an explicit example see [6, Section 4].
- There are maximal coclass trees for which the depth in the conjectured periodic part is different from the depth in the non-periodic part at the top.

**Fig. 5** Conjectured features of the maximal coclass trees in  $\mathcal{G}_{\mathbb{F}_2}(4)$

rank	depth	period	#trees
2	1	1	35
2	2	1	7
2	2	2	16
2	3	2	9
2	4	2	4
2	4	4	6
3	1	1	461

rank	depth	period	#trees
3	2	1	50
3	2	2	45
3	3	2	5
4	1	1	69
4	2	1	1
4	2	2	1
5	1	1	1

**Coclass 4**

The coclass graph  $\mathcal{G}_{\mathbb{F}_2}(4)$  conjecturally contains 710 maximal coclass trees. We collect its conjectured features in Fig. 5.

**References**

1. C. Bertram, Ein Algorithmus zur Klassifikation nilpotenter Algebren. Master’s thesis, TU Braunschweig, 2011
2. W.A. de Graaf, Classification of nilpotent associative algebras of small dimension. arXiv:1009.5339 (2010)
3. B. Eick, Computing automorphism groups and testing isomorphism for modular group algebras. J. Algebra **320**(11), 3895–3910 (2008)
4. B. Eick, T. Moede, ccalgs version 1.0 - a GAP4 package (2015). <http://www.icm.tu-bs.de/~tobmoede/ccalgs/>
5. B. Eick, T. Moede, Nilpotent associative algebras and coclass theory. J. Algebra **434**, 249–260 (2015)
6. B. Eick, T. Moede, Coclass theory for finite nilpotent associative algebras: algorithms and a periodicity conjecture. Exp. Math. **26** (2016). <http://dx.doi.org/10.1080/10586458.2016.1162229>
7. B. Eick, E.A. O’Brien, Enumerating  $p$ -groups. J. Aust. Math. Soc. Ser. A **67**(2), 191–205 (1999). Group theory
8. R. Kruse, D. Price, *Nilpotent Rings* (Gordon and Breach, New York, 1969)
9. The GAP Group, GAP – Groups, Algorithms, and Programming, Version 4.7.9 (2015)

# Desingularization of Arithmetic Surfaces: Algorithmic Aspects



Anne Frühbis-Krüger and Stefan Wewers

**Abstract** The quest for regular models of arithmetic surfaces allows different viewpoints and approaches: using valuations or a covering by charts. In this article, we sketch both approaches and then show in a concrete example, how surprisingly beneficial it can be to exploit properties and techniques from both worlds simultaneously.

**Keywords** Arithmetic surfaces • Regular models • Valuations • Desingularization

**Subject Classifications** Primary: 14E15

## 1 Introduction

Resolution of singularities in dimension 2 was first proved by Jung in 1908 [18], but it was not until Hironaka's work in 1964 [17] that this could also be mastered in dimensions beyond 3. However, Hironaka's result only applies to characteristic zero, but not to positive or mixed characteristic. There the general question is still wide open with partial results for low dimensions. In particular, Lipman gave a construction for 2-dimensional schemes in full generality in [19].

Lipman's result includes the case of an *arithmetic surface*, i.e. integral models of curves over number fields. In fact, the existence of (minimal) regular models of curves over number fields is a cornerstone of modern arithmetic geometry. Important

---

A. Frühbis-Krüger (✉)

Institut für Algebraische Geometrie, Leibniz Universität Hannover, Hannover, Germany  
e-mail: [anne@math.uni-hannover.de](mailto:anne@math.uni-hannover.de)

S. Wewers

Institut für Reine Mathematik, Universität Ulm, Ulm, Germany  
e-mail: [stefan.wewers@uni-ulm.de](mailto:stefan.wewers@uni-ulm.de)

early results are for instance the existence of a minimal regular model of an elliptic curve by Néron [25] and Tate's algorithm [32] for computing it explicitly.

In this paper we study a particular series of examples of surface singularities which is a special case of a construction due to Lorenzini [21, 22]. The singularity in question is a *wild quotient singularity*. By this we mean the following: the singular point lies on an arithmetic surface of mixed characteristic  $(0, p)$  which is the quotient of a regular surface by a cyclic group of prime order  $p$ , such that the group action has isolated fixed points. We prove that in our example one obtains a series of rational determinantal singularities of multiplicity  $p$ , and we are able to write down explicit equations for these (see Proposition 3.4).

Determinantal rings (of expected codimension) are well-studied objects in commutative algebra: the free resolution is the Eagon-Northcott complex and hence many invariants of the ring such as projective dimension, depth, Castelnuovo-Mumford regularity, etc. are known (see e.g. [5, 10]). Beyond that, such singularities (in the geometric case) are an active area of current research in singularity theory studying e.g. classification questions, invariants, notions of equivalence and topological properties, see e.g. [12, 26, 34]. We show, by a direct computation, that the resolution in our arithmetic setting is completely analogous to the geometric case.

Both for deriving the equations of our singularities and for resolving them, we employ and mix two rather different approaches to represent and to compute with arithmetic surfaces. The first approach is more standard and consists in representing a surfaces as a finite union of affine charts, and the coordinate ring of each affine chart as a finitely generated algebra over the ground ring. From this point of view, computations with arithmetic surfaces can be performed with standard tools from computer algebra, like standard bases (e.g. in SINGULAR [8]). However, these techniques are not yet as mature in the arithmetic case as they are in the geometric case.

The second approach uses valuations as its main tool. We work over a discrete valuation ring  $R$ . An arithmetic surface  $X$  over  $\text{Spec } R$  is considered as an  $R$ -model of its generic fiber  $X_K$  (a smooth curve over  $K = \text{Frac}(R)$ ). Then any (normal)  $R$ -model  $X$  of  $X_K$  is determined by a finite set  $V(X)$  of discrete valuations on the function field of  $X_K$  corresponding to the irreducible components of the special fiber of  $X$ . A priori, it is not clear how to extract useful information about the model  $X$  from the set  $V(X)$ . Nevertheless, in joint work with J. Rüdth the second named author has used this technique successfully for computing semistable reduction of curves (see e.g. [30]).

The paper is structured as follows. In Sect. 2 we give some general definitions concerning arithmetic surfaces, and we present our two approaches for representing them explicitly. Section 3 then presents our series of wild quotient singularities. In the final section, we compute, in one concrete example of our wild quotient singularities, an explicit desingularization.

We thank Tudor Micu and the referee for careful reading of a previous version and for numerous comments which helped us to improve the article.

## 2 Arithmetic Surfaces and Models of Curves

### 2.1 General Definitions

**Definition 2.1** By a *surface* we mean an integral and noetherian scheme  $X$  of dimension 2. An *arithmetic surface* is a surface  $X$  together with a faithfully flat morphism  $f : X \rightarrow S = \text{Spec}(R)$  of finite type, where  $R$  is a Dedekind domain. To avoid technicalities, we always assume that  $R$  (and hence  $X$ ) is excellent. Moreover, we will assume in addition that  $X$  is normal, unless we explicitly say otherwise.

A common situation where arithmetic surfaces occur is the following. Let  $R$  be a Dedekind domain,  $K = \text{Frac}(R)$  and  $X_K$  a smooth and projective curve over  $K$ . An  $R$ -*model* of  $X_K$  is an arithmetic surface  $X \rightarrow \text{Spec}(R)$ , together with an identification of  $X_K$  with the generic fiber of  $X$ , i.e.  $X_K = X \otimes_R K$ .

For the following discussion we fix an arithmetic surface  $X \rightarrow \text{Spec}(R)$ . We write  $X^{\text{sing}}$  for the subset of points whose local ring is not regular. Since we assume that  $X$  is normal,  $X^{\text{sing}}$  is closed of codimension 2 and hence consists of a finite set of closed points of  $X$ . A point  $\xi \in X^{\text{sing}}$  is called a *singularity* of  $X$ . (If we drop the normality condition, then  $X^{\text{sing}}$  may also have components of codimension 1.)

By a *modification* of  $X$  we mean a proper birational map  $f : X' \rightarrow X$ . A modification is an isomorphism outside a finite set of closed points. If  $f$  is an isomorphism away from a single point  $\xi \in X$ , then  $\xi$  is called the *center* of the modification and  $E := f^{-1}(\xi) \subset X'$  the *exceptional fiber* or *exceptional locus* (we endow  $E$  with the reduced subscheme structure). Note that  $E$  is a connected scheme of dimension one. We will use the notation

$$E = \cup_{i=1}^n C_i,$$

where the  $C_i$  are the irreducible components. Each of them is a projective curve over the residue field  $k = k(\xi)$ . If the modification changes more than a single point, we will still denote the exceptional locus by  $E$ , but  $E$  obviously does not need to be connected any more.

**Definition 2.2** Let  $p : X \rightarrow S$  be an arithmetic surface and  $\xi \in X^{\text{sing}}$  a singularity. A *desingularization* of  $\xi \in X$  is a modification  $f : X' \rightarrow X$  with center  $\xi$  and exceptional fiber  $E = f^{-1}(\xi)$  such that every point  $\xi' \in E$  is a regular point of  $X'$ . A desingularization of  $X$  is a modification consisting of desingularizations at all points of  $X^{\text{sing}}$ .

By a theorem of Lipman [19], a desingularization of  $X$  always exists by means of a sequence of normalizations and blow-ups. Depending on the situation we often want  $f$  to satisfy further conditions. We list some of them:

- (a) The exceptional divisor  $E$  is a normal crossing divisor of  $X'$ .
- (b) Let  $s := p(x)$ . Then the fiber  $X'_s$  of  $X'$  over  $s$  is a normal crossing divisor on  $X'$  (when endowed with the reduced subscheme structure).

- (c) The desingularization  $f : X' \rightarrow X$  is minimal (among all desingularizations of  $\xi \in X$ ).
- (d)  $f : X' \rightarrow X$  is minimal among all desingularizations satisfying (a) (resp. (b)).

Choosing a different approach than Lipman and avoiding normalizations completely, Cossart, Janssen and Saito proved a desingularization algorithm relying only on blow-ups at regular centers in [7], see also [6]. The approach allows us to additionally satisfy yet another rather common condition:

- (e) If  $X \subset W$  for some regular scheme,<sup>1</sup> then desingularization of  $X$  can be achieved by modifications of  $W$  which are isomorphisms outside  $X^{\text{sing}}$ .

## 2.2 Presentation by Affine Charts

We are interested in the problem of computing a desingularization  $f : X' \rightarrow X$  of a given singularity  $\xi \in X$  on an arithmetic surface explicitly. Before we can even state this problem precisely, we have to say something about the way in which the surface  $X$  is represented.

The most obvious way<sup>2</sup> to present  $X$  is to write it as a union of affine charts,

$$X = \cup_{j=1}^r U_j, \quad U_j = \text{Spec } A_j.$$

Here each  $A_j$  is a finitely generated  $R$ -algebra whose fraction field is the function field  $F(X)$  of  $X$ . After choosing a set of generators of  $A_j/R$ , we can obtain a presentation ‘by generators and relations’. This means that

$$A_j = R[\underline{x}]/I_j,$$

where  $\underline{x} = (x_1, \dots, x_{n_j})$  is a set of indeterminates and  $I_j \triangleleft R[\underline{x}]$  is an ideal. Choosing a list of generators of  $I_j$ , we obtain a presentation

$$R[\underline{x}]^{m_j} \rightarrow R[\underline{x}] \rightarrow A_j \rightarrow 0.$$

Taking into account the relations among the generators of the ideal  $I_j$  this presentation extends to

$$R[\underline{x}]^{n_j} \rightarrow R[\underline{x}]^{m_j} \rightarrow R[\underline{x}] \rightarrow A_j \rightarrow 0,$$

where the matrix describing the left-most map is usually referred to as the first syzygy matrix of  $I_j$  or  $A_j$  respectively. Iteratively forming higher syzygies, this leads to free resolutions, i.e. exact sequences of free  $R[\underline{x}]$ -modules. As  $R[\underline{x}]$  is a polynomial

<sup>1</sup>As before  $W$  should be excellent, noetherian, integral.

<sup>2</sup>Thanks to Grothendieck.

ring over a Dedekind domain, it has global dimension  $n_j + 1$  and hence  $A_j$  possesses a free resolution of length at most  $n_j + 1$ . Working locally at a maximal ideal  $\mathfrak{m} \subset R[x]$ , this allows e.g. the calculation of the  $\mathfrak{m}$ -depth of  $A_j$  by the Auslander-Buchsbaum formula.

In the subsequent sections, we shall encounter examples placing us in a particular situation, for which free resolutions are well understood: determinantal varieties corresponding to maximal minors. For these,  $I_j$  is generated by the maximal minors of an  $m \times n$  matrix defining a variety of codimension  $(m - t + 1)(n - t + 1)$ , where  $t = \min\{m, n\}$ . Most prominently, the Hilbert-Burch theorem (see for instance [10]) relates Cohen-Macaulay codimension 2 varieties to the  $t$ -minors of their first syzygy matrix, which is of size  $t \times (t + 1)$ , and ensures the map given by this matrix to be injective.

### 2.3 Presentation Using Valuations

An alternative way<sup>3</sup> to present an arithmetic surface is the following. To describe it is convenient to assume that  $R$  is a local ring. Then  $R$  is actually the valuation ring of a discrete valuation  $v_K : K^\times \rightarrow \mathbb{Q}$  of its fraction field  $K = \text{Frac}(R)$ . We choose a uniformizer  $\pi$  of  $v_K$  (i.e. a generator of the maximal ideal  $\mathfrak{p} \triangleleft R$ ) and normalize  $v_K$  such that  $v_K(\pi) = 1$ . We denote the residue field of  $v_K$  by  $k$ . In addition we make the following assumption<sup>4</sup>:

**Assumption 2.3** The valuation  $v_K$  is either Henselian, or its residue field  $k$  is algebraic over a finite field.

We fix a smooth projective curve  $X_K$  over  $K$ . Note that  $X_K$  is uniquely determined by its function field  $F_X$ , and conversely every finitely generated field extension  $F/K$  of transcendence degree 1 is the function field of a smooth projective curve  $X_K$ .

Let  $X$  be an  $R$ -model of  $X_K$ ,  $X_s$  its special fiber and

$$X_s = \cup_i \bar{X}_i$$

its decomposition into irreducible components. Then each component  $\bar{X}_i$  is a prime divisor on the surface  $X$ . Because  $X$  is normal,  $\bar{X}_i$  gives rise to a discrete valuation  $v_i$  on  $F_X$  such that  $v_i(\pi) > 0$ . We normalize  $v_i$  such that  $v_i(\pi) = 1$ , i.e. such that  $v_i|_K = v_K$ . By definition, the residue field  $k(v_i)$  of  $v_i$  is the function field of the component  $\bar{X}_i$ . In particular,  $k(v_i)$  is a function field over  $k$  of transcendence degree 1.

A discrete valuation  $v$  on the function field  $F_X$  is called *geometric* if  $v|_K = v_K$  and the residue field  $k(v)$  is a finitely generated extension of  $k$  of transcendence

<sup>3</sup>Historically, this was actually the first method, pioneered by Deuring [9] more than 10 years before the invention of schemes.

<sup>4</sup>More generally, we could have assumed that  $(K, v_K)$  satisfies the *local Skolem property*, see [15].

degree 1. Let  $V(F_X)$  denote the set of geometric valuations. Given a model  $X$  of  $X_K$ , we write

$$V(X) := \{v_1, \dots, v_r\} \subset V(F_X)$$

for the set of geometric valuations corresponding to the components of the special fiber of  $X$ .

**Theorem 2.4** *The map*

$$X \mapsto V(X)$$

*is a bijection between the set of isomorphism classes of  $R$ -models of  $X_K$  and the set of finite nonempty subsets of  $V(F_X)$ .*

*Furthermore, given two models  $X, X'$  of  $X_K$ , there exists a map  $X' \rightarrow X$  which is the identity on  $X_K$  (and which is then automatically a modification) if and only if  $V(X) \subset V(X')$ .*

*Proof* See [14] or [28]. □

By the above theorem models of a given smooth projective curve  $X_K$  over a valued field  $(K, v_K)$  can be defined simply by specifying a finite list of valuations. An obvious drawback of this approach is that it is not obvious how to extract detailed information on the model  $X$  from the set  $V(X)$ . A priori,  $V(X)$  only gives ‘birational’ information on the special fiber  $X_s$ . For instance, it is not immediate to see whether the model  $X$  is regular.

So far, the above approach based on valuations has proved to be very useful for the computation of semistable models (see [30]). We intend to extend it to other problems in the future. In Sect. 4.2 we will see a first attempt to use it for desingularization.

## 2.4 Computational Tools

In this section we report on some ongoing work to implement computational tools for dealing with arithmetic surfaces and their desingularization.

### 2.4.1 Valuation Based Approach

As we have explained in Sect. 2.3, it is in principle possible to describe arithmetic surfaces over a local field purely in terms of valuations. In order to use this approach for explicit computations, one needs a way to write down, manipulate and compute with geometric valuations. Fortunately, such methods are available (but maybe not as widely known as they should). Our approach goes back to work of MacLane



[23, 24]. In the present context (i.e. for describing models of curves over local fields) it has been developed systematically in Julian R uth's PhD thesis [28].

We will not go into details, but for later use we need to introduce the notion of an *inductive valuation*. Let  $K$  be a field with a discrete valuation  $v_K$  and valuation ring  $R$  as before. Let  $v$  be an extension of  $v_K$  to a geometric valuation on the rational function field  $K(x)$ . We assume in addition that  $v(x) \geq 0$  (i.e. that  $R[x]$  is contained in the valuation ring of  $v$ ). Let  $\phi \in R[x]$  be a monic integral polynomial, and let  $\lambda \in \mathbb{Q}$  be a rational number satisfying  $\lambda > v(\phi)$ . If  $\phi$  satisfies a technical condition with respect to  $v$  (being a *key polynomial*, see [28, Definition 4.7]) then we can define a new geometric valuation  $v'$  (called an *augmentation* of  $v$ ) with the property that

$$v'(\phi) = \lambda, \quad v'(f) = v(f) \text{ for } f \in K[x] \text{ with } \deg(f) < \deg(\phi).$$

The definition of  $v'$  is easy and explicit: for an arbitrary polynomial  $f \in K[x]$  we compute its  $\phi$ -development

$$f = f_0 + f_1\phi + \dots + f_m\phi^m,$$

where  $\deg(f_i) < \deg(\phi)$ . Then

$$v'(f) := \min_i v(f_i) + i \cdot \lambda.$$

For a rational function  $f/g \in K(x)$  we set  $v'(f/g) := v'(f) - v'(g)$ . The condition that  $\phi$  is a key polynomial for  $v$  then implies that the map  $v' : K(x) \rightarrow \mathbb{Q} \cup \{\infty\}$  defined above is indeed a valuation. See [28, §4], for more details. We write

$$v' = [v, v'(\phi) = \lambda].$$

The process of augmenting a given geometric valuation can be iterated. A geometric valuation  $v$  on  $K(x)$  which is obtained by a sequence of augmentations, starting from the Gauss valuation with respect to  $x$ , is called an *inductive valuation*. It can be written as

$$v = v_n = [v_0, v_1(\phi_1) = \lambda_1, \dots, v_n(\phi_n) = \lambda_n]. \quad (1)$$

Here  $v_0$  is the Gauss valuation,  $\lambda_i \in \mathbb{Q}$  and  $\phi_i \in R[x]$  is monic. Furthermore,  $\phi_i$  is a key polynomial for  $v_{i-1}$  and  $\lambda_i > v_{i-1}(\phi_i)$ . By R uth [28, Theorem 4.31], every geometric valuation  $v$  on  $K(x)$  with  $v(x) \geq 0$  can be written as an inductive valuation.

The notion of inductive valuation can be extended in several ways. Firstly, by replacing  $x$  with  $x^{-1}$  if necessary, we can drop the condition  $v(x) \geq 0$ . Hence we can write every geometric valuation on  $K(x)$  as an inductive valuation. Secondly, for the last augmentation step in (1) we can allow the value  $\lambda_n = \infty$ . The resulting

$v_n$  is then only a *pseudo-valuation* and induces a true valuation on the quotient ring  $L := K[x]/(\phi_n)$  (which is a field because key polynomials are irreducible). Thirdly, given an arbitrary finite extension  $L/K$ , we can compute the (finite) set of extensions  $w$  of  $v_K$  to  $L$  as follows. We write  $L = K[x]/(f)$  for an irreducible polynomial  $f \in K[x]$ . If  $f$  is irreducible over the completion  $\hat{K}$  of  $K$  with respect to  $v_K$ , then there exists a unique extension  $w$  of  $v$  to  $L$  which can be written as an inductive pseudo-valuation on  $K[x]$  (with  $\phi_n = f$ ). In general, let  $f = \prod_i f_i$  be the factorization into irreducibles over  $\hat{K}$ . Then each factor  $f_i$  gives rise to an extension  $w_i$  of  $v$  to  $L$ . Considering  $w_i$  as a pseudo-valuation on  $K[x]$ , MacLane shows that  $w_i$  can be written as a *limit valuation* of a chain of inductive valuations  $v_n$ . By this we mean that  $v_n$  is an augmentation of  $v_{n-1}$ , and for every  $\alpha = (g(x) \bmod (f)) \in L$  there exists  $n \geq 0$  such that  $w_i(\alpha) = v_n(g) = v_{n+1}(g) = \dots$

MacLane's theory is constructive and can be used to implement algorithms for dealing with discrete valuations on a fairly large class of fields. A Sage package written by Julian R uth called `mac_lane` [29] is available under [github.com/saraedum/mac\\_lane](https://github.com/saraedum/mac_lane). It can be used to define and compute with discrete valuations of the following kind:

- $p$ -adic valuations on number fields.
- Geometric valuations  $v$  on function fields  $F/K$  (of dimension 1) whose restriction to  $K$  is either trivial, or can be defined by this package.

Given a valuation  $v$  on a field  $K$  of the above kind and a finite separable extension  $L/K$ , it is possible to compute the set of all extension of  $v$  to  $L$ .

These algorithms are used in a crucial way in [3, 30] and [4].

## 2.4.2 Chart Based Approach

On the other hand, a description by affine charts as in Sect. 2.2 not only emphasizes the similarity to the geometric setting, it also allows the use of computational techniques such as standard bases (whenever a suitably powerful arithmetic for computations in  $R$  is available). This, in turn, opens up a whole portfolio of algorithms ranging from basic functionality like elimination or ideal quotients to more sophisticated algorithms such as blowing up and normalization, which eventually permit to practically implement the above mentioned algorithms of Lipman and of Cossart-Janssen-Saito for desingularization of 2-dimensional schemes. Note at this point that neither of the two algorithms imposes the condition of normality on the surfaces to be resolved.

In a nutshell, the desingularization problem for 2-dimensional schemes is the problem of finding suitable centers which improve the singularity without introducing new complications. In this context, 0-dimensional centers for blow-ups usually do not pose any major problems: such blow-ups at different centers may be interchanged, as they are isomorphisms outside their respective centers and hence do not interact. However, even resolving a 0-dimensional singular point in the

geometric case may already require the use of 1-dimensional centers to achieve a regular model and normal crossing divisors. These curves can exhibit significantly more structure than sets of points, e.g. they can possess intersecting components or non-regular branches. So the central problems in resolving the singularities of 2-dimensional schemes are ensuring improvement in each step and treating 1-dimensional loci which need to be improved. In particular for the latter, the two aforementioned approaches differ significantly.

The key idea behind Lipman’s algorithm [19] is that normal varieties are regular in codimension 1, i.e. that their singular locus is 0-dimensional. Thus a normalization step can always ensure that only sets of points will be required for subsequent blowing up:

**Theorem 2.5 ([19])** *Let  $X$  be an excellent, noetherian, reduced scheme of dimension 2, then  $X$  possesses a desingularization by a finite sequence of birational morphisms of the form*

$$X_r \xrightarrow{\pi_r \circ n_r} \dots \xrightarrow{\pi_2 \circ n_2} X_1 \xrightarrow{\pi_1 \circ n_1} X_0 = X,$$

where  $\pi_i$  denotes a blow up at a finite number of points,  $n_i$  a normalization and  $X_r$  is regular.

While blowing up is algorithmically straightforward e.g. using an elimination (see e.g. [11]), the hard step is the normalization. Although there has been significant improvement in the efficiency of Grauert-Remmert style normalization algorithms in the last decade (see e.g. [2, 16]), this is still a bottleneck when working over a Dedekind domain  $R$  instead of a field. The crucial step here is the choice of a suitable test ideal, i.e. a radical ideal contained in the ideal of the non-normal locus and containing a non-zerodivisor. In the geometric case, the ideal of the singular locus—generated by the original set of generators and the appropriate minors of the Jacobian matrix—is well-suited for this task, but in the current setting it also sees fibre singularities which do not contribute to the non-regular locus. Hence the approximation of the non-normal locus by this test ideal is rather coarse and significantly impedes efficiency. In practice, a better approximation of the non-normal locus is achieved by constructing a test ideal following an idea of Hironaka’s termination criterion: we use the locus where Hironaka’s invariant  $\nu^*$ , i.e. the tuple of orders (in the sense of orders of power series) of the elements of a local standard basis, sorted by increasing order, is lexicographically greater than a tuple of ones.

The approach of Cossart-Janssen-Saito [7] (CJS for short) on the other hand, avoids normalization completely and allows well-chosen 1-dimensional centers, whenever necessary; when choosing centers, it takes into account the full history of blowing ups leading to the current situation. In contrast to Lipman’s approach, this algorithm yields an embedded desingularization. Nevertheless, a key step is again the use of the locus where  $\nu^*$  lexicographically exceeds a tuple of ones. But then, no normalization follows, instead the singularities of this locus are first resolved before it is itself used as a 1-dimensional center. Each arising exceptional curve in this process remembers when it was created and whether its center was

of dimension 0 or 1, because this information is crucial in the choice of center for ensuring improvement as well as normal crossing of exceptional curves.

A beta version of the first algorithm is available as SINGULAR-library `reslip-man.lib` and is planned to become part of the distribution in the near future. A prototype implementation of the CJS-algorithm has been implemented and is closely related to an ongoing PhD-project on a parallel approach to resolution of singularities using the `gpi-space` parallelization environment (for recent progress along this train of thought see [1, 27]).

### 3 Explicit Construction of Wild Quotient Singularities

In this section we describe a series of examples for arithmetic surfaces with interesting singularities. The general construction is due to Lorenzini (see [21] and [22]). Our contribution is to explicitly describe the (local) ring of the singularity by generators and relations. In the next section we also describe the desingularization in an equally explicit way.

Let  $R$  be a discrete valuation ring, with maximal ideal  $\mathfrak{p}$ , residue field  $k = R/\mathfrak{p}$  and fraction field  $K$ . Let  $v_K$  denote the corresponding discrete valuation on  $K$ . We assume that  $k$  has positive characteristic  $p$  and that  $v_K$  is Henselian (in particular, Assumption 2.3 holds).

Let  $X_K$  be a smooth, projective and absolutely irreducible curve over  $K$ , of genus  $g$ . We assume that  $X_K$  has potentially good reduction with respect to  $v_K$ . This means that there exists a finite extension  $L/K$  and a smooth model  $Y$  of  $X_L := X_K \otimes_K L$  over the integral closure  $R_L$  of  $R$  in  $L$ . Note that  $R_L$  is a discrete valuation ring corresponding to the unique extension  $v_L$  of  $v_K$  to  $L$ . We assume in addition that  $L/K$  is a Galois extension, and that the natural action of  $G := \text{Gal}(L/K)$  on  $X_L$  extends to an action on  $Y$ . Under this assumption, we can form the quotient scheme  $X_Y/G$ . It is an  $R$ -model of  $X_K$ .

The model  $Y$  is regular because  $Y \rightarrow \text{Spec}(R)$  is smooth by assumption. However, the quotient scheme  $X = Y/G$  may have singularities. In fact, let  $\xi \in X_s$  be a closed point on the special fiber of  $X$ , and let  $\eta \in Y_s$  be a point above  $\xi$ . Let  $I_\eta \subset G$  denote the inertia subgroup of  $\eta$  in  $G$ . If  $I_\eta = 1$  then the map  $Y \rightarrow X$  is étale in  $\eta$ . It follows that  $X$  is regular in  $\xi$  because  $Y$  is regular in  $\eta$ .

In general, the locus of points with  $I_\eta \neq 1$  may consist of the entire closed fiber  $Y_s$  and hence be a subset of codimension 1 on  $Y$ . To obtain isolated quotient singularities we impose the following condition:

**Assumption 3.1** The action of  $G$  on the special fiber  $Y_s$  is generically free.

Under this assumption, there are at most a finite number of points  $\eta \in Y_s$  with nontrivial inertia  $I_\eta \neq 1$ . Let  $\xi_1, \dots, \xi_r \in X_s$  be the images of the points  $\eta \in Y_s$  with  $I_\eta \neq 1$ . Then  $\xi_1, \dots, \xi_r$  are precisely the singularities of the model  $X$ .

*Remark 3.2* In Lorenzini’s original setting, Assumption 3.1 holds automatically because the curve  $Y$  has genus  $g(Y) \geq 2$ . In our series of examples we have  $g(Y) = 0$ , but the assumption holds nevertheless.

### 3.1 An Explicit Example

Let  $p$  be a prime number and  $K$  a number field. We denote by  $\mathcal{O}_K$  the ring of integers of  $K$  and fix a prime ideal  $\mathfrak{p} \triangleleft \mathcal{O}_K$  lying over  $p$  (i.e. such that  $p \in \mathfrak{p}$ ). Let  $v_K$  denote the discrete valuation on  $K$  corresponding to  $\mathfrak{p}$  and  $R$  the valuation ring of  $v_K$ . Let  $L/K$  be a Galois extension of degree  $p$  which is totally ramified at  $\mathfrak{p}$ . This means that  $v_K$  has a unique extension  $v_L$  to  $L$ . Let  $\sigma$  be a generator of the cyclic group  $G = \text{Gal}(L/K)$ . Let  $\pi_L$  be a uniformizer for  $v_L$ . We normalize  $v_L$  such that  $v_L(\pi_L) = 1/p$ . Then  $v_L|_K = v_K$ . Set

$$m := p \cdot v_L(\sigma(\pi_L) - \pi_L).$$

Then  $m \geq 2$  is the first and only break in the filtration of  $G$  by higher ramification groups. We let  $u \in k^\times$  denote the image of the element  $(\sigma(\pi_L) - \pi_L)/\pi_L^m \in R^\times$ .

Let  $X_K := \mathbb{P}_K^1$  be the projective line over  $K$ . We identify the function field  $F_X$  with the rational function field  $K(x)$  in the indeterminate  $x$ . Then  $L(x)$  is the function field of  $X_L = \mathbb{P}_L^1$ . We define an element

$$y := \frac{x - \pi_L}{\pi_L^m} \in L(x).$$

Clearly,  $L(x) = L(y)$ , and so  $y$ , considered as a rational function on  $X_L$ , gives rise to an isomorphism  $X_L \cong \mathbb{P}_L^1$ . We let  $Y$  denote the smooth  $R_L$ -model of  $X_L$  such that  $y$  extends to an isomorphism  $Y \cong \mathbb{P}_{R_L}^1$ . By an easy calculation we see that  $\sigma(y) = ay + b$ , with  $a \in R_L^\times$  and  $b \in R_L$ . Furthermore,

$$\sigma(y) \equiv y + u \pmod{\pi_L}.$$

In geometric terms this means that the action of  $G$  on  $X_L$  extends to the smooth model  $Y$ , and that the restriction of this action to the special fiber  $Y_s \cong \mathbb{P}_k^1$  is generically free (and hence Assumption 3.1 holds). In fact, the action of  $G$  is fix point free on the affine line  $\text{Spec } k[y]$ , and if  $\eta \in Y_s$  denotes the point corresponding to  $y = \infty$  then  $I_\eta = G$ .

Let  $\xi \in X_s$  denote the image of  $\eta$ . By construction,  $\xi$  is a wild quotient singularity (see the introduction, p. 232), and it is the only singular point on  $X$ . Our goal is to write down explicitly an affine chart  $U = \text{Spec } A \subset X$  containing  $\xi$ .

To state our result we need some more notation. Let  $\phi \in K[x]$  denote the minimal polynomial of  $\pi_L$  over  $K$ . Then

$$\phi = x^p + \sum_{i=0}^{p-1} a_i x^i = \prod_{k=0}^{p-1} (x - \sigma^k(\pi_L)),$$

where  $a_0, \dots, a_{p-1} \in \mathfrak{p}$ . The constant coefficient

$$\pi_K := a_0 = N_{L/K}(\pi_L)$$

is actually a prime element of  $R$ , i.e.  $\phi$  is an Eisenstein polynomial.

The following lemma gives a characterization of the model  $X$  in terms of the set  $V(X)$  of valuations corresponding to the irreducible components of the special fiber (as in Theorem 2.4).

**Lemma 3.3** *We have*

$$V(X) = \{v\}$$

where  $v$  is the inductive valuation on  $K(x)$  extending  $v_K$  given by

$$v := [v_0, v_1(x) = 1/p, v_2(\phi) = m].$$

(See Sect. 2.3 and (1) for the relevant notation.)

*Proof* It is clear that  $V(Y) = \{w\}$ , where  $w$  is the Gauss valuation on  $F(X_L) = L(y)$  with respect to the parameter  $y$  and the valuation  $v_L$ . Since  $Y \rightarrow X = Y/G$  is a finite morphism between (normal) models of their respective generic fibers, we have  $V(X) = \{v\}$ , where  $v$  is the restriction of  $w$  to the subfield  $F(X_K) = K(x) \subset F(X_L) = L(y)$ . It remains to identify  $v$  with the inductive valuation given in the statement of the lemma.

We will use the characterization of an inductive valuation which is implicit in [28, §4.4]. Let  $v'$  be a valuation on  $K(x)$  which extends  $v_K$  and satisfies

$$v'(x) \geq 0, \quad v'(\phi) \geq m.$$

Then we claim that  $v(f) \leq v'(f)$  for all  $f \in K[x]$ . By Rüth [28, Theorem 4.56], the claim implies that

$$v = [v_0, v_1(x) = 1/p, v_2(\phi) = m].$$

To prove the claim, we choose an extension  $w'$  of  $v'$  to the overfield  $L(y)$ . Then

$$m \leq v'(\phi) = \sum_{i=0}^{p-1} w'(x - \sigma^i(\pi_L)) = \sum_{i=0}^{p-1} w'(\pi_L^m y + \pi_L - \sigma^i(\pi_L)). \tag{2}$$

By definition we have

$$w'(\pi_L) = v_L(\pi_L) = 1/p, \quad w'(\pi_L - \sigma^i(\pi_L)) = v_L(\pi_L - \sigma^i(\pi_L)) \geq m/p. \quad (3)$$

Combining (2), (3) and the strong triangle inequality we conclude that  $w'(y) \geq 0$ . The valuation  $w$  being the Gauss valuation with respect to  $y$  and  $v_L$  this implies  $w(f) \leq w'(f)$  for all  $f \in L[y]$ . But  $K[x] \subset L[y]$ , and therefore  $v(f) \leq v'(f)$  for all  $f \in K[x]$ . This proves the claim and also the lemma.  $\square$

Let  $D_K \subset X_K$  be the divisor of zeroes of  $\phi$ , and let  $D \subset X$  be the closure of  $D_K$ . Let  $U := X - D$  denote the complement.

**Proposition 3.4**

1. We have  $U = \text{Spec } A$ , where  $A \subset F_X = K(x)$  is the sub- $R$ -algebra generated by the elements  $x_0, \dots, x_{p-1}$ , where

$$x_i := \pi_K^m x^i \phi^{-1}, \quad i = 0, \dots, p - 1.$$

The point  $\xi$  lies on  $U$  and corresponds to the maximal ideal

$$\mathfrak{m} := (\pi_K, x_0, \dots, x_{p-1}) \triangleleft A.$$

2. The ideal of relations between the generators  $x_0, \dots, x_{p-1}$  is generated by the  $2 \times 2$  minors of the matrix

$$M := \begin{pmatrix} x_0 & x_1 \\ x_1 & x_2 \\ \vdots & \vdots \\ x_{p-2} & x_{p-1} \\ x_{p-1} & z \end{pmatrix}, \quad \text{with } z := \pi_K^m - \sum_{i=0}^{p-1} a_i x_i.$$

*Proof* It follows from [20, Corollary 5.3.24], that the divisor  $D \subset X$  is ample, and hence  $U := X - D = \text{Spec}(A)$  is affine. Since  $X$  is normal, the ring  $A$  consists precisely of all rational functions  $f \in K(x)$  with  $\text{ord}_Z(f) \geq 0$ , for any prime divisor  $Z \subset X$  distinct from  $D$ .

A prime divisor  $Z \subset X$  is either horizontal (i.e. the closure of a closed point on  $X_K$ ) or equal to  $X_s$ . By Lemma 3.3,  $X_s$  is a prime divisor with corresponding valuation  $v$  on  $K(x)$ . It follows that

$$A = \{f \in A_K \mid v(f) \geq 0\},$$

where

$$A_K = K[\phi^{-1}, x\phi^{-1}, \dots, x^{p-1}\phi^{-1}].$$

In order to make the condition  $v(f) \geq 0$  more explicit, we write  $f \in A_K$  in the form

$$f = c_0 + \sum_{i=0}^{r-1} \sum_{j=0}^{p-1} c_{i,j} x^j \phi^{i-r},$$

with  $c_0, c_{i,j} \in K$ . Then Lemma 3.3 shows that

$$v(f) = \min\{v_K(c_0), v_K(c_{i,j}) + j/p - m(r - i)\}.$$

So the condition  $v(f) \geq 0$  is equivalent to

$$v_K(c_{i,j}) + j/p \geq m(r - i),$$

for  $i = 0, \dots, r - 1$  and  $j = 0, \dots, p - 1$ . It follows that

$$A = R[x_0, \dots, x_{p-1}], \quad \text{where } x_j := \pi_K^m x^j \phi^{-1}.$$

This is the first part of Statement (i); the second part is obvious.

To prove Statement (ii) we let  $I$  be the ideal in the polynomial ring  $R[x] = R[x_0, \dots, x_{p-1}]$  generated by the  $2 \times 2$ -minors of the matrix  $M$ . It is easy to check that the generators of  $A$  satisfy these relations. Therefore, we have a surjective map  $A' := R[x_0, \dots, x_{p-1}]/I \rightarrow A$ . We want to prove that  $A' = A$ .

Let  $A'' := A'[x_0^{-1}]$  and consider the matrix  $M$  with entries in  $A''$ . By definition we have  $\text{rk} M \leq 1$ , and the upper left entry  $x_0$  is a unit. An elementary argument shows that there exists  $t \in A''$  such that

$$x_0 \phi(t) = \pi_K^m, \quad x_i = t^i x_0, \quad i = 1, \dots, p - 1.$$

It follows that

$$A'' = R[x_0, x_0^{-1}, t \mid x_0 \phi(t) = \pi_K^m].$$

In particular  $A''/R[x_0, x_0^{-1}]$  is a finite flat and generically étale extension of degree  $p$ . We deduce that  $A''$  is an integral domain of dimension 2. Looking at the equations defining  $A'$ , it is easy to see that

$$(x_0)^{\text{rad}} = (x_0, \dots, x_{p-1})$$

and that  $A'/(x_0)^{\text{rad}} \cong R$  has dimension 1. Together with  $\dim A'' = 2$  this implies that  $\dim A' = 2$ . Therefore,  $A'$  is a determinantal ring of the ‘expected’ codimension  $(p - 2 + 1)(2 - 2 + 1) = p - 1$ . Now a theorem of Eagon and Hochster shows that  $A'$  is Cohen-Macaulay (see [10, Theorem 18.18] for a textbook reference). Every associated prime of a Cohen-Macaulay ring is minimal [10, Corollary 18.10]. Since  $A'' = A'[x_0^{-1}]$  is an integral domain, it follows that  $A'$  is an integral domain as well.



The analysis of  $A''$  from above also shows that

$$A''_K = A'_K[x_0^{-1}] = A_K[x_0^{-1}] = K[x, \phi^{-1}].$$

It follows that  $J = \ker(A' \rightarrow A)$  is an ideal of codimension  $\geq 1$ . But  $A, A'$  have the same dimension, so  $J$  consists of zero divisors. On the other hand, we have shown above that  $A'$  is an integral domain. Hence  $J = 0$ . This completes the proof of Proposition 3.4.  $\square$

*Example 3.5* The simplest special case of Proposition 3.4 where the resulting singularity is not a complete intersection is for  $p = 3$ . To make this even more explicit, we set  $K := \mathbb{Q}$  and let  $v_K$  denote the 3-adic valuation on  $K$  and  $R := \mathbb{Z}_{(3)}$  the valuation ring (the localization of  $\mathbb{Z}$  at 3). Moreover, we set

$$\phi := x^3 - 3x^2 + 3.$$

The splitting field  $L/K$  of  $\phi$  is a Galois extension of degree 3 which is totally ramified at  $p = 3$ . Indeed, we can factor  $\phi$  as

$$\phi = (x - \pi)(x - \sigma(\pi))(x - \sigma^2(\pi)) = (x - \pi)(x - \pi - \pi^2 + 3\pi)(x - \pi + \pi^2 - 3),$$

where  $\pi$  is prime elements for the unique extension  $v_L$  of  $v_K$  to  $L$ . We see that

$$m := 3 \cdot v_L(\pi - \sigma(\pi)) = 2.$$

The resulting singularity  $\xi$  of the model  $X$  of  $X_K = \mathbb{P}_K^1$  constructed above is a rational triple point.

*Remark 3.6* The generic fiber  $X_K$  of our model  $X$  is a curve of genus zero and so is not, strictly speaking, an example of the situation studied by Lorenzini. But we can easily modify our construction to get examples with arbitrary high genus. For instance, choose  $m > 1$ ,  $p \nmid m$  and consider the Kummer cover  $Y_K \rightarrow X_K$  of smooth projective curves with generic equation

$$Y_K : y^m = \phi(x).$$

Then  $g(Y_K) \geq 2$  (except for  $p = 3$  and  $m = 2$  when  $g(Y_K) = 1$ ). Let  $Y$  denote the normalization of the  $R$ -model  $X$  inside the function field of  $Y_K$ . Then  $Y$  is a (normal)  $R$ -model of  $Y_K$ . It can easily be shown that  $Y$  has a unique singular point  $\eta$  (which is the unique point in the inverse image of  $\xi \in X$ ), and that  $\eta \in Y$  is a wild quotient singularity in the sense of [22]. We intend to study this situation in a subsequent paper.

## 4 An Explicit Resolution

To keep the construction of a desingularization in an explicit example as concise as possible we now focus on the specific Example 3.5. This case already illustrates the general situation quite well, but is still sufficiently small to avoid lengthy explicit computations.

Set  $K := \mathbb{Q}$  and let  $v_K$  denote the 3-adic valuation on  $K$  and  $R := \mathbb{Z}_{(3)}$  the valuation ring (the localization of  $\mathbb{Z}$  at 3). Let  $v_0$  denote the Gauss valuation on  $K(x)$  with respect to  $x$ . We define an inductive valuation  $v$  on  $K(x)$  as follows:

$$v := [v_0, v(x) = 1/3, v(x^3 - 3x^2 + 3) = 2].$$

Let  $X$  be the model of  $X_K := \mathbb{P}_K^1$  with  $V(X) = \{v\}$ . We have shown in the preceding section that  $X$  has a unique singularity  $\xi$  with an affine open neighborhood  $U = \text{Spec } A$ , where

$$A = R[x, y, z]/I,$$

and where  $I$  is the ideal generated by the 2-minors of the matrix

$$M = \begin{pmatrix} x & y \\ y & z \\ z & 3x - 3z - 9 \end{pmatrix}.$$

The singular point  $\xi$  corresponds to the maximal ideal  $\mathfrak{m} = (3, x, y, z) \triangleleft A$ .

### 4.1 Explicit Blowups and Tjurina Modifications

Our goal is to construct explicitly a desingularization  $f : X' \rightarrow X$  of  $\xi$ . For ease of notation we replace the projective scheme  $X$  by the affine open subset  $U = \text{Spec } A$ .

We not only know that  $A$  is Cohen-Macaulay of codimension 2, we are in an even better setting, the situation of the Hilbert-Burch theorem, which then implies that a free resolution of  $A$  is of the form

$$0 \longrightarrow R[x, y, z]^2 \xrightarrow{M} R[x, y, z]^3 \longrightarrow R[x, y, z] \longrightarrow A \longrightarrow 0,$$

i.e. the Eagon-Northcott complex of  $M$ .

At first glance this seems to be unrelated to our task of desingularizing  $A$ . However, these structural observations point us to well known results in the complex geometric case: In the late 1960s, Galina Tjurina classified the rational triple point singularities over the complex numbers in [33] and constructed minimal desingularizations thereof in a direct way. Our given matrix  $M$  structurally corresponds to a

singularity of type  $H_5$  in Tjurina’s article, which we will refer to as  $Y$  here and for which a presentation matrix (over  $\mathbb{C}[x, y, z, w]$ ) is of the form

$$N = \begin{pmatrix} x & y \\ y & z \\ z & wx - w^2 \end{pmatrix}.$$

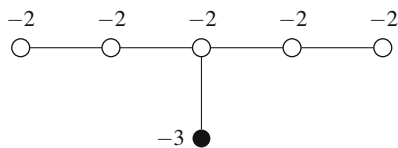
The last entry can be replaced by  $wx - wz - w^2$  without changing the analytic type of the singularity as is shown in the classification of simple Cohen-Macaulay codimension 2 singularities in [12]. This similarity suggests to try and mimic the philosophy of Tjurina’s choice of centers for the desingularization of  $X$ .

Tjurina’s first step towards a resolution of singularities is nowadays called a Tjurina modification and is based on the observation that at each point of  $Y$  except the origin the row space of the presentation matrix defines a unique direction in  $\mathbb{C}^2$  and hence a point in the Grassmanian of lines in 2-space. Resolving indeterminacies of this rational map into the Grassmanian then yields the Tjurina transform which can then be described by the equations

$$N \cdot \begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

(For a more detailed treatment of Tjurina modifications see the first section of [13].) Three further blow-ups, each at the (0-dimensional) singular locus, which happens to be the non-normal crossing locus of the exceptional curves in the second and third blow-up, then lead Tjurina to a desingularization. The exceptional locus of this sequence of blow-ups consists of six curves of genus zero, where the one originating from the Tjurina modification is the only one with self-intersection  $-3$ ; all others have self-intersection  $-2$ . The dual graph of the resolution is of the form (Fig. 1):

**Fig. 1** Tjurina’s intersection graph  $H_5$



Returning to our setting, we can mimic these steps, obtaining the following as ideal of the Tjurina transform:

$$I_{X_1} = \langle sx - ty, sy - tz, sz - t(3z - 3x - 9) \rangle$$

By direct computation, it is easy to see that  $X_1$  is regular except above 3 and that above 3 the non-regular locus is contained in the chart  $t \neq 0$ . The exceptional curve  $C_0$  which arose in this blow-up is a  $\mathbb{P}^1$  and corresponds to the ideal  $\langle x, y, z, 3 \rangle$ .

Passing to the chart  $t \neq 0$ , we can harmlessly eliminate the variables  $y$  and  $z$  according to the first two generators. This essentially leaves a hypersurface described by the ideal

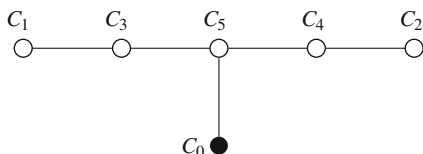
$$I_{X_{1,\text{new}}} = \langle s^3x - 3s^2x + 3x + 9 \rangle \subset R[x, s]$$

and an exceptional curve  $I_{C_0} = \langle x, 3 \rangle$ . The non-regular locus of this hypersurface corresponds to  $\langle x, s, 3 \rangle$  as a direct computation shows; this is the center of the upcoming blow-up, which leads to three charts, two of which only contain regular points and only see normal crossing divisors. In the remaining chart ( $y_1 \neq 0$ ), the strict transform is given by

$$I_{X_2} = \langle 3 - y_2s, s^2y_0 - s^2y_0y_2 + y_0y_2 + y_2^2 \rangle,$$

the strict transform of the exceptional curve  $C_0$  by  $\langle 3, y_0, y_2 \rangle$  and the two components  $C_1$  and  $C_2$  of the new exceptional curve  $E_2$  by  $\langle 3, s, y_2(y_0 + y_2) \rangle$ . As the non-regular locus is given by  $\langle 3, s, y_0, y_2 \rangle$  and the non-normal crossing locus of the exceptional curves is the same point, analogous to Tjurina’s setting, this point has to be chosen as upcoming center. After blowing up this point of  $X_2$ , we see in one chart that each of the two components  $C_1$  and  $C_2$  of the preceding exceptional curve  $E_2$  meets one component of the new exceptional curve  $E_3$ ; more precisely,  $C_1$  meets  $C_3$  and  $C_2$  meets  $C_4$ . In another chart, we see that the transform of  $C_0$  meets both  $C_3$  and  $C_4$  at the origin, which is also the only singular point. Blowing up this point then introduces yet another exceptional curve  $C_5$  meeting  $C_0$ ,  $C_3$  and  $C_4$ ; at this stage, the strict transform is regular and the exceptional divisor is normal crossing. All exceptional curves are  $-2$ -curves except the  $-3$ -curve  $C_0$ . Hence we obtained the dual graph:

**Fig. 2** The intersection graph of the desingularization of  $X$



An explicit comparison of the computations of Tjurina and of the one presented in our setting shows that all computational steps as well as the final result are analogous in both cases. This certainly raises the question whether other singularities from Tjurina’s list also have an analogue arising from the construction of Sect. 3 and what geometric properties the singularities corresponding to the matrices of the previous section might exhibit.

*Remark 4.1*

1. In the above calculation, we saw that we could safely replace the matrix  $N$ , which is the normal form in the classification of simple Cohen-Macaulay codimension

2 singularities [12], by a matrix say  $N'$  which directly corresponds to the original matrix  $M$ , differing only by using a variable  $w$  instead of  $\pi_k = 3$ . The isomorphism of the local rings of the singularities represented by  $N$  and  $N'$  does not involve any change of  $w$ , whence we could hope for an equivalent isomorphism for  $M$ . This, however, does not exist, as the isomorphism over  $\mathbb{C}$  involves the multiplicative inverse of 3.

2. As in the explicit example here, all the determinantal singularities from Proposition 3.4 allow a Tjurina modification at the origin of the respective chart at the beginning of the desingularization; this provides an exceptional curve  $C_0$ . After this step, we see only one singular point, an  $A_{pm-1}$  singularity. This latter singularity is well known to have a dual graph of resolution which is a chain with  $pm - 1$  vertices and  $pm - 2$  edges, where the middle vertex corresponds to the youngest exceptional curve. At this middle vertex we additionally find the connecting edge to the vertex  $C_0$  originating from the Tjurina transform.

### 4.2 A Posteriori Description via Valuations

We return to our original notation, i.e.  $X$  denotes the  $R$ -model of  $X_K = \mathbb{P}_K^1$  with  $V(X) = \{v\}$  (and not its affine subset  $\text{Spec}A$ ). Also,  $x$  again denotes the original coordinate function on  $X_K$ .

The computation of the previous section shows that there exists a desingularization  $f : X' \rightarrow X$  of  $\xi$  such that the exceptional fiber  $E := f^{-1}(\xi)$  is a normal crossing divisor and consists of 6 smooth rational curves, with an intersection graph given in Fig. 2. The arithmetic surface  $X'$  is itself an  $R$ -model of  $X_K$  and is hence completely determined by the set  $V(X')$  of geometric valuations of  $K(x)$  corresponding to the irreducible components of the special fiber  $X'_s$ . But  $X'_s$  consists precisely of the strict transform  $C_6$  of  $X_s$  (which corresponds to the valuation  $v_6 := v$ ) and the 6 components  $C_0, \dots, C_5$  of the exceptional divisor.

The obvious question is: what are the valuations corresponding to the components  $C_i, i = 0, \dots, 5$ ?

**Proposition 4.2** *Let  $v_i$  denote the valuation on  $K(x)$  corresponding to the component  $C_i$ , for  $i = 0, \dots, 5$ . We normalize  $v_i$  such that  $v_i(3) = 1$  (i.e. such that  $v_i|_K = v_K$ ). Then  $v_0$  is the Gauss valuation with respect to the coordinate  $x$ . For  $i = 1, 3, 5$ ,*

$$v_i = [v_0, v_i(x) = r_i], \quad r_i = \begin{cases} 1/3, & i = 5, \\ 1/2, & i = 3, \\ 1, & i = 1. \end{cases}$$

For  $i = 2, 4$  we have

$$v_i = [v_0, v_i(x) = 1/3, v_i(\phi) = s_i], \quad s_i = \begin{cases} 4/3, & i = 4, \\ 5/3, & i = 2. \end{cases}$$

*Proof* This can be checked by a direct (but somewhat involved) computation, using the explicit description of the desingularization by affine charts in Sect. 4.1. As an illustration of the general method let us convince ourselves that the Gauss valuation  $v_0$  corresponds to the component  $C_0$ .

It suffices to consider the first step of the desingularization, the Tyurina modification  $X_1 \rightarrow X$ . We use the notation from p. 247. The affine chart of  $X_1$  defined by  $t \neq 0$  has the form

$$\text{Spec } R[x_0, s \mid s^3 x_0 - 3s^2 x_0 + 3x_0 + 9 = 0]$$

and the exceptional divisor  $E_1 \subset X_1$  is given on this chart by  $I_{E_1} = (x_0, 3)$ . So  $\text{Spec } \mathbb{F}_3[s]$  is an affine open of  $E_1$ , and hence  $E_1$  is a projective line. We claim that  $E_1$ , as a prime divisor on  $X$ , gives rise to the valuation  $v_0$  (the Gauss valuation with respect to  $x$ ).

We write  $x_0, s$  as rational functions in  $x$ :

$$x_0 = 9\phi^{-1}, \quad s = \frac{x_1}{x_0} = x.$$

Now we see that the generators of the ideal  $I_{E_1}$  have positive valuation ( $v_0(3) = 1$ ,  $v_0(x_0) = 2$ ) and  $s$  is a  $v_0$ -unit and is a generator of its residue field. This shows that the prime divisor  $E_1 \subset X_1$  corresponds to the valuation  $v_0$ . As the component  $C_0$  of the desingularization  $X' \rightarrow X$  is simply the strict transform of  $E_1$  under the map  $X' \rightarrow X_1$ , we have proved the proposition for  $i = 0$ . For  $i = 1, \dots, 5$  one can proceed in a similar way.  $\square$

*Remark 4.3*

1. We have found the set  $V(X') = \{v_0, \dots, v_6\}$  after computing the desingularization  $X' \rightarrow X$ . By Theorem 2.4,  $X'$  is determined by  $V(X')$ . Could we have found  $V(X')$  by some other method, and would this give an alternative way to compute desingularization? In this simple case it is indeed possible to check the regularity of  $X'$  (and the fact that  $X'_s$  is a normal crossing divisor) purely in terms of the set of valuations  $\{v_0, \dots, v_6\}$ .
2. If we accept that  $X'$  is regular and  $X'_s$  is a normal crossing divisor, it is easy to compute the self intersection numbers of the irreducible components  $C_i$ , as follows. Let

$$\tilde{E} := (3) = \sum_{i=0}^6 m_i C_i \in \text{Div}(X)$$

be the principal divisor of the prime 3. For each  $i$  the integer  $m_i$  (the *multiplicity* of the component  $C_i$ ) is equal to the ramification index of the extension  $K(x)/K$  with respect to  $v_i$ . It is easy to read off  $m_i$  from the explicit description of the  $v_i$  in Proposition 4.2:

$$m_0 = 1, m_1 = 1, m_2 = 3, m_3 = 2, m_4 = 3, m_5 = 3, m_6 = 3.$$

Since  $\tilde{E}$  is a principal divisor, we have

$$0 = (C_i \cdot \tilde{E}) = \sum_{j=0}^6 m_j (C_i \cdot C_j),$$

for  $i = 0, \dots, 6$ , see e.g. [31, §IV.7]. The component graph from Fig. 2 tells us what  $(C_i \cdot C_j)$  is for  $i \neq j$  (either 1 or 0). Now the self intersection numbers  $(C_i \cdot C_i)$  can be computed easily. We find that

$$(C_i \cdot C_i) = \begin{cases} -3, & i = 0, \\ -2, & i = 1, \dots, 5, \\ -1, & i = 6. \end{cases}$$

## References

1. J. Böhm, A. Frühbis-Krüger, A smoothness test for higher codimension. *J. Symb. Comput.* **86**, 153–165 (2018)
2. J. Böhm, W. Decker, S. Laplagne, G. Pfister, A. Steenpaß, S. Steidel, Parallel algorithms for normalization. *J. Symbolic Comput.* **51**, 99–114 (2013)
3. M. Börner, I. Bouw, S. Wewers, The functional equation for l-functions of hyperelliptic curves. *Exp. Math.* 1–16 (2016). <http://dx.doi.org/10.1080/10586458.2016.1189860>
4. M. Börner, I. Bouw, S. Wewers, Picard curves with small conductor (2017). Preprint, arXiv:1701.01986
5. W. Bruns, U. Vetter, *Determinantal Rings*. Lecture Notes in Mathematics, vol. 1327 (Springer, New York, 1988)
6. V. Cossart, B. Schober, A strictly decreasing invariant for resolution of singularities in dimension two. Preprint, arXiv:1411.4452
7. V. Cossart, U. Jannsen, S. Saito, Canonical embedded and non-embedded resolution of singularities for excellent two-dimensional schemes (2009). Preprint, arXiv:0905.2191
8. W. Decker, G.M. Greuel, G. Pfister, H. Schönemann, SINGULAR 4-1-0 — A computer algebra system for polynomial computations (2016). <http://www.singular.uni-kl.de>
9. M. Deuring, Reduktion algebraischer Funktionenkörper nach Primdivisoren des Konstantenkörpers. *Math. Z.* **47**(1), 643–654 (1942)
10. D. Eisenbud, *Commutative Algebra with a View Towards Algebraic Geometry* (Springer, New York, 1995)
11. A. Frühbis-Krüger, Computational aspects of singularities, in *Singularities in Geometry and Topology*, ed. by J. Brasselet, J. Damon (World Scientific Publishing, Singapore, 2007), pp. 253–327

12. A. Frühbis-Krüger, A. Neumer, Simple Cohen-Macaulay codimension 2 singularities. *Commun. Algebra* **38**(2), 454–495 (2010)
13. A. Frühbis-Krüger, M. Zach, On the vanishing topology of isolated Cohen-Macaulay codimension 2 singularities (2015). [arXiv:1501.01915](https://arxiv.org/abs/1501.01915)
14. B. Green, On curves over valuation rings and morphisms to  $\mathbb{P}^1$ . *J. Number Theory* **59**(2), 262–290 (1996)
15. B. Green, M. Matignon, F. Pop, On the local Skolem property. *J. Reine Angew. Math.* **458**, 183–200 (1995)
16. G.M. Greuel, C. Lossen, F. Seelisch, Normalization of rings. *J. Symb. Comput.* **45**, 887–901 (2010)
17. H. Hironaka, Resolution of singularities of an algebraic variety over a field of characteristic zero: II. *Ann. Math.* **79**(2), 205–326 (1964)
18. H.W. Jung, Darstellung der Funktionen eines algebraischen Körpers zweier unabhängigen Veränderlichen  $x, y$  in der Umgebung einer Stelle  $x = a, y = b$ . *J. Reine Angew. Math.* **133**, 289–314 (1908)
19. J. Lipman, Desingularization of two-dimensional schemes. *Ann. Math. (2)* **107**(1), 151–207 (1978)
20. Q. Liu, *Algebraic Geometry and Arithmetic Curves* (Oxford University Press, Oxford, 2002)
21. D. Lorenzini, Models of curves and wild ramification. *Pure Appl. Math. Q* **6**(1), 41–82 (2010)
22. D. Lorenzini, Wild models of curves. *Algebra Number Theory* **8**(2), 331–367 (2014)
23. S. MacLane, A construction for absolute values in polynomial rings. *Trans. Am. Math. Soc.* **40**(3), 363–395 (1936)
24. S. MacLane, A construction for prime ideals as absolute values of an algebraic field. *Duke Math. J.* **2**(3), 492–510 (1936)
25. A. Néron, Modèles minimaux des variétés abéliennes sur les corps locaux et globaux. *Publ. Math. IHES* **21**, 5–128 (1964)
26. J. Nuno-Ballesteros, B. Oréface-Okamoto, J. Tomazella, The vanishing euler characteristic of an isolated determinantal singularity. *Isr. J. Math.* **197**, 475–495 (2013)
27. F.J. Pfreundt, M.E.A. Rahn, Gpi-space. Technical Report, Fraunhofer ITWM Kaiserslautern (2014). <http://www.gpi-space.de/>
28. J. Rüth, Models of curves and valuations. Ph.D. Thesis, Universität Ulm, 2014
29. J. Rüth, Mac Lane infrastructure for discrete valuations in Sage. [https://github.com/saraedum/mac\\_lane](https://github.com/saraedum/mac_lane)
30. J. Rüth, S. Wewers, Semistable reduction of superelliptic curves of degree  $p$  (in preparation)
31. J. Silverman, *Advanced Topics in the Arithmetic of Elliptic Curves* (Springer, New York, 1994)
32. J. Tate, Algorithm for determining the type of a singular fiber in an elliptic pencil, in *Modular Functions of One Variable IV: Proceedings of the International Summer School, University of Antwerp, RUCA, July 17 – August 3, 1972*, ed. by B.J. Birch, W. Kuyk (Springer, Berlin/Heidelberg, 1975), pp. 33–52
33. G. Tjurina, Absolute isolatedness of rational singularities and triple rational points. *Func. Anal. Appl.* **2**, 324–333 (1968)
34. M. Zach, Vanishing cycles of smoothable isolated Cohen-Macaulay codimension 2 singularities of type 2. Preprint, [arXiv:1607.07527](https://arxiv.org/abs/1607.07527)



# Moduli Spaces of Curves in Tropical Varieties



Andreas Gathmann and Dennis Ochse

**Abstract** We describe a framework to construct tropical moduli spaces of rational stable maps to a smooth tropical hypersurface or curve. These moduli spaces will be tropical cycles of the expected dimension, corresponding to virtual fundamental classes in algebraic geometry. As we focus on the combinatorial aspect, we take the weights on certain basic 0-dimensional local combinatorial curve types as input data, and give a compatibility condition in dimension 1 to ensure that this input data glues to a global well-defined tropical cycle. As an application, we construct such moduli spaces for the case of lines in surfaces, and in a subsequent paper for stable maps to a curve [Gathmann et al., Tropical moduli spaces of stable maps to a curve, in *Algorithmic and Experimental Methods in Algebra, Geometry, and Number Theory*, ed. by G. Böckle, W. Decker, G. Malle (Springer, Heidelberg, 2018). [https://doi.org/10.1007/978-3-319-70566-8\\_12](https://doi.org/10.1007/978-3-319-70566-8_12)].

**Keywords** Tropical geometry • Enumerative geometry • Gromov-Witten theory

**Subject Classifications** 14T05, 14N35, 51M20

## 1 Introduction

Moduli spaces of stable maps to a smooth projective variety are one of the most important tools in modern enumerative geometry [5, 15]. Intersection theory on these spaces has been used successfully to solve many enumerative problems, such as e.g., determining the numbers of plane curves of fixed genus and degree through given points, or the numbers of rational curves of fixed degree in a general quintic threefold [9, 20].

In recent times, tropical geometry has also been proven to be very useful for attacking enumerative problems, starting with Mikhalkin's famous Correspondence

---

A. Gathmann (✉) • D. Ochse  
Fachbereich Mathematik, Technische Universität Kaiserslautern, Postfach 3049, 67653  
Kaiserslautern, Germany  
e-mail: [andreas@mathematik.uni-kl.de](mailto:andreas@mathematik.uni-kl.de); [ochse@mathematik.uni-kl.de](mailto:ochse@mathematik.uni-kl.de)

© Springer International Publishing AG, part of Springer Nature 2017  
G. Böckle et al. (eds.), *Algorithmic and Experimental Methods  
in Algebra, Geometry, and Number Theory*,  
[https://doi.org/10.1007/978-3-319-70566-8\\_11](https://doi.org/10.1007/978-3-319-70566-8_11)

Theorem that provided the first link between such problems in algebraic and tropical geometry [21]. Accordingly, it is an important goal in tropical enumerative geometry to construct tropical analogues of moduli spaces of stable maps, i.e., tropical cycles whose points parametrize curves with certain properties in a given tropical variety. This has been achieved for rational curves in toric varieties (corresponding to tropical curves in a real vector space) in [16, 22], and (tropical) intersection theory on these spaces has been used in many cases to attack and solve enumerative problems from a purely combinatorial point of view.

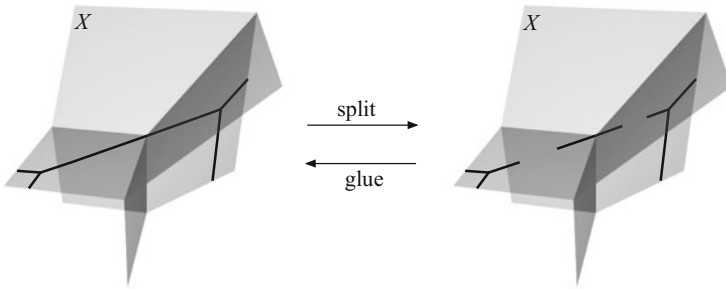
Of course, it would be very desirable to have such moduli spaces of tropical stable maps also for other target varieties. However, there is currently no general known method to construct such spaces, mainly for the following two reasons:

- (a) Already in algebraic geometry, the moduli spaces of stable maps may have bigger dimension than expected from deformation theory. One can solve this problem by introducing virtual fundamental classes, i.e., cycle classes in the moduli spaces which are of the expected dimension and replace the ordinary fundamental classes for intersection-theoretic purposes [3, 4]. These classes can usually be constructed as certain Chern classes of vector bundles. However, there is no corresponding counterpart of this theory in tropical geometry yet.
- (b) Tropical curves (in the sense of: metric graphs of the given degree and genus) in the given tropical variety might not be tropicalizations of actual algebraic curves inside the algebraic variety. Consequently, the naive tropical moduli space may not capture the situation from algebraic geometry appropriately, and it might have too big dimension even if the algebraic moduli space does not. This already happens for lines in cubic surfaces: Whereas each smooth algebraic cubic contains exactly 27 lines, there are smooth tropical cubics with infinitely many lines on them [31]. In general, already this question whether a tropical curve is realizable by an algebraic one inside the given ambient variety is an unsolved problem. It is known that the space of realizable tropical curves is a polyhedral set [32], but there is currently no explicit way to describe it.

In this paper, we will therefore study these tropical moduli spaces from an axiomatic and purely combinatorial point of view. Given a smooth tropical hypersurface or curve  $X$ , we will describe a set of (essentially 0-dimensional) input data and (essentially 1-dimensional) compatibility conditions that allow to construct from them tropical moduli spaces  $\mathcal{M}_0(X, \Sigma)$  of rational curves in  $X$  that are cycles of the expected dimension (where the subscript 0 denotes the genus and  $\Sigma$  the degree of the curves). We can therefore consider such cycles as tropical analogues of both the algebraic moduli spaces of stable maps and their virtual fundamental classes.

A central concept for this construction is the *resolution dimension*  $\text{rdim}(V)$  of a vertex  $V$  of a curve in  $X$ . It is an integer determined by the local combinatorial type of the curve and  $X$  at  $V$  that describes the expected dimension of the moduli space of such local curves when the vertex of the curve is resolved, modulo the local lineality space of  $X$ . For example, a 4-valent vertex in the plane  $\mathbb{R}^2$  has resolution dimension 1, since it can be resolved to two 3-valent vertices in a 1-dimensional

family (modulo translations in  $\mathbb{R}^2$ ). The origin of the curve in the following picture has resolution dimension 0 since this curve piece cannot be resolved or moved in  $X$ .



As a first rule, vertices of negative resolution dimension are not allowed in our moduli spaces—in the case of target curves this is known as the Riemann-Hurwitz condition [6, 7]. To construct the maximal cells of  $\mathcal{M}_0(X, \Sigma)$  together with their weights, the general idea is then that splitting and gluing of the curves allows to reduce this question to vertices of resolution dimension 0. For example, the picture above on the left shows a line in a plane in  $\mathbb{R}^3$ , which can vary in a 2-dimensional moduli space by moving its two vertices along the direction of its bounded edge. We can split the curve in three pieces as in the picture on the right, all of which have resolution dimension 0. If we have weights for the moduli spaces for these three local pieces, we can glue them back together using tropical intersection theory to obtain a moduli space for our original situation on the left. Technically, this means that we consider the curve pieces to have bounded ends, and that we impose the intersection-theoretic condition that corresponding endpoints map to the same point in  $X$  by suitable evaluation maps.

In this way, we can make  $\mathcal{M}_0(X, \Sigma)$  into a weighted polyhedral complex of the expected dimension by just giving weights for vertices of resolution dimension 0 as input data (we will refer to them as *moduli data* in Definition 3.9). However, as this input data can a priori be arbitrary, we need a certain compatibility condition for the resulting polyhedral complex to be balanced. A central result of our paper is that checking this condition in resolution dimension 1 is enough to ensure that the gluing process then works for all dimensions of the moduli spaces (Corollaries 3.17 and 3.18).

We check this condition for the moduli spaces of lines in surfaces in  $\mathbb{R}^3$ , leading e.g., to a well-defined 0-dimensional moduli cycle of lines on an arbitrary smooth tropical cubic surface, even if the actual number of such lines is infinite. In a particular example from [31] of such a cubic with infinitely many lines, we verify that this 0-cycle still has degree 27 as expected. In a subsequent paper, we use our methods to obtain well-defined moduli spaces of rational stable maps (of any degree) to an arbitrary target curve [17]. In any case, the initial 0-dimensional input data is obtained by tropicalization from the algebraic situation. For example, for the vertex in the origin in the picture above the weight is just the number of lines in  $\mathbb{P}^3$  through  $L_1 \cap L_2$  and  $L_3 \cap L_4$  for any four general lines  $L_1, L_2, L_3, L_4 \subset \mathbb{P}^3$ , which is 1.

The organization of this paper is as follows. In Sect. 2 we give the necessary background from tropical geometry. While most of this material is well-known, there are three techniques that go beyond the usual theory: partially open versions of tropical varieties in Sect. 2.1, quotient maps (and their intersection-theoretic properties) in Sect. 2.2, and pull-backs of diagonals of smooth varieties in Appendix. Section 3 then describes the gluing process for curves and constructs the tropical moduli spaces from the given input data. Finally, in Sect. 4 we study the case of lines in surfaces.

This paper is based on parts of the Ph.D. thesis of the second author [24]. It would not have been possible without extensive computations of examples which enabled us to establish and prove conjectures about polyhedra and their weights in our moduli spaces. For this we used the polymake extension a-tint [18, 19] and GAP [29]. We thank an anonymous referee for useful suggestions on the exposition. The work of the first author was partially funded by the DFG grant GA 636/4-2, as part of the Priority Program 1489.

## 2 Preliminaries

### 2.1 Partially Open Tropical Varieties

Although most of the spaces occurring in this paper will be tropical varieties, some of our intermediate constructions also involve “partially open” versions of them. In these more general spaces the boundary faces of some polyhedra can be missing, and thus the balancing condition is required to hold at fewer places. The constructions in this introductory chapter are adapted to this setting and thus sometimes slightly more general than usual. However, since all constructions relevant to us are local, the required changes are minimal and straightforward.

For more details on the notions of tropical cycles and fans, see e.g., [2, 16].

**Notation 2.1 (Polyhedra)** Let  $\Lambda$  denote a lattice isomorphic to  $\mathbb{Z}^N$  for some  $N \geq 0$ , and let  $V := \Lambda \otimes_{\mathbb{Z}} \mathbb{R}$  be the corresponding real vector space. A *partially open (rational) polyhedron* in  $V$  is a subset  $\sigma \subset V$  that is the intersection of finitely many open or closed half-spaces  $\{x \in V : f(x) < c\}$  resp.  $\{x \in V : f(x) \leq c\}$  for  $c \in \mathbb{R}$  and  $f$  in  $\Lambda^\vee$ , the dual of  $\Lambda$ . We call  $\sigma$  a (closed) polyhedron if it can be written in this way with only closed half-spaces.

The *(relative) interior*  $\sigma^\circ$  of a partially open polyhedron is the topological interior of  $\sigma$  in its affine span. We denote by  $V_\sigma \subset V$  the linear space which is the shift of the affine span of  $\sigma$  to the origin, and set  $\Lambda_\sigma := V_\sigma \cap \Lambda$ . The *dimension* of  $\sigma$  is defined to be the dimension of  $V_\sigma$ .

A *face*  $\tau$  of a partially open polyhedron  $\sigma$  is a non-empty subset of  $\sigma$  that can be obtained by changing some of the non-strict inequalities  $f(x) \leq c$  defining  $\sigma$  into equalities. We write this as  $\tau \leq \sigma$ , or  $\tau < \sigma$  if in addition  $\tau \neq \sigma$ . If  $\dim \tau = \dim \sigma - 1$  we call  $\tau$  a *facet* of  $\sigma$ . In this case we denote by  $u_{\sigma/\tau} \in \Lambda_\sigma / \Lambda_\tau$  the

*primitive normal vector* of  $\sigma$  relative to  $\tau$ , i.e., the unique generator of  $\Lambda_\sigma/\Lambda_\tau$  lying in the half-line of  $\sigma$  in  $V_\sigma/V_\tau \cong \mathbb{R}$ .

**Definition 2.2 (Polyhedral Complexes and Tropical Varieties)** A *partially open polyhedral complex* in  $V$  is a collection  $X$  of partially open polyhedra in a vector space  $V = \Lambda \otimes_{\mathbb{Z}} \mathbb{R}$ , also called cells, such that

- (a) if  $\sigma \in X$  and  $\tau$  is a face of  $\sigma$  then  $\tau \in X$ ; and
- (b) if  $\sigma_1, \sigma_2 \in X$  then  $\sigma_1 \cap \sigma_2$  is empty or a face of both  $\sigma_1$  and  $\sigma_2$ .

It is called *pure-dimensional* if each inclusion-maximal cell has the same dimension. The *support* of  $X$ , denoted by  $|X|$ , is the union of all  $\sigma \in X$  in  $V$ .

A *weighted partially open polyhedral complex* is a pair  $(X, \omega_X)$ , where  $X$  is a purely  $k$ -dimensional partially open polyhedral complex, and  $\omega_X$  is a map associating a *weight*  $\omega_X(\sigma) \in \mathbb{Z}$  to each  $k$ -dimensional cell  $\sigma \in X$ . If there is no risk of confusion we will write  $\omega_X$  as  $\omega$ , and  $(X, \omega_X)$  just as  $X$ . A *partially open tropical cycle*  $X$  in  $V$  is a weighted polyhedral complex such that for each cell  $\tau$  of dimension  $k - 1$  the *balancing condition*

$$\sum_{\sigma:\sigma>\tau} \omega(\sigma) \cdot u_{\sigma/\tau} = 0 \quad \in V/V_\tau$$

holds. It is called a *partially open tropical variety* if all weights are non-negative. If all polyhedra in  $X$  are closed, we omit the attribute “partially open” and speak e.g., of tropical cycles and tropical varieties. A *tropical fan* is a tropical variety all of whose polyhedra are cones.

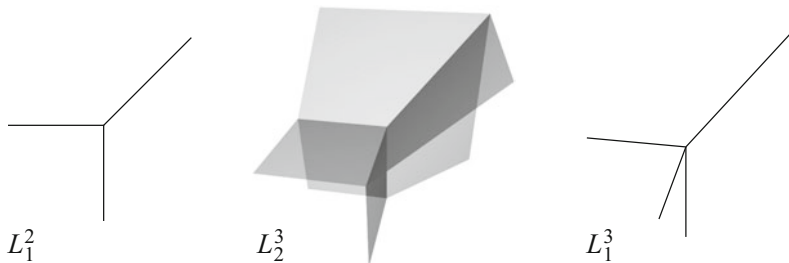
Often, the exact polyhedral complex structure of our cycles is not important. We call two (partially open) cycles equivalent if they allow a common refinement (where a refinement is required to respect the weights, and every polyhedron of a refinement must be closed in its corresponding cell of the original cycle). By abuse of notation, the corresponding equivalence classes will again be called (partially open) tropical cycles.

A *morphism*  $f : X \rightarrow Y$  between (partially open) tropical cycles  $X$  and  $Y$  is a locally affine linear map  $f : |X| \rightarrow |Y|$ , with the linear part induced by a map between the underlying lattices. It is called an *isomorphism* if it has a two-sided inverse (up to refinements) and respects the weights.

*Example 2.3 (Linear Spaces)* Let  $V = \mathbb{R}^k$ , denote by  $e_i$  for  $i = 1, \dots, k$  the negatives of the vectors of the standard basis, and set  $e_0 = -e_1 - \dots - e_k$ . For  $r < k$  we denote by  $L_r^k$  the tropical fan whose simplicial cones are indexed by subsets  $I \subset \{0, \dots, k\}$  with at most  $r$  elements and given by the cones generated by all  $e_i$  with  $i \in I$ . The weights of the top-dimensional cones, corresponding to subsets of size  $r$ , are all set to 1. This is the tropicalization of a general  $r$ -dimensional linear space over the Puiseux series with constant coefficient equations [11, Proposition 2.5 and Theorem 4.1].

The following pictures illustrate these spaces, where all displayed cones are thought to be extending to infinity. If instead we interpret the pictures as bounded

spaces they represent partially open tropical varieties obtained by intersecting  $L_r^k$  with an open bounded polyhedron.



*Remark 2.4* In our Definition 2.2 it is allowed that two partially open polyhedra in a complex do not intersect although their closures do. For example, in the pictures above we could replace all polyhedra by their relative interiors. This would give us weighted partially open polyhedral complexes with the same support, and whose face relations and balancing conditions are trivial. However, spaces of this type will not occur in our constructions in this paper—we will always have partially open polyhedral complexes  $X$  such that  $\sigma \cap \tau = \emptyset$  for given  $\sigma, \tau \in X$  implies  $\overline{\sigma} \cap \overline{\tau} = \emptyset$ .

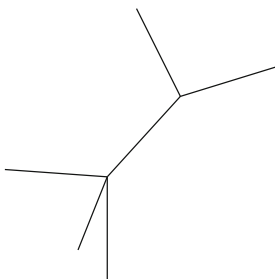
**Definition 2.5 (Smooth Varieties)** For simplicity, in this paper we will follow [1] and call a tropical variety  $X$  *smooth* if it is locally isomorphic to some  $L_r^k \times \mathbb{R}^m$  around 0 at each point (where  $r + m = \dim X$  is fixed, but otherwise  $k, r, m$  may depend on the chosen point). This is more special than the usual definition of smoothness which allows any polyhedral complex locally isomorphic to a matroid fan as in Appendix. We expect that our results would hold in this more general setting as well.

For a smooth variety  $X$ , following [22, Section 5.3] the *canonical divisor*  $K_X$  of  $X$  is defined to be the weighted polyhedral complex given by the codimension-1 skeleton of  $X$ , where the weight of a codimension-1 cell of  $X$  is the number of adjacent maximal cells minus 2.

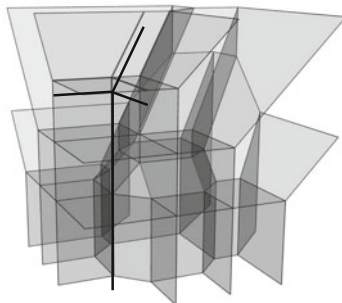
*Example 2.6 (Smooth Curves and Hypersurfaces)* In this paper, the following two cases of smooth varieties will be of particular importance.

- (a) Consider a connected 1-dimensional tropical variety  $X$  in  $\mathbb{R}^N$  which is a tree—we will refer to such a space as a *rational curve*. The smoothness condition then means that for any vertex of  $X$  the primitive integer vectors in the directions of the adjacent edges can be mapped to  $e_0, \dots, e_k$  in  $\mathbb{R}^k$  for some  $k \leq N$  by a  $\mathbb{Z}$ -linear map, and that all edges have weight 1. Such a vertex occurs in the canonical divisor  $K_X$  with weight  $k - 1$ .
- (b) Consider a hypersurface  $X$  in  $\mathbb{R}^N$ , i.e., a tropical variety given by a *tropical polynomial* in  $N$  variables [25]. The coefficients of this polynomial determine a subdivision of its Newton polytope, and the weighted polyhedral complex structure of  $X$  is induced by this subdivision. Smoothness then means that this

subdivision is unimodular [21, Proposition 3.11], and hence  $K_X$  contains each codimension-1 cell of  $X$  with weight 1. Most important for us will be the case of a cubic surface in  $\mathbb{R}^3$ , i.e., of  $X$  being dual to a unimodular subdivision of the lattice polytope in  $\mathbb{Z}^3$  with vertices  $(0, 0, 0)$ ,  $(3, 0, 0)$ ,  $(0, 3, 0)$ , and  $(0, 0, 3)$ .



(a) a smooth curve in  $\mathbb{R}^3$



(b) a cubic surface in  $\mathbb{R}^3$  with a line on it

## 2.2 Tropical Quotients

In this section we will define quotients of partially open tropical cycles by vector spaces in certain cases.

**Definition 2.7 (Lineality Space)** Let  $X$  be a partially open polyhedral complex in a vector space  $V = \Lambda \otimes_{\mathbb{Z}} \mathbb{R}$ , and let  $L \subset V$  be a vector subspace defined over  $\mathbb{Q}$ . We say that  $L$  is a *lineality space* for  $X$  if, for a suitable complex structure of  $X$ , for all  $\sigma \in X$  and  $x \in \sigma$  the intersection  $\sigma \cap (x + L)$  is open in  $x + L$  and equal to  $|\sigma| \cap (x + L)$ .

*Remark 2.8* If  $X$  is a tropical variety, i.e.,  $\sigma$  is closed in  $V$  for all  $\sigma \in X$ , then  $\sigma \cap (x + L)$  can only be open and non-empty if it is all of  $x + L$ . So in this case we arrive at the usual notion of lineality space found in the literature (although note that most authors only call a maximal subspace with this property a lineality space).

**Lemma 2.9** Let  $X$  be a partially open polyhedral complex in  $V$  with a lineality space  $L$ . Denote by  $q : V \rightarrow V/L$  the quotient map, where  $V/L$  is considered to have the underlying lattice  $\Lambda/(\Lambda \cap L)$ . Then for all  $\sigma, \tau \in X$  we have:

- (a)  $q(\sigma)$  is a partially open polyhedron of dimension  $\dim q(\sigma) = \dim \sigma - \dim L$ .
- (b) If  $\tau \leq \sigma$  then  $q(\tau) \leq q(\sigma)$ .
- (c)  $q(\sigma \cap \tau) = q(\sigma) \cap q(\tau)$ .
- (b) If  $q(\sigma) = q(\tau)$  then  $\sigma = \tau$ .
- (d)  $\Lambda_{q(\sigma)} = \Lambda_{\sigma}/(\Lambda \cap L)$ ; and if  $\tau$  is a facet of  $\sigma$  then  $u_{q(\sigma)/q(\tau)} = \overline{u_{\sigma/\tau}}$  with this identification.

*Proof* By induction it suffices to prove the statements for  $\dim L = 1$ . We choose coordinates  $(x, y) \in \mathbb{R}^{\dim V - 1} \times \mathbb{R} \cong V$  such that  $L = \{(x, y) : x = 0\}$ , and consider  $x$  as coordinates on  $V/L$ .

- (a) In the defining inequalities for  $\sigma$  we may assume that all of them that contain  $y$  are strict: if one of the non-strict defining inequalities  $f(x, y) \leq c$  containing  $y$  was satisfied as an equality at a point  $(x_0, y_0) \in \sigma$  it could not be satisfied in a neighborhood of  $y_0 \in \mathbb{R}$ , contradicting the openness of  $\sigma \cap ((x_0, y_0) + L)$ . Hence  $\sigma$  can be written as

$$\sigma = q^{-1}(\sigma_0) \cap \{(x, y) : y > f_i(x) + a_i \text{ and } y < g_j(x) + b_j \text{ for all } i, j\} \quad (1)$$

for some linear forms  $f_i, g_j$ , constants  $a_i, b_j$ , and a partially open polyhedron  $\sigma_0$  given by the defining inequalities of  $\sigma$  that do not contain  $y$ . But then

$$q(\sigma) = \sigma_0 \cap \{x : f_i(x) + a_i < g_j(x) + b_j \text{ for all } i, j\}, \quad (2)$$

since for  $x$  satisfying these conditions we can always find  $y \in \mathbb{R}$  with  $f_i(x) + a_i < y < g_j(x) + b_j$  for all  $i, j$ . Hence  $q(\sigma)$  is a partially open polyhedron. Moreover, the openness of  $\sigma \cap ((x, y) + L)$  means that all non-empty fibers of  $q|_\sigma$  have dimension  $\dim L$ , so that  $\dim q(\sigma) = \dim \sigma - \dim L$ .

- (b) If  $\tau \leq \sigma$  is obtained from  $\sigma$  by changing some non-strict inequalities to equalities this means that  $\tau$  can be written as in (1) for some  $\tau_0 \leq \sigma_0$  and with the same  $f_i, g_j$ . Then (2) holds for  $\tau$  and  $\tau_0$  as well, and we conclude that  $q(\tau)$  is a face of  $q(\sigma)$ .
- (c) The inclusion “ $\subset$ ” is obvious. Conversely, if  $x \in q(\sigma) \cap q(\tau)$  then there are  $y, y' \in \mathbb{R}$  with  $(x, y) \in \sigma$  and  $(x, y') \in \tau$ . Hence

$$(x, y) \in \sigma \cap ((x, y') + L) \subset |X| \cap ((x, y') + L),$$

which implies  $(x, y) \in \tau \cap ((x, y') + L)$  by definition of a lineality space. Hence  $(x, y) \in \sigma \cap \tau$ , i.e.,  $x = q(x, y) \in q(\sigma \cap \tau)$ .

- (d) By (c), the equality  $q(\sigma) = q(\tau)$  implies  $q(\sigma \cap \tau) = q(\sigma)$ . In particular,  $\sigma \cap \tau \neq \emptyset$ . Hence  $\sigma \cap \tau$  is a face of  $\sigma$ , which by the dimension statement of (a) must be of the same dimension as  $\sigma$ . But this means that  $\sigma \cap \tau = \sigma$ , and by symmetry thus also  $\sigma \cap \tau = \tau$ .
- (e) From the definition of a lineality space it follows that  $L \subset V_\sigma$ ; hence  $V_{q(\sigma)} = V_\sigma/L$  and thus  $\Lambda_{q(\sigma)} = \Lambda_\sigma/(\Lambda \cap L)$ . In particular, for  $\tau$  a facet of  $\sigma$  we have  $\Lambda_\sigma/\Lambda_\tau = \Lambda_{q(\sigma)}/\Lambda_{q(\tau)}$  (with the isomorphism given by taking quotients by  $\Lambda \cap L$ ), from which the statement about the primitive normal vector follows.  $\square$

**Corollary 2.10 (Quotients)** *Let  $X$  be an  $n$ -dimensional partially open tropical variety in  $V = \Lambda \otimes_{\mathbb{Z}} \mathbb{R}$  with a lineality space  $L$ , and let  $q : V \rightarrow V/L$  be the quotient morphism. Then*

$$X/L := \{q(\sigma) : \sigma \in X\}$$



together with the weights  $\omega_{X/L}(q(\sigma)) := \omega_X(\sigma)$  is a partially open tropical variety of dimension  $n - \dim L$  in the vector space  $V/L$  with lattice  $\Lambda/(\Lambda \cap L)$ . We will also denote it by  $q(X)$ .

*Proof* By Lemma 2.9 (a) we see that  $X/L$  is a collection of partially open polyhedra which by (d) are in bijection to the polyhedra in  $X$ . The statements (b) and (c) of the lemma now imply that  $X/L$  is a partially open polyhedral complex as in Definition 2.2. Moreover, the dimension statement of part (a) of the lemma means that  $X/L$  is of pure dimension  $n - \dim L$ , and that  $\omega_{X/L}(q(\sigma)) := \omega_X(\sigma)$  defines a weight function on the top-dimensional cones. Finally, part (e) of the lemma shows that the images of the balancing conditions for  $X$  in  $V$  give us the balancing conditions for  $X/L$  in  $V/L$ . □

Our main examples for this quotient construction are (cycles in) the moduli spaces of tropical curves, which we introduce in Sect. 2.4.

### 2.3 Tropical Intersection Theory

We now briefly recall some constructions and results of tropical intersection theory. For a detailed introduction we refer to [2, 26, 27]. Although the theory is only developed for closed tropical cycles there, the extension to the case of partially open tropical cycles stated below follows immediately since all constructions involved are local.

**Construction 2.11 (Rational Functions and Divisors)** For a partially open pure-dimensional tropical cycle  $X$ , a (*non-zero*) rational function on  $X$  is a continuous function  $\varphi : |X| \rightarrow \mathbb{R}$  that is affine linear on each cell (for a suitable polyhedral complex structure), with linear part given by an element of  $\Lambda^\vee$ . We can associate to such a rational function  $\varphi$  a (*Weil*) divisor, denoted by  $\varphi \cdot X$ . It is a partially open tropical subcycle of  $X$  (i.e. a partially open cycle whose support is a subset of  $|X|$ ) of codimension 1. Its support is contained in the subset of  $|X|$  where  $\varphi$  is not locally affine linear [2, Construction 3.3].

Two important examples for rational functions and divisors in this paper are:

- (a) On  $X = \mathbb{R}^N$ , a tropical polynomial  $\varphi$  defines a rational function, and its divisor is just the tropical hypersurface defined by  $\varphi$ .
- (b) If  $X$  is one-dimensional, the divisor  $\varphi \cdot X$  consists of finitely many points. We define its *degree* to be the sum of the weights of the points in  $\varphi \cdot X$ . By abuse of notation, we sometimes write this degree as  $\varphi \cdot X$  as well.

Multiple intersection products  $\varphi_1 \cdot \dots \cdot \varphi_r \cdot X$  are commutative by Allermann and Rau [2, Proposition 3.7].

*Remark 2.12 (Pull-Backs and Push-Forwards)* Rational functions on a (partially open) tropical cycle  $Y \subset \Lambda' \otimes_{\mathbb{Z}} \mathbb{R}$  can be *pulled back* along a morphism  $f : X \rightarrow Y$  to rational functions  $f^* \varphi = \varphi \circ f$  on  $X$ . Also, we can *push forward* subcycles  $Z$  of  $X$

to subcycles  $f_*Z$  of  $Y$  of the same dimension [2, Proposition 4.6 and Corollary 7.4], where in the partially open case we will always restrict ourselves to injective maps  $f$  so that no problems can arise from two partially open polyhedra with different boundary behavior that are mapped by  $f$  to an overlapping image. In any case, by picking a suitable refinement of  $X$  we can ensure that the partially open image polyhedra  $f(\sigma)$  for  $\sigma \in X$  form a partially open polyhedral complex  $f_*Z$ . For a top-dimensional cell  $\sigma' \in f_*Z$  its weight  $\omega_{f_*Z}(\sigma')$  is given by

$$\omega_{f_*Z}(\sigma') := \sum_{\sigma} \omega_X(\sigma) \cdot |\Lambda'_{\sigma'} / f(\Lambda_{\sigma})|,$$

where the sum goes over all top-dimensional cells  $\sigma \in Z$  with  $f(\sigma) = \sigma'$ . Of course, in the partially open case, there will be at most one such  $\sigma$  in each sum since  $f$  is assumed to be injective. As expected, push-forwards and pull-backs satisfy the projection formula [2, Proposition 4.8 and Corollary 7.7].

**Construction 2.13 (Pull-Backs Along Quotient Maps)** Let  $X$  be a partially open tropical variety with a lineality space  $L$ , so that there is a quotient variety  $X/L$  as in Corollary 2.10 with quotient map  $q : X \rightarrow X/L$ . Moreover, let  $Z$  be a partially open subcycle of  $X/L$ . Then the collection of polyhedra  $q^{-1}(\sigma)$  for  $\sigma \in Z$ , together with the weight function  $\omega(q^{-1}(\sigma)) = \omega_Z(\sigma)$ , is a partially open subcycle of  $X$  of dimension  $\dim Z + \dim L$  (in fact, the balancing conditions follows from Lemma 2.9 (e)). We denote it by  $q^*Z$ .

We conclude this short excursion into tropical intersection theory with two compatibility statements between the quotient construction and pull-backs of rational functions resp. push-forward of cycles.

**Lemma 2.14** *Let  $X$  be a partially open tropical cycle with lineality space  $L$  and quotient map  $q : X \rightarrow X/L$ . Then for any rational function  $\varphi$  on  $X/L$  we have*

$$(q^*\varphi \cdot X)/L = \varphi \cdot (X/L).$$

*Proof* This is obvious from the definitions. □

**Lemma 2.15** *Let  $f : X \rightarrow Y$  be a morphism between partially open tropical cycles. Assume that  $X$  and  $Y$  have lineality spaces  $L$  and  $L'$ , respectively, and that the linear part of  $f$  on each cell maps  $\Lambda_X \cap L$  isomorphically to  $\Lambda_Y \cap L'$ . If  $g : X/L \rightarrow Y/L'$  is the morphism giving a commutative diagram*

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ \downarrow q & & \downarrow q' \\ X/L & \xrightarrow{g} & Y/L' \end{array}$$

*then  $f_*(X)/L' = g_*(X/L)$ .*

*Proof* Assume that the polyhedral complex structure of all cycles is sufficiently fine to be compatible with all of the morphisms. Let  $\sigma$  be a maximal cell in  $X$ , and set  $\tau = f(\sigma)$  and  $\rho = q'(\tau)$ . Applying a suitable translation, we may assume that  $f$  (and thus also  $g$ ) is linear on  $\sigma$ .

If  $\dim \rho < \dim \sigma - \dim L$  then  $\sigma$  does not contribute to either cycle in the statement of the lemma. Otherwise, the assumption implies that  $f$  maps  $\Lambda_X \cap L$ , and hence every saturated sublattice of  $\Lambda_X \cap L$  such as  $\Lambda_\sigma \cap L$ , to a saturated sublattice of  $\Lambda_Y$ . Hence, in the inclusion

$$f(\Lambda_\sigma \cap L) \subset f(\Lambda_\sigma) \cap f(L) = f(\Lambda_\sigma) \cap L' \subset \Lambda_\tau \cap L'$$

we must have equality since both sides are saturated lattices in  $\Lambda_Y$  of the same rank. By the last equality in this chain it then follows from the weight formulas of Corollary 2.10 and Remark 2.12 that the contribution of  $\sigma$  to the cell  $\rho$  in the two cycles of the lemma is

$$\begin{aligned} \omega_{f_*(X)/L'}(\rho) &= \omega_X(\sigma) \cdot |\Lambda_\tau / f(\Lambda_\sigma)| \\ &= \omega_X(\sigma) \cdot |(\Lambda_\tau / (\Lambda_\tau \cap L')) / (f(\Lambda_\sigma) / (f(\Lambda_\sigma) \cap L'))| \\ &= \omega_X(\sigma) \cdot |q'(\Lambda_\tau) / q'(f(\Lambda_\sigma))| \\ &= \omega_X(\sigma) \cdot |q'(\Lambda_\tau) / g(q(\Lambda_\sigma))| \\ &= \omega_{g_*(X/L)}(\rho). \end{aligned} \quad \square$$

### 2.4 Tropical Moduli Spaces of Curves

We will now come to the construction of the moduli spaces  $\mathcal{M}_0(\mathbb{R}^N, \Sigma)$  of tropical curves in  $\mathbb{R}^N$ . Analogously to the algebraic case, elements of these spaces will be given by a rational  $n$ -marked abstract tropical curve, together with a map to  $\mathbb{R}^N$  whose image is a (not necessarily smooth) tropical curve in  $\mathbb{R}^N$  as in Sect. 2.1, with unbounded directions as determined by  $\Sigma$ .

For later purposes we will also need a version  $\mathcal{M}'_0(\mathbb{R}^N, \Sigma)$  of these spaces where the ends  $x_i$  have bounded lengths. In contrast to the original spaces  $\mathcal{M}_0(\mathbb{R}^N, \Sigma)$  they will only be partially open tropical varieties since there is no limit curve when the length of such a bounded end approaches zero.

Let us start with the discussion of abstract tropical curves.

**Construction 2.16 (Abstract Curves)** A (rational) abstract tropical curve [16, Definition 3.2] is a metric tree graph  $\Gamma$  with all vertices of valence at least 3. Unbounded ends (with no vertex there) are allowed and labeled by  $x_1, \dots, x_n$ . The tuple  $(\Gamma, x_1, \dots, x_n)$  will be referred to as an  $n$ -marked abstract tropical curve. Two such curves are called isomorphic (and will from now on be identified) if there is an isometry between them that respects the markings. The set of all  $n$ -marked abstract tropical curves (modulo isomorphisms) is denoted  $\mathcal{M}_{0,n}$ .

For  $n \geq 3$  it follows from [28, Theorem 3.4], [23, Section 2], or [16, Theorem 3.7] that  $\mathcal{M}_{0,n}$  can be given the structure of a tropical fan as follows: Consider the map

$$d : \mathcal{M}_{0,n} \rightarrow \mathbb{R}^{\binom{n}{2}}, \quad (\Gamma, x_1, \dots, x_n) \mapsto (\text{dist}_\Gamma(x_i, x_j))_{i < j}$$

where  $\text{dist}_\Gamma(x_i, x_j)$  denotes the distance between the two marked ends  $x_i$  and  $x_j$  in the metric graph  $\Gamma$ . Moreover, consider the linear map

$$\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^{\binom{n}{2}}, \quad (a_i)_i \mapsto (a_i + a_j)_{i,j},$$

let  $u_i = \varphi(e_i)$  be the images of the unit vectors, let  $U_n := \varphi(\mathbb{R}^n) = \langle u_1, \dots, u_n \rangle$ , and

$$q_n : \mathbb{R}^{\binom{n}{2}} \rightarrow Q_n := \mathbb{R}^{\binom{n}{2}} / U_n$$

be the quotient map. For a subset  $I \subset [n]$  with  $2 \leq |I| \leq n - 2$  define  $v_I$  to be the image under  $q_n \circ d : \mathcal{M}_{0,n} \rightarrow Q_n$  of a tree  $\Gamma$  with exactly one bounded edge of length one, the marked leaves  $x_i$  with  $i \in I$  on one side and the leaves  $x_i$  for  $i \notin I$  on the other. Let  $\Lambda_n := \langle v_I \rangle_{\mathbb{Z}} \subset Q_n$  be the lattice in  $Q_n$  generated by the vectors  $v_I$ . By Speyer and Sturmfels [28, Theorem 4.2] the map  $q_n \circ d : \mathcal{M}_{0,n} \rightarrow Q_n$  is injective, and its image is a purely  $(n - 3)$ -dimensional simplicial tropical fan in  $Q_n = \Lambda_n \otimes_{\mathbb{Z}} \mathbb{R}$ , with all top-dimensional cones having weight 1. In the following we will always consider  $\mathcal{M}_{0,n}$  with this structure of a tropical fan.

The cones of  $\mathcal{M}_{0,n}$  are labeled by the *combinatorial types* of marked curves, i.e., by the homeomorphism classes of the curves relative to their ends. The vectors  $v_I$  generate the rays of  $\mathcal{M}_{0,n}$ .

**Construction 2.17 (Abstract Curves with Bounded Ends)** We will now adapt Construction 2.16 to the case when the ends also have bounded lengths. So we say that an  $n$ -marked *abstract tropical curve with bounded ends* is a metric tree graph as above with the unbounded ends replaced by bounded intervals, i.e., a metric graph  $\Gamma$  without 2-valent vertices, and the 1-valent vertices labeled by  $x_1, \dots, x_n$ . The notions of isomorphisms and combinatorial types carry over from the case with unbounded ends. The set of all  $n$ -marked curves (modulo isomorphisms) with bounded ends is denoted  $\mathcal{M}'_{0,n}$ .

To make  $\mathcal{M}'_{0,n}$  for  $n \geq 3$  into a partially open tropical variety we consider the distance map  $d : \mathcal{M}'_{0,n} \rightarrow \mathbb{R}^{\binom{n}{2}}$  as above, which in this case however includes the lengths of the bounded ends. Then  $d$  is injective: the vectors  $u_i = \varphi(e_i) \in Q_n$  in the notation of Construction 2.16 correspond exactly to a change of the length of the bounded edge at  $x_i$ —so by Construction 2.16 the image point under  $q_n \circ d : \mathcal{M}'_{0,n} \rightarrow Q_n$  allows to reconstruct the combinatorial type of the graph as well as the lengths of all edges not adjacent to the markings, whereas the full vector in  $\mathbb{R}^{\binom{n}{2}}$  then allows to reconstruct the lengths of the ends as well.

Note that the combinatorial types of these curves with bounded ends are in one-to-one correspondence with the types of curves with unbounded ends, and that  $d$  maps  $\mathcal{M}'_{0,n}$  to a partially open polyhedral complex (with cones in bijection to the

combinatorial types), which by abuse of notation we will also denote by  $\mathcal{M}'_{0,n}$ . Taking again the lattice in  $\mathbb{R}^{\binom{n}{2}}$  generated by all graphs with integer lengths, and giving all cells of  $\mathcal{M}'_{0,n}$  weight 1, we get in fact a partially open tropical variety in  $\mathbb{R}^{\binom{n}{2}}$  of dimension  $2n - 3$ . It has lineality space  $\mathcal{Q}_n$  in the sense of Definition 2.7, and we have  $\mathcal{M}'_{0,n}/\mathcal{Q}_n = \mathcal{M}_{0,n}$  as in Corollary 2.10.

**Construction 2.18 (Parametrized Curves in  $\mathbb{R}^N$ )** A (rational, parametrized) curve in  $\mathbb{R}^N$  [16, Definition 4.1] is a tuple  $(\Gamma, x_1, \dots, x_n, h)$ , with  $(\Gamma, x_1, \dots, x_n)$  a rational  $n$ -marked abstract tropical curve and  $h : \Gamma \rightarrow \mathbb{R}^N$  a continuous map satisfying:

- (a) On each edge  $e$  of  $\Gamma$ , with metric coordinate  $t$ , the map  $h$  is of the form  $h(t) = a + t \cdot v$  for some  $a \in \mathbb{R}^N$  and  $v \in \mathbb{Z}^N$ . If  $V \in e$  is a vertex and we choose  $t$  positive on  $e$ , the vector  $v$  will be denoted  $v(e, V)$  and called the *direction* of  $e$  (at  $V$ ). If  $e$  is an end and  $t$  pointing in its direction, we write  $v$  as  $v(e)$ .
- (b) For every vertex  $V$  of  $\Gamma$  we have the *balancing condition*

$$\sum_{e:V \in e} v(e, V) = 0.$$

Two such curves in  $\mathbb{R}^N$  are called isomorphic (and will be identified) if there is an isomorphism of the underlying abstract  $n$ -marked curves commuting with the maps to  $\mathbb{R}^N$ .

The *degree* of a curve  $(\Gamma, x_1, \dots, x_n, h)$  in  $\mathbb{R}^N$  as above is the  $n$ -tuple  $\Sigma = (v(x_1), \dots, v(x_n)) \in (\mathbb{Z}^N)^n$  of directions of its ends. We denote the space of all curves in  $\mathbb{R}^N$  of a given degree  $\Sigma$  by  $\mathcal{M}_0(\mathbb{R}^N, \Sigma)$ .

Note that  $\Sigma$  may contain zero vectors, corresponding to contracted ends that can be thought of as marked points in the algebraic setting (see Construction 2.19).

If  $n \geq 3$  and there is at least one end  $x_i$  with  $v(x_i) = 0$  we can use the bijection

$$\mathcal{M}_0(\mathbb{R}^N, \Sigma) \rightarrow \mathcal{M}_{0,n} \times \mathbb{R}^N, \quad (\Gamma, x_1, \dots, x_n, h) \mapsto ((\Gamma, x_1, \dots, x_n), h(x_i))$$

(which forgets  $h$  except for its image on  $x_i$ ) to give it the structure of a tropical variety [16, Proposition 4.7].

**Construction 2.19 (Evaluation Maps)** For each  $i$  with  $v(x_i) = 0$  there is an *evaluation map*

$$ev_i : \mathcal{M}_0(\mathbb{R}^N, \Sigma) \rightarrow \mathbb{R}^N$$

assigning to a tropical curve  $(\Gamma, x_1, \dots, x_n, h)$  the position  $h(x_i)$  of its  $i$ -th marked end (note that this is well-defined since the marked end  $x_i$  is contracted to a point). By Gathmann et al. [16, Proposition 4.8], these maps are morphisms of tropical fans.

**Construction 2.20 (Parametrized Curves in  $\mathbb{R}^N$  with Bounded Ends)** Constructions 2.17 and 2.18 can obviously be combined to obtain moduli spaces  $\mathcal{M}'_0(\mathbb{R}^N, \Sigma)$

of curves in  $\mathbb{R}^N$  with bounded ends, as pull-backs under the quotient maps that forget the lengths of the bounded ends. They are partially open tropical varieties and admit evaluation maps to  $\mathbb{R}^N$  as in Construction 2.19 at all ends (i.e., not just at the contracted ones).

*Remark 2.21 (Parametrized Curves in  $\mathbb{R}^N$  with Few Markings)* In the following, we will also need the moduli spaces  $\mathcal{M}_0(\mathbb{R}^N, \Sigma)$  of parametrized curves  $(\Gamma, x_1, \dots, x_n, h)$  for the special cases when  $n < 3$  or there is no  $x_i$  with  $v(x_i) = 0$ , so that the above bijection with  $\mathcal{M}_{0,n} \times \mathbb{R}^N$  is not available. In order to overcome this technical problem and still give  $\mathcal{M}_0(\mathbb{R}^N, \Sigma)$  the structure of a tropical variety, there are two possibilities:

- (a) One can use *barycentric coordinates*, taking a certain weighted average of the vertices of the curve.
- (b) One can combine evaluation maps at several non-contracted ends, where at each such end the evaluations are only taken modulo the direction of the edge to make them well-defined.

Details on these alternative construction can be found in [24, Section 1.2]. In the following, we will just assume that  $\mathcal{M}_0(\mathbb{R}^N, \Sigma)$  has the structure of a tropical variety in any case.

*Remark 2.22 (Degree of the Canonical Divisor on Curves)* As in Example 2.6, let  $X \subset \mathbb{R}^N$  be a smooth rational curve or a smooth hypersurface. Consider an  $n$ -marked curve  $(\Gamma, x_1, \dots, x_n, h) \in \mathcal{M}_0(\mathbb{R}^N, \Sigma)$  with  $h(\Gamma) \subset X$ , so that  $h$  can also be viewed as a morphism from the (abstract) tropical curve  $\Gamma$  to  $X$ .

We can then consider  $K_X$  as a divisor on  $X$  as in Construction 2.11 and compute its pull-back  $h^*K_X$  according to Remark 2.12. Its degree (as in Construction 2.11 (b)) depends only on  $X$  and  $\Sigma$ , and not on  $h$ :

- (a) If  $X$  is a rational curve note that any two points in  $X$  are rationally equivalent divisors in the sense of [2]. Hence the degree of the divisor  $h^*P$  for a point  $P \in X$  does not depend on  $P$ ; for a general point  $P$  it is just the sum of the weights of the direction vectors for all edges of  $\Gamma$  that map some point to  $P$  under  $h$ . In particular, taking for  $P$  a point far out on an unbounded edge of  $X$ , we see that this degree depends only on  $\Sigma$  and not on  $h$ . It will be called the *degree*  $\deg h$  of  $h$  and is the tropical counterpart of the notion of degree of a morphism between smooth curves in algebraic geometry. The degree of  $h^*K_X$  is now just  $\deg h \cdot \deg K_X$ ; in particular by the above it depends only on  $\Sigma$  and not on  $h$ .
- (b) If  $X$  is a hypersurface in  $\mathbb{R}^N$  a local computation shows that  $K_X = X \cdot X$ , and hence by the projection formula the degree of  $h^*K_X$  on  $\Gamma$  is the same as the degree of  $h_*\Gamma \cdot X$  on  $\mathbb{R}^N$ . But the 1-cycle  $h_*\Gamma$  in  $\mathbb{R}^N$  is rationally equivalent to its so-called *recession fan*, i.e., the fan obtained by shrinking all bounded edges to zero length. As this fan is determined by  $\Sigma$  it follows also in this case that the degree of  $h^*K_X$  depends only on  $\Sigma$  and not on  $h$ .

We will therefore denote the degree of  $h^*K_X$  also by  $K_X \cdot \Sigma$ .

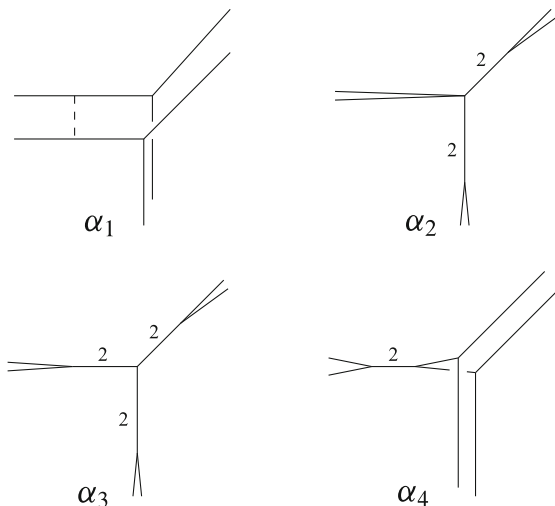
### 3 Gluing Moduli Spaces

Throughout this section, let  $X \subset \mathbb{R}^N$  be a smooth rational curve or a smooth hypersurface as in Example 2.6. Aiming at enumerative applications, we want to construct tropical analogues of the algebraic moduli spaces of stable maps to a variety, or in other words generalizations of the tropical moduli spaces  $\mathcal{M}_0(\mathbb{R}^N, \Sigma)$  of Construction 2.18 to other target spaces than  $\mathbb{R}^N$ . The naive approach would simply be to use the subset

$$\tilde{\mathcal{M}}_0(X, \Sigma) := \{(\Gamma, x_1, \dots, x_n, h) \in \mathcal{M}_0(\mathbb{R}^N, \Sigma) : h(\Gamma) \subset X\}$$

of tropical curves mapping to  $X$ . As  $K_X \cdot \Sigma$  is independent of the curves in this space by Remark 2.22, the algebro-geometric analogue tells us that we expect  $\tilde{\mathcal{M}}_0(X, \Sigma)$  to be of dimension  $\dim X + |\Sigma| - 3 - K_X \cdot \Sigma$  (in fact, this independence is the reason why we restrict to the curve and hypersurface cases in this paper). However, just as in the algebraic case, the actual dimension of this space might be bigger, as the following example shows.

*Example 3.1* Let  $X = L_1^2$  be the tropical line in  $\mathbb{R}^2$  as in Example 2.3, and consider degree-2 covers of  $X$ , i.e.,  $\Sigma = (e_0, e_0, e_1, e_1, e_2, e_2)$ . The following pictures show combinatorial types of tropical curves in  $\mathcal{M}_0(X, \Sigma)$ . Edges are labeled with their weight if it is not 1, and their directions in the picture indicate which cell of  $X$  they are mapped to. The edge drawn with a dashed line is contracted to a point.



Note that in  $\alpha_1$  and  $\alpha_4$  the lengths of the bounded edges are not independent, since in both cases the two horizontal bounded edges adjacent to the origin must have the same length. Hence the types  $\alpha_1, \alpha_2$ , and  $\alpha_4$  are described by 2-dimensional cells

in the moduli spaces, whereas  $\alpha_3$  is 3-dimensional (with  $\alpha_2$  as one of its faces). As we expect our tropical moduli space to have dimension 2 (equal to the space of algebraic degree-2 covers of  $\mathbb{P}^1$ ), we see that  $\tilde{\mathcal{M}}_0(X, \Sigma)$  has too big dimension.

Our first aim is therefore to define a suitable subset  $\mathcal{M}_0(X, \Sigma)$  of  $\tilde{\mathcal{M}}_0(X, \Sigma)$  of the expected dimension. We fix the polyhedral complex structure on  $X$  to be the unique coarsest one (which exists since  $X$  is smooth), and choose the polyhedral complex structure on  $\tilde{\mathcal{M}}_0(X, \Sigma)$  as follows.

**Notation 3.2 (Curves in  $X$ )** Let  $(\Gamma, x_1, \dots, x_n, h) \in \tilde{\mathcal{M}}_0(X, \Sigma)$ . In the following, all isolated points of  $h^{-1}(\sigma)$  for a cell  $\sigma \in X$  will be considered as (possibly additional 2-valent) vertices of  $\Gamma$ . The interior of every edge of  $\Gamma$  then maps to the relative interior of a unique cell of  $X$ , and we include this information in the combinatorial type of a curve in  $X$ .

For such a combinatorial type  $\alpha$ , we denote the set of all curves in  $\tilde{\mathcal{M}}_0(X, \Sigma)$  of this type by  $\mathcal{M}(\alpha)$ . These are partially open polyhedra, and their closures  $\overline{\mathcal{M}(\alpha)}$  give  $\tilde{\mathcal{M}}_0(X, \Sigma)$  the structure of a polyhedral complex. We write  $\beta \geq \alpha$  for two combinatorial types with  $\overline{\mathcal{M}(\beta)} \supset \overline{\mathcal{M}(\alpha)}$ , and say in this case that  $\alpha$  is a *face* of  $\beta$ . If in addition  $\beta \neq \alpha$  we call  $\beta$  a *resolution* of  $\alpha$ .

**Notation 3.3 (Local Curves)** Let  $C = (\Gamma, x_1, \dots, x_n, h) \in \tilde{\mathcal{M}}_0(X, \Sigma)$  be a tropical curve in  $X$ , and fix a vertex  $V$  of  $\Gamma$ . We can restrict  $C$  to the local situation around  $V$  and obtain the following data, all written with an index  $V$ :

- (a) The collection of all vectors  $v(e, V)$  for  $e \ni V$  as in Construction 2.18 is called the *local degree*  $\Sigma_V$  of the curve at  $V$ .
- (b) The *star* of  $X$  at  $h(V)$  will be denoted  $X_V$ ; it is a shifted tropical fan.
- (c) By our assumption on  $X$ , we know that  $X_V$  is isomorphic to  $L_{r_V}^{k_V} \times \mathbb{R}^{m_V}$  for unique  $k_V, r_V, m_V$  (where  $0 < r_V < k_V$  unless  $(k_V, r_V) = (0, 0)$ ). Note that the degree of the canonical divisor as in Remark 2.22 then splits up as

$$K_X \cdot \Sigma = \sum_V K_{X_V} \cdot \Sigma_V,$$

with the sum taken over all vertices of  $\Gamma$ .

- (d) Let  $C_V \in \tilde{\mathcal{M}}_0(X_V, \Sigma_V)$  be the curve in  $X_V$  with one vertex mapping to  $h(V)$ , and unbounded ends of directions  $\Sigma_V$ . We refer to  $C_V$  as a *local curve*. Its combinatorial type  $\alpha_V$  will be called the *trivial combinatorial type* of  $V$  in  $\tilde{\mathcal{M}}_0(X_V, \Sigma_V)$ . We will refer to a resolution of  $\alpha_V$  also as a *resolution* of  $V$ .

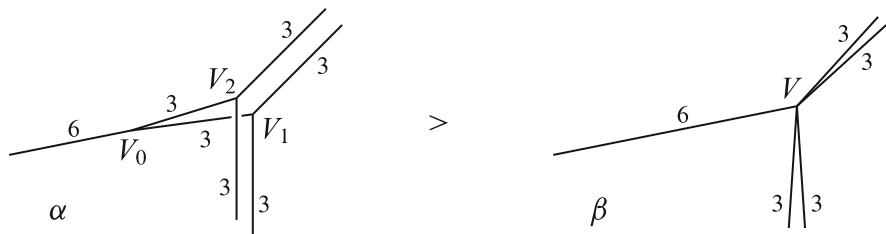
**Definition 3.4** Let  $V \in \Gamma$  be a vertex of a curve  $(\Gamma, x_1, \dots, x_n, h) \in \tilde{\mathcal{M}}_0(X, \Sigma)$ . Using Notation 3.3, we define

- (a) the *virtual dimension* of  $V$  as  $\text{vdim}(V) = \text{val}(V) - K_{X_V} \cdot \Sigma_V + \dim X - 3$ ,
- (b) the *resolution dimension* of  $V$  as  $\text{rdim}(V) = \text{val}(V) - K_{X_V} \cdot \Sigma_V + r_V - 3$ ,
- (c) the *classification number* of  $V$  as  $c_V = \text{val}(V) + r_V$ .

*Example 3.5* Let  $X = L_3^4$ , and consider the combinatorial types  $\alpha > \beta$  of curves in  $X$  as shown in the picture below. In the type  $\beta$ , the vertex  $V$  is mapped to the origin, and the unbounded ends have directions  $e_0 + e_1 + e_2$  (with weight 6), and twice  $e_3$



and  $e_4$  (with weight 3). The type  $\alpha$  has the same directions of the ends,  $V_1$  and  $V_2$  map to the origin, and consequently  $V_0$  to a positive multiple of  $e_0 + e_1 + e_2$ .



We then have  $\text{vdim}(V) = \text{rdim}(V) = -1$ ,  $\text{vdim}(V_1) = \text{rdim}(V_1) = \text{vdim}(V_2) = \text{rdim}(V_2) = 0$ ,  $\text{vdim}(V_0) = 3$ , and  $\text{rdim}(V_0) = 0$ . Moreover,  $c_V = 8$ ,  $c_{V_1} = c_{V_2} = 6$ , and  $c_{V_0} = 3$ .

The virtual dimension of a vertex  $V$  can be thought of as the expected dimension of the moduli space of curves in  $X_V \cong L_{r_V}^{k_V} \times \mathbb{R}^{m_V}$  of degree  $\Sigma_V$ . It includes the dimension of a lineality space coming from translations in  $\mathbb{R}^{m_V}$ , which is subtracted from the virtual dimension to obtain the resolution dimension (as can be seen in the example above for  $V_0$ , which can locally be moved in its 3-dimensional cell spanned by  $e_0, e_1, e_2$ ). The classification number of Definition 3.4 (c) is a useful number for inductive proofs because it is always non-negative and becomes smaller in resolutions, as the following lemma shows.

**Lemma 3.6** *Let  $\alpha$  be a resolution of a vertex  $V$  of a curve in  $X$ . Then the classification number of every vertex  $W$  of  $\alpha$  is smaller than  $c_V$ .*

*Proof* Note first that  $c_W \leq c_V$  since both summands in the definition of the classification number cannot get bigger when passing from  $V$  to  $W$ :

- (a) As  $\alpha$  is a tree, shrinking an edge to zero length will merge two vertices into one, whose valence is the sum of the original valences minus 2. In particular, since the trivial combinatorial type of  $V$  is obtained from  $\alpha$  by a sequence of such processes, we conclude that  $\text{val}(V) \geq \text{val}(W)$ , with equality if and only if all vertices of  $\alpha$  except  $W$  have valence 2.
- (b) If  $W$  is mapped to the relative interior of a cone  $\sigma_W$ , we have  $r_W = \dim X - \dim \sigma_W$ . The analogous statement holds for  $V$ , and hence

$$r_V = \dim X - \dim \sigma_V \geq \dim X - \dim \sigma_W = r_W$$

since  $\sigma_V$  is a face of  $\sigma_W$ .

If we had equality for both numbers, all vertices  $W'$  of  $\alpha$  except  $W$  must have valence 2 by (a). Moreover,  $W$  and  $V$  have to lie in the same cell by (b). Hence  $\alpha$  is the same combinatorial type as the trivial type  $V$  after removing all 2-valent vertices  $W'$ . But this means that all two-valent vertices lie in the interior of an edge that is completely

mapped to one cell of  $X$ . As this is excluded by definition, there can actually be no 2-valent vertices. Hence  $\alpha$  is the trivial type  $V$ , in contradiction to  $\alpha$  being a resolution of  $V$ .  $\square$

The idea to construct the desired moduli spaces of curves in  $X$  is now as follows.

- (a) Vertices of negative resolution dimension should not be admitted, since they correspond (locally, and modulo their lineality space) to an algebraic moduli space of curves of negative virtual dimension. We will exclude them in Definition 3.8. For the case of curves, this corresponds to the Riemann-Hurwitz condition as e.g., in [6–8, 10].
- (b) Vertices of resolution dimension 0 correspond (again locally and modulo their lineality space) to a 0-dimensional algebraic moduli space. Hence the curves in the corresponding tropical moduli spaces should not allow any resolutions, i.e., these moduli spaces will consist of only one cell, whose weight is the degree of the algebraic moduli space.

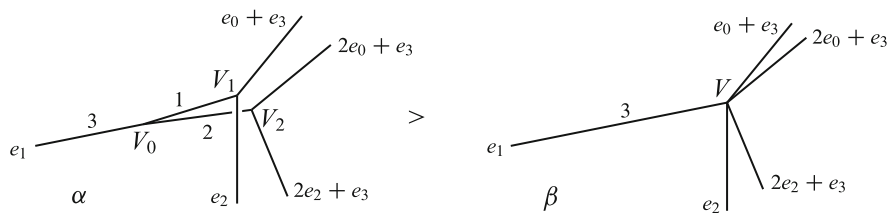
In this paper, we will only consider the tropical situation. The weights of the cells for vertices of resolution dimension 0 are therefore considered as initial input data for our constructions, as in Definition 3.9 below.

- (c) Vertices of positive resolution dimension will lead to curves that allow resolutions, and whose moduli spaces therefore consist of several cells. These spaces can be obtained recursively by gluing from the initial ones of (b) using Construction 3.12, so that no additional input data is required for these cases.

However, as for this paper the initial weights of (b) are arbitrary numbers a priori, they need to satisfy some compatibility conditions in order for the gluing process to lead to a well-defined space. These conditions are encoded in the notion of a good vertex in Definition 3.11. They are originally formulated in every resolution dimension; we will see in Corollary 3.18 however that only the conditions in resolution dimension 1 are relevant since the others follow from them.

In this recursive construction of the moduli spaces, the following example shows that the resolution dimension is not necessarily strictly increasing. We will therefore use the classification number for these purposes.

*Example 3.7* Consider again two combinatorial types  $\alpha > \beta$  as in Example 3.5, however in  $X = L_2^3$  and with directions as indicated in the following picture.



Then  $\text{rdim}(V) = \text{rdim}(V_0) = \text{rdim}(V_1) = 1$  and  $\text{rdim}(V_2) = 0$ , i.e., all resolution dimensions are non-negative and hence will be admitted. The weight for the type  $\alpha$  will be defined by a gluing procedure over its three vertices in Construction 3.12. However,  $V_1$  has the same resolution dimension as  $V$ , so that a recursive construction over the resolution dimension would not work.

**Definition 3.8 (The Moduli Space  $\mathcal{M}_0(X, \Sigma)$  as a Polyhedral Complex)**

- (a) An *admissible* combinatorial type of a curve in  $X$  is a combinatorial type  $\alpha$  such that for all vertices  $V$  in  $\alpha$  we have  $\text{rdim}(V) \geq 0$ .
- (b) We denote by  $\mathcal{M}_0(X, \Sigma)$  the polyhedral subcomplex of  $\tilde{\mathcal{M}}_0(X, \Sigma)$  consisting of all closed cells  $\mathcal{M}(\alpha)$  such that  $\alpha$  and all its faces are admissible, and

$$\dim \mathcal{M}(\alpha) = \text{vdim} \mathcal{M}_0(X, \Sigma) := |\Sigma| - K_X \cdot \Sigma + \dim X - 3.$$

Note that this dimension is just  $\text{vdim}(V)$  if  $\alpha$  is a combinatorial type with just one vertex  $V$ .

- (c) The *neighborhood* of a combinatorial type  $\alpha$  in  $\mathcal{M}_0(X, \Sigma)$  is defined as

$$\mathcal{N}(\alpha) := \bigcup_{\beta \geq \alpha} \mathcal{M}(\beta),$$

where the union is taken over all combinatorial types  $\beta \geq \alpha$  in  $\mathcal{M}_0(X, \Sigma)$ .

In the following, we will apply this definition also to the case of local curves as in Notation 3.3, in order to obtain moduli spaces  $\mathcal{M}_0(X_V, \Sigma_V)$ .

Example 3.5 shows that faces of admissible types need not be admissible again, so that the condition of all faces of  $\alpha$  being admissible in Definition 3.8 (b) is not vacuous.

**Definition 3.9 (Moduli Data)** *Moduli data* for a smooth variety  $X$  are a collection  $(\omega_V)_V$  of weights in  $\mathbb{Q}$  for every vertex  $V$  of a curve in  $X$  with  $\text{rdim}(V) = 0$ . All subsequent constructions and results in this section will depend on the choice of such moduli data.

**Construction 3.10 (Vertices of Resolution Dimension 0)** Let  $V$  be a vertex of a local curve in  $\mathcal{M}_0(X_V, \Sigma_V)$ . As  $X_V \cong L_{rv}^{k_V} \times \mathbb{R}^{m_V}$ , the trivial combinatorial type of  $V$  in  $\mathcal{M}_0(X_V, \Sigma_V)$  has dimension  $m_V$ , corresponding to a translation along  $\mathbb{R}^{m_V}$ . But  $\text{vdim}(V) = \text{rdim}(V) + m_V$ , and thus the trivial combinatorial type is maximal in  $\mathcal{M}_0(X_V, \Sigma_V)$  if and only if  $\text{rdim}(V) = 0$ . In this case, the moduli space  $\mathcal{M}_0(X_V, \Sigma_V)$  consists of only one cell, and we equip it with the weight  $\omega_V$  from our moduli data.

To define the weights on  $\mathcal{M}_0(X, \Sigma)$  in general we need Definition 3.11 and Construction 3.12, which depend on each other and work in a combined recursion on the classification number of vertices. The following definition of a good vertex of a certain classification number thus assumes that good vertices of smaller classification number have already been defined. Moreover, for every combinatorial type  $\alpha$  in a local moduli space  $\mathcal{M}_0(X_V, \Sigma_V)$  all of whose vertices have smaller

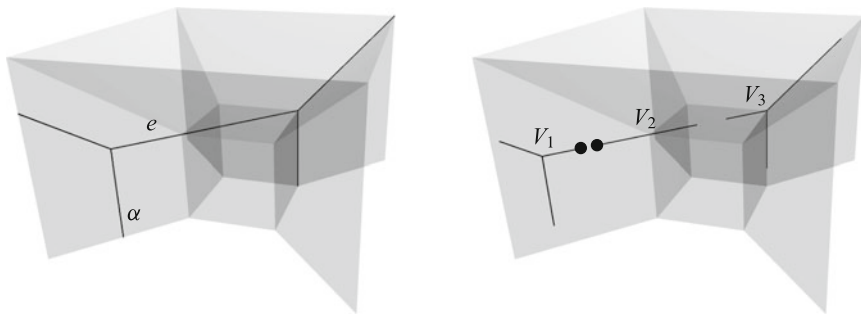
classification number and are good it assumes from Construction 3.12 below that there is a *gluing cycle*  $\mathcal{Z}(\alpha)$  on the neighborhood  $\mathcal{N}(\alpha)$ .

**Definition 3.11 (Good Vertices)** Let  $V$  be a vertex of a (local) curve in  $\mathcal{M}_0(X_V, \Sigma_V)$ , of classification number  $c_V$ . By recursion on  $c_V$ , we define  $V$  to be a *good* vertex if the following three conditions hold.

- (a) Every vertex of every resolution  $\alpha$  of  $V$  in  $\mathcal{M}_0(X_V, \Sigma_V)$  (which has classification number smaller than  $c_V$  by Lemma 3.6) is good. (There is then a gluing cycle  $\mathcal{Z}(\alpha)$  on  $\mathcal{N}(\alpha)$  by Construction 3.12).
- (b)  $\mathcal{M}_0(X_V, \Sigma_V)$  is a tropical cycle with the following weights:
  - If  $\text{rdim}(V) = 0$  we equip the unique cell of  $\mathcal{M}_0(X_V, \Sigma_V)$  with the weight from the moduli data as in Construction 3.10.
  - If  $\text{rdim}(V) > 0$  the maximal types  $\alpha$  in  $\mathcal{M}_0(X_V, \Sigma_V)$  are not the trivial one, i.e., they are resolutions of  $V$ . The weights on the corresponding cells  $\mathcal{M}(\alpha) = \mathcal{N}(\alpha)$  are then the ones of the gluing cycles  $\mathcal{Z}(\alpha)$  as in (a).
- (c) For every resolution  $\alpha$  of  $V$  in  $\mathcal{M}_0(X_V, \Sigma_V)$  and every maximal type  $\beta \geq \alpha$  in  $\mathcal{M}_0(X_V, \Sigma_V)$  (which is then also a resolution of  $V$ ), the weight of the cell  $\mathcal{M}(\beta)$  is the same in the gluing cycles  $\mathcal{Z}(\alpha)$  and  $\mathcal{Z}(\beta)$ .

**Construction 3.12 (The Gluing Cycle  $\mathcal{Z}(\alpha)$ )** Fix a (not necessarily maximal) combinatorial type  $\alpha$  in a moduli space  $\mathcal{M}_0(X, \Sigma)$  as in the picture below on the left, and assume that all its vertices are good. We will now construct a cycle  $\mathcal{Z}(\alpha)$  of dimension  $\text{vdim} \mathcal{M}_0(X, \Sigma)$  on the neighborhood  $\mathcal{N}(\alpha)$ . An important example of this is when  $\alpha$  is a resolution of a vertex in a local moduli space. More technical details on this construction can be found in [24, Construction 1.5.13].

For each vertex  $V$  of  $\alpha$  let  $\sigma_V^\circ$  be the open cell of  $X$  in which  $V$  lies, and let  $X(V) = \bigcup_{\sigma \supset \sigma_V} \sigma^\circ \subset X$ ; it is an open neighborhood of  $\sigma_V$ . Similarly, for each edge  $e$  of  $\alpha$  let  $\sigma_e^\circ$  be the open cell of  $X$  in which  $e$  lies, and set  $X(e) = \bigcup_{\sigma \supset \sigma_e} \sigma^\circ$ .



First we will construct local moduli spaces  $\mathcal{M}_V$  for each vertex  $V$  of  $\alpha$ . As  $V$  is good the local moduli space  $\mathcal{M}_0(X_V, \Sigma_V)$  is a tropical cycle by Definition 3.11 (b). Let  $\mathcal{M}'_0(X_V, \Sigma_V)$  be the corresponding moduli cycle of curves with bounded ends,

i.e., the pull-back of  $\mathcal{M}_0(X_V, \Sigma_V)$  under the quotient map that forgets the lengths of the ends as in Construction 2.20. We denote by

$$\mathcal{M}_V = \{(\Gamma, x_1, \dots, x_n, h) \in \mathcal{M}'_0(X_V, \Sigma_V) : h(\Gamma) \subset X(V)\}$$

its partially open polyhedral subcomplex of all curves with bounded ends that lie entirely in  $X(V)$ . A typical element of  $\mathcal{M}_V$  is a curve piece as shown in the picture above on the right. (Note that these pieces might also be resolutions of the corresponding vertices.)

Now we glue these pieces  $\mathcal{M}_V$  together. For each bounded edge  $e$  of  $\alpha$ , joining two vertices  $V_1$  and  $V_2$  (as in the picture), there are corresponding bounded ends in  $\mathcal{M}_{V_1}$  and  $\mathcal{M}_{V_2}$ . Let

$$\text{ev}_e : \prod_V \mathcal{M}_V \rightarrow X(e) \times X(e)$$

be the evaluation map at these two ends, where the product is taken over all vertices  $V$  of  $\alpha$ . The product

$$\prod_e \text{ev}_e^* \Delta_{X(e)}$$

over all pull-backs of the diagonals  $\Delta_{X(e)}$  along these evaluation maps as in Appendix can also be written as the product  $\prod_e \text{ev}_e^* \Delta_X$  along the extended evaluation maps to  $X \times X$  by Remark 4.11. We will therefore abbreviate it by  $\text{ev}^* \Delta_X$ ; it is a cocycle on  $\prod_V \mathcal{M}_V$ . The support of the cycle  $\text{ev}^* \Delta_X \cdot \prod_V \mathcal{M}_V$  then consists of points corresponding to curve pieces that can be glued together to a curve in  $X$ , i.e., so that e.g., the positions of the dots in the picture above coincide.

However, these curve pieces still carry the information about the position of the gluing points. In order to forget these positions, we take a quotient as follows. For each bounded edge  $e$  of  $\alpha$  between two vertices  $V_1$  and  $V_2$  as above, there are two vectors  $u_{V_1}$  and  $u_{V_2}$  in the lineality spaces of  $\mathcal{M}_{V_1}$  and  $\mathcal{M}_{V_2}$ , respectively, that parametrize the lengths of these ends as in Constructions 2.16 and 2.17. The vector  $u_e := u_{V_1} - u_{V_2}$  is then in the lineality space of  $\prod_V \mathcal{M}_V$ , and we denote by  $L_\alpha$  the vector space spanned by the vectors  $u_e$  for all bounded edges  $e$  of  $\alpha$ . Let  $q$  be the quotient map by  $L_\alpha$ ; the quotient  $q(\text{ev}^* \Delta_X \cdot \prod_V \mathcal{M}_V)$  then does not contain the information about the gluing points any more.

From this cycle we now obtain an injective morphism that considers the curve pieces as a glued curve in  $X$ . It can be defined as

$$f : q(\text{ev}^* \Delta_X \cdot \prod_V \mathcal{M}_V) \rightarrow \mathcal{M}'_{0,n}(\mathbb{R}^N, \Sigma)$$

$$((d_{i,j})_{\{i,j\} \in R_V}, a^V)_V \mapsto \left( \left( \sum_{\{k,l\} \in R_{i,j}} d_{k,l} \right)_{\{i,j\} \in R}, a^{V_0} \right),$$

where

- $a^V$  denotes the coordinates of the root vertex in the local moduli space  $\mathcal{M}_V$  for  $V$ , of which we choose  $V_0$  as the root vertex in  $\mathcal{M}'_{0,n}(\mathbb{R}^N, \Sigma)$ ;
- $d_{i,j}$  denotes the distance coordinates on the moduli spaces, with  $R_V$  and  $R$  the index sets of all pairs of ends in the local moduli spaces  $\mathcal{M}_V$  and the full moduli space  $\mathcal{M}'_{0,n}(\mathbb{R}^N, \Sigma)$ , respectively;
- $R_{i,j} \subset \bigcup_V R_V$  for  $\{i, j\} \in R$  is the set of all pairs of ends of the local curves in the moduli spaces  $\mathcal{M}_V$  that lie on the unique path between the ends  $x_i$  and  $x_j$ .

The push-forward cycle

$$\mathcal{Z}'(\alpha) := f_* q \left( \text{ev}^* \Delta_X \cdot \prod_V \mathcal{M}_V \right) := f_* \left( q \left( \text{ev}^* \Delta_X \cdot \prod_V \mathcal{M}_V \right) \right)$$

in  $\mathcal{M}'_{0,n}(\mathbb{R}^N, \Sigma)$  will be called the *gluing cycle* (with bounded ends) for  $\alpha$ . Finally, taking the quotient by the lineality space corresponding to the ends of  $\alpha$ , we obtain a gluing cycle  $\mathcal{Z}(\alpha)$  (with unbounded ends) in  $\mathcal{M}_{0,n}(\mathbb{R}^N, \Sigma)$ . The cells of maximal dimension come with a natural weight in this construction, which we will call the *gluing weight*. It is not clear a priori that this weight is independent of the choice of  $\alpha$ , but it will turn out to be so in Theorem 3.16.

**Lemma 3.13** *Assume that all vertices occurring in a combinatorial type  $\alpha$  in  $\mathcal{M}_0(X, \Sigma)$  are good. Then*

$$\dim \mathcal{Z}(\alpha) = \dim X + |\Sigma| - 3 - K_X \cdot \Sigma.$$

*Proof* By definition we have

$$\mathcal{Z}'(\alpha) = f_* q \left( \text{ev}^* \Delta_X \cdot \prod_V \mathcal{M}_V \right).$$

Let  $s$  denote the number of vertices  $V$  in  $\alpha$ , so that  $s - 1$  is the number of bounded edges. As the push-forward preserves dimensions, we only have to compute the dimension of  $q(\text{ev}^* \Delta_X \cdot \prod_V \mathcal{M}_V)$ . We have that

$$\begin{aligned} \dim \prod_V \mathcal{M}_V &= \sum_V (2 \text{val}(V) - K_{X_V} \cdot \Sigma_V + \dim X - 3) \\ &= s \dim X - K_X \cdot \Sigma + \sum_V (\text{val}(V) - 3) + \sum_V \text{val}(V) \\ &= s \dim X - K_X \cdot \Sigma + 2|\Sigma| - 4 + s, \end{aligned}$$

where we used that for a tree the number of vertices equals  $|\Sigma| - 2 - \sum_V (\text{val}(V) - 3)$ . The cycle  $\text{ev}^*(\Delta_X) \cdot \prod_V \mathcal{M}_V$  has codimension  $(s - 1) \dim X$ , and taking the quotient

$q$  eliminates another  $s - 1$  dimensions. Passing from  $\mathcal{Z}'(\alpha)$  to  $\mathcal{Z}(\alpha)$  reduces the dimension by  $|\Sigma|$ , so in total we get

$$\begin{aligned} \dim \mathcal{Z}(\alpha) &= s \dim X - K_X \cdot \Sigma + 2|\Sigma| - 4 + s - (s - 1) \dim X - (s - 1) - |\Sigma| \\ &= \dim X + |\Sigma| - 3 - K_X \cdot \Sigma. \end{aligned} \quad \square$$

**Lemma 3.14** *Assume that all vertices occurring in a combinatorial type  $\alpha$  in  $\mathcal{M}_0(X, \Sigma)$  are good. If the gluing cycle  $\mathcal{Z}(\alpha)$  is not zero, the support of  $\mathcal{Z}(\alpha)$  contains  $\mathcal{M}(\alpha)$  and is contained in  $\mathcal{N}(\alpha)$  (so in particular also in  $\mathcal{M}_0(X, \Sigma)$ ).*

*Proof* In the notation of Construction 3.12, consider the set  $q^{-1}(f^{-1}(\mathcal{Z}'(\alpha)))$  of all curve pieces that glue to the combinatorial type  $\alpha$ . Under each evaluation map  $ev_e$ , this set maps to the lineality space of  $\Delta_{X(e)}$  in  $X(e) \times X(e)$ . The functions used for cutting out the diagonal in Construction 4.8 all contain this lineality space in their own lineality space, so their pull-backs are linear on the above set and hence do not subdivide it. These properties are preserved under push-forward with  $f$  and taking the quotient  $q$ , and thus  $\mathcal{Z}(\alpha)$  contains all of  $\mathcal{M}(\alpha)$  if it is non-zero.

By construction, the other points in the gluing cycle  $\mathcal{Z}(\alpha)$  correspond to curves obtained by resolving each vertex of  $\alpha$ , and thus to resolutions of  $\alpha$ . Hence,  $\mathcal{Z}(\alpha)$  is contained in the union of all neighboring cells of  $\mathcal{M}(\alpha)$ , and thus by Lemma 3.13 in  $\mathcal{N}(\alpha)$ . □

**Definition 3.15 (The Moduli Space  $\mathcal{M}_0(X, \Sigma)$  as a Weighted Polyhedral Complex)** Assume that all vertices occurring in curves in  $\mathcal{M}_0(X, \Sigma)$  are good. By Lemmas 3.13 and 3.14, each maximal cell  $\alpha$  of  $\mathcal{M}_0(X, \Sigma)$  is also a maximal cell in the gluing cycle  $\mathcal{Z}(\alpha)$ . We define the weight of this cell of  $\mathcal{M}_0(X, \Sigma)$  to be the corresponding gluing weight of Construction 3.12.

The aim of this section is to show that these weights make  $\mathcal{M}_0(X, \Sigma)$  into a tropical variety, i.e., balanced, if all vertices of resolution dimension 1 are good. Examples can be found in Sect. 4 and [17].

**Theorem 3.16** *Assume that all vertices in a combinatorial type  $\alpha$  in  $\mathcal{M}_0(X, \Sigma)$  are good. If  $\beta$  is a resolution of  $\alpha$  in  $\mathcal{M}_0(X, \Sigma)$  (so that in particular  $\mathcal{N}(\beta) \subset \mathcal{N}(\alpha)$ ) then the weight of every maximal cell in the gluing cycle  $\mathcal{Z}(\beta)$  in  $\mathcal{N}(\beta)$  agrees with the weight of the same cell in  $\mathcal{Z}(\alpha)$  in  $\mathcal{N}(\alpha)$ .*

*In particular, if  $\alpha$  is of virtual codimension 1 then the cycle  $\mathcal{M}_0(X, \Sigma)$  with the weights of Definition 3.15 is balanced at  $\alpha$ .*

*Proof* This is basically a straight-forward reduction proof; however, we have to pay attention to several intersection-theoretical details. We start by describing the gluing cycle  $\mathcal{Z}'(\beta)$  as in Construction 3.12. For every vertex  $V$  of  $\alpha$ , let  $J_V$  be the set of vertices of  $\beta$  that degenerate to  $V$  in  $\alpha$ , so that  $\cup_V J_V$  is the set of all vertices of  $\beta$ . We denote by  $EV^* \Delta_X$  the product over all pull-backs of the diagonals along evaluation maps belonging to the edges of  $\beta$ , by  $Q$  the quotient map on  $EV^* \Delta_X \cdot \prod_V \prod_{W \in J_V} \mathcal{M}_W$  forgetting the gluing points along the bounded edges, and by  $F$  the

morphism embedding the resulting cycle to  $\mathcal{M}'_0(X, \Sigma)$ . Then by Construction 3.12 the gluing cycle  $\mathcal{L}'(\beta)$  is given by

$$\mathcal{L}'(\beta) = F_*Q\left(\text{EV}^* \Delta_X \cdot \prod_V \prod_{W \in J_V} \mathcal{M}_W\right),$$

where the first product is taken over all vertices  $V$  of  $\alpha$ . We will now decompose the maps  $Q, F$ , and  $\text{EV}$  into contributions coming from the vertices of  $\alpha$  as follows. For each vertex  $V$  of  $\alpha$ , let  $I_V$  be the set of bounded edges of  $\beta$  contracting to  $V$  in  $\alpha$ . We denote by  $q_V$  the quotient map that forgets the gluing points on these edges, and by  $f_V$  the morphism that embeds the cycle  $\prod_{g \in I_V} \text{ev}_g^* \Delta_X \cdot \prod_{W \in J_V} \mathcal{M}_W$  in the local moduli space  $\mathcal{M}'_0(X_V, \Sigma_V)$ . Furthermore, denote by  $q$  and  $f$  the quotient and embedding maps for the gluing cycle for  $\alpha$ , respectively. With  $\tilde{q} = \prod_V q_V$  and  $\tilde{f} = \prod_V f_V$  we can then write

$$\mathcal{L}'(\beta) = f_*q\left(\tilde{f}_*\tilde{q}\left(\prod_e \text{ev}_e^* \Delta_X \cdot \prod_V \prod_{g \in I_V} \text{ev}_g^* \Delta_X \cdot \prod_V \prod_{W \in J_V} \mathcal{M}_W\right)\right),$$

by Lemma 2.15, with the product over  $e$  running over all edges of  $\alpha$ . By the compatibility of push-forwards and quotient maps with diagonal pull-backs (see Lemma 4.9), we may rewrite this as

$$\begin{aligned} \mathcal{L}'(\beta) &= f_*q\left(\prod_e \text{ev}_e^* \Delta_X \cdot \tilde{f}_*\tilde{q}\left(\prod_V \prod_{f \in I_V} \text{ev}_f^* \Delta_X \cdot \prod_V \prod_{W \in J_V} \mathcal{M}_W\right)\right) \\ &= f_*q\left(\text{ev}^* \Delta_X \cdot \prod_V \underbrace{f_{V*}q_V\left(\prod_{f \in I_V} \text{ev}_f^* \Delta_X \cdot \prod_{W \in J_V} \mathcal{M}_W\right)}_{=: \tilde{\mathcal{M}}_V}\right), \end{aligned} \tag{*}$$

where  $\text{ev}^* \Delta_X$  is the product over all pull-backs of the diagonals along evaluation maps belonging to edges of  $\alpha$ . But now all vertices  $V$  of  $\alpha$  are good by assumption, and therefore by Definition 3.11 the weights of all maximal cells in the gluing cycle  $\tilde{\mathcal{M}}_V$  agree with those of  $\mathcal{M}_V$ . Hence  $\tilde{\mathcal{M}}_V$  is an open subcycle of  $\mathcal{M}_V$ , and as (\*) is just the gluing construction for  $\alpha$  we conclude that every maximal cell of  $\mathcal{L}'(\beta)$  has the same weight in  $\mathcal{L}'(\alpha)$ . Of course, this property is preserved when passing to unbounded ends in  $\mathcal{L}(\beta)$  and  $\mathcal{L}(\alpha)$ .

In particular, if  $\alpha$  is of virtual codimension 1 and thus  $\beta$  of virtual codimension 0 in  $\mathcal{M}_0(X, \Sigma)$ , we see that the cycle  $\mathcal{M}_0(X, \Sigma)$  is locally around  $\alpha$  given by the cycle  $\mathcal{L}(\alpha)$ , and thus balanced.  $\square$

**Corollary 3.17** *If all vertices that can appear in combinatorial types in  $\mathcal{M}_0(X, \Sigma)$  are good, then  $\mathcal{M}_0(X, \Sigma)$  is a tropical variety of dimension*

$$\dim \mathcal{M}_0(X, \Sigma) = \dim X + |\Sigma| - 3 - K_X \cdot \Sigma$$

with the weights of Definition 3.15.



*Proof* Apply Theorem 3.16 to all combinatorial types of codimension 1. □

So in order to obtain a tropical variety  $\mathcal{M}_0(X, \Sigma)$  from gluing we will have to show that all vertices are good with respect to our given moduli data. The following result tells us that we only have to do this in resolution dimension 1.

**Corollary 3.18** *If all vertices  $V$  with  $\text{rdim}(V) = 1$  in a given moduli space are good, then all vertices are good.*

*Proof* Let  $V$  be a vertex in a given moduli space, with  $X_V \cong L_{rv}^{k_V} \times \mathbb{R}^{m_V}$ . We will prove the statement of the lemma by induction on the classification number  $c_V$ .

If  $\text{rdim}(V) = 0$  then  $V$  does not admit any resolution in  $\mathcal{M}_0(X, \Sigma)$  by Construction 3.10, and hence  $V$  is good. If  $\text{rdim}(V) = 1$  then  $V$  is good by assumption. We can therefore assume that  $\text{rdim}(V) > 1$ . By Definition 3.4 this means that  $\text{vdim}(V)$  is at least by 2 bigger than the dimension  $m_V$  of the lineality space in  $\mathcal{M}_0(X_V, \Sigma_V)$  coming from translations, and hence all combinatorial types  $\alpha$  of virtual codimension 1 in  $\mathcal{M}_0(X_V, \Sigma_V)$  correspond to (non-trivial) resolutions of  $V$ .

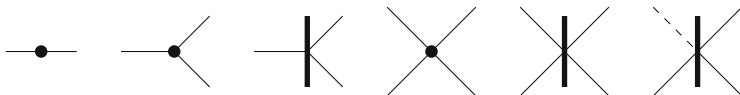
Condition (a) of Definition 3.11 of a good vertex now follows by induction on  $c_V$ , condition (b) by Theorem 3.16 applied to all these cells of virtual codimension 1, and condition (c) by Theorem 3.16 applied to all resolutions  $\alpha$  and maximal cells  $\beta > \alpha$ . □

Taking Corollaries 3.17 and 3.18 together, we thus see that in order to obtain a well-defined moduli space  $\mathcal{M}_0(X, \Sigma)$  we only have to check that all vertices of resolution dimension 1 are good.

### 4 Moduli Spaces of Lines in Surfaces

In this section we want to construct the moduli spaces  $\mathcal{M}_0(X, 1)$  of lines in a surface  $X \subset \mathbb{R}^3$ . By Corollary 3.17, the dimension of these spaces is the same as in the classical case, namely  $3 - \text{deg } X$ . So it is empty for  $\text{deg } X > 3$ , and we obtain a finite number of lines counted with multiplicities for  $\text{deg } X = 3$ .







Let us consider all possible local situations in such a surface. We want to use decorations on the graph of the line to describe the vertices, as introduced in [30]: A bold dot indicates that the line passes through a vertex of  $X$  and a bold line indicates that the line passes through an edge of  $X$ . This leads to the following combinatorial possibilities.



Here the difference between the last two decorations is that either one edge of the line lies on the edge of  $X$  (which is indicated by the dashed line in the last picture), or all edges of the line point into maximal cells of  $X$  (which is indicated by the second picture from the right). Note that the pictures above do not specify the

combinatorial type completely, as there are in general several possibilities for the directions of the ends.

*Remark 4.1 (Local Degrees)* If  $d = \deg X$  then  $K_X \cdot \Sigma = d$  for the degree  $\Sigma$  of a line. This means that every local degree  $K_{X_V} \cdot \Sigma_V$  at a vertex  $V$  can be at most  $d \leq 3$ . Moreover, as  $V$  has to be admissible, i.e., must satisfy  $\text{rdim}(V) \geq 0$ , Definition 3.4 (b) implies that  $\text{val}(V) \geq K_{X_V} \cdot \Sigma_V + 1$  for the bold dot decorations and  $\text{val}(V) \geq K_{X_V} \cdot \Sigma_V + 2$  for the bold line decorations. This leaves us with the following table, which lists the resolution dimensions of the possible types, and a name of the type in brackets. Impossible types are marked with “X”. (The case above type (H) is excluded since there would have to be a maximal cell of  $X$  containing two of the four edges, and hence  $K_{X_V} \cdot \Sigma_V$  cannot be 1.)

$K_{X_V} \cdot \Sigma_V$						
1	0 (A)	1 (B)	0 (D)	2 (E)	X	1 (I)
2	X	0 (C)	X	1 (F)	0 (H)	0 (J)
3	X	X	X	0 (G)	X	X

**Construction 4.2 (Moduli Data for Resolution Dimension 0)** For the vertices  $V$  of resolution dimension 0 in Remark 4.1, we have to define moduli data as in Definition 3.9. We will fix this according to the situation in algebraic geometry as follows. Assume first that  $V$  lies on a vertex of  $X$ , so that  $X_V \cong L_2^3$  after an integer linear isomorphism. Let  $\Sigma = (\sum_{i=0}^3 \alpha_i^j e_i)_{j=1, \dots, n}$  with  $\alpha_i^j \geq 0$  be the degree of  $\Sigma$ , where  $n = \text{val}(V)$ .

We then consider four planes  $H_0, H_1, H_2, H_3$  in  $\mathbb{P}^3$  in general position and count rational algebraic stable maps  $(C, x_1, \dots, x_n, f)$  relative to these planes with intersection profiles  $(\alpha_i^j)_{j=1, \dots, n}$  at  $H_i$  for all  $i$ . Their (finite) number will be the weight that we assign to the vertex  $V$ . In more complicated cases when this number is infinite, we expect that the corresponding (virtual) relative Gromov-Witten invariant would be the correct choice here, but for our situation at hand this problem does not occur.

If  $V$  lies on an edge of  $X$ , we assign a weight to  $V$  analogously after projecting  $X_V \cong L_1^2 \times \mathbb{R}$  to  $L_1^2$ .

Here are two examples:

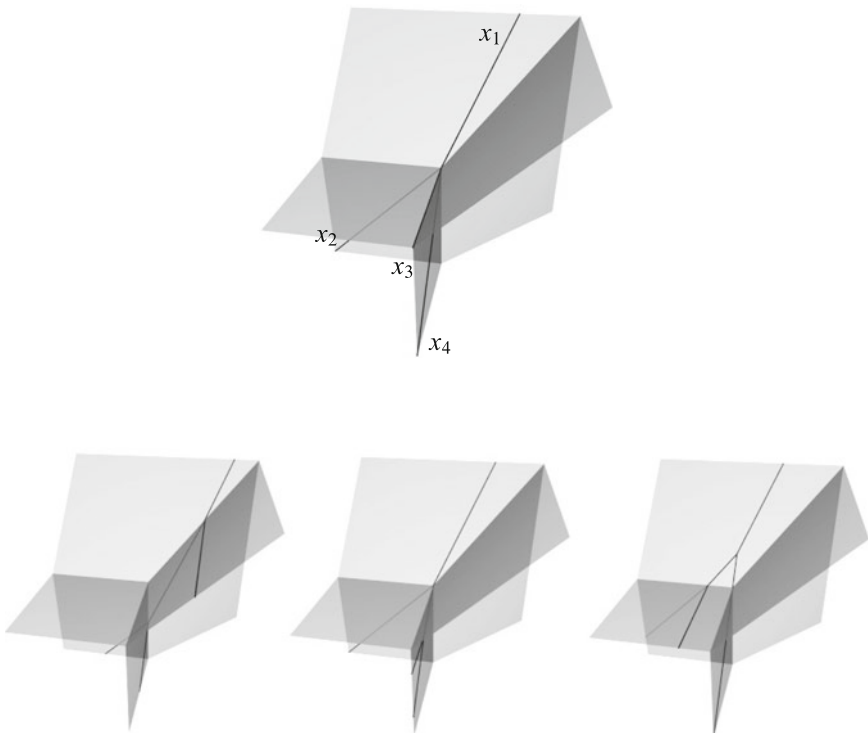
- (a) For type (A) in the table above the only possible degree is  $\Sigma = (e_0 + e_1, e_2 + e_3)$  up to permutations, corresponding to lines in  $\mathbb{P}^3$  passing to the two points  $H_0 \cap H_1$  and  $H_2 \cap H_3$ . As there is exactly one such line, we assign to (A) the weight 1.
- (b) For type (G) there are several possible degrees; as an example we will consider  $\Sigma = (3e_0 + 2e_1, e_1 + e_2, e_2 + e_3, e_2 + 2e_3)$ , and thus count maps with  $f^*H_0 = 3x_1, f^*H_1 = 2x_1 + x_2, f^*H_2 = x_2 + x_3 + x_4, f^*H_3 = x_3 + 2x_4$  (in the Chow groups of the corresponding  $f^{-1}(H_i)$ ). Such a map would have to send  $x_1$  to  $H_0 \cap H_1$ ,

and  $x_3$  and  $x_4$  to  $H_2 \cap H_3$ . As  $f^*H_0$  only contains  $x_1$  and  $f^*H_3$  only contains  $x_3$  and  $x_4$ , the curve would have to lie completely over the line through those two points. But then  $x_2$  would have to map to both of these points simultaneously, which is impossible. Hence we assign the weight 0 to this type.

*Remark 4.3 (Conditions in Resolution Dimension 1)* As the next step, we have to verify that, with the given moduli data, all vertices of resolution dimension 1 are good. To show the general procedure will sketch this here for type (F), more details and the other cases can be found in [24, Section 3.3].

Note that the rays in this type must satisfy exactly the same linear relation as the rays of a tropical line. Also, none of the three possible resolutions is allowed to have a bounded edge of higher weight, as this does not occur for lines. It is checked immediately that this leaves only the degree  $\Sigma = (2e_0 + e_1, e_1 + e_3, e_2, e_2 + e_3)$ , up to isomorphism. It is shown in the picture below, together with its three resolutions in  $X$ .

In order to embed the local moduli space into  $\mathcal{M}_{0,4} \times \mathbb{R}^3$ , we evaluate the position of  $x_3$  in  $\mathbb{R}^2 \cong \mathbb{R}^3 / \langle e_2 \rangle$  with coordinate directions  $e_1$  and  $e_3$ , and the position of  $x_2$  in  $\mathbb{R} \cong \mathbb{R}^3 / \langle e_1, e_3 \rangle$  with coordinate direction  $e_2$  (see Remark 2.21). Then the rays of the local moduli space are spanned by the vectors  $v_{\{1,3\}} + e_1 - e_3$ ,  $v_{\{1,4\}} + e_3$ , and  $v_{\{1,2\}} - e_1$  (in the notation of Construction 2.16) for the three resolutions below, respectively. This is balanced with weights one, which are actually the gluing weights.



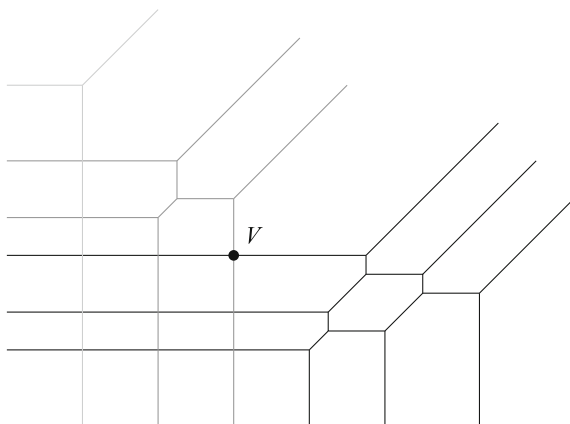
From Corollaries 3.17 and 3.18 we therefore conclude:

**Corollary 4.4** *With the moduli data of Construction 4.2, the moduli space  $\mathcal{M}_0(X, 1)$  of lines in a tropical surface  $X \subset \mathbb{R}^3$  is a tropical variety of dimension*

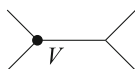
$$\dim \mathcal{M}_0(X, 1) = 3 - \deg X$$

*(and empty if  $\deg X > 3$ ). In particular, this moduli space consists of finitely many (weighted) points if  $\deg X = 3$ .*

*Example 4.5 (Infinitely Many Lines in a Tropical Cubic Surface [31])* Consider a floor-decomposed generic cubic surface where the three walls (represented by a line, a conic and a cubic) have the following relative position to each other, projected in the  $e_3$ -direction:



Such a cubic surface contains exactly 27 isolated lines that count with multiplicity 1. In addition, it has a 1-dimensional family of lines not containing any of the others. All lines in this family have a vertex mapping to the point  $V$  shown above, while the rest of them is mapped to maximal cells of  $X$ . General lines in this family are therefore decorated as in the following picture.



If the vertex mapped to  $V$  is 3-valent, it has to be of resolution dimension  $-1$ . Therefore the only admissible line in this family, i.e., the only line in the family actually in  $\mathcal{M}_0(X, 1)$ , is the one with no bounded edge, i.e., with a 4-valent vertex mapping to  $V$ . This vertex is then of resolution dimension 0. To determine its type, we map the four rays of the cubic at  $V$ , which are  $-e_1 + e_3, e_1 - 2e_3, -e_2 - 2e_3$ , and

$e_2 + 3e_3$ , by an integer linear isomorphism to the four unit vectors. This maps the rays of the line to the degree  $\Sigma = (3e_0 + 2e_1, e_1 + e_2, e_2 + e_3, e_2 + 2e_3)$ . As this line has weight 0 by Construction 4.2 (b), we conclude that the whole family does not contribute to the virtual number, and the degree of the 0-cycle  $\mathcal{M}_0(X, 1)$  is again 27.

We therefore conjecture:

*Conjecture 4.6* For every smooth cubic surface  $X \subset \mathbb{R}^3$  the 0-cycle  $\mathcal{M}_0(X, 1)$  has degree 27.

## Appendix: Pulling Back the Diagonal of a Smooth Variety

Let  $X$  be a partially open tropical cycle, and let  $Y$  be a smooth tropical variety. In order to glue moduli spaces in Sect. 3 we need the pull-back of the diagonal  $\Delta_Y$  of  $Y$  by some morphism  $f : X \rightarrow Y \times Y$ . But although the diagonal is locally a product of Cartier divisors in this case, tropical intersection theory unfortunately does not yet provide a well defined pull-back for it. This appendix therefore contains the technical details necessary to construct a well-defined pull-back cycle  $f^* \Delta_Y$ .

First we briefly review some facts about matroids and matroid fans from [14]. Let  $M$  be a matroid of rank  $r$  on a finite ground set  $E$ . To every flat  $F$  of  $M$  we associate a vector  $e_F := \sum_{i \in F} e_i \in \mathbb{R}^E$ , where the  $e_i$  are the negative standard basis vectors. Correspondingly, to every chain of flats  $\emptyset \subsetneq F_1 \subsetneq \dots \subsetneq F_s = E$  we assign a cone, spanned by  $e_{F_1}, \dots, e_{F_s}$ , and  $-e_{F_s}$ . Let  $B(M)$  denote the collection of all these cones, where the maximal ones are equipped with weight 1. This simplicial fan defines a tropical variety  $B(M)$  whose dimension is the rank of  $M$ . We call it the *matroid fan* associated to  $M$ .

Of special interest to us is the *uniform matroid*  $U_{r,k}$  on a ground set  $E$  of cardinality  $k$ , with rank function  $r(A) = \min(|A|, r)$ . Its associated matroid fan is  $B(U_{r,k}) \cong L_{r-1}^{k-1} \times \mathbb{R}$ .

**Construction 4.7** By François and Rau [14, Section 4] the diagonal  $\Delta_{B(M)}$  in  $B(M) \times B(M)$  can be cut out by a product of rational functions: we have  $\Delta_{B(M)} = \varphi_1 \cdot \dots \cdot \varphi_r \cdot (B(M) \times B(M))$  for the rational functions  $\varphi_i$  linear on the cones of  $B(M)$  determined by

$$\varphi_i(e_A, e_B) = \begin{cases} -1 & \text{if } r_M(A) + r_M(B) - r_M(A \cup B) \geq i, \\ 0 & \text{else} \end{cases}$$

for flats  $A, B$  of  $M$ , where  $r_M$  is the rank function of  $M$ . Moreover, recursively intersecting with the  $\varphi_i$  yields a matroid fan in each intermediate step, hence a locally irreducible tropical variety; this will be important in the construction. If we want to specify the matroid  $M$  in the notation, we will write  $\varphi_i$  also as  $\varphi_i^M$ .

We can now give the construction to pull back the diagonal from a smooth tropical fan.

**Construction 4.8** Consider a morphism  $f : X \rightarrow Y \times Y$  where  $Y \cong L_{r-1}^{k-1} \times \mathbb{R}^m$ . Then there is a (non-canonical) isomorphism  $\theta : Y \times \mathbb{R} \rightarrow \mathbf{B}(M) \times \mathbb{R}^m$ , where  $M = U_{r,k}$ , and  $\theta$  maps the additional factor  $\mathbb{R}$  onto the lineality space of the matroid fan. Associated to  $f$  we denote by  $\tilde{f}$  the composition map

$$X \times \mathbb{R}^2 \xrightarrow{f \times \text{id}} Y \times Y \times \mathbb{R}^2 \cong (Y \times \mathbb{R}) \times (Y \times \mathbb{R}) \xrightarrow{\theta \times \theta} \mathbf{B}(M)^2 \times (\mathbb{R}^m)^2.$$

Let  $\psi_1, \dots, \psi_m$  denote functions which cut out the diagonal  $\Delta_{\mathbb{R}^m}$ , and consider the cocycle

$$\Phi_Y := \varphi_1 \cdot \dots \cdot \varphi_r \cdot \psi_1 \cdot \dots \cdot \psi_m \quad \text{on } \mathbf{B}(M)^2 \times (\mathbb{R}^m)^2,$$

where  $\varphi_i$  are the functions on  $\mathbf{B}(M)^2$  from Construction 4.7 above. One verifies immediately that the pull-back  $\tilde{f}^* \Phi_Y \cdot (X \times \mathbb{R}^2)$  has the lineality space  $L_X := 0 \times \Delta_{\mathbb{R}}$  in  $X \times \mathbb{R}^2$ . So we can take the quotient by  $L_X$  and use the projection  $p_X : (X \times \mathbb{R}^2)/L \rightarrow X$  to define

$$f^* \Delta_Y := f^* \Delta_Y \cdot X := p_{X*} [(\tilde{f}^* \Phi_Y \cdot (X \times \mathbb{R}^2))/L_X].$$

As the intermediate steps in Construction 4.7 are locally irreducible, it follows from [12, Lemma 3.8.13] that the support of  $\tilde{f}^* \Phi_Y \cdot (X \times \mathbb{R}^2)$  lies over the diagonal of  $\mathbb{R}$  in  $\mathbb{R}^2$  (so that  $p_X$  is injective on  $(\tilde{f}^* \Phi_Y \cdot (X \times \mathbb{R}^2))/L_X$ , in accordance with our convention in Remark 2.12), and that the support of  $f^* \Delta_Y$  lies in  $f^{-1}(\Delta_Y)$ .

Moreover, it can be shown that this definition depends neither on the choice of  $\psi_1, \dots, \psi_m$  [13, Theorem 2.25] nor on the choice of isomorphism  $\theta$  [24, Lemma 1.4.3]. However, it is not known whether it depends on the choice of the rational functions  $\varphi_1, \dots, \varphi_r$  cutting out the diagonal of  $\mathbf{B}(M)$ .

Let us briefly state the main properties of this definition that follow from the compatibilities between the various intersection-theoretic constructions.

**Lemma 4.9**

(a) (Projection formula) For two morphisms  $Z \xrightarrow{g} X \xrightarrow{f} Y \times Y$ , where  $g$  is injective and  $Y$  a smooth fan, we have

$$g_* [(f \circ g)^* \Delta_Y \cdot Z] = f^* \Delta_Y \cdot g_* Z.$$

(b) (Quotients) Let  $X$  be a partially open tropical variety with lineality space  $L$  and quotient map  $q : X \rightarrow X/L$ , and let  $f : X/L \rightarrow Y \times Y$  be a morphism for a smooth fan  $Y$ . Then

$$q((f \circ q)^* \Delta_Y \cdot X) = f^* \Delta_Y \cdot (X/L).$$

(c) (Commutativity) For two morphisms  $f : X \rightarrow Y \times Y$  and  $g : X \rightarrow Z \times Z$  to smooth fans  $Y$  and  $Z$  we have

$$f^* \Delta_Y \cdot (g^* \Delta_Z \cdot X) = g^* \Delta_Z \cdot (f^* \Delta_Y \cdot X).$$

(d) (Projections) Let  $X$  and  $Z$  be partially open tropical varieties, and denote by  $p : X \times Z \rightarrow X$  the projection. For any morphism  $f : X \rightarrow Y \times Y$  for a smooth fan  $Y$  we have

$$(f \circ p)^* \Delta_Y \cdot (X \times Z) = (f^* \Delta_Y \cdot X) \times Z.$$

(e) (Products) Let  $f : X \rightarrow Y \times Y$  and  $f' : X' \rightarrow \mathbb{R}^k \times \mathbb{R}^k$  be two morphisms, for a smooth fan  $Y$ . Then

$$(f \times f')^* \Delta_{Y \times \mathbb{R}^k} \cdot (X \times X') = (f^* \Delta_Y \cdot X) \times (f'^* \Delta_{\mathbb{R}^k} \cdot X').$$

*Proof* All these statements can be checked immediately, see [24, Section 1.4] for details. As an example, we show part (a): We have

$$\begin{aligned} g_* [(f \circ g)^* \Delta_Y \cdot Z] &= g_* p_{Z*} \left[ (\widetilde{f \circ g}^* \Phi_Y \cdot (Z \times \mathbb{R}^2)) / L_Z \right] \\ &= p_{X*} (g \times \text{id})_* \left[ ((g \times \text{id})^* \widetilde{f}^* \Phi_Y \cdot (Z \times \mathbb{R}^2)) / L_Z \right] \quad (\text{functoriality}) \\ &= p_{X*} \left[ ((g \times \text{id})_* (g \times \text{id})^* \widetilde{f}^* \Phi_Y \cdot (Z \times \mathbb{R}^2)) / L_X \right] \quad (\text{Lemma 2.15}) \\ &= p_{X*} \left[ (\widetilde{f}^* \Phi_Y \cdot (g_*(Z \times \mathbb{R}^2))) / L_X \right] \quad (\text{projection formula}) \\ &= f^* \Delta_Y \cdot g_*(Z), \end{aligned}$$

where we have used the projection formula for cocycles as in [13, Proposition 2.24 (3)]. □

So far  $Y \cong L_{r-1}^{k-1} \times \mathbb{R}^m$  was assumed to be a smooth fan. In order to generalize this to smooth varieties we need the following compatibility statement. Let  $\sigma$  be a cell of  $Y$  in the coarsest subdivision, with relative interior  $\sigma^\circ$ . We consider both  $\sigma$  and  $\sigma^\circ$  as subsets of the diagonal  $\Delta_Y \subset Y \times Y$ . Let  $Y(\sigma)$  be the union of all open cells in the matroid subdivision of  $Y \times Y$  whose closure intersects  $\sigma^\circ$ , which is then an open neighborhood of  $\sigma^\circ$  in  $Y \times Y$ . It is also contained in  $Y_\sigma \times Y_\sigma$ , where  $Y_\sigma$  is the star of  $Y$  at  $\sigma$ . As  $Y_\sigma$  is again a smooth fan, we can regard the restriction of  $f$  to  $X_\sigma := f^{-1}(Y(\sigma))$  also as a morphism  $f_\sigma : X_\sigma \rightarrow Y_\sigma \times Y_\sigma$ , and apply Construction 4.8 to this map. As expected, we will now show that over the cell  $\sigma$  this gives the same result as for  $f : X \rightarrow Y \times Y$ .

**Lemma 4.10 (Compatibility)** *Let  $f : X \rightarrow Y \times Y$  be a morphism, with  $Y$  a smooth fan. Moreover, let  $\sigma$  be a cell in the coarsest subdivision of  $Y$ . With notations as above, the weights of  $f^* \Delta_Y \cdot X$  and  $f_\sigma^* \Delta_{Y_\sigma} \cdot X_\sigma$  then agree on all cells of  $X$  whose interior is mapped by  $f$  to the cell  $\sigma^\circ$  in the diagonal  $\Delta_Y$ .*

*Proof* Let  $Y \cong L_{r-1}^{k-1} \times \mathbb{R}^m$ , and let  $M = U_{r,k}$  be the corresponding uniform matroid on  $E = \{1, \dots, k\}$ , so that  $Y \times \mathbb{R} \cong B(M) \times \mathbb{R}^m$ . The cell  $\sigma$  then corresponds to a subset  $S \subset E$ , i.e., it consists of all cells in the matroid subdivision for chains of flats in  $S$ . If  $\dim \sigma = s$  then  $Y_\sigma \cong L_{r-s-1}^{k-s-1} \times \mathbb{R}^{m+s}$ , or more precisely  $Y_\sigma \times \mathbb{R} \cong B(M_S) \times \mathbb{R}^s \times \mathbb{R}^m$ , where  $M_S$  is the uniform matroid of rank  $r - s$  on  $E \setminus S$ .

As Construction 4.8 is local, it suffices to show that the rational functions cutting out the diagonal in this construction are the same for the spaces  $Y \times Y \times \mathbb{R}^2 \cong B(M)^2 \times (\mathbb{R}^m)^2$  and  $Y_\sigma \times Y_\sigma \times \mathbb{R}^2 \cong B(M_S)^2 \times (\mathbb{R}^s)^2 \times (\mathbb{R}^m)^2$  when restricted to the common open neighborhood  $Y(\sigma) \times \mathbb{R}^2$  of the cell  $\sigma^\circ \times \mathbb{R}$  in the diagonal. For this we need to prove that these rational functions agree on all rays of  $Y(\sigma)$ . As  $(e_S, e_S)$  is the only interior ray of  $\sigma^\circ \times \mathbb{R}$ , the maximal cells of  $Y(\sigma)$  correspond to maximal chains of flats in  $M \oplus M$  containing  $(S, S)$ , and hence we have to compare the rational functions of Constructions 4.7 and 4.8 on all rays  $(e_A, e_B)$  for flats  $A, B$  of  $M$  with  $A, B \subset S$  or  $A, B \supset S$ .

The functions cutting out the diagonal of  $\mathbb{R}^m$  can obviously be chosen to be the same in both cases. By Construction 4.7, the others are  $\varphi_1^M, \dots, \varphi_k^M$  for  $B(M)^2$ , and  $\varphi_1^{M_S}, \dots, \varphi_{k-s}^{M_S}$  and  $\varphi_1^S, \dots, \varphi_s^S$  for  $B(M_S)^2 \times (\mathbb{R}^s)^2$  (where  $S$  stands for the uniform matroid of full rank on  $S$ , so that  $\varphi_1^S, \dots, \varphi_s^S$  can be used to cut out the diagonal of  $\mathbb{R}^s$ ). Now, on the rays  $(e_A, e_B)$  mentioned above ...

- $\varphi_i^M$  agrees with  $\varphi_{i-s}^{M_S}$  for  $i = s + 1, \dots, k$ :  
 If  $A, B \subset S$  then both  $r_M(A) + r_M(B) - r_M(A \cup B) \geq i$  and  $r_{M_S}(A \setminus S) + r_{M_S}(B \setminus S) - r_{M_S}((A \cup B) \setminus S) \geq i - s$  are never satisfied.  
 If  $A, B \supset S$  then  $r_M(A) = r_{M_S}(A \setminus S) + s$ , and similarly for  $B$  and  $A \cup B$ . Hence  $r_M(A) + r_M(B) - r_M(A \cup B) \geq i$  is equivalent to  $r_{M_S}(A \setminus S) + r_{M_S}(B \setminus S) - r_{M_S}((A \cup B) \setminus S) \geq i - s$ .
- $\varphi_i^M$  agrees with  $\varphi_i^S$  for  $i = 1, \dots, s$ :  
 If  $A, B \subset S$  then  $r_M(A) + r_M(B) - r_M(A \cup B) \geq i$  is equivalent to  $r_S(A \cap S) + r_S(B \cap S) - r_S((A \cup B) \cap S) \geq i$ .  
 If  $A, B \supset S$  then both  $r_M(A) + r_M(B) - r_M(A \cup B) \geq i$  and  $r_S(A \cap S) + r_S(B \cap S) - r_S((A \cup B) \cap S) \geq i$  are always satisfied. □

*Remark 4.11 (Pullbacks of Diagonals of Smooth Varieties)* Lemma 4.10 implies that we cannot only pull back diagonals of smooth fans, but also of smooth varieties: Let  $f : X \rightarrow Y \times Y$  be a morphism from a partially open tropical cycle to a smooth tropical variety.

To assign a weight to a cell  $\tau$  of dimension  $\dim X - \dim Y$  over the diagonal  $\Delta_Y$ , let  $\sigma$  be the cell in  $Y \cong \Delta_Y$  so that the relative interior of  $\tau$  maps to the relative interior of  $\sigma$ . Choose any face  $\sigma'$  of  $\sigma$  (which might be  $\sigma$  itself), replace  $Y$  by the star  $Y_{\sigma'}$  at this face and  $X$  by the open subcycle of  $X$  consisting of all points mapping to  $\sigma'$  or any of its adjacent open cells in both components of  $Y \times Y$ , and assign to  $\tau$  its weight in the cycle  $f_{\sigma'}^* \Delta_{Y_{\sigma'}} \cdot X_{\sigma'}$ . By Lemma 4.10 the result does not depend on the choice of  $\sigma'$ .

Using the same local procedure for a cell  $\tau$  of dimension  $\dim X - \dim Y - 1$ , we obtain a balanced cycle  $f_{\sigma'}^* \Delta_{Y_{\sigma'}} \cdot X_{\sigma'}$  including all adjacent cells of dimension  $\dim X - \dim Y$ . Hence our local construction glues to give a well-defined cycle  $f^* \Delta_Y \cdot X$ .



## References

1. L. Allermann, Tropical intersection products on smooth varieties. *J. Eur. Math. Soc.* **14**(1), 107–126 (2012)
2. L. Allermann, J. Rau, First steps in tropical intersection theory. *Math. Z.* **264**(3), 633–670 (2010)
3. K. Behrend, Gromov-Witten invariants in algebraic geometry. *Invent. Math.* **127**(3), 601–617 (1997)
4. K. Behrend, B. Fantechi, The intrinsic normal cone. *Invent. Math.* **128**, 45–88 (1997)
5. K. Behrend, Y. Manin, Stacks of stable maps and Gromov-Witten invariants. *Duke Math. J.* **85**(1), 1–60 (1996)
6. B. Bertrand, E. Brugallé, G. Mikhalkin, Tropical open Hurwitz numbers. *Rend. Semin. Math. Univ. Padova* **125**, 157–171 (2011)
7. E. Brugallé, H. Markwig, Deformation of tropical Hirzebruch surfaces and enumerative geometry. *J. Algebraic Geom.* **25**(4), 633–702 (2016)
8. L. Caporaso, Gonality of algebraic curves and graphs. *Springer Proc. Math. Stat.* **71**, 77–108 (2014)
9. L. Caporaso, J. Harris, Counting plane curves of any genus. *Invent. Math.* **131**, 345–392 (1998)
10. R. Cavalieri, H. Markwig, D. Ranganathan, Tropicalizing the space of admissible covers. *Math. Ann.* **364**, 1275–1313 (2016)
11. E.M. Feichtner, B. Sturmfels, Matroid polytopes, nested sets and Bergman fans. *Port. Math.* **62**, 437–468 (2005)
12. G. François, Tropical intersection products and families of tropical curves. Ph.D. Thesis, TU Kaiserslautern, 2012
13. G. François, Cocycles on tropical varieties via piecewise polynomials. *Proc. Am. Math. Soc.* **141**, 481–497 (2013)
14. G. François, J. Rau, The diagonal of tropical matroid varieties and cycle intersections. *Collect. Math.* **64**(2), 185–210 (2013)
15. W. Fulton, R. Pandharipande, Notes on stable maps and quantum cohomology, in *Algebraic Geometry, Santa Cruz 1995*, ed. by J.K. et al. Proceedings of Symposia in Pure Mathematics, vol. 62 (American Mathematical Society, Providence, 1997), pp. 45–96
16. A. Gathmann, M. Kerber, H. Markwig, Tropical fans and the moduli space of rational tropical curves. *Compos. Math.* **145**(1), 173–195 (2009)
17. A. Gathmann, H. Markwig, D. Ochse, Tropical moduli spaces of stable maps to a curve, in *Algorithmic and Experimental Methods in Algebra, Geometry, and Number Theory*, ed. by G. Böckle, W. Decker, G. Malle (Springer, Heidelberg, 2018). [https://doi.org/10.1007/978-3-319-70566-8\\_12](https://doi.org/10.1007/978-3-319-70566-8_12)
18. E. Gawrilow, M. Joswig, polymake: a framework for analyzing convex polytopes, in *Polytopes: Combinatorics and Computation*, ed. by G. Kalai, G.M. Ziegler (Birkhäuser, Basel, 2000), pp. 43–74
19. S. Hampe, A-tint: a polymake extension for algorithmic tropical intersection theory. *Eur. J. Comb.* **36C**, 579–607 (2014)
20. M. Kontsevich, Enumeration of rational curves via torus actions, in *The Moduli Space of Curves (Texel Island 1994)* (Birkhäuser, Basel, 1995), pp. 335–368
21. G. Mikhalkin, Enumerative tropical geometry in  $\mathbb{R}^2$ . *J. Am. Math. Soc.* **18**, 313–377 (2005)
22. G. Mikhalkin, Tropical geometry and its applications, in *International Congress of Mathematicians*, vol. II (European Mathematical Society, Zürich, 2006), pp. 827–852
23. G. Mikhalkin, Moduli spaces of rational tropical curves, in *Proceedings of Gökova Geometry-Topology Conference (GGT) 2006* (2007), pp. 39–51
24. D. Ochse, Moduli spaces of rational tropical stable maps into smooth tropical varieties. Ph.D. Thesis, TU Kaiserslautern, 2013
25. J. Richter-Gebert, B. Sturmfels, T. Theobald, First steps in tropical geometry, in *Idempotent Mathematics and Mathematical Physics, Proceedings Vienna* (2003)

26. K. Shaw, Tropical intersection theory and surfaces. Ph.D. Thesis, Université de Genève, 2011
27. K. Shaw, A tropical intersection product in matroidal fans. *SIAM J. Discrete Math.* **27**(1), 459–491 (2013)
28. D. Speyer, B. Sturmfels, The tropical Grassmannian. *Adv. Geom.* **4**, 389–411 (2004)
29. The GAP Group, GAP – Groups, Algorithms, and Programming, Version 4.8.6 (2016). <http://www.gap-system.org>
30. M. Vigeland, Tropical lines on smooth tropical surfaces (2007). ArXiv: 0708.3847
31. M.D. Vigeland, Smooth tropical surfaces with infinitely many tropical lines. *Ark. Mat.* **48**(1), 177–206 (2010)
32. T.Y. Yu, Tropicalization of the moduli space of stable maps. *Math. Z.* **281**(3), 1035–1059 (2015)

# Tropical Moduli Spaces of Stable Maps to a Curve



Andreas Gathmann, Hannah Markwig, and Dennis Ochse

**Abstract** We construct moduli spaces of rational covers of an arbitrary smooth tropical curve in  $\mathbb{R}^r$  as tropical varieties. They are contained in the balanced fan parametrizing tropical stable maps of the appropriate degree to  $\mathbb{R}^r$ . The weights of the top-dimensional polyhedra are given in terms of certain lattice indices and local Hurwitz numbers.

**Keywords** Tropical geometry • Enumerative geometry • Gromov-Witten theory

**Subject Classifications** 14T05, 14N35, 51M20

## 1 Introduction

Tropical enumerative geometry has developed from interesting applications following so-called correspondence theorems which settle the equality of certain enumerative numbers in algebraic geometry to their tropical counterparts [22]. There is an ongoing effort to put the striking similarities between algebro-geometric and tropical enumerative geometry onto a more solid ground.

Modern enumerative algebraic geometry is based on the moduli spaces  $\overline{M}_{g,n}(X, \beta)$  of  $n$ -pointed stable maps of genus  $g$  and class  $\beta$  to a smooth projective variety  $X$  [15], together with their virtual fundamental classes [3, 4] that resolve the issues arising when these spaces are not of the expected dimension. Hence a key ingredient for the further development of tropical enumerative geometry is the

---

A. Gathmann • D. Ochse

Fachbereich Mathematik, Technische Universität Kaiserslautern, Postfach 3049,  
67653 Kaiserslautern, Germany

e-mail: [andreas@mathematik.uni-kl.de](mailto:andreas@mathematik.uni-kl.de); [ochse@mathematik.uni-kl.de](mailto:ochse@mathematik.uni-kl.de)

H. Markwig (✉)

Fachbereich Mathematik, Eberhard Karls Universität Tübingen, Auf der Morgenstelle 10,  
72076 Tübingen, Germany

e-mail: [hannah@math.uni-tuebingen.de](mailto:hannah@math.uni-tuebingen.de)

construction of tropical analogues of these concepts. If  $g = 0$  and  $X$  is a toric variety, corresponding to rational tropical curves in  $\mathbb{R}^r$ , such tropical spaces have been constructed as balanced fans in [17]. In this case, ideas relating to virtual fundamental classes are not needed, and the intersection theory of the resulting spaces recovers the correspondence theorems for rational tropical curves in  $\mathbb{R}^r$  [19].

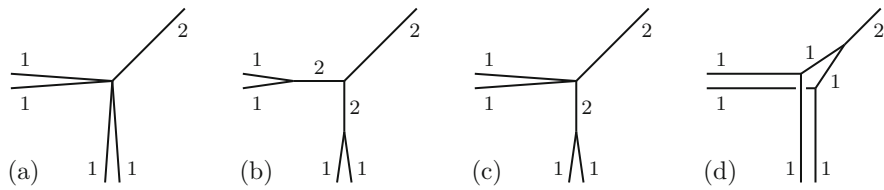
For more general target spaces, we run into the same problems as in algebraic geometry: the naively defined spaces of tropical curves in a tropical variety are usually not of the expected dimension, maybe not even pure-dimensional. However, as there is no general theory of virtual fundamental classes in tropical geometry yet, the tropical approach to this problem is different: right from the start we have to construct the moduli spaces as balanced polyhedral complexes of the expected dimension—which necessarily means that they are *not* just the spaces of maps from a tropical curve to the given target. From an algebro-geometric point of view, one could say that this constructs the moduli space and its virtual fundamental class at the same time, with the additional benefit that (in accordance with the general philosophy of tropical intersection theory) we actually obtain a virtual *cycle* and not just a *cycle class*.

A general approach how this idea might be realized has been presented in [16]. Here, we will restrict ourselves to the case when  $g = 0$  and the target is a smooth (rational) tropical curve  $L$  in  $\mathbb{R}^r$ . The resulting moduli spaces  $\mathcal{M}_{0,n}(L, \Sigma)$  (where  $\Sigma$  is a degree of tropical curves as in Definition 2.4) then describe rational covers of a rational smooth tropical curve.

Tropical covers and tropical Hurwitz numbers (i.e. enumerative numbers counting covers with prescribed properties [5, 10]) are useful e.g. for the study of the structural behavior of Hurwitz numbers [11] and in the tropical enumeration of Zeuthen numbers [6]. Spaces of tropical (admissible) covers have been studied in [12] as tropicalizations of corresponding algebro-geometric spaces, in terms of a tropicalization map on the Berkovich analytification. The space of tropical covers of  $\mathbb{R}$  has been described in [13] as tropicalization of the open part of a suitable space of relative stable maps (whose compactification is then realized as a tropical compactification defined by the tropical moduli space). The present work complements this point of view by fixing a rational smooth tropical curve  $L \subset \mathbb{R}^r$ , restricting to genus 0 covers, and embedding the abstract polyhedral subcomplex of the abstract cone complex described in [12] as a balanced polyhedral subcomplex. In this way, we make these moduli spaces accessible to the current state of the art of tropical intersection theory.

As mentioned above, to construct the moduli spaces  $\mathcal{M}_{0,n}(L, \Sigma)$  we cannot just take the subset of  $\mathcal{M}_{0,n}(\mathbb{R}^r, \Sigma)$  consisting of all stable maps whose image lies in  $L$ , as this would yield a non-pure subcomplex with strata of too big dimension. Instead, we have to incorporate the so-called Riemann-Hurwitz condition (see Definition 3.2), which implies the algebraic realizability of the corresponding maps. For an example, let  $L \subset \mathbb{R}^2$  be the standard tropical line, let  $\Sigma$  be the degree consisting of the directions  $(-1, 0)$ ,  $(0, -1)$ ,  $(0, -1)$ ,  $(0, -1)$ ,  $(2, 2)$ , and set  $n = 0$ . A fan curve in  $L$  of this degree—in fact representing the origin of the fan  $\mathcal{M}_{0,0}(L, \Sigma)$ —is shown in picture (a) below. It is given by a map from an abstract

star curve with 5 ends to  $L$ , with the directions and weights on the ends as indicated in the picture.



Possible resolutions of this curve in  $L$  are shown in (b), (c), and (d). However, case (b) is excluded in  $\mathcal{M}_{0,0}(L, \Sigma)$  as its central vertex violates the Riemann-Hurwitz condition: it would correspond to an algebraic degree-2 cover of the projective line by itself with three ramification points of order 2, which does not exist. In contrast, the combinatorial types (c) and (d) are allowed, and represent two rays in  $\mathcal{M}_{0,0}(L, \Sigma)$  since they describe 1-dimensional families of curves. They both have a similar type obtained by symmetry: in (c) the bounded weight-2 edge could also be on the horizontal edge of  $L$ , and in (d) there are two choices how to group the weight-1 ends. In total, this means that  $\mathcal{M}_{0,0}(L, \Sigma)$  is a 1-dimensional fan with four rays. The weights that we will construct on these rays incorporate the triple Hurwitz numbers corresponding to the local degrees of the maps at each point mapping to the vertex of  $L$ ; they all turn out to be 1 here. In this example, it is then easy to check explicitly that  $\mathcal{M}_{0,0}(L, \Sigma) \subset \mathcal{M}_{0,0}(\mathbb{R}^2, \Sigma) \cong \mathcal{M}_{0,5} \times \mathbb{R}^2$  is indeed balanced. Our main result on the moduli spaces  $\mathcal{M}_{0,n}(L, \Sigma)$  is that this construction works in general:

**Theorem 1.1** *Let  $L$  be a smooth tropical curve in  $\mathbb{R}^r$  and  $\Sigma$  a degree of tropical stable maps to  $L$  (see Definitions 2.4 and 3.3). Then the space  $\mathcal{M}_{0,n}(L, \Sigma)$  (with weights defined in terms of local Hurwitz numbers) is a balanced weighted polyhedral subcomplex of  $\mathcal{M}_{0,n}(\mathbb{R}^r, \Sigma)$  of pure dimension*

$$|\Sigma| - \deg(\Sigma) \cdot \left( \sum_{W \in L} (\text{val}(W) - 2) \right) - 2.$$

We expect that  $\mathcal{M}_{0,n}(L, \Sigma)$  is in fact the tropicalization of (relevant parts) of the corresponding algebro-geometric moduli space.

Theorem 1.1 is proved in two major steps: the first being the treatment of 1-dimensional moduli spaces of the form above (see Theorem 4.3), and the second the generalization to arbitrary dimension. For the generalization to arbitrary dimension, we use a general gluing construction for tropical moduli spaces which was developed by the first and last author in [16] and has further applications to other target spaces.

This paper is organized as follows. In Sect. 2 we review the necessary preliminaries. The tropical moduli spaces  $\mathcal{M}_{0,n}(L, \Sigma)$  are then defined in Sect. 3. More

precisely, we define their structure as a polyhedral subcomplex of  $\mathcal{M}_{0,n}(\mathbb{R}^r, \Sigma)$  in Sect. 3.1, and the weights of their maximal cells in Sect. 3.2. The definition of the weights relies on the gluing construction of [16], which we therefore review in Sect. 3.2, together with the main result of [16] allowing a gluing construction of tropical moduli spaces under some requirements. In our case, these requirements are satisfied if all one-dimensional tropical moduli spaces  $\mathcal{M}_{0,n}(L, \Sigma)$  are balanced fans. We prove this fact in Sect. 4 (see Theorem 4.3). Theorem 1.1 is then an immediate consequence of the foundational work on the gluing construction in [16].

## 2 Preliminaries

### 2.1 Background on Tropical Varieties and Intersection Theory

To fix notation, we quickly review notions of tropical intersection theory. Some of our constructions involve partially open versions of tropical varieties, i.e. varieties containing polyhedra that are open at some faces. We adapt the usual conventions to this situation. For a more detailed survey of the relevant preliminaries, see [16, section 2.1].

We let  $\Lambda$  be a lattice in an  $r$ -dimensional real vector space  $V$ . A (*partially open*) (*rational*) *polyhedron* in  $V$  is a finite intersection of (open or) closed affine half-spaces given by (strict or) non-strict inequalities whose linear parts are given by elements in the dual of  $\Lambda$ . We denote by  $V_\sigma$  the linear space obtained by shifting the affine span of  $\sigma$  to the origin and define  $\Lambda_\sigma := V_\sigma \cap \Lambda$ . A *face*  $\tau \leq \sigma$  (also written as  $\tau < \sigma$  if it is proper) is a non-empty subset of  $\sigma$  that can be obtained by changing some of the defining non-strict inequalities into equalities. If  $\dim \tau = \dim \sigma - 1$  we call  $\tau$  a *facet* of  $\sigma$ . In this case we denote by  $u_{\sigma/\tau} \in \Lambda_\sigma / \Lambda_\tau$  the *primitive normal vector* of  $\sigma$  relative to  $\tau$ , i.e. the unique generator of  $\Lambda_\sigma / \Lambda_\tau$  lying in the half-line of  $\sigma$  in  $V_\sigma / V_\tau \cong \mathbb{R}$ . The well-known notion of a (pure-dimensional) *weighted polyhedral complex*  $X$  (formed by cells  $\sigma$  as above, and with integer weights on maximal cells), its dimension and support are easily adapted to the case of partially open polyhedral complexes. Such a (partially open) weighted polyhedral complex  $(X, \omega)$  is called a (*partially open*) *tropical variety* (or *cycle*, if negative weights occur) if it satisfies the balancing condition, i.e. for each cell  $\tau$  of codimension 1 we have

$$\sum_{\sigma:\sigma>\tau} \omega(\sigma) \cdot u_{\sigma/\tau} = 0 \quad \in V/V_\tau.$$

For intersection-theoretic purposes, the exact polyhedral complex structure is often not important, and we fix it only up to refinements respecting the weights.

*Example 2.1 (Smooth Curves)* Let  $V = \mathbb{R}^q$ . We let  $L_1^q$  denote the 1-dimensional tropical variety containing the origin and rays spanned by  $-e_i$  (where  $e_i$  denotes

the canonical basis vectors) and  $-e_0 := \sum e_i$ , with all weights one. This is the tropicalization of a general line over the Puiseux series with constant coefficient equations [14, proposition 2.5 and theorem 4.1]. A one-dimensional tropical variety  $L \subset \mathbb{R}^r$  with all weights one is called a *rational smooth curve* if its underlying polyhedral complex is rational (i.e. combinatorially a tree), and if it locally at each vertex equals  $L_1^q$  up to a *unimodular transformation*, i.e. up to an isomorphism of vector spaces which is also an isomorphism of the underlying lattices [1].

Some of our constructions involve *quotients*  $X/W$  of partially open tropical varieties  $X$  by a *lineality space*  $W$ . We say that a vector subspace  $W$  of  $V$  is a lineality space for  $X$  if for all  $\sigma \in X$  and  $x \in \sigma$  the intersection  $\sigma \cap (x + L)$  is open in  $x + L$  and equal to  $|X| \cap (x + L)$ . Note that for the case of a closed polyhedral complex this generalizes the usual notion of a lineality space (which is commonly the maximal subspace with this property). For more details on such quotients, see [16, section 2.2].

A *morphism* between (partially open) tropical cycles  $X$  and  $Y$  is a map  $f : |X| \rightarrow |Y|$  which is locally affine linear, with the linear part induced by a map between the underlying lattices [2, definition 7.1]. A *rational function* on a tropical variety  $X$  is a continuous function  $\varphi : |X| \rightarrow \mathbb{R}$  that is affine linear on each cell, and whose linear part is integer, i.e. in the dual of the lattice. We associate a *divisor*  $\varphi \cdot X$  to a rational function; a cycle of codimension 1 in  $X$  support on the cells at which  $\varphi$  is not locally linear [2, construction 3.3]. Multiple intersection products  $\varphi_1 \cdot \dots \cdot \varphi_m \cdot X$  are commutative by Allermann and Rau [2, proposition 3.7].

*Remark 2.2 (Weights of Intersections as Lattice Indices)* Often, the weight of a cell of a multiple intersection product can be computed locally in terms of a lattice index. To do this, we write locally  $\varphi_i = \max\{h_i, 0\}$  for linearly independent integer linear functions  $h_1, \dots, h_m$ , and let  $H$  be a matrix representing the integer linear map  $\Lambda \rightarrow \mathbb{Z}^m : x \mapsto (h_1(x), \dots, h_m(x))$ . Then the local weight of  $\varphi_1 \cdot \dots \cdot \varphi_m \cdot X$  equals the greatest common divisor of the maximal minors of  $H$  [21, lemma 5.1].

Rational functions can be pulled back along a morphism  $f : X \rightarrow Y$  to rational functions  $f^*(\varphi) = \varphi \circ f$  on  $X$ . We can push forward a subvariety  $Z$  of  $X$  to a subvariety  $f_*(Z)$  of  $Y \subset A' \otimes_{\mathbb{Z}} \mathbb{R}$  [2, proposition 4.6 and corollary 7.4]: For suitable refinements of the polyhedral structures of  $X$  and  $Y$ , we obtain  $f(\sigma) \in Y$  for all  $\sigma \in X$ , and define the weight of the push-forward to be

$$\omega_{f_*(Z)}(\sigma') := \sum_{\sigma} \omega_X(\sigma) \cdot |\Lambda'_{\sigma'} / f(\Lambda_{\sigma})|,$$

where the sum goes over all top-dimensional cells  $\sigma \in Z$  with  $f(\sigma) = \sigma'$ . In the partially open case, we will restrict ourselves to injective morphisms in order to avoid problems with overlapping cells with different boundary behavior.

## 2.2 Tropical Moduli Spaces of Curves

An (abstract)  $N$ -marked rational tropical curve is a tuple  $(\Gamma, x_1, \dots, x_N)$ , where  $\Gamma$  is a metric tree with  $N$  unbounded edges labeled  $x_1, \dots, x_N$  (also called *marked ends*) that have infinite length, and such that the valence of each vertex is at least 3. The set of all  $N$ -marked tropical curves is denoted  $\mathcal{M}_{0,N}$ . It follows from [25, theorem 3.4], [23, section 2], or [17, theorem 3.7] that  $\mathcal{M}_{0,N}$  can be embedded as a tropical variety via the distance map, more precisely, as a balanced, simplicial,  $(N - 3)$ -dimensional fan whose top-dimensional cones all have weight one. The distance map sends a tropical curve to the vector of distances of its ends in  $\mathbb{R}^{\binom{N}{2}}$ . We mod out an  $N$ -dimensional lineality space  $U_N$ , identifying vectors corresponding to trees whose metrics only differ on the ends. For a tree with only one bounded edge of length one, the ends with markings  $I \subset \{1, \dots, N\}$ ,  $1 < |I| < N - 1$ , on one side and the ends with markings  $I^c$  on the other, we denote the equivalence class of its image under the distance map in  $\mathbb{R}^{\binom{N}{2}}/U_N$  by  $v_I$ . The vectors  $v_I$  generate the rays of  $\mathcal{M}_{0,N}$  and the lattice we fix for  $\mathbb{R}^{\binom{N}{2}}/U_N$ .

For local computations, we sometimes use a finite index set  $I$  instead of  $\{1, \dots, N\}$  as labels for the markings, and denote the corresponding moduli spaces by  $\mathcal{M}_{0,I}$ . Also, we can modify the definition above by assigning bounded lengths in  $\mathbb{R}_{>0}$  to the ends, corresponding to not taking the quotient by  $U_N$ . In this case we obtain a partially open moduli space which we will denote by  $\mathcal{M}'_{0,N}$ . There is then a map  $\mathcal{M}'_{0,N} \rightarrow \mathcal{M}_{0,N}$  forgetting the lengths of the bounded ends, which is just the quotient by  $U_N$ .

For every subset  $I \subset \{1, \dots, N\}$  of cardinality at least three, there is a *forgetful map*  $\text{ft}_I : \mathcal{M}_{0,N} \rightarrow \mathcal{M}_{0,|I|}$  which maps  $(\Gamma, x_1, \dots, x_N)$  to the tree where we remove all ends  $x_i$  with labels  $i \notin I$  (and possibly straighten 2-valent vertices). Forgetful maps are morphisms by Gathmann et al. [17, proposition 3.9]. In coordinates, we project to distances of ends in  $I$ .

**Lemma 2.3** *A vector  $x$  in  $\mathbb{R}^{\binom{N}{2}}/U_N$  is zero if and only if  $\text{ft}_I(x) = 0$  for all  $I \subset \{1, \dots, N\}$  with  $|I| = 4$ .*

*Proof* As  $\text{ft}_I$  is linear, the “only if” direction is obvious. For the other direction, denote the standard basis vectors of  $\mathbb{R}^{\binom{N}{2}}$  by  $e_{ij}$  for  $i < j$ . Let  $\tilde{x} = \sum_{i < j} \lambda_{ij} e_{ij} \in \mathbb{R}^{\binom{N}{2}}$  be a representative of  $x$ . For any  $I$  with  $|I| = 4$ , the assumption  $\text{ft}_I(x) = 0$  means that the projection  $\sum_{i,j \in I; i < j} \lambda_{ij} e_{ij}$  is in  $U_4$ . By definition of  $U_4$ , it follows that there is a vector  $\mu \in \mathbb{R}^I$  such that  $\lambda_{ij} = \mu_i + \mu_j$  for all  $i < j$  in  $I$ , and thus that  $\lambda_{ik} + \lambda_{jl} = \lambda_{ij} + \lambda_{kl}$  if  $I = \{i, j, k, l\}$ .

But this means that for all  $i = 1, \dots, N$  the assignment

$$\lambda_i := \frac{1}{2}(\lambda_{ij} + \lambda_{ik} - \lambda_{jk}) \quad \text{for arbitrary } j, k \neq i$$



is well-defined, because if  $m$  is another index we have

$$\begin{aligned} \frac{1}{2}(\lambda_{ij} + \lambda_{ik} - \lambda_{jk}) &= \frac{1}{2}(\lambda_{im} + \lambda_{ik} - \lambda_{mk}) + \frac{1}{2}(\lambda_{ij} - \lambda_{im} + \lambda_{mk} - \lambda_{jk}) \\ &= \frac{1}{2}(\lambda_{im} + \lambda_{ik} - \lambda_{mk}). \end{aligned}$$

As the definition of  $\lambda_i$  also implies that  $\lambda_{ij} = \lambda_i + \lambda_j$  for all  $i < j$ , we conclude that  $\tilde{x} \in U_N$ , and hence  $x = 0$ .

**Definition 2.4 (Tropical Stable Maps)** Let  $n \in \mathbb{N}$  and  $N \geq n$ . Consider a tuple  $(\Gamma, x_1, \dots, x_N, h)$ , where  $(\Gamma, x_1, \dots, x_N)$  is an  $N$ -marked abstract rational tropical curve and  $h : \Gamma \rightarrow \mathbb{R}^r$  is a continuous map that is integer linear on each edge. For an edge  $e$  starting at a vertex  $V$  of  $\Gamma$ , we denote the tangent vector of  $h|_e$  at  $V$  by  $v(e, V) \in \mathbb{Z}^r$  and call it the *direction* of  $e$  at  $V$ . If  $e$  is an end and  $V$  its only neighboring vertex we write  $v(e, V)$  also as  $v(e)$  for simplicity.

We say that  $(\Gamma, x_1, \dots, x_N, h)$  is an  $n$ -marked (rational) tropical stable map to  $\mathbb{R}^r$ , also called a (parameterized)  $n$ -marked curve in  $\mathbb{R}^r$  [17, definition 4.1], if

- $h$  satisfies the balancing condition  $\sum_{e \ni V} v(e, V) = 0$  at each vertex  $V$  of  $\Gamma$ ;
- $v(x_i) = 0$  for  $i = 1, \dots, n$  (i.e. each of the first  $n$  ends is contracted by  $h$ ), whereas  $v(x_i) \neq 0$  for  $i > n$  (i.e. the remaining  $N - n$  ends are “non-contracted ends”).

Two  $n$ -marked tropical stable maps  $(\Gamma, x_1, \dots, x_N, h)$  and  $(\tilde{\Gamma}, \tilde{x}_1, \dots, \tilde{x}_N, \tilde{h})$  in  $\mathbb{R}^r$  are isomorphic (and will from now on be identified) if there is an isomorphism  $\varphi$  of the underlying  $N$ -marked abstract curves such that  $\tilde{h} \circ \varphi = h$ .

The *degree* of an  $n$ -marked tropical stable map is the  $N$ -tuple

$$\Sigma = (v(x_1), \dots, v(x_N)) \in (\mathbb{Z}^r)^N$$

of directions of its ends, including the zero directions at the first  $n$  ends. Its *combinatorial type* is given by the data of the combinatorial type of the underlying abstract marked tropical curve  $(\Gamma, x_1, \dots, x_N)$  (i.e. where we drop the metrization data) together with the directions of all its edges.

The space of all  $n$ -marked rational tropical stable maps of a given degree  $\Sigma$  in  $\mathbb{R}^r$  is denoted by  $\mathcal{M}_{0,n}(\mathbb{R}^r, \Sigma)$ .

Since  $n$  equals the number of zero-entries in  $\Sigma$  and thus can be deduced from  $\Sigma$ , we sometimes drop the subscript and write only  $\mathcal{M}_0(\mathbb{R}^r, \Sigma)$ . While all  $N$  ends come with markings  $x_1, \dots, x_N$ , only the ends with markings  $x_1, \dots, x_n$  are contracted (i.e. have zero direction) and are thus highlighted in the notation.

*Remark 2.5 ( $\mathcal{M}_{0,n}(\mathbb{R}^r, \Sigma)$  as a Tropical Variety)* We assume  $n \geq 1$ . Then by Gathmann et al. [17, proposition 4.7],  $\mathcal{M}_{0,n}(\mathbb{R}^r, \Sigma)$  is a tropical variety, identified with  $\mathcal{M}_{0,N} \times \mathbb{R}^r$  via the map

$$\mathcal{M}_{0,n}(\mathbb{R}^r, \Sigma) \rightarrow \mathcal{M}_{0,N} \times \mathbb{R}^r, \quad (\Gamma, x_1, \dots, x_N, h) \mapsto ((\Gamma, x_1, \dots, x_N), h(x_1))$$

which forgets  $h$ , but records the image  $h(x_1)$  of a root vertex. It thus inherits the fan structure of  $\mathcal{M}_{0,N}$ . In particular, it can be embedded via this map into  $\mathbb{R}^{\binom{N}{2}}/U_N \times \mathbb{R}^r$ . When we work with an element of  $\mathcal{M}_{0,n}(\mathbb{R}^r, \Sigma)$  in coordinates, we usually give its coordinates in  $\mathbb{R}^{\binom{N}{2}} \times \mathbb{R}^r$ , i.e. its image under the distance map and the position of the root vertex. If  $n = 0$  it is still possible to find suitable coordinates for  $\mathcal{M}_{0,n}(\mathbb{R}^r, \Sigma)$  as  $\mathcal{M}_{0,N} \times \mathbb{R}^r$ , not by evaluating a marked end but by evaluating for example a barycenter [24, construction 1.2.21].

For each  $i = 1, \dots, n$ , we have the *evaluation map*

$$\text{ev}_i : \mathcal{M}_{0,n}(\mathbb{R}^r, \Sigma) \rightarrow \mathbb{R}^r$$

assigning to a tropical stable map  $(\Gamma, x_1, \dots, x_n, h)$  the position  $h(x_i)$  of its  $i$ -th marked end. It is shown in [17, proposition 4.8] that these maps are morphisms of tropical fans.

As above, we will also allow curves in  $\mathbb{R}^r$  where some of the non-contracted ends are bounded, and write the corresponding moduli spaces as  $\mathcal{M}'_{0,n}(\mathbb{R}^r, \Sigma)$ .

In the following, we will compute several intersection products in cells of tropical moduli spaces. Since we are often interested in a local situation, we can restrict to curves of a given combinatorial type  $\alpha$ . Local coordinates for the cell of curves of type  $\alpha$  are given by the coordinates of the root vertex and the lengths of each bounded edge. The map sending a unit vector in these local coordinates to a vector  $v_i$  as above is a unimodular transformation to the vector space spanned by the corresponding cell in the moduli space. Therefore we can compute lattice indices also in these local coordinates.

### 3 The Polyhedral Complex $\mathcal{M}_{0,n}(L, \Sigma)$ and Its Gluing Weights

For the whole section, let  $L \subset \mathbb{R}^r$  be a smooth tropical curve as in Example 2.1, and let  $\Sigma$  be the degree of a tropical  $n$ -marked stable map to  $\mathbb{R}^r$ . We want to define a moduli space  $\mathcal{M}_{0,n}(L, \Sigma)$  of tropical  $n$ -marked stable maps to  $L$  as a tropical variety. Let us first construct this space as a polyhedral complex, and then define its weights in the next subsection.

#### 3.1 The Polyhedral Complex $\mathcal{M}_{0,n}(L, \Sigma)$

We have already mentioned that not all stable maps with image in  $L$  will be allowed in  $\mathcal{M}_{0,n}(L, \Sigma)$ . Instead, we have to impose the so-called Riemann-Hurwitz condition that we introduce now. As we will see in Construction 3.11, it corresponds to a local realizability condition.

**Notation 3.1 (Covering Degrees)** Let  $(\Gamma, x_1, \dots, x_N, h) \in \mathcal{M}_{0,n}(\mathbb{R}^r, \Sigma)$  satisfy  $h(\Gamma) \subset L$  as sets. As  $L$  is irreducible we have  $h_*(\Gamma) = d \cdot L$  for some integer  $d$  (which depends only on  $\Sigma$ ). We call  $d$  the *covering degree* of the stable map and denote it by  $\text{deg}(\Sigma)$ .

For a vertex  $V$  of  $\Gamma$ , the *local degree*  $\Sigma_V$  at  $V$  is the collection of the directions of its adjacent edges, labeled in an arbitrary way starting with the zero directions. We let  $N_V = |\Sigma_V|$  and  $n_V$  the number of zero directions in  $\Sigma_V$  (which may come from marked ends or contracted bounded edges). The local covering degree will be denoted  $d_V = \text{deg}(\Sigma_V)$ .

**Definition 3.2 (Riemann-Hurwitz Number)** Let  $(\Gamma, x_1, \dots, x_N, h) \in \mathcal{M}_{0,n}(\mathbb{R}^r, \Sigma)$  satisfy  $h(\Gamma) \subset L$ . We define the *Riemann-Hurwitz number* of a vertex  $V$  of  $\Gamma$  with image  $W = h(V)$  as

$$\text{RH}(V) = N_V - n_V - d_V \cdot (\text{val}(W) - 2) - 2$$

(where  $\text{val } W = 2$  if  $W$  lies in the interior of an edge of  $L$ ). Note that it depends only on the combinatorial type of the stable map.

The Riemann-Hurwitz number gives a realizability condition for tropical stable maps to smooth curves. It appears e.g. in [5, definition 2.2], [9, proposition 2.4], [12, section 3.2.2], and [7, definition 3.11].

**Definition 3.3 ( $\mathcal{M}_{0,n}(L, \Sigma)$  as a Polyhedral Complex)** Let  $\alpha$  be a combinatorial type of tropical stable maps in  $\mathcal{M}_{0,n}(\mathbb{R}^r, \Sigma)$ . We denote the subset of maps  $(\Gamma, x_1, \dots, x_N, h)$  of type  $\alpha$  and satisfying  $h(\Gamma) \subset L$  by  $\mathcal{M}(\alpha)$ ; this is easily seen to be a partially open polyhedron. Let  $\mathcal{M}_{0,n}(L, \Sigma)$  be the set of all such cells  $\mathcal{M}(\alpha)$  with  $\text{RH}(V) \geq 0$  for all vertices  $V$  in  $\alpha$ ; this is a polyhedral complex [8].

Note that this definition of  $\mathcal{M}_{0,n}(L, \Sigma)$  formally differs from the one used in [16] in order to make it compatible with the literature mentioned above. In [16, definition 3.8], more cells are included a priori, but they obtain weight zero in the gluing construction of Sect. 3.2.

*Remark 3.4 (Dimension of  $\mathcal{M}_{0,n}(L, \Sigma)$ )* By an easy generalization of [8, lemma 2.14], it follows that  $\mathcal{M}_{0,n}(L, \Sigma)$  is pure of dimension  $|\Sigma| - \text{deg}(\Sigma) \cdot \sum_{W \in L} (\text{val}(W) - 2) - 2$ . The maximal cells correspond to combinatorial types such that

- each vertex mapping to a vertex of  $L$  satisfies  $\text{RH}(V) = 0$ ,
- each vertex mapping to an edge of  $L$  is 3-valent, and
- no edge is contracted to a vertex.

More precisely, we have:

**Lemma 3.5 (Dimension of Cells of  $\mathcal{M}_{0,n}(L, \Sigma)$ )** *Let  $\alpha$  be a combinatorial type in  $\mathcal{M}_{0,n}(L, \Sigma)$ . The dimension of the corresponding cell  $\mathcal{M}(\alpha)$  equals the number of vertices mapping to edges of  $L$  plus the number of bounded edges mapping to vertices of  $L$ .*

Intuitively, this holds true since we can independently vary the length of each bounded edge mapping to a vertex without leaving the cell of a combinatorial type, as well as the lengths of edges adjacent to a vertex mapping to an edge, in the appropriate way that “moves” the vertex along the edge.

### 3.2 The Gluing Construction for Moduli Spaces

In this section, we want to equip  $\mathcal{M}_{0,n}(L, \Sigma)$  with weights satisfying the balancing condition, to make it a tropical variety. To do this, we review the general technique developed in [16], adapted to the case when the target of the stable maps is a smooth curve. The idea is to construct the tropical moduli spaces by a gluing procedure from local moduli spaces for the vertices. This construction depends on a condition: all vertices appearing in a combinatorial type of the moduli space are required to be “good”. We start by repeating the relevant definitions in the case of smooth curves.

**Notation 3.6 (Links of Vertices)** Let  $(\Gamma, x_1, \dots, x_n, h) \in \mathcal{M}_{0,n}(L, \Sigma)$ , and let  $V$  be a vertex of  $\Gamma$ . We denote by  $L_V$  the link of  $L$  around  $h(V)$ . Generalizing the notation of Example 2.1, we denote a point by  $L_V^0$ , so that  $L_V$  is (an affine shift of a unimodular transformation of)  $L_r^q \times \mathbb{R}^s$ , where  $r + s = 1$  and  $q = 0$  if  $r = 0$ . Hence we have  $(r, s) = (1, 0)$  if  $V$  maps to a vertex of  $L$  (of valence  $q + 1$ ), and  $(r, s) = (0, 1)$  if  $V$  maps to an edge. Note that there is an associated local moduli space  $\mathcal{M}_0(L_V, \Sigma_V)$ .

**Definition 3.7 (Resolution Dimension)** For a tropical stable map  $(\Gamma, x_1, \dots, x_n, h) \in \mathcal{M}_{0,n}(L, \Sigma)$ , let  $V$  be a vertex of  $\Gamma$  with image  $W = h(V) \in L$ . As in Notation 3.6, we have  $L_V \cong L_r^q \times \mathbb{R}^s$  with  $r + s = 1$  and  $q = 0$  if  $r = 0$ . Treating again a point on an edge of  $L$  as a 2-valent vertex, we define the *resolution dimension* of  $V$  as

$$\text{rdim}(V) = N_V - d_V \cdot (\text{val}(W) - 2) + r - 3$$

and the *classification number* as

$$c_V = N_V + r \in \mathbb{N}.$$

*Remark 3.8 (Dimension of Local Moduli Spaces)* By the dimension formula, we see that the local moduli space at  $V$  has dimension  $\dim \mathcal{M}_0(L_V, \Sigma_V) = \text{rdim}(V) + s$ , where again  $L_V \cong L_r^q \times \mathbb{R}^s$ . As this moduli space has an  $s$ -dimensional lineality space coming from shifting the curves along  $\mathbb{R}^s$ , the resolution dimension of  $V$  is just the dimension of the local moduli space at  $V$  modulo its lineality space.

*Remark 3.9 (Dimension of  $\mathcal{M}_{0,n}(L, \Sigma)$  in Terms of Resolution Dimensions)* Let  $\alpha$  be a combinatorial type in  $\mathcal{M}_{0,n}(L, \Sigma)$ , and assume that  $\alpha$  has  $s$  vertices mapping to an edge in  $L$  (i.e. so that the corresponding link is  $L_0^0 \times \mathbb{R}$ ). Adding up the resolution dimensions of all vertices in  $\alpha$ , we obtain by Remark 3.4

$$\sum_V \text{rdim}(V) + s = \dim \mathcal{M}_{0,n}(L, \Sigma).$$

*Remark 3.10* Note that  $\text{rdim}(V)$  and  $\text{RH}(V)$  are very similar: in fact,  $\text{rdim}(V)$  is just  $\text{RH}(V)$  with additional contributions

- (a)  $n_V$  of the number of contracted edges at  $V$ , and
- (b)  $-1$  if  $V$  maps to an edge of  $L$ .

In particular, the condition  $\text{RH}(V) \geq 0$  of Definition 3.3 also implies  $\text{rdim}(V) \geq 0$  (otherwise we would have  $\text{RH}(V) = 0$  and  $\text{rdim}(V) = -1$ , i.e.  $V$  maps to an edge,  $N_V = 2$ , and  $n_V = 0$ , which is a contradiction since we do not allow 2-valent vertices).

The reason to introduce the numbers of Definition 3.7 is that they are used in the recursive definition of good vertices and the weights of  $\mathcal{M}_{0,n}(L, \Sigma)$  below. For this construction we start with the case of resolution dimension 0 and pass to the general case by gluing. The initial case is obtained by passing to the corresponding situation in algebraic geometry and considering (algebraic) Hurwitz numbers.

**Construction 3.11 (Algebraic Moduli Spaces for a Vertex)** Let  $V$  be a vertex of a combinatorial type in  $\mathcal{M}_{0,n}(L, \Sigma)$  such that  $L_V \cong L_1^q$ . Up to unimodular transformation,  $\Sigma_V = (\delta_1, \dots, \delta_{N_V})$  is a degree of tropical stable maps to  $\mathbb{R}^q$  with ends in the directions of  $L_1^q$ . We decompose  $\{1, \dots, N_V\}$  into a partition  $\eta_0, \dots, \eta_q$  and  $\eta$ , where

$$\eta_i = \{j \mid \delta_j = -m_j e_i \text{ for some } m_j \in \mathbb{N}_{>0}\}$$

and  $\eta = \{j \mid \delta_j = 0\}$ . This also uniquely defines the values  $m_j$  as the weights of the edges adjacent to  $V$ .

To construct an algebraic moduli space for  $V$ , fix  $q + 1$  distinct points  $P_0, \dots, P_q$  on the complex projective line  $\mathbb{P}^1$ . Inside the well-known moduli stack  $\overline{M}_{0,N_V}(\mathbb{P}^1, d_V)$  of  $N_V$ -marked degree- $d_V$  rational stable maps to  $\mathbb{P}^1$ , consider the substack  $M(\Sigma_V)$  of all smooth stable maps  $\mathcal{C} = (C, x_1, \dots, x_{N_V}, \pi)$  such that  $\pi^* P_i = \sum_{j \in \eta_i} m_j x_j$  for all  $i = 0, \dots, q$ , i.e. such that the ramification profile of  $\pi$  over  $P_0, \dots, P_q$  is as specified by  $\Sigma_V$ . We denote its closure inside  $\overline{M}_{0,N_V}(\mathbb{P}^1, d_V)$  by  $\overline{M}(\Sigma_V)$ , and its boundary by  $\partial M(\Sigma_V) = \overline{M}(\Sigma_V) \setminus M(\Sigma_V)$ . Its dimension is

$$\dim M(\Sigma_V) = 2d_V - 2 + N_V - d_V \cdot (q + 1) = \text{rdim}(V).$$

**Construction 3.12 (The Case of Resolution Dimension 0)** Let  $V$  be a vertex of a combinatorial type in  $\mathcal{M}_{0,n}(L, \Sigma)$  with  $\text{rdim}(V) = 0$ , where  $L_V \cong L_r^q \times \mathbb{R}^s$  as above. Then  $\dim \mathcal{M}_0(L_V, \Sigma_V) \cong \mathbb{R}^s$  by Remark 3.8, i.e. the local moduli space at  $V$  consists of only one cell. We make it into a tropical variety by giving it the following local weight  $\omega_V$ , depending on whether  $V$  maps to a vertex or to an edge of  $L$ .

- (a) If  $L_V \cong L_1^q$ , the algebraic moduli space  $\overline{M}(\Sigma_V)$  of Construction 3.11 has dimension zero. We define the *local weight* of  $V$  to be  $\omega_V := \deg \overline{M}(\Sigma_V)$ ; i.e. the number of points in  $\overline{M}(\Sigma_V)$ , counted with weight  $|\text{Aut}(\pi)|^{-1}$  as we work with a stack. This number is also called the (*marked*) *Hurwitz number* and denoted  $H(\Sigma_V)$ .

- (b) If  $L_V \cong L_0^0 \times \mathbb{R}$ , the dimension condition implies  $N_V = 3$ . In this case, we set  $\omega_V := 1$ .

In fact, the second case could be treated similarly to the first one by introducing a rubber variant of the moduli space  $\overline{M}(\Sigma_V)$ . We avoid this formulation for the sake of simplicity.

Let us now describe the gluing construction that gives the local moduli space  $\mathcal{M}_0(L_V, \Sigma_V)$  of a vertex  $V$  the structure of a tropical variety if  $\text{rdim}(V) > 0$ . In the following, any combinatorial type occurring in  $\mathcal{M}_0(L_V, \Sigma_V)$  will be called a *resolution* of  $V$ . For a combinatorial type  $\alpha$  occurring in a moduli space we denote by  $\mathcal{N}(\alpha)$  the “neighborhood of  $\alpha$ ”, i.e. the union of all cells  $\mathcal{M}(\beta)$  whose closure intersects  $\mathcal{M}(\alpha)$ .

Definition 3.13 of a good vertex and the following gluing Construction 3.14 depend on each other and work in a combined recursion on the classification number of vertices. The following definition of a good vertex thus assumes that good vertices of lower classification number are already defined recursively. Moreover, for every combinatorial type  $\alpha$  in a local moduli space  $\mathcal{M}_0(L_V, \Sigma_V)$  all of whose vertices have smaller classification number and are good it assumes that there is a gluing cycle in the neighborhood  $\mathcal{N}(\alpha)$  from Construction 3.14.

**Definition 3.13 (Good Vertices [16, definition 3.13])**

Let  $V$  be a vertex of a (local) tropical stable map in  $\mathcal{M}_0(L_V, \Sigma_V)$ , so that in particular  $\text{rdim}(V) \geq 0$ . The vertex  $V$  is called *good* if the following holds:

- (a) Every vertex of every resolution  $\alpha$  of  $V$  in  $\mathcal{M}_0(L_V, \Sigma_V)$  (which has classification number smaller than  $c_V$  by Gathmann and Ochse [16, lemma 3.6]) is good (so that a gluing cycle is defined on  $\mathcal{N}(\alpha)$  by Construction 3.14).
- (b) If  $\text{rdim}(V) > 0$  the maximal types in  $\mathcal{M}_0(L_V, \Sigma_V)$  are resolutions of  $V$ . We let  $\mathcal{M}_0(L_V, \Sigma_V)$  be a weighted polyhedral complex by defining the weights on maximal cells  $\mathcal{M}(\alpha) = \mathcal{N}(\alpha)$  using the gluing Construction 3.14. If  $\text{rdim}(V) = 0$ , we equip the unique cell of  $\mathcal{M}_0(L_V, \Sigma_V)$  with the weight of Construction 3.12. We require that the space  $\mathcal{M}_0(L_V, \Sigma_V)$  is a tropical cycle with these weights.
- (c) For every resolution  $\alpha$  of  $V$  in  $\mathcal{M}_0(L_V, \Sigma_V)$  and every maximal type  $\beta$  such that  $\mathcal{M}(\beta)$  contains  $\mathcal{M}(\alpha)$  in  $\mathcal{M}_0(L_V, \Sigma_V)$  ( $\beta$  is then also a resolution of  $V$ ), the weight of  $\beta$  is the same in the gluing cycles  $\mathcal{N}(\alpha)$  and  $\mathcal{N}(\beta)$ .

In the following review of the gluing construction from [16, construction 3.12], we omit some of the technical details for the sake of clarity.

**Construction 3.14 (The Gluing Construction for a Combinatorial Type  $\alpha$ )** Fix a (not necessarily maximal) combinatorial type  $\alpha$  of curves in  $\mathcal{M}_{0,n}(L, \Sigma)$  and assume that all its vertices are good. We will construct weights on the maximal cells of the neighborhood  $\mathcal{N}(\alpha)$  such that this partially open polyhedral complex becomes a tropical cycle. In particular, if  $\alpha$  is already maximal this defines a weight on  $\mathcal{M}(\alpha) = \mathcal{N}(\alpha)$ .

We cut each bounded edge of  $\alpha$  at some point in its interior, and in addition introduce lengths for all ends. This yields a set of connected components  $\alpha_V$ , each

containing only one vertex  $V$ , edges of directions  $\Sigma_V$ , and (now bounded) ends labeled by an index set  $I_V$ .

For every such vertex  $V$ , consider the local moduli space  $\mathcal{M}_0(L_V, \Sigma_V)$ , which is a tropical variety since  $V$  is good. We introduce lengths on all ends of  $\Sigma_V$ , obtaining a moduli space  $\mathcal{M}'_0(L_V, \Sigma_V)$  (of which  $\mathcal{M}_0(L_V, \Sigma_V)$  is a quotient) as in Sect. 2.2. Each bounded end  $i \in I_V$  is mapped to an edge or vertex of  $L$  that we denote by  $\sigma_i$ . We consider the open subcomplex of  $\mathcal{M}'_0(L_V, \Sigma_V)$  of all curves for which the evaluation at  $i$  still lies in  $\sigma_i$ , i.e. the partially open tropical subvariety

$$\mathcal{M}_V := \bigcap_{i \in I_V} \text{ev}_i^{-1}(\sigma_i)$$

of  $\mathcal{M}'_0(L_V, \Sigma_V)$ .

Now we want to glue these pieces  $\mathcal{M}_V$  back together. Consider a bounded edge  $e$  of  $\alpha$  adjacent to two vertices  $V_1(e)$  and  $V_2(e)$ , and denote the two bounded ends produced by cutting  $e$  by  $i_1(e) \in I_{V_1(e)}$  and  $i_2(e) \in I_{V_2(e)}$ , where  $\sigma_{i_1(e)} = \sigma_{i_2(e)} =: \sigma_e$ . There is a corresponding evaluation map

$$\text{ev}_e := (\text{ev}_{i_1(e)} \times \text{ev}_{i_2(e)}) : \prod_V \mathcal{M}_V \longrightarrow \sigma_e \times \sigma_e$$

at the endpoints of these two bounded ends in the factors for  $V_1$  and  $V_2$ . To impose the condition that these ends fit together to form the edge  $e$  we need to pull back the diagonal  $\Delta_{\sigma_e}$  via  $\text{ev}_e$  [16, appendix]. We abbreviate all these pull-backs by

$$\text{ev}^*(\Delta_L) \cdot \prod_V \mathcal{M}_V := \prod_e \text{ev}_e^* \Delta_{\sigma_e} \cdot \prod_V \mathcal{M}_V,$$

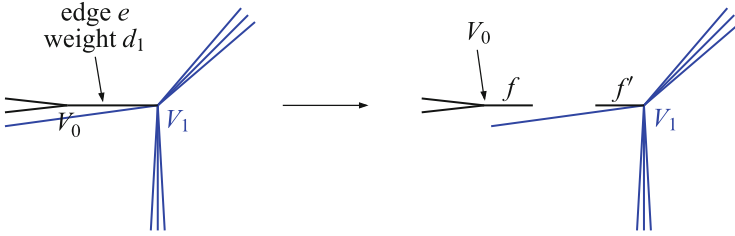
where  $e$  runs over all bounded edges  $e$  of  $\alpha$ . By construction, this cycle consists of stable map pieces that glue back to a stable map in  $\mathcal{M}_{0,n}(L, \Sigma)$ . However, it also carries the superfluous information on the position of the gluing points. To get rid of this we apply the quotient map  $q$  by the lineality space generated by the appropriate differences of vectors taking care of the lengths of the bounded ends, and by the vectors taking care of ends which should be unbounded. We finally use a morphism  $f$  identifying a stable map glued from pieces with the corresponding element in  $\mathcal{M}_{0,n}(L, \Sigma)$ , where we use the distance and barycentric coordinates mentioned in Remark 2.5. Hence we get a partially open tropical cycle

$$f_*q \left[ \text{ev}^*(\Delta_L) \cdot \prod_V \mathcal{M}_V \right] \quad \text{in} \quad \mathcal{M}_{0,n}(L, \Sigma).$$

Its weights on the maximal cells of  $\mathcal{M}_{0,n}(L, \Sigma)$  will be called the *gluing weights*. It is easy to see that the gluing morphism  $f$  is unimodular and induces a bijection of cells. In particular, the weight of a maximal cell in  $f_*q \left[ \text{ev}^*(\Delta_L) \cdot \prod_V \mathcal{M}_V \right]$  is equal to the

weight of  $\text{ev}^*(\Delta_L) \cdot \prod_V \mathcal{M}_V$  in the corresponding cell of  $\prod_V \mathcal{M}_V$ . By Remark 2.2, it can be computed as the greatest common divisor of the maximal minors of a matrix whose rows represent the differences  $\text{ev}_{i_1(e)} - \text{ev}_{i_2(e)}$  in local coordinates.

*Example 3.15* Let  $L = L_1^2$  be a tropical line in  $\mathbb{R}^2$  and let  $\alpha$  be a combinatorial type of degree- $\Sigma$  curves in  $L_1^2$  as shown below on the left (where the directions of the edges indicate their images in  $\mathbb{R}^2$ ). Then  $\text{rdim}(V_0) = 0$ . We assume in addition that  $\text{rdim}(V_1) = 0$ .



We cut the unique bounded edge  $e$  of weight  $d_1$ , obtaining two bounded ends that we denote  $f$  and  $f'$ . By the assumption on the resolution dimension, the local moduli spaces for  $V_0$  and  $V_1$  consist of only one cell each, and we can explicitly describe isomorphisms to open polyhedra in some  $\mathbb{R}^k$  as follows. The space  $\mathcal{M}_{V_0}$  is isomorphic to  $\mathbb{R}_{>0}^2$ , where one coordinate that we denote by  $l_f$  corresponds to the length of the bounded end, and the other that we call  $x_{V_0}$  to the position of the image of  $V_0$  on the corresponding ray of  $L$ . The space  $\mathcal{M}_{V_1}$  is  $\mathbb{R}_{>0}$  with coordinate  $l_{f'}$  corresponding to the length of its bounded end. By Construction 3.12, the weight of  $\mathcal{M}_{V_1}$  is the Hurwitz number  $\omega_{V_1} = H(\Sigma_{V_1})$ , whereas  $\mathcal{M}_{V_0}$  has weight 1. Using these coordinates, we can pull back the diagonal of  $L$  as  $\text{ev}_e^* \max\{x - y, 0\} = \max\{\text{ev}_f - \text{ev}_{f'}, 0\}$ , where  $x, y$  are the coordinates of  $L^2$  on the left ray. By Remark 2.2, the weight of  $\text{ev}^* \Delta_L \cdot (\mathcal{M}_{V_0} \times \mathcal{M}_{V_1})$  equals the weight of  $\mathcal{M}_{V_0} \times \mathcal{M}_{V_1}$  times the greatest common divisor of the maximal minors of the matrix

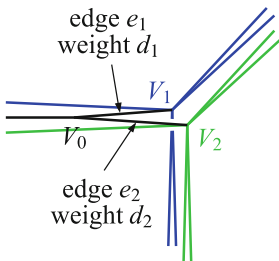
$$\frac{\begin{array}{c|ccc} & x_{V_0} & l_f & l_{f'} \\ \hline \text{ev}_f - \text{ev}_{f'} & 1 & -d_1 & -d_1, \end{array}}$$

which is 1. Hence the cell corresponding to  $\alpha$  in  $\mathcal{M}_0(L, \Sigma)$  has weight  $\omega_\alpha = H(\Sigma_{V_1})$ . The analogous result holds for  $L = L_1^q$  for all  $q$ .

*Example 3.16* Let  $L = L_1^2$  be a tropical line in  $\mathbb{R}^2$  again, and let  $\alpha$  be the combinatorial type of degree- $\Sigma$  curves mapping to  $L_1^2$  depicted below, with  $V_1$  and  $V_2$  mapping to the vertex of  $L$ . As above, we then have  $\text{rdim}(V_0) = 0$ , and assume in addition that  $\text{rdim}(V_1) = \text{rdim}(V_2) = 0$ .

We cut the two edges  $e_1$  of weight  $d_1$  and  $e_2$  of weight  $d_2$ , obtaining four new bounded ends that we denote by  $f_i$  and  $f'_i$  for  $i = 1, 2$ . As before, each local





moduli space consists of only one cell. The space  $\mathcal{M}_{V_0}$  is isomorphic to  $\mathbb{R}_{>0}^3$ , where two coordinates ( $l_{f_1}$  and  $l_{f_2}$ ) correspond to the lengths of the bounded ends and one ( $x_{V_0}$ ) to the position of the image of  $V_0$  on the corresponding ray of  $L$ . By Construction 3.12, it is equipped with weight  $\omega_{V_0} = 1$ . Similarly,  $\mathcal{M}_{V_i}$  for  $i = 1, 2$  is isomorphic to  $\mathbb{R}_{>0}$ , where the coordinate  $l_{f'_i}$  is given by the length of the bounded end, and equipped with the appropriate Hurwitz number  $\omega_{V_i} = H(\Sigma_{V_i})$  as weight. As in the previous example, pulling back the diagonal of  $L^2$  twice and using Remark 2.2, we deduce that the weight of  $\text{ev}^* \Delta_L \cdot (\mathcal{M}_{V_0} \times \mathcal{M}_{V_1} \times \mathcal{M}_{V_2})$  equals the weight of  $\mathcal{M}_{V_0} \times \mathcal{M}_{V_1} \times \mathcal{M}_{V_2}$  times the greatest common divisor of the maximal minors of the matrix

$$\begin{array}{c|ccccc} & x_{V_0} & l_{f_1} & l_{f_2} & l_{f'_1} & l_{f'_2} \\ \hline \text{ev}_{f_1} - \text{ev}_{f'_1} & 1 & -d_1 & 0 & -d_1 & 0 \\ \text{ev}_{f_2} - \text{ev}_{f'_2} & 1 & 0 & -d_2 & 0 & -d_2, \end{array}$$

which is  $\text{gcd}(d_1, d_2)$ . Thus the weight of the cell corresponding to  $\alpha$  in  $\mathcal{M}_0(L, \Sigma)$  equals

$$\omega_\alpha = \text{gcd}(d_1, d_2) \omega_{V_0} \omega_{V_1} \omega_{V_2} = \text{gcd}(d_1, d_2) \cdot H(\Sigma_{V_1}) \cdot H(\Sigma_{V_2}).$$

As in Example 3.15, the same result holds for  $L = L_1^q$  for all  $q$ .

We end this section by stating the main result of [16], together with a lemma that provides a major simplification for checking the requirements of the following theorem:

**Theorem 3.17 (The Gluing Theorem [16, corollary 3.17])** *Assume that all vertices  $V$  that can possibly occur in combinatorial types of the moduli space  $\mathcal{M}_{0,n}(L, \Sigma)$  are good. Then the gluing construction is well-defined for all these combinatorial types. In particular,  $\mathcal{M}_{0,n}(L, \Sigma)$  is a tropical variety.*

**Lemma 3.18 (Restriction to Resolution Dimension One [16, corollary 3.18])** *If all vertices  $V$  of combinatorial types of  $\mathcal{M}_{0,n}(L, \Sigma)$  with  $\text{rdim}(V) = 1$  are good, then all vertices are good.*

## 4 One-Dimensional Moduli Spaces of Rational Covers of Smooth Tropical Curves

Throughout this section, let  $V$  be a vertex of a combinatorial type in  $\mathcal{M}_{0,n}(L, \Sigma)$  with  $\text{rdim}(V) = 1$ . Our aim is to show that  $V$  is good, so that we can apply Lemma 3.18 and the gluing Theorem 3.17 to deduce Theorem 1.1. We continue to use the notation of Sect. 3. Moreover, let  $I_V$  be the set of labels of the ends in the local moduli space  $\mathcal{M}_0(L_V, \Sigma_V)$ , so that  $\mathcal{M}_0(L_V, \Sigma_V) = \mathcal{M}_{0,I_V}(L_V, \Sigma_V)$ . As in Construction 3.11, let  $m_j \in \mathbb{N}_{>0}$  be the weight of the end  $j \in I_V$ .

To prove that  $V$  is good, we have to show by Definition 3.13 that

- (1) every vertex appearing in a non-trivial resolution in  $\mathcal{M}_0(L_V, \Sigma_V)$  is good;
- (2)  $\mathcal{M}_0(L_V, \Sigma_V)$  is a tropical variety with the gluing weights; and
- (3) for every non-trivial resolution  $\alpha$  of  $V$ , the weight of each maximal cell in the neighborhood  $\mathcal{N}(\alpha)$  is the same no matter if we apply the gluing construction for  $\alpha$  or just for this maximal cell.

Assume first that  $V$  maps to an edge of  $L$ , so that  $L_V \cong L_0^0 \times \mathbb{R}$ . Then  $\text{rdim}(V) = 1$  implies  $N_V = 4$ , hence the possible resolutions are just the usual resolutions of a 4-valent vertex. Also, any gluing weight is just 1, and the balancing condition is satisfied—this is just the usual balancing condition of  $\mathcal{M}_{0,4}$ . It follows that  $V$  is good.

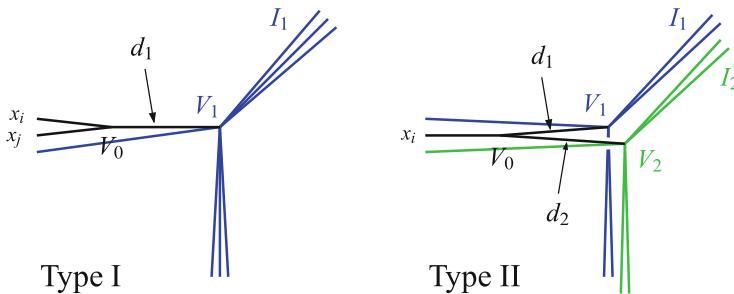
We can thus assume now that  $V$  maps to a vertex of  $L$ , so that  $L_V \cong L_1^q$ . By Remark 3.8, this means that  $\dim \mathcal{M}_0(L_V, \Sigma_V) = 1$ . In particular, every resolution of  $V$  corresponds already to a maximal cell of the local moduli space, which implies that condition (3) above is trivially satisfied. Moreover, Lemma 3.5 implies that every non-trivial resolution of  $V$  has at least one vertex mapping to an edge of  $L$ , or a bounded edge contracted to a vertex. In the former case, Remark 3.9 then shows that all vertices in this resolution must have resolution dimension 0 and are thus good, and the latter case is an immediate contradiction to Remark 3.4. Hence condition (1) is always satisfied as well, and it only remains to check the balancing condition (2).

Next, since  $1 - n_V = \text{rdim}(V) - n_V = \text{RH}(V) \geq 0$ , we can either have  $n_V = 1$  and  $\text{RH}(V) = 0$ , or  $n_V = 0$  and  $\text{RH}(V) = 1$ . In the first case, there is one contracted end, say with the marking 1, adjacent to the vertex. In the possible resolutions, this contracted end is adjacent to any other of the non-contracted ends, leading to a generating vector of the form  $v_{\{1,i\}}$  for the corresponding ray in  $\mathcal{M}_0(L_V, \Sigma_V)$ . As in Example 3.15, we can see that any gluing weight equals  $H(\Sigma_V \setminus \{0\})$ . We have  $\sum_{i=2}^{N_V} v_{\{1,i\}} = 0$  in  $\mathcal{M}_0(\mathbb{R}^q, \Sigma_V)$ , and hence the balancing condition is satisfied in this case.

So the only thing left to be done is to study the remaining case, where we have a vertex  $V$  mapping to a vertex of  $L$ , without contracted ends and having  $\text{rdim}(V) = 1$ , and to prove the balancing condition (2) for the 1-dimensional local moduli space  $\mathcal{M}_{0,n}(L, \Sigma)$  in this situation. We start by listing the possible resolutions of such a vertex, i.e. the maximal cones of  $\mathcal{M}_{0,n}(L, \Sigma)$ .

**Construction 4.1 (Resolutions of a Vertex with  $\text{rdim}(V) = 1$ )** Let  $V$  be a vertex of a combinatorial type in  $\mathcal{M}_{0,n}(L, \Sigma)$ . Assume that  $V$  maps to  $L_V \cong L_1^q$  and satisfies  $\text{rdim}(V) = 1$  and  $n_V = 0$ .

As  $\dim \mathcal{M}_0(L_V, \Sigma_V) = 1$ , it follows from Remark 3.4 and Lemma 3.5 that in each (necessarily maximal) resolution of  $V$ , there is one (necessarily 3-valent) vertex  $V_0$  mapping to an edge of  $L_1^q$ . This vertex can either join two ends or split an end, so that we obtain the following two types of resolutions:



- (I) There is exactly one vertex  $V_1$  mapping to the vertex of  $L_V$ . The vertex  $V_0$  is adjacent to two ends  $i, j \in I_V$  and a bounded edge of weight  $d_1 = m_i + m_j$  connecting  $V_0$  to  $V_1$ . The ends in  $I_1 := I_V \setminus \{i, j\}$  are adjacent to  $V_1$ .

Such a type exists for all choices of ends  $i$  and  $j$  of the same (primitive) direction.

- (II) There are exactly two vertices  $V_1, V_2$  mapping to the vertex of  $L_V$ . The vertex  $V_0$  is adjacent to an end  $i \in I_V$  and two bounded edges of weights  $d_1, d_2$  with  $d_1 + d_2 = m_i$  connecting  $V_0$  to  $V_1$  and  $V_2$ , respectively. The two vertices  $V_1$  and  $V_2$  are adjacent to ends in  $I_1$  and  $I_2$ , respectively, where  $I_1 \cup I_2 \cup \{i\} = I_V$ .

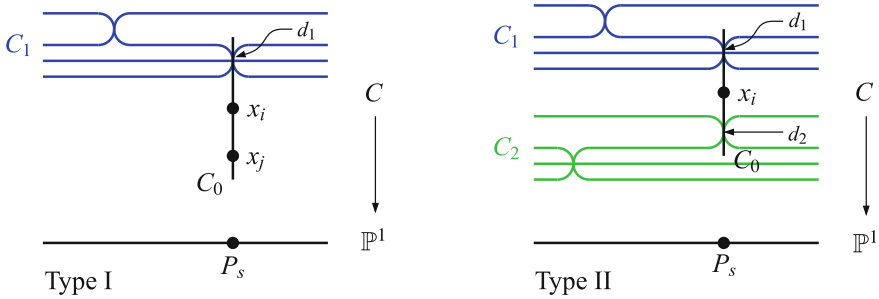
Such a type exists for all choices of  $i$  and all partitions of  $I_V \setminus \{i\}$  into  $I_1$  and  $I_2$  for which there is a stable map with the above conditions.

With the notations of Sect. 2, these types correspond to rays of  $\mathcal{M}_0(L_V, \Sigma_V)$  generated by the vectors  $v_{\{i,j\}}$  for type I and  $d_2 v_{I_1} + d_1 v_{I_2}$  for type II (where the latter does not need to be primitive).

Let us now consider the corresponding algebraic situation, i.e. the 1-dimensional algebraic moduli space  $\overline{M}(\Sigma_V)$  of Construction 3.11. By the Riemann-Hurwitz condition, a point in the open part  $M(\Sigma_V)$  corresponds to a cover with precisely one simple ramification which is not marked, and whose image does not coincide with one of the points  $P_0, \dots, P_q$  at which we fixed the ramification imposed by  $\Sigma_V$ . The boundary points correspond to degenerate covers that we obtain when the additional branch point runs into a point  $P_s$  for  $s \in \{0, \dots, q\}$ .

As deformations of covers are always local around special fibers [27, proposition 1.1], we see that a cover in  $\partial M(\Sigma_V)$  must have exactly one collapsed component, which then has exactly three special points. So we have the following

two types for the curves in the boundary  $\partial M(\Sigma_V)$ , which are exactly dual to the tropical picture above (see [8, proposition 3.12] for a related statement):

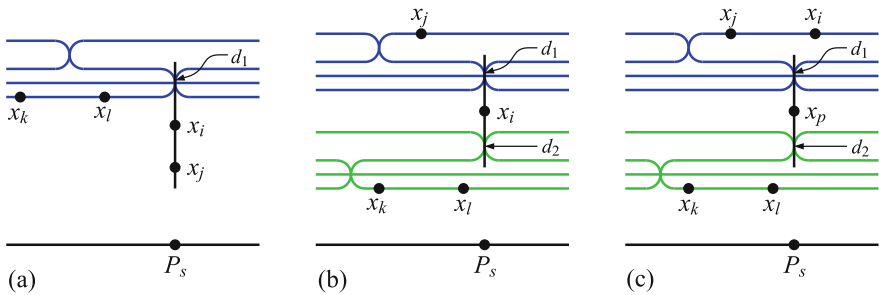


Here,  $C_0$  is the collapsed component, and  $C_k$  for  $k \in \{1, 2\}$  denotes the at most two non-collapsed irreducible components. In the type I case, the map  $\pi|_{C_1}$  has order  $d_1 := m_i + m_j$  at the singular point of  $C$ . In the type II case, the orders  $d_1$  and  $d_2$  of  $\pi|_{C_1}$  and  $\pi|_{C_2}$  at the singular points of  $C$  add up to  $m_i$ .

To check the balancing condition in the 1-dimensional fan  $\mathcal{M}_0(L_V, \Sigma_V)$ , it suffices by Lemma 2.3 to consider the situation after applying the various forgetful maps to  $\mathcal{M}_{0,4}$ . We will do this first in the algebraic and then in the tropical case.

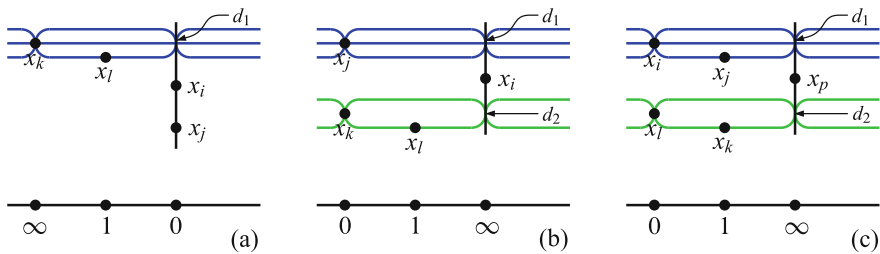
**Lemma 4.2 (The Pull-Back of the Forgetful Map)** *Let  $\mathcal{C} \in \partial M(\Sigma_V)$  be a stable map in the boundary of the local moduli space of a vertex  $V$  as in Construction 4.1. Consider the forgetful map  $\text{ft}_I : \overline{M}(\Sigma_V) \rightarrow \overline{M}_{0,I} \cong \mathbb{P}^1$  for a choice of four-element subset  $I = \{i, j, k, l\} \subset I_V$ . Then the multiplicity  $\text{ord}_{\mathcal{C}} \text{ft}_I^*(ij|kl)$  of the pullback of the divisor  $(ij|kl)$  on  $\overline{M}_{0,I}$  at  $\mathcal{C}$  equals*

- (a) 1 if  $\mathcal{C}$  is of type I, with  $x_i, x_j \in C_0$  and  $x_k, x_l \in C_1$  or vice versa;
- (b)  $d_1$  if  $\mathcal{C}$  is of type II, with  $x_i \in C_0$ , and  $x_j \in C_1$  and  $x_k, x_l \in C_2$  or vice versa;
- (c)  $d_1 + d_2$  if  $\mathcal{C}$  is of type II, with  $x_i, x_j \in C_1$  and  $x_k, x_l \in C_2$  or vice versa.



These are all cases in which we have a non-zero multiplicity.

*Proof* Since  $x_i, x_j$  and  $x_k, x_l$  must lie on different components after applying the forgetful map and  $\mathcal{C}$  has at least two and at most three components, it is obvious that we can only have the three cases stated in the lemma. We want to determine the multiplicity of  $\mathcal{C}$  in  $\text{ft}_l^*(ij|kl)$  for each case. By Vakil [27, proposition 1.1], we may replace our family  $\overline{M}(\Sigma_V)$  of curves around  $\mathcal{C}$  by another family  $\overline{M}$  of curves étale locally isomorphic to the original ones around the collapsed component. The following picture illustrates the new curve  $\mathcal{C}$  after this replacement in each case; the corresponding families are described below.



**Case (a)** Let  $M$  be the moduli space of all smooth covers  $(C, x_i, x_j, x_k, x_l, \pi)$  of  $\mathbb{P}^1$  of degree  $d_1 = m_i + m_j$  satisfying

$$\pi^*0 = m_jx_j + m_ix_i, \quad \pi^*\infty = d_1x_k, \quad \text{and} \quad \pi(x_l) = 1.$$

On the source curve  $C \cong \mathbb{P}^1$ , we set  $x_i = 0, x_j = \infty, x_k = 1$ , and  $x_l = (1 : w)$  with  $w \in \mathbb{C}^* \setminus \{1\}$ . Then every element in  $M$  can be written as

$$\pi(z_0 : z_1) = ((z_0 - z_1)^{d_1} : \lambda z_0^{m_j} z_1^{m_i})$$

for  $\lambda \in \mathbb{C}^*$  satisfying  $\lambda w^{m_i} = (1 - w)^{d_1}$ . Thus, the 1-dimensional space  $M$  is parameterized by those  $(\lambda, w) \in \mathbb{C}^* \times (\mathbb{C}^* \setminus \{1\})$  with  $\lambda w^{m_i} = (1 - w)^{d_1}$ . The non-marked branch point of  $\pi$  can be computed to be at  $P = (d_1^{d_1} : (-1)^{m_i} m_j^{m_j} m_i^{m_i} \cdot \lambda)$ , since the equation  $\pi(z_0 : z_1) = P$  has a double root at  $(m_j : -m_i)$ .

Hence, in this family the singular curve  $\mathcal{C}$  in the picture above corresponds to the coordinates  $(\lambda, w) = (0, 1)$ . After inserting this point into the family, we obtain  $\overline{M} \cong \mathbb{C}^*$  via  $(\lambda, w) \mapsto w$ . The divisor  $\text{ft}_l^*(ij|kl)$  is given by the function  $w - 1$ , which vanishes to order 1 at  $\mathcal{C}$ . As  $\mathcal{C}$  has no automorphisms due to the marked point  $x_l$ , we obtain  $\text{ord}_{\mathcal{C}} \text{ft}_l^*(ij|kl) = 1$  as claimed.

**Case (b)** Now let  $M$  be the space of those smooth covers  $(C, x_i, x_j, x_k, x_l, \pi)$  of  $\mathbb{P}^1$  of degree  $d = m_i$  such that

$$\pi^*0 = d_1x_j + d_2x_k, \quad \pi^*\infty = dx_i, \quad \text{and} \quad \pi(x_l) = 1$$

for fixed  $d_1, d_2$  with  $d_1 + d_2 = d$ . We set  $x_k = 1, x_i = \infty, x_j = 0$ , and  $x_l = (1 : w)$ , where  $w \in \mathbb{C}^* \setminus \{1\}$ . Then every element of  $M$  can be written as

$$\pi(z_0 : z_1) = (\lambda z_0^d : (z_0 - z_1)^{d_2} z_1^{d_1}),$$

where  $\lambda \in \mathbb{C}^*$  satisfies  $\lambda = (1 - w)^{d_1} w^{d_2}$ . The non-marked branch point of a cover  $\pi$  can be computed to be at  $P = (\lambda \cdot d^d : d_1^{d_1} d_2^{d_2})$ , since the equation  $\pi(z_0 : z_1) = P$  has a double root at  $(d : d_1)$ .

Again, as in the picture above we want to insert the special fiber  $\mathcal{C}$  over  $(\lambda, w) = (0, 1)$  to obtain the space  $\overline{M}$ . As before,  $\overline{M} \cong \mathbb{C}^*$  via  $(\lambda, w) \mapsto w$ , and the divisor  $\text{ft}_l^*(ij|kl)$  is given by the function  $w - 1$ , which vanishes to order 1 at  $\mathcal{C}$ . Since  $\mathcal{C}$  has  $d_1$  automorphisms on  $C_1$  (which is totally ramified over 0 and  $\infty$ ), we obtain  $\text{ord}_{\mathcal{C}} \text{ft}_l^*(ij|kl) = d_1$ .

**Case (c)** In this case, we use the previous computations and the WDVV equations. Denote by  $x_p$  the marked point of  $\mathcal{C}$  on the collapsed component. We consider the moduli space  $\overline{M}$  which is the closure of all smooth  $(C, x_i, x_j, x_k, x_l, x_p, \pi)$  of degree  $d = m_p$  such that

$$\pi^*0 = d_1x_i + d_2x_l, \quad \pi^*\infty = dx_p, \quad \text{and} \quad \pi(x_j) = \pi(x_k) = 1$$

for fixed  $d_1, d_2$  with  $d_1 + d_2 = d$ . Again, by the Riemann-Hurwitz formula this is a 1-dimensional space, with one non-marked ramification for a smooth curve in  $\overline{M}$ . By letting the additional branch point run into 0, 1 and  $\infty$ , we can see that  $\partial M$  contains the following reducible curves:

- (1) a degree- $d_1$  component with  $x_i, x_j$  connected to a degree- $d_2$  component with  $x_k, x_l$  via a collapsed component over  $\infty$  with  $x_p$  (this is the curve in the picture above);
- (2) a degree- $d_1$  component with  $x_i, x_k$  connected to a degree- $d_2$  component with  $x_j, x_l$  via a collapsed component over  $\infty$  with  $x_p$ ;
- (3) one collapsed component over 0 with  $x_i, x_l$  and one degree- $d$  component with  $x_j, x_k, x_p$ ;
- (4) one collapsed component over 1 with  $x_j, x_k$  and one degree- $d$  component with  $x_i, x_l, x_p$ .

The non-collapsed components in types (1)–(3) are all completely ramified over two points. In types (1) and (2), exactly one point with no ramification is marked, killing the automorphisms. Hence, for each of these types (1) and (2) we have one corresponding boundary point in  $\overline{M}$ . In type (3), the point  $x_j$  fixes the automorphisms, but then we have a choice to mark any preimage of 1 but  $x_j$  to be  $x_k$ . Hence there are  $d - 1$  boundary points corresponding to a cover of type (3). For type (4), a computation of the corresponding Hurwitz number shows that there is a unique such cover, so that we have one such boundary point in  $\overline{M}$ .

By the WDVV equations for  $\text{ft}_l : M \rightarrow \overline{M}_{0,4}$ , we have  $\text{ft}_l^*(ij|kl) = \text{ft}_l^*(il|kj)$ . The left side of this equation is obviously supported on the boundary point of type (1)

that we are interested in, whereas the right side is supported on all boundary points of type (3) or (4). The multiplicity of  $\text{ft}_I^*(il|kj)$  is 1 at each such boundary point by our former computation. As there are  $d$  such boundary points, in total we obtain  $\text{ord}_{\mathcal{C}} \text{ft}_I^*(ij|kl) = d$ .

**Theorem 4.3 (One-Dimensional Moduli Spaces  $\mathcal{M}_0(L_V, \Sigma_V)$ )** *Let  $V$  be a vertex as in Construction 4.1: mapping to  $L_V \cong L_1^q$  and satisfying  $\text{rdim}(V) = 1$  and  $n_V = 0$ . Then  $\mathcal{M}_0(L_V, \Sigma_V)$  with the weights obtained from the gluing Construction 3.14 is a one-dimensional balanced fan. In particular,  $V$  is good.*

*Proof* The rays of  $\mathcal{M}_0(L_V, \Sigma_V)$  are given by the combinatorial types  $\alpha$  of Construction 4.1. With the notation used there, we can take as spanning vectors for these rays  $u_\alpha = v_{\{i,j\}}$  in a type I case and  $u_\alpha = d_2v_{I_1} + d_1v_{I_2}$  in a type II case. As the integer length of these vectors is 1 and  $\text{gcd}(d_1, d_2)$ , respectively, it follows from Examples 3.15 and 3.16 that the gluing weight times the primitive vector in direction of the ray corresponding to  $\alpha$  equals  $H_\alpha u_\alpha$ , where  $H_\alpha$  denotes the Hurwitz number of  $V_1$  for type I, and the product of the Hurwitz numbers of  $V_1$  and  $V_2$  for type II. Hence we have to show that  $\sum_\alpha H_\alpha u_\alpha = 0$ .

By Lemma 2.3, it suffices to prove that  $\sum_\alpha H_\alpha \text{ft}_I(u_\alpha) = 0$  for all four-element subsets  $I = \{i, j, k, l\}$  of  $I_V$ . The combinatorial types  $\alpha$  for which  $\text{ft}_I(u_\alpha)$  is a multiple of  $v_{\{i,j\}}$  are exactly the ones corresponding to the three cases in Lemma 4.2. Due to the definition of  $u_\alpha$ , this multiple is 1,  $d_1$ , and  $d_1 + d_2$ , respectively, and hence always equal to  $\text{ord}_{\mathcal{C}} \text{ft}_I^*(ij|kl)$  for a stable map  $\mathcal{C}$  of this type. As the number of such stable maps is exactly  $H_\alpha$ , it follows that  $\sum_\alpha H_\alpha \text{ft}_I(u_\alpha)$  contains the vector  $v_{\{i,j\}}$  with a factor of  $\text{deg} \text{ft}_I^*(ij|kl)$ . But the same holds for the other two splittings of  $I$ , and thus we conclude as desired that

$$\sum_\alpha H_\alpha \text{ft}_I(u_\alpha) = \text{deg} \text{ft}_I^*(ij|kl) v_{\{i,j\}} + \text{deg} \text{ft}_I^*(ik|jl) v_{\{i,k\}} + \text{deg} \text{ft}_I^*(il|jk) v_{\{i,l\}} = 0$$

since these three divisors are linearly equivalent and  $v_{\{i,j\}} + v_{\{i,k\}} + v_{\{i,l\}} = 0$  in  $\mathcal{M}_{0,I}$ .

*Proof (Proof of Theorem 1.1)* Theorem 4.3 together with the arguments at the beginning of Sect. 4 shows that all vertices  $V$  of combinatorial types of  $\mathcal{M}_{0,n}(L, \Sigma)$  with  $\text{rdim}(V) = 1$  are good. By Lemma 3.18 we conclude that all vertices are good. Hence  $\mathcal{M}_{0,n}(L, \Sigma)$  is a tropical variety by Theorem 3.17, with the weights given in Constructions 3.12 and 3.14. The claim about the dimension follows from Sect. 3.1.

**Remark 4.4** By Lemma 3.18, the case of one-dimensional moduli spaces of tropical stable maps to a curve represents a main building block for the proof of Theorem 1.1 stating that arbitrary-dimensional moduli spaces of tropical stable maps to a curve are balanced. It was also a natural starting point for the investigation of the balancing condition for tropical moduli spaces of stable maps to a curve. In collaboration with Simon Hampe, the polymake extension a-tint [18, 20] was used to compute—for a large series of relevant examples—the generating vectors of rays for such one-dimensional moduli spaces. GAP [26] was used to compute

conjectural local weights in terms of Hurwitz numbers, and to check the balancing condition. These experiments with one-dimensional moduli spaces helped us to form a precise conjecture for the weights. Finally, the computation of a series of one-dimensional balanced examples led to the proof of the balancing condition in the one-dimensional case, and thus also in the general case. This work thus heavily relies on the examples computed with the help of a-tint and GAP.

**Acknowledgements** We would like to thank Erwan Brugallé, Renzo Cavalieri, Simon Hampe, and Diane Maclagan for helpful discussions.

This work would not have been possible without extensive computations of example classes which enabled us to establish and prove conjectures about polyhedra and their weights in our moduli spaces. We used the polymake-extension a-tint [18, 20] and GAP [26].

Part of this work was accomplished at the Mittag-Leffler Institute in Stockholm, during the semester program in spring 2011 on Algebraic Geometry with a View towards Applications. The authors would like to thank the institute for hospitality.

The first and second author were partially funded by DFG grant GA 636/4-2 resp. MA 4797/3-2, as part of the DFG Priority Program 1489.

We thank an anonymous referee for helpful comments on an earlier version of this paper.

## References

1. L. Allermann, Tropical intersection products on smooth varieties. *J. Eur. Math. Soc.* **14**(1), 107–126 (2012)
2. L. Allermann, J. Rau, First steps in tropical intersection theory. *Math. Z.* **264**(3), 633–670 (2010). arXiv:0709.3705
3. K. Behrend, Gromov-Witten invariants in algebraic geometry. *Invent. Math.* **127**(3), 601–617 (1997)
4. K. Behrend, B. Fantechi, The intrinsic normal cone. *Invent. Math.* **128**, 45–88 (1997)
5. B. Bertrand, E. Brugallé, G. Mikhalkin, Tropical open Hurwitz numbers. *Rend. Semin. Mat. Univ. Padova* **125**, 157–171 (2011)
6. B. Bertrand, E. Brugallé, G. Mikhalkin, Genus 0 characteristic numbers of tropical projective plane. *Compos. Math.* **150**(1), 46–104 (2014). arXiv:1105.2004
7. E. Brugallé, H. Markwig, Deformation of tropical Hirzebruch surfaces and enumerative geometry. *J. Algebraic Geom.* (2013, preprint). arXiv:1303.1340
8. A. Buchholz, H. Markwig, Tropical covers of curves and their moduli spaces. *Commun. Contemp. Math.* (2013). <https://doi.org/10.1142/S0219199713500454>
9. L. Caporaso, Gonality of algebraic curves and graphs. *Springer Proc. Math. Stat.* **71**, 77–108 (2014)
10. R. Cavalieri, P. Johnson, H. Markwig, Tropical Hurwitz numbers. *J. Algebr. Comb.* **32**(2), 241–265 (2010). arXiv:0804.0579. <https://doi.org/10.1007/s10801-009-0213-0>
11. R. Cavalieri, P. Johnson, H. Markwig, Wall crossings for double Hurwitz numbers. *Adv. Math.* **228**(4), 1894–1937 (2011). arXiv:1003.1805
12. R. Cavalieri, H. Markwig, D. Ranganathan, Tropicalizing the space of admissible covers. *Math. Ann.* (2014, preprint). arXiv:1401.4626
13. R. Cavalieri, H. Markwig, D. Ranganathan, Tropical compactification and the Gromov–Witten theory of  $\mathbb{P}^1$ . *Sel. Math.* **1**(34) (2016). arXiv:1410.2837. <https://doi.org/10.1007/s00029-016-0265-7>
14. E.M. Feichtner, B. Sturmfels, Matroid polytopes, nested sets and Bergman fans. *Port. Math.* **62**, 437–468 (2005). arXiv:math.CO:0411260



15. W. Fulton, R. Pandharipande, Notes on stable maps and quantum cohomology, in *Algebraic Geometry: Santa Cruz 1995*, ed. by J. Kollár et al. Proceedings of Symposia in Pure Mathematics, vol. 62(2) (American Mathematical Society, Providence, RI, 1997), pp. 45–96
16. A. Gathmann, D. Ochse, Moduli spaces of curves in tropical varieties, in *Algorithmic and Experimental Methods in Algebra Geometry, and Number Theory*, ed. by G. Böckle, W. Decker, G. Malle (Springer, Heidelberg, 2018). [https://doi.org/10.1007/978-3-319-70566-8\\_11](https://doi.org/10.1007/978-3-319-70566-8_11)
17. A. Gathmann, M. Kerber, H. Markwig, Tropical fans and the moduli space of rational tropical curves. *Compos. Math.* **145**(1), 173–195 (2009). arXiv:0708.2268
18. E. Gawrilow, M. Joswig, Polymake: a framework for analyzing convex polytopes, in *Polytopes – Combinatorics and Computation*, ed. by G. Kalai, G.M. Ziegler (Birkhäuser, Boston, 2000), pp. 43–74
19. A. Gross, Correspondence theorems via tropicalizations of moduli spaces. *Commun. Contemp. Math.* (2014, to appear). arXiv:1406.1999
20. S. Hampe, a-tint: a polymake extension for algorithmic tropical intersection theory. *Eur. J. Comb.* **36C**, 579–607 (2014). arXiv:1208.4248
21. H. Markwig, J. Rau, Tropical descendant Gromov-Witten invariants. *Manuscripta Math.* **129**(3), 293–335 (2009). arXiv:0809.1102. <https://doi.org/10.1007/s00229-009-0256-5>
22. G. Mikhalkin, Enumerative tropical geometry in  $\mathbb{R}^2$ . *J. Am. Math. Soc.* **18**, 313–377 (2005). arXiv:math.AG/0312530
23. G. Mikhalkin, Moduli spaces of rational tropical curves, in *Proceedings of Gökova Geometry-Topology Conference GGT 2006*, pp. 39–51 (2007). arXiv:0704.0839
24. D. Ochse, Moduli spaces of rational tropical stable maps into smooth tropical varieties. Ph.D. thesis, TU Kaiserslautern (2013)
25. D. Speyer, B. Sturmfels, The tropical Grassmannian. *Adv. Geom.* **4**, 389–411 (2004)
26. The GAP Group: GAP – Groups, Algorithms, and Programming, Version 4.8.6 (2016). <http://www.gap-system.org>
27. R. Vakil, The enumerative geometry of rational and elliptic curves in projective space. *J. Reine Angew. Math. (Crelle’s Journal)* **529**, 101–153 (2000)

# Invariant Bilinear Forms on $W$ -Graph Representations and Linear Algebra Over Integral Domains



Meinolf Geck and Jürgen Müller

**Abstract** Lie-theoretic structures of type  $E_8$  (e.g., Lie groups and algebras, Iwahori–Hecke algebras and Kazhdan–Lusztig cells, . . .) are considered to serve as a “gold standard” when it comes to judging the effectiveness of a general algorithm for solving a computational problem in this area. Here, we address a problem that occurred in our previous work on decomposition numbers of Iwahori–Hecke algebras, namely, the computation of invariant bilinear forms on so-called  $W$ -graph representations. We present a new algorithmic solution which makes it possible to produce and effectively use the main results in further applications.

**Keywords** Iwahori-Hecke algebras • Balanced representations •  $W$ -graph representations • Invariant forms • Integral linear algebra • Linear algebra over polynomial rings • MeatAxe philosophy

**Subject Classifications** 20C08, 20C40

## 1 Introduction

This paper is concerned with the representation theory of Iwahori–Hecke algebras. Such an algebra  $\mathcal{H}$  is a certain deformation of the group algebra of a finite Coxeter group  $W$ . In [6], the notion of “balanced representations” of  $\mathcal{H}$  was introduced, which has turned out to be useful in several applications. We mention here the construction of cellular structures on  $\mathcal{H}$  (see, e.g., [8, Chap. 2]), the determination

---

M. Geck (✉)

IAZ-Lehrstuhl für Algebra, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany  
e-mail: [meinolf.geck@mathematik.uni-stuttgart.de](mailto:meinolf.geck@mathematik.uni-stuttgart.de)

J. Müller

Arbeitsgruppe Algebra und Zahlentheorie, Bergische Universität Wuppertal, Gauß-Straße 20,  
42119 Wuppertal, Germany  
e-mail: [juergen.mueller@math.uni-wuppertal.de](mailto:juergen.mueller@math.uni-wuppertal.de)

of decomposition numbers of  $\mathcal{H}$  (see [9]), and the computation of Lusztig's function  $\mathbf{a}: W \rightarrow \mathbb{Z}$  (see [7, §4]). To check whether a given representation of  $\mathcal{H}$  is balanced or not is a computationally hard problem; it involves the construction of a certain invariant bilinear form on the underlying  $\mathcal{H}$ -module. It has been conjectured in [6] that so-called “ $W$ -graph representations” of  $\mathcal{H}$  are always balanced. But even if such a theoretical result were known to be true, certain applications (e.g., the determination of decomposition numbers) would still require the explicit knowledge of the Gram matrices of the invariant bilinear forms. In this paper, we discuss algorithms for the construction of these Gram matrices for  $W$  of exceptional type. The biggest challenge—by far—is the case where  $W$  is of type  $E_8$ . (The distinguished role of  $E_8$  when it comes to performing explicit computations is highlighted in various recent survey articles; see, e.g., Garibaldi [5], Lusztig [18], Vogan [25].)

In the situations of interest to us, the algebra  $\mathcal{H}$  is defined over the field of rational functions  $K = \mathbb{Q}(v)$  (where  $v$  is an indeterminate); it has a natural basis  $\{T_w \mid w \in W\}$ . Explicit models for the irreducible representations of  $\mathcal{H}$  are known by the work of Naruse [22], Howlett and Yin [14, 15]. Now let us fix an irreducible matrix representation  $\mathfrak{X}: \mathcal{H} \rightarrow K^{d \times d}$ . In order to show that  $\mathfrak{X}$  is balanced, one needs to determine a non-zero symmetric matrix  $P \in K^{d \times d}$  such that

$$P \mathfrak{X}(T_w) = \mathfrak{X}(T_{w^{-1}})^t P \quad \text{for all } w \in W;$$

this matrix  $P$  then has to satisfy certain additional properties. Thus, the computation of  $P$  essentially amounts to solving a system of linear equations; for theoretical reasons, we know that this system has a unique solution up to multiplication by a scalar. Rescaling a given solution by a suitable non-zero polynomial in  $\mathbb{Q}[v]$ , we can assume that all entries of  $P$  are in  $\mathbb{Z}[v]$  and that their greatest common divisor is  $\pm 1$ ; then  $P$  is unique up to sign and is called a “primitive Gram matrix”. The general theory also shows that a particular solution is given by

$$P_0 = \sum_{w \in W} \mathfrak{X}(T_w)^t \mathfrak{X}(T_w) \in K^{d \times d}.$$

Thus, if the matrices  $\mathfrak{X}(T_w)$  ( $w \in W$ ) are known and if  $|W|$  is not too large, then we can simply perform the above summation and obtain  $P_0$ ; rescaling  $P_0$  yields a primitive Gram matrix  $P$ . This procedure works for types  $F_4$ ,  $E_6$ , for example (and this is easily implemented in CHEVIE [11]).

Already for type  $E_7$ , one needs to use a more sophisticated approach as described in [9, §4.3], based on Parker's “standard basis algorithm” [23], in combination with interpolation and modular techniques. This also works for type  $E_8$ , but it is efficient only for irreducible representations of dimension up to about 2500. In our previous work on decomposition numbers, this was sufficient to obtain the desired results for type  $E_8$ ; see [9, Remark 4.10]. In principle, one could have run the above procedure on all irreducible representations of type  $E_8$ , but experiments showed that this would have needed a total of nearly 1 year of CPU time. On the other hand, from a strictly

logical point of view, one does not need to know exactly how the Gram matrices have been obtained, because as an independent verification one can simply check that they form a solution to the above system of linear equations. However, to store the various primitive Gram matrices requires about 28 GB of disk space, and even the verification alone is a major task as it involves the computation of products of (large) matrices with polynomial entries. In any case, this raises a serious issue of making sure that our results are reliable and reproducible.

In our view, the solution to deal with this issue is to develop better mathematical tools which make it possible to reproduce the results efficiently as needed, and this is what we will do in this paper. Indeed, for example, in order to deal with the irreducible representation of largest dimension for type  $E_8$  (which is 7168), the old approach would have needed roughly 7 weeks of CPU time, while the one described here requires only about 20 h, which amounts to a factor of almost 60. (See Sect. 9.1 for more details.) In view of the complexity of the task, and the experiences made elsewhere with explicit computations in type  $E_8$  (see the references cited above), it was clear that developing efficient methods would not be a standard, let alone press-button application of existing tools from computer algebra. Maier et al. [19] proposed an approach based on parallel techniques, but type  $E_8$  still seems to be a major challenge there. Hence one of the purposes of this paper is to give a systematic description of the (serial) methods we have used for the computation of Gram matrices of invariant bilinear forms for Iwahori–Hecke algebras.

The basic strategy in our approach is to reduce computational linear algebra over the Laurent polynomial ring  $\mathbb{Q}[v, v^{-1}]$  to linear algebra over the integers. Thus, generally speaking, we are faced with the problem of devising efficient tools to do computational linear algebra over integral domains, not just over fields. In order to do so, we build on general ideas from computational representation theory, more precisely on the celebrated so-called **MeatAxe** philosophy [23], which comprises of specially tailored, highly efficient techniques for computational linear algebra over (small) finite fields. Attempts to generalize these ideas to linear algebra over the (infinite) field of rational numbers, and further to linear algebra over the integers have been coined the **IntegralMeatAxe** [24]. The last word on this has not been said yet, and in this paper we are trying to contribute here as well. (As future work, we are planning to develop a full **IntegralMeatAxe** package along the present lines.) But we are additionally going one step further by setting out to extend these ideas to linear algebra over the univariate polynomial rings over the rationals or the integers.

To do so, the basic idea is to reduce to linear algebra over the integers by evaluating polynomials with rational coefficients at integral places, where we are using as few “small” places as possible, and to recover the polynomials in question by a Chinese remainder technique. Hence this strategy, fitting nicely into the **IntegralMeatAxe** philosophy, differs from those known to the literature, inasmuch we are neither using modular methods (which would mean to go over to polynomial rings over finite fields), nor are we in a position to use interpolation (which would mean to use lots of places to evaluate at). Thus another purpose of this paper is to give a detailed description of the new computational tasks arising in pursuing this strategy, and how we have accomplished them. Although the choice of the material

presented is governed by our application to Iwahori–Hecke algebras, it is exhibited with a view towards general applicability.

Here is an outline of the paper: In Sect. 2 we recall some basic facts about representations of finite Coxeter groups and Iwahori–Hecke algebras, in particular the notions of  $W$ -graphs, balancedness, and invariant bilinear forms. We conclude with Theorem 2.5 saying that for the representations afforded by the  $W$ -graphs given by Naruse [22], Howlett and Yin [14, 15] are actually balanced, and in Tables 1 and 2 we list some numerical data associated with their primitive Gram matrices.

In the subsequent sections we describe our general approach towards linear algebra over integral domains, which consists of a cascade of steps: In Sect. 3 we first deal with linear algebra over  $\mathbb{Z}$ . We discuss the key tasks of rational number recovery and of finding integral linear dependencies. Both tasks are known to the literature, but for the former we provide a variant containing a new feature, while for the latter we proceed along another strategy, within the `IntegralMeatAxe` philosophy. Subsequently, we apply this to computing nullspaces, inverses, and the so-called “exponents” of matrices over  $\mathbb{Z}$ . In Sect. 4 we then describe our general approach to deal with polynomials, in view of our aim to do linear algebra over polynomial rings. The key task is to recover a polynomial with rational coefficients from some of its evaluations at integral places. Here, we are aiming at using as few “small” places as possible, whence we are not in a position to apply interpolation, but we are using a Chinese remainder technique instead. Moreover, we devise a method to recover a polynomial from some of its evaluations where the latter are “rescaled” by unknown scalars; the necessity of being able to solve this task is closely related to our use of the `IntegralMeatAxe`, hence to our knowledge this method is new as well. In Sect. 5 we proceed to show how linear algebra over  $\mathbb{Z}$  and polynomial recovery, as discussed in earlier sections, can now be combined to do linear algebra over  $\mathbb{Z}[X]$  and  $\mathbb{Q}[X]$ , by devising methods to computing nullspaces, inverses, exponents and products of matrices using this new approach. In Sect. 6 we finally recall the “standard basis algorithm” originally developed in [23] for computations over finite fields. We present a general variant for absolutely irreducible matrix representations over an arbitrary field, show how this can be used to compute homomorphisms between such representations, and discuss how the necessary computations are facilitated over the fields  $\mathbb{Q}$  and  $\mathbb{Q}(X)$ , using the tools we have developed.

Having the general tools in place, in Sect. 7 we return to our particular application of computing Gram matrices of invariant bilinear forms for  $W$ -graph representations  $\mathfrak{X}$  of Iwahori–Hecke algebras. We proceed along the strategy which has already been indicated in [9, Section 4.3], where here we take the opportunity to provide full details. We begin by computing standard bases for the representations  $\mathfrak{X}$  and  $\mathfrak{X}'$ , where the latter is given by  $\mathfrak{X}'(T_w) := \mathfrak{X}(T_{w^{-1}})^{\text{tr}}$ , for  $w \in W$ . In order to find suitable seed vectors to start with, we use an observation on restrictions of representations of Iwahori–Hecke algebras to parabolic subalgebras, which naturally leads to certain distinguished elements of  $\mathcal{H}$  having actions of co-rank one on  $\mathfrak{X}$  and  $\mathfrak{X}'$ . To actually run the standard basis algorithm subsequently, we again revert to a specialization technique. In Sect. 8 we proceed by collecting a few observations on the standard

bases  $B$  and  $B'$  of the representations  $\mathfrak{X}$  and  $\mathfrak{X}'$  thus obtained. Indeed, the matrix entries occurring seem to be much less arbitrary than expected from general principles, but this has only been verified experimentally for the representations under consideration here, while a priori proofs are largely missing (so far). The final computational step then essentially is to determine the product  $B^{-1} \cdot B'$ , which up to rescaling is a Gram matrix as desired. To do this efficiently, apart from the general tools developed above, we make heavy use of the special form of the matrix entries of  $B^{-1} \cdot B'$  just mentioned. In the concluding Sect. 9 we provide running times and workspace requirements for our computations in types  $E_7$  and  $E_8$ , and present an explicit (tiny) example for type  $E_6$ .

It should be clear from the above description that to pursue our novel approach we had to solve quite a few tasks for which there was no pre-existing implementation, let alone in one and the same computer algebra system. To develop the necessary new code, as our computational platform we have chosen the computer algebra system **GAP** [4]. This system provides efficient arithmetics for the various basic objects we need: (1) rational integers and rational numbers, which in turn are handled by the **GMP** library [12]; (2) row vectors and matrices over the integers, the rationals or (small) finite fields, where in this context the entries of row vectors are actually treated as immediate objects; (3) floating point numbers, where the limited built-in facilities are sufficient for our purposes. Moreover, the necessary input data on Iwahori–Hecke algebras and their representations is provided by the computer algebra system **CHEVIE** [20], which conveniently is a branch of **GAP**.

## 2 Iwahori–Hecke Algebras and Balanced Representations

We begin by recalling some basic facts about representations of finite Coxeter groups and Iwahori–Hecke algebras; see [8, 10, 17] for further details.

### 2.1 Iwahori–Hecke Algebras

We fix a finite Coxeter group  $W$  with set of simple reflections  $S$ ; for  $w \in W$ , we denote by  $l(w)$  the length of  $w$  with respect to  $S$ . Let  $L: W \rightarrow \mathbb{Z}$  be a weight function as in [17], that is, we have  $L(w w') = L(w) + L(w')$  whenever  $w, w' \in W$  satisfy  $l(w w') = l(w) + l(w')$ . Such a weight function is uniquely determined by its values  $L(s)$  for  $s \in S$ . We will assume throughout that

$$L(s) > 0 \quad \text{for all } s \in S.$$

Let  $R \subseteq \mathbb{C}$  be a subring and  $A = R[v, v^{-1}]$  be the ring of Laurent polynomials over  $R$  in the indeterminate  $v$ . Let  $\mathcal{H} = \mathcal{H}_A(W, L)$  be the corresponding generic Iwahori–Hecke algebra. Thus,  $\mathcal{H}$  is an associative  $A$ -algebra which is free over  $A$

with a basis  $\{T_w \mid w \in W\}$ ; the multiplication is given by the following rule, where  $s \in S$  and  $w \in W$ :

$$T_s T_w = \begin{cases} T_{sw} & \text{if } l(sw) = l(w) + 1, \\ T_{sw} + (v^{L(s)} - v^{-L(s)})T_s & \text{if } l(sw) = l(w) - 1. \end{cases}$$

### 2.2 Modules for Iwahori-Hecke Algebras

Let  $F \subseteq \mathbb{C}$  be the field of fractions of  $R$  and assume that  $F$  is a splitting field for  $W$ . (For example, we could take  $R = F = \mathbb{R}$  since  $\mathbb{R}$  is known to be a splitting field for  $W$ .) Let  $\text{Irr}(W)$  be the set of simple  $F[W]$ -modules (up to isomorphism); we shall use the following notation:

$$\text{Irr}(W) = \{E^\lambda \mid \lambda \in \Lambda\} \quad \text{and} \quad d_\lambda = \dim E^\lambda \quad (\lambda \in \Lambda),$$

where  $\Lambda$  is a finite index set. Let  $K = F(v)$  be the field of fractions of  $A$  and  $\mathcal{H}_K = K \otimes_A \mathcal{H}$  be the  $K$ -algebra obtained by extension of scalars from  $A$  to  $K$ . Then  $\mathcal{H}_K$  is a split semisimple algebra and there is a bijection between  $\text{Irr}(W)$  and  $\text{Irr}(\mathcal{H}_K)$ , the set of simple  $\mathcal{H}_K$ -modules (up to isomorphism). Given  $\lambda \in \Lambda$ , we denote by  $E_v^\lambda$  a simple  $\mathcal{H}_K$ -module corresponding to  $E^\lambda$ . Then  $E_v^\lambda$  is uniquely determined (up to isomorphism) by the following property. For  $w \in W$ , we have

$$\text{trace}(T_w, E_v^\lambda) \in F[v, v^{-1}] \quad \text{and} \quad \text{trace}(w, E^\lambda) = \text{trace}(T_w, E_v^\lambda)|_{v \rightarrow 1}.$$

### 2.3 Iwahori-Hecke Algebras as Symmetric Algebras

The algebra  $\mathcal{H}_K$  is symmetric, with trace form  $\tau: \mathcal{H}_K \rightarrow K$  given by  $\tau(T_1) = 1$  and  $\tau(T_w) = 0$  for  $1 \neq w \in W$ . The basis dual to  $\{T_w \mid w \in W\}$  is given by  $\{T_{w^{-1}} \mid w \in W\}$ . By the general theory of symmetric algebras, there are well-defined elements  $0 \neq \mathbf{c}_\lambda \in A$  ( $\lambda \in \Lambda$ ) such that the following orthogonality relations hold for  $\lambda, \mu \in \Lambda$ :

$$\sum_{w \in W} \text{trace}(T_w, E_v^\lambda) \text{trace}(T_{w^{-1}}, E_v^\mu) = \begin{cases} d_\lambda \mathbf{c}_\lambda & \text{if } \lambda = \mu, \\ 0 & \text{if } \lambda \neq \mu. \end{cases}$$

As observed by Lusztig, we can write each  $\mathbf{c}_\lambda$  uniquely in the form

$$\mathbf{c}_\lambda = f_\lambda v^{-2a_\lambda} + \text{linear combination of larger powers of } v,$$

where  $f_\lambda$  is a strictly positive real number and  $\mathbf{a}_\lambda$  is a non-negative integer. The “ $a$ -invariants”  $\mathbf{a}_\lambda$  will play a major role in the sequel; these numbers are explicitly known for all types of  $W$  and all choices of  $L$  (see [8, §1.3], [17, Chap. 22]). Alternatively,  $\mathbf{a}_\lambda$  can be characterized as follows:

$$\mathbf{a}_\lambda = \min\{i \geq 0 \mid v^i \text{trace}(T_w, E_v^\lambda) \in F[v] \text{ for all } w \in W\}.$$

### 2.4 Balanced Representations

Let  $\mathcal{O} \subseteq K$  be the localization of  $F[v]$  in the prime ideal  $(v)$ , that is,  $\mathcal{O}$  consists of all fractions of the form  $f/g \in K$  where  $f, g \in F[v]$  and  $g(0) \neq 0$ . Let  $\mathfrak{X}^\lambda: \mathcal{H}_K \rightarrow K^{d_\lambda \times d_\lambda}$  be a matrix representation afforded by  $E_v^\lambda$ . Following [6], we say that  $\mathfrak{X}^\lambda$  is balanced if

$$v^{\mathbf{a}_\lambda} \mathfrak{X}^\lambda(T_w) \in \mathcal{O}^{d_\lambda \times d_\lambda} \quad \text{for all } w \in W.$$

This concept plays a crucial role in the study of “cellular structures” on  $\mathcal{H}$  (see [6]) and the determination of Kazhdan–Lusztig cells (see [7, §4]). It is known that every  $E_v^\lambda$  affords a balanced representation. Note that, given some matrix representation afforded by  $E_v^\lambda$ , the above condition is hard to verify since it involves representing matrices for *all*  $w \in W$ . Much better for practical purposes is the following condition.

**Proposition 2.1** (See [6, Prop. 4.3, Remark 4.4]) *Assume that  $F \subseteq \mathbb{R}$ . Let  $\lambda \in \Lambda$  and  $\mathfrak{X}^\lambda: \mathcal{H}_K \rightarrow K^{d_\lambda \times d_\lambda}$  be a matrix representation afforded by  $E_v^\lambda$ . Then  $\mathfrak{X}^\lambda$  is balanced if and only if there exists a symmetric matrix  $\Omega^\lambda \in \text{GL}_{d_\lambda}(\mathcal{O})$  such that*

$$\Omega^\lambda \mathfrak{X}^\lambda(T_s) = \mathfrak{X}^\lambda(T_s)^{\text{tr}} \Omega^\lambda \quad \text{for all } s \in S. \tag{*}$$

*Remark 2.2* Note that, if a matrix  $\Omega^\lambda$  satisfies (\*), then it immediately follows that

$$\Omega^\lambda \mathfrak{X}^\lambda(T_{w^{-1}}) = \mathfrak{X}^\lambda(T_w)^{\text{tr}} \Omega^\lambda \quad \text{for all } w \in W.$$

Thus,  $\Omega^\lambda$  is the Gram matrix of a symmetric bilinear form  $\langle \cdot, \cdot \rangle_\lambda: E_v^\lambda \times E_v^\lambda \rightarrow K$  which is  $\mathcal{H}_K$ -invariant in the sense that

$$\langle T_w \cdot e, e' \rangle_\lambda = \langle e, T_{w^{-1}} \cdot e' \rangle_\lambda \quad \text{for all } e, e' \in E_v^\lambda \text{ and } w \in W.$$

*Remark 2.3* Assume that  $F \subseteq \mathbb{R}$ . Let  $\lambda \in \Lambda$  and  $\mathfrak{X}^\lambda: \mathcal{H}_K \rightarrow K^{d_\lambda \times d_\lambda}$  be a matrix representation afforded by  $E_v^\lambda$ . Let  $\mathcal{E}(\mathfrak{X}^\lambda)$  be the set of all  $P \in K^{d_\lambda \times d_\lambda}$  such that  $P \mathfrak{X}^\lambda(T_s) = \mathfrak{X}^\lambda(T_s)^{\text{tr}} P$  for  $s \in S$ . Since  $\mathfrak{X}^\lambda$  is irreducible, Schur’s Lemma implies that all matrices in  $\mathcal{E}(\mathfrak{X}^\lambda)$  are scalar multiples of each other. By Geck and Jacon



[8, Remark 1.4.9], there is a specific element  $P_0 \in \mathcal{E}(\mathfrak{X}^\lambda)$  given by

$$P_0 := \sum_{w \in W} \mathfrak{X}^\lambda(T_w)^{\text{tr}} \mathfrak{X}^\lambda(T_w) \in K^{d_\lambda \times d_\lambda};$$

furthermore, we have  $\det(P_0) \neq 0$ . By the Schur Relations (see [10, 7.2.1]), we have

$$\sum_{w \in W} \mathfrak{X}^\lambda(T_{w^{-1}}) P_0^{-1} \mathfrak{X}^\lambda(T_w) = \text{trace}(P_0^{-1}) \mathbf{c}_\lambda I_{d_\lambda}.$$

Using the relation  $P_0 \mathfrak{X}^\lambda(T_{w^{-1}}) = \mathfrak{X}^\lambda(T_w)^{\text{tr}} P_0$  for all  $w \in W$ , we deduce that

$$\text{trace}(P_0^{-1}) \mathbf{c}_\lambda = 1.$$

This provides a direct criterion for checking if a given matrix  $P \in \mathcal{E}(\mathfrak{X}^\lambda)$  equals  $P_0$ . Furthermore, if  $P \neq 0$  is an element of  $\mathcal{E}(\mathfrak{X}^\lambda)$ , then  $P = cP_0$  for some  $0 \neq c \in K$  and so  $\mathbf{c}_\lambda \text{trace}(P^{-1})P = \mathbf{c}_\lambda \text{trace}(P_0^{-1})P_0 = P_0$ .

The following concept was introduced by Kazhdan–Lusztig [16] in the equal parameter case (where  $L(s) = 1$  for all  $s \in S$ ); for the general case see [8, §1.4].

**Definition 2.4** Let  $V$  be an  $\mathcal{H}_K$ -module with  $d := \dim V < \infty$ . We say that  $V$  is *afforded by a  $W$ -graph* if there exist

- a basis  $\{e_1, \dots, e_d\}$  of  $V$ ,
- subsets  $I_i \subseteq S$  for  $1 \leq i \leq d$ ,
- and elements  $m_{ij}^s \in A$ , where  $1 \leq i, j \leq d$  and  $s \in I_i \setminus I_j$ ,

such that the following hold. First, we require that

$$v^{L(s)} m_{ij}^s \in vR[v] \quad \text{and} \quad m_{ij}^s = m_{ij}^s|_{v \mapsto v^{-1}} \quad \text{for all } 1 \leq i, j \leq d, s \in I_i \setminus I_j.$$

Furthermore, for  $s \in S$ , the action of  $T_s$  on  $V$  is given by

$$T_s \cdot e_j = \begin{cases} v^{L(s)} e_j + \sum_{1 \leq i \leq d: s \in I_i} m_{ij}^s e_i & \text{if } s \notin I_j, \\ -v^{-L(s)} e_j & \text{if } s \in I_j. \end{cases}$$

Thus, if  $V$  is afforded by a  $W$ -graph representation, then the action of  $T_s$  on  $V$  is given by matrices of a particularly simple form.

It has been conjectured in [6] (see also [8, 1.4.14]) that, if the simple  $\mathcal{H}_K$ -module  $E_v^\lambda$  is afforded by a  $W$ -graph, then the corresponding matrix representation is balanced. We now turn to the problem of explicitly verifying if a given irreducible matrix representation of  $\mathcal{H}_K$  is balanced or not.

### 2.5 Explicit Results

We shall assume from now that  $W$  is a finite Weyl group and that we are in the equal parameter case where  $L(s) = 1$  for all  $s \in S$ ; we may take  $R = \mathbb{Z}$ ,  $F = \mathbb{Q}$  in the above discussion. (The remaining cases have been dealt with in [6, Examples 4.5, 4.6].) It is known that every simple  $\mathcal{H}_K$ -module  $E_v^\lambda$  is afforded by a  $W$ -graph; see [8, Theorem 2.7.2] and the references there. As far as  $W$  of exceptional type is concerned, such  $W$ -graphs have been determined explicitly, by Naruse [22], Howlett and Yin [14, 15]. They are available in electronic form through Michel's development version of the CHEVIE system; see [20]. Now let us fix  $\lambda \in \Lambda$  and assume that  $\mathfrak{X}^\lambda: \mathcal{H}_K \rightarrow K^{d_\lambda \times d_\lambda}$  is a corresponding representation afforded by a  $W$ -graph. Concretely, this will mean that we are given the collection of matrices  $\{X_s := \mathfrak{X}^\lambda(T_s) \mid s \in S\}$ . Our aim is to find a matrix  $P = (p_{ij})_{1 \leq i, j \leq d_\lambda}$  such that

$$PX_s = X_s^t P \quad \text{for all } s \in S. \tag{*}$$

This is a system of  $|S|d_\lambda^2$  homogeneous linear equations for the  $d_\lambda(d_\lambda + 1)/2$  unknown entries of  $P$ . (Recall that  $P$  is symmetric.) We know that  $P$  is uniquely determined up to scalar multiples. Rescaling a given solution by a suitable non-zero polynomial in  $\mathbb{Q}[v]$ , we can assume that all entries of  $P$  are in  $\mathbb{Z}[v]$  and that their greatest common divisor is  $\pm 1$ ; then  $P$  is unique up to a sign. Such a solution  $P$  will be called a *primitive Gram matrix* for  $\mathfrak{X}^\lambda$ . As in Remark 2.3, a specific solution  $P_0$  can be singled out by the condition that  $\text{trace}(P_0^{-1})\mathbf{c}_\lambda = 1$ . We claim that

- the matrix  $P'_0 := v^{2l(w_0)}P_0$  has entries in  $\mathbb{Z}[v]$ , and
- the non-zero entries of  $P'_0$  have degree at most  $2l(w_0)$ .

Here,  $w_0$  denotes the longest element of  $W$ . Indeed, since all the entries of the matrices  $X_s$  ( $s \in S$ ) are in  $\mathbb{Z}[v, v^{-1}]$ , the same will be true for  $P_0$  as well. The formulae in Definition 2.4 show that each matrix  $vX_s$  ( $s \in S$ ) has entries in  $\mathbb{Z}[v]$ . Hence, all matrices  $v^{l(w_0)}\mathfrak{X}^\lambda(T_w)$  have entries in  $\mathbb{Z}[v]$  and so  $P'_0$  has entries in  $\mathbb{Z}[v]$ . Furthermore, the non-zero entries of each matrix  $vX_s$  have degree 0, 1 or 2. This yields the degree bound for the entries of  $P'_0$ .

Since the entries of  $P'_0$  are integer polynomials of bounded degree, we can determine  $P'_0$  by interpolation and modular techniques (Chinese remainder). Combining this with the techniques described in [9, §4.3], one obtains an algorithm which can be implemented in GAP in a straightforward way. Rescaling these matrices by suitable non-zero polynomials in  $\mathbb{Q}[v]$ , we obtain primitive Gram matrices as solutions of (\*). This approach readily produces primitive Gram matrices for  $W$  of type  $F_4, E_6$  and  $E_7$  in a few hours of computing time. As was already advertised in Sect. 1, we also succeeded in obtaining primitive Gram matrices for type  $E_8$ , where it is one of the purposes of this paper to describe the methods involved.

Tables 1 and 2 contain some information about these primitive Gram matrices  $P$ :

- 1st column: usual names of the irreducible representations.
- 2nd column: maximum degree of a non-zero entry of  $P$ .
- 3rd column: maximum absolute value of a coefficient of an entry of  $P$ .
- 4th column: is the specialized matrix  $P|_{v \rightarrow 0}$  diagonal?
- 5th column: prime divisors of the determinant of  $P|_{v \rightarrow 0}$ .  
(No entry means that this determinant is  $\pm 1$ .)

We note that the primes in the 5th column are so-called “bad primes” for  $W$  (as in [8, 1.5.11]). In particular, the fact that  $P|_{v \rightarrow 0}$  always has a non-zero determinant means that  $\det(P) \in \mathcal{O}^\times$  (see Proposition 2.1). Thus, we can conclude:

**Table 1** Information on primitive Gram matrices for type  $F_4, E_6, E_7$ ; cf. 2.5

$F_4$	Deg.	Abs. val.	Diag.	Det	$E_6$	Deg.	Abs. val.	Diag.	Det	$E_7$	Deg.	Abs. val.	Diag.	Det	$E_7$	Deg.	Abs. val.	Diag.	Det
1 <sub>1</sub>	0	1	y		1 <sub>p</sub>	0	1	y		1 <sub>a</sub>	0	1	y		168 <sub>a</sub>	10	35	y	
1 <sub>2</sub>	0	1	y		1' <sub>p</sub>	0	1	y		1' <sub>a</sub>	0	1	y		168' <sub>a</sub>	26	2193	y	
1 <sub>3</sub>	0	1	y		10 <sub>s</sub>	6	3	y		7 <sub>a</sub>	12	6	y		189 <sub>a</sub>	12	56	y	
1 <sub>4</sub>	0	1	y		6 <sub>p</sub>	2	1	y		7' <sub>a</sub>	2	1	y		189' <sub>a</sub>	16	112	y	
2 <sub>1</sub>	2	1	y		6' <sub>p</sub>	10	5	y		15 <sub>a</sub>	8	8	y		189 <sub>b</sub>	28	7498	y	
2 <sub>2</sub>	2	1	y		20 <sub>s</sub>	6	3	y		15' <sub>a</sub>	6	3	y		189' <sub>b</sub>	10	42	y	
2 <sub>3</sub>	2	1	y		15 <sub>p</sub>	4	2	y		21 <sub>a</sub>	4	2	y		189 <sub>c</sub>	22	454	y	
2 <sub>4</sub>	2	1	y		15' <sub>p</sub>	8	4	y		21' <sub>a</sub>	10	7	y		189' <sub>c</sub>	10	38	y	
4	4	2	y		15 <sub>q</sub>	6	3	y		21 <sub>b</sub>	16	19	y		210 <sub>a</sub>	10	35	y	
9 <sub>1</sub>	4	2	y		15' <sub>q</sub>	8	8	y		21' <sub>b</sub>	4	2	y		210' <sub>a</sub>	22	973	y	
9 <sub>2</sub>	6	8	y	2	20 <sub>p</sub>	4	2	y		27 <sub>a</sub>	4	2	y		210 <sub>b</sub>	14	253	y	
9 <sub>3</sub>	6	12	y	2	20' <sub>p</sub>	22	61	y		27' <sub>a</sub>	32	164	y		210' <sub>b</sub>	16	468	y	
9 <sub>4</sub>	10	6	y		24 <sub>p</sub>	6	5	y		35 <sub>a</sub>	8	6	y		216 <sub>a</sub>	22	1596	y	
6 <sub>1</sub>	4	2	y		24' <sub>p</sub>	12	20	y		35' <sub>a</sub>	6	3	y		216' <sub>a</sub>	14	227	y	
6 <sub>2</sub>	4	4	y	2	30 <sub>p</sub>	6	6	y	2	35 <sub>b</sub>	6	3	y		280 <sub>a</sub>	22	1836	n	3
12	8	54	y	2,3	30' <sub>p</sub>	18	304	y	2	35' <sub>b</sub>	22	144	y		280' <sub>a</sub>	12	58	n	3
4 <sub>1</sub>	2	2	y	2	60 <sub>s</sub>	10	26	y		56 <sub>a</sub>	26	1082	y	2	280 <sub>b</sub>	14	241	y	
4 <sub>2</sub>	2	1	y		80 <sub>s</sub>	14	711	y	2,3	56' <sub>a</sub>	6	6	y	2	280' <sub>b</sub>	20	2368	y	
4 <sub>3</sub>	2	1	y		90 <sub>s</sub>	12	58	n	3	70 <sub>a</sub>	12	56	y		315 <sub>a</sub>	26	47,277	y	2,3
4 <sub>4</sub>	6	4	y	2	60 <sub>p</sub>	10	21	y		70' <sub>a</sub>	10	26	y		315' <sub>a</sub>	14	4122	y	2,3
8 <sub>1</sub>	4	2	y		60' <sub>p</sub>	12	44	y		84 <sub>a</sub>	10	26	y		336 <sub>a</sub>	20	892	y	
8 <sub>2</sub>	6	3	y		64 <sub>p</sub>	8	12	y		84' <sub>a</sub>	16	148	y		336' <sub>a</sub>	14	175	y	
8 <sub>3</sub>	4	2	y		64' <sub>p</sub>	20	192	y		105 <sub>a</sub>	22	377	y		378 <sub>a</sub>	24	7310	y	
8 <sub>4</sub>	6	3	y		81 <sub>p</sub>	10	24	y		105' <sub>a</sub>	8	12	y		378' <sub>a</sub>	14	453	y	
16	8	16	y	2	81' <sub>p</sub>	12	32	y		105 <sub>b</sub>	10	21	y		405 <sub>a</sub>	14	637	y	2
										105' <sub>b</sub>	20	504	y		405' <sub>a</sub>	26	46,878	y	2
										105 <sub>c</sub>	12	38	y		420 <sub>a</sub>	16	1332	y	2
										105' <sub>c</sub>	12	44	y		420' <sub>a</sub>	20	4148	y	2
										120 <sub>a</sub>	8	24	y	2	512 <sub>a</sub>	20	6036	y	
										120' <sub>a</sub>	30	7516	y	2	512' <sub>a</sub>	20	6036	y	

**Table 2** Information on primitive Gram matrices for type  $E_8$ ; cf. 2.5

Repr.	Deg.	Abs. val.	Diag.	Det	Repr.	Deg.	Abs. val.	Diag.	Det	Repr.	Deg.	Abs. val.	Diag.	Det	Repr.	Deg.	Abs. val.	Diag.	Det
$1_x$	0	1	y	538	700 <sub>x</sub>	12	700 <sub>x</sub>	2	2	2835 <sub>x</sub>	24	1,344,484	y	840 <sub>x</sub>	26	8048	y		
$1'_x$	0	1	y	16,489,188	700' <sub>x</sub>	54	16,489,188	y	2	2835' <sub>x</sub>	32	5,391,418	y	1008 <sub>x</sub>	42	156	n	3	
$28_x$	4	2	y	22,286	1400 <sub>x</sub>	22	22,286	n	2,3	5670 <sub>x</sub>	30	10,762,741	n	2,3,5	1008' <sub>x</sub>	40	66,780	n	3
$28'_x$	12	10	y	6044	840 <sub>x</sub>	16	6044	y		3200 <sub>x</sub>	24	266,284	y		2016 <sub>w</sub>	28	797,422	y	
$35_x$	4	2	y	37,603	840' <sub>x</sub>	26	37,603	y		3200' <sub>x</sub>	30	587,345	y		1296 <sub>x</sub>	14	345	y	
$35'_x$	38	377	y	3447	1680 <sub>x</sub>	22	3447	n	2,5	4096 <sub>x</sub>	22	531,634	y		1296' <sub>x</sub>	34	23,195	y	
$70_y$	8	6	y	2098	972 <sub>x</sub>	16	2098	y		4096' <sub>x</sub>	44	234,956,568	y		1400 <sub>z</sub>	16	10,042	y	
$50_x$	8	6	y	185,342	972' <sub>x</sub>	36	185,342	y		4200 <sub>x</sub>	24	5,413,484	y	2	1400' <sub>z</sub>	34	358,379	y	
$50'_x$	22	257	y	3792	1050 <sub>x</sub>	16	3792	y		4200' <sub>x</sub>	36	129,331,224	y	2	1400 <sub>z</sub>	14	8148	y	2,3
$84_x$	6	3	y	390,765	1050' <sub>x</sub>	34	390,765	y		6075 <sub>x</sub>	26	894,864	y		1400' <sub>z</sub>	50	60,122,676	y	2,3
$84'_x$	38	675	y	5561	2100 <sub>x</sub>	22	5561	y		6075' <sub>x</sub>	34	10,488,013	y		2400 <sub>x</sub>	22	6380	y	
$168_y$	16	340	y	1140	1344 <sub>x</sub>	14	1140	y		8 <sub>z</sub>	2	1	y		2400' <sub>x</sub>	28	55,922	y	
$175_x$	12	52	y	381,082	1344' <sub>x</sub>	40	381,082	y		8' <sub>z</sub>	14	6	y		2800 <sub>x</sub>	20	38,038	y	2
$175'_x$	20	992	y	169,180	2688 <sub>x</sub>	24	169,180	y		56 <sub>z</sub>	6	3	y		2800' <sub>x</sub>	30	882,222	y	2
$210_x$	8	24	y	41,820	1400 <sub>x</sub>	16	41,820	y	2,3	56' <sub>z</sub>	10	7	y		5600 <sub>w</sub>	26	372,230	n	3
$210'_x$	42	95,780	y	763,453,596	1400' <sub>x</sub>	48	763,453,596	y	2,3	112 <sub>z</sub>	6	6	y	2	3240 <sub>x</sub>	16	25,586	y	
$420_y$	16	1432	y	783	1575 <sub>x</sub>	14	783	n	3	112' <sub>z</sub>	54	20,790	y	2	3240' <sub>x</sub>	48	33,653,538	y	
$300_x$	10	41	y	850,956	1575' <sub>x</sub>	44	850,956	n	3	160 <sub>z</sub>	8	12	y		3360 <sub>x</sub>	20	29,722	y	
$300'_x$	40	12,710	y	6,166,994	3150 <sub>x</sub>	26	6,166,994	y	2	160' <sub>z</sub>	32	400	y		3360' <sub>x</sub>	32	775,084	y	
$350_x$	12	56	y	3514	2100 <sub>x</sub>	20	3514	y		448 <sub>w</sub>	16	128	y		7168 <sub>w</sub>	32	1,190,470,476	y	2,3
$350'_x$	20	290	y	12,511	2100' <sub>x</sub>	26	12,511	y		400 <sub>z</sub>	12	132	y		4096 <sub>x</sub>	22	531,634	y	
$525_x$	12	76	y	58,249,760	4200 <sub>x</sub>	28	58,249,760	n	2	400' <sub>z</sub>	38	58,368	y		4096' <sub>x</sub>	44	234,956,568	y	
$525'_x$	24	1946	y	1,878,156	2240 <sub>x</sub>	20	1,878,156	y	2	448 <sub>z</sub>	12	290	y		4200 <sub>x</sub>	26	728,053	y	
$567_x$	10	54	y	60,390,945	2240' <sub>x</sub>	42	60,390,945	y	2	448' <sub>z</sub>	32	17,290	y		4200' <sub>x</sub>	28	1,298,612	y	
$567'_x$	42	57,812	y	85,556,320,920	4480 <sub>x</sub>	32	85,556,320,920	y	2,3,5	560 <sub>z</sub>	10	73	y		4536 <sub>x</sub>	24	2,728,756	y	
$1134_y$	22	8739	y	5948	2268 <sub>x</sub>	16	5948	y	2	560' <sub>z</sub>	46	408,409	y		4536' <sub>x</sub>	38	50,779,421	y	
$700_{xx}$	18	1399	y	6,442,224	2268' <sub>x</sub>	40	6,442,224	y	2	1344 <sub>w</sub>	24	177,956	y		5600 <sub>x</sub>	26	3,115,126	y	2
$700'_{xx}$	20	5982	y	3,887,856	4536 <sub>y</sub>	28	3,887,856	n	2	840 <sub>z</sub>	14	643	y		5600' <sub>x</sub>	30	3,848,044	y	2

**Theorem 2.5** *Let  $W$  be of type  $F_4, E_6, E_7$  or  $E_8$  and  $L(s) = 1$  for all  $s \in S$ . Then the  $W$ -graph representations of Naruse [22], Howlett and Yin [14, 15] are balanced.*

### 3 Linear Algebra Over the Integers

As was already mentioned in Sect. 1, the basic strategy of our approach to determine Gram matrices of invariant bilinear forms for representations of Iwahori–Hecke algebras is to reduce computational linear algebra over the polynomial rings  $\mathbb{Z}[X]$  or  $\mathbb{Q}[X]$ , where from now on  $X$  denotes our favorite indeterminate, to computational linear algebra over the integers  $\mathbb{Z}$ . Thus in this section we begin by describing how we deal with matrices over  $\mathbb{Z}$ , where we restrict ourselves to the aspects needed for our present application.

Let us fix the following convention: For  $x, y \in \mathbb{Z}$ , not both zero, let  $\gcd(x, y) \in \mathbb{Z}$  denote the positive greatest common divisor of  $x$  and  $y$ . A vector  $0 \neq v \in \mathbb{Q}^m$ , where  $m \in \mathbb{N}$ , is called *primitive*, if actually  $v \in \mathbb{Z}^m$ , and for the greatest common divisor  $\gcd(v)$  of its entries we have  $\gcd(v) = 1$ . Clearly greatest common divisor computations in  $\mathbb{Z}$  yield a  $\mathbb{Q}$ -multiple of  $v$  which is primitive. Similarly, a matrix  $0 \neq A \in \mathbb{Z}^{m \times n}$ , where  $m, n \in \mathbb{N}$ , is called *primitive*, if actually  $A \in \mathbb{Z}^{m \times n}$ , and for the greatest common divisor  $\gcd(A)$  of its entries we have  $\gcd(A) = 1$ .

#### 3.1 Continued Fractions and the Euclidean Algorithm

The first computational task we are going to discuss, in Sect. 3.2 below, is rational number recovery. This has been discussed in the literature at various places, see for example [3, 21, 24] or [26, Section 5.10]. (We also gratefully acknowledge additional private discussions with R. Parker on this topic.) Although the ideas pursued in these references are closely related to ours, none of them completely coincides with our approach, and proofs (if given at all) are not too elucidating. Hence we present our approach in detail, for which we need a few preparations first:

**Continued Fraction Expansions** We recall a few notions from the theory of continued fraction expansions; as a general reference see for example [13, Chapter 10]: Given  $\rho \in \mathbb{R}$  such that  $\rho \geq 0$ , let

$$\text{cf}[q_1, q_2, \dots] = q_1 + \frac{1}{q_2 + \frac{1}{\ddots}}$$

be its (*regular*) *continued fraction* expansion, where  $q_1 \in \mathbb{N}_0$  and  $q_i \in \mathbb{N}$  for  $i \geq 2$ . This is obtained by letting  $q_1 := \lfloor \rho \rfloor$ , and, as long as  $\rho \neq q_1$ , proceeding recursively with  $\frac{1}{\rho - q_1}$  instead of  $\rho$ . This process terminates, after  $l \geq 1$  steps say, if and only

if  $\rho \in \mathbb{Q}$ ; otherwise we let  $l := \infty$ . Truncating at  $i \leq l$  yields the  $i$ -th *convergent*  $\rho_i := \text{cf}[q_1, \dots, q_i] \in \mathbb{Q}$  of  $\rho$ , hence we may write  $\rho_i := \frac{\sigma_i}{\tau_i}$ , where  $\sigma_i, \tau_i \in \mathbb{N}_0$  such that  $\tau_i \geq 1$  and  $\text{gcd}(\sigma_i, \tau_i) = 1$ . Letting additionally  $\sigma_{-1} := 0$  and  $\tau_{-1} := 1$ , as well as  $\sigma_0 := 1$  and  $\tau_0 := 0$ , for  $i \geq 1$  we get by induction

$$\sigma_i = q_i \sigma_{i-1} + \sigma_{i-2} \quad \text{and} \quad \tau_i = q_i \tau_{i-1} + \tau_{i-2}.$$

Hence the sequences  $[\sigma_1, \sigma_2, \dots, \sigma_l]$  and  $[\tau_2, \tau_3, \dots, \tau_l]$  are strongly increasing.

Now let  $\rho = \frac{a}{b} \in \mathbb{Q}$ , where  $a, b \in \mathbb{N}$ . Then the continued fraction expansion of  $\rho$  can be computed by the extended Euclidean algorithm, see [1, Algorithm 1.3.6], as follows: Setting  $r_0 := a$  and  $r_1 := b$ , for  $1 \leq i \leq l$  let recursively  $q_i \in \mathbb{N}_0$  and

$$r_{i+1} := r_{i-1} - q_i r_i \in \mathbb{N}_0 \quad \text{such that} \quad r_{i+1} < r_i,$$

where  $l \geq 1$  is defined by  $r_l > 0$  but  $r_{l+1} = 0$ ; actually we have  $q_i \geq 1$  for  $i \geq 2$ , and of course  $r_l = \text{gcd}(a, b)$ . Hence the sequence  $[r_1, \dots, r_{l+1}]$  has non-negative entries and is strongly decreasing. Moreover, setting  $s_0 := 1$  and  $t_0 := 0$ , as well as  $s_1 := 0$  and  $t_1 := 1$ , and for  $1 \leq i \leq l$  letting recursively

$$s_{i+1} := s_{i-1} - q_i s_i \quad \text{and} \quad t_{i+1} := t_{i-1} - q_i t_i,$$

we get  $r_i = s_i a + t_i b$ . Then it is immediate by induction that  $\sigma_i = (-1)^i \cdot t_{i+1}$  and  $\tau_i = (-1)^{i+1} \cdot s_{i+1}$ , for  $i \geq 1$ , and hence

$$\rho_i = -\frac{t_{i+1}}{s_{i+1}}, \quad \text{where} \quad \text{gcd}(s_{i+1}, t_{i+1}) = 1, \quad \text{for} \quad 1 \leq i \leq l.$$

Hence the sequences  $[-s_3, s_4, -s_5, \dots, \pm s_{l+1}]$  and  $[-t_2, t_3, -t_4, \dots, \pm t_{l+1}]$  have positive entries and are strongly increasing. Finally, a direct computation yields

$$\rho - \rho_i = \frac{a}{b} - \frac{\sigma_i}{\tau_i} = \frac{\tau_i a - \sigma_i b}{\tau_i b} = \frac{s_{i+1} a + t_{i+1} b}{s_{i+1} b} = \frac{r_{i+1}}{b s_{i+1}}, \quad \text{for} \quad 1 \leq i \leq l.$$

**Another View on the Euclidean Algorithm** For  $a, b \in \mathbb{N}$  we consider the  $\mathbb{Z}$ -lattice

$$L_{a,b} := \langle [1, a], [0, b] \rangle_{\mathbb{Z}} \subseteq \mathbb{Z}^2.$$

Then we have  $|\det(L_{a,b})| = b$ , and it is immediate that  $[x, y] \in \mathbb{Z}^2$  is an element of  $L_{a,b}$  if and only if  $y \equiv ax \pmod{b}$ . Note that if  $0 \neq [x, y] \in L_{a,b}$  is primitive, then we necessarily have  $\text{gcd}(x, b) = 1$ . Moreover, the extended Euclidean algorithm shows that  $L_{a,b} = \langle [s_i, r_i], [s_{i+1}, r_{i+1}] \rangle_{\mathbb{Z}}$ , for all  $0 \leq i \leq l$ . We collect a few properties of  $L_{a,b}$ :

**Lemma 3.1**

- (a) For all  $0 \leq i \leq l + 1$  we have  $\langle [s_i, r_i] \rangle_{\mathbb{Q}} \cap L_{a,b} = \langle [s_i, r_i] \rangle_{\mathbb{Z}}$ .
- (b) We have  $\langle [s_i, r_i] \rangle_{\mathbb{Q}} = \langle [s_j, r_j] \rangle_{\mathbb{Q}}$ , where  $1 \leq i, j \leq l + 1$ , if and only if  $i = j$ .

*Proof* We first show that whenever  $[x, y] \in L_{a,b}$  such that  $0 < |y| < r_i$ , for some  $0 \leq i \leq l$ , then  $|x| \geq |s_{i+1}|$ : We may assume that  $i \geq 2$ . Let  $c, d \in \mathbb{Z}$  such that

$$[x, y] = [c, d] \cdot \begin{bmatrix} s_i & r_i \\ s_{i+1} & r_{i+1} \end{bmatrix},$$

where we may assume that  $c \neq 0$ , which entails  $d \neq 0$  as well. Since  $r_i > r_{i+1} \geq 0$ , this implies  $c \cdot d < 0$ . Since the sequence  $[s_2, -s_3, s_4, -s_5 \dots, \pm s_{l+1}]$  has positive entries, we get  $|x| = |cs_i + ds_{i+1}| = |c| \cdot |s_i| + |d| \cdot |s_{i+1}| \geq |s_{i+1}|$ , as asserted.

- (a) We may assume that  $i \geq 2$ . Moreover, for  $i = l + 1$  letting  $[x, 0] \in L_{a,b}$ , it is immediate from  $ax \equiv 0 \pmod{b}$  that  $|s_{l+1}| = \frac{b}{r_l} = \frac{b}{\gcd(a,b)}$  divides  $x$ . Hence we may assume  $i \leq l$ , too. Then let  $d \neq 1$  be a divisor of  $\gcd(s_i, r_i)$  such that  $\frac{1}{d} \cdot [s_i, r_i] \in L_{a,b}$ . Then we have  $0 < |\frac{r_i}{d}| < r_i$  and  $|\frac{s_i}{d}| < |s_i| \leq |s_{i+1}|$ , contradicting the statement above.
- (b) It follows from (a) that there are  $c, d \in \mathbb{Z}$  such that  $[s_j, r_j] = c \cdot [s_i, r_i]$  and  $[s_i, r_i] = d \cdot [s_j, r_j]$ . Hence we get  $cd = 1$ , and since the sequence  $[r_1, \dots, r_{l+1}]$  has non-negative entries and is strongly decreasing, we infer  $r_i = r_j$  and  $i = j$ . □

Note that the statement in (b) is trivial if  $[s_i, r_i]$  is primitive, that is  $\gcd(s_i, r_i) = 1$ . But this is not always fulfilled, as the example in [26, Example 5.27] shows.

**Proposition 3.2**

- (a) Let  $[x, y] \in L_{a,b}$  such that  $x \neq 0$  and  $|x| \cdot |y| \leq \frac{b}{2}$ . Then we have  $[x, y] \in \langle [s_i, r_i] \rangle_{\mathbb{Z}}$ , for a unique  $2 \leq i \leq l + 1$ . In particular, if  $[x, y]$  is primitive then we have  $[x, y] = [s_i, r_i]$  or  $[x, y] = -[s_i, r_i]$ .
- (b) Assume there is  $0 \neq [x, y] \in L_{a,b}$  such that  $\|[x, y]\| := \sqrt{x^2 + y^2} < \sqrt{b}$ . Then there is a unique  $2 \leq i \leq l + 1$  such that  $\|[s_i, r_i]\| < \sqrt{b}$ , and the shortest non-zero elements of  $L_{a,b}$  are precisely  $[s_i, r_i]$  and  $-[s_i, r_i]$ .

*Proof*

- (a) Since  $[x, y] \in L_{a,b}$  there is  $z \in \mathbb{Z}$  such that  $y = xa - zb$ . Then we have

$$\left| \frac{a}{b} - \frac{z}{x} \right| = \frac{|y|}{b \cdot |x|} = \frac{|x| \cdot |y|}{b \cdot |x|^2} \leq \frac{1}{2 \cdot |x|^2}.$$

Thus by Legendre’s Theorem, see [13, Section 10.15, Theorem 184], we infer that  $\frac{z}{x}$  occurs as a convergent in the continued fraction expansion of  $\rho = \frac{a}{b}$ , that is, there is  $2 \leq i \leq l + 1$  such that  $\frac{z}{x} = \rho_{i-1}$ . This yields

$$\frac{y}{x} = \frac{xa - zb}{x} = a - \frac{zb}{x} = a - b\rho_{i-1} = b(\rho - \rho_{i-1}) = \frac{r_i}{s_i}.$$

Hence we have  $[x, y] \in \langle [s_i, r_i] \rangle_{\mathbb{Q}}$ , and thus from Lemma 3.1 we get  $[x, y] \in \langle [s_i, r_i] \rangle_{\mathbb{Z}}$ , together with the uniqueness statement.

(b) Assume first that  $x = 0$ , then by Lemma 3.1 we infer that  $b$  divides  $y$ , and hence  $\|[x, y]\| \geq b \geq \sqrt{b}$ , a contradiction. Hence we have  $x \neq 0$ . Moreover, from  $(x - y)^2 = x^2 + y^2 - 2xy \geq 0$  we get  $2 \cdot |x| \cdot |y| \leq x^2 + y^2 = \|[x, y]\|^2 < b$ , hence from (a) we see that there is  $2 \leq i \leq l + 1$  such that  $[x, y] = \langle [s_i, r_i] \rangle_{\mathbb{Z}}$ . Thus in particular we have  $\|[s_i, r_i]\| < \sqrt{b}$ .

In order to show uniqueness, and the statement on shortest elements, let  $0 \neq [x', y'] \in L_{a,b}$  such that  $\|[x', y']\| < \sqrt{b}$ . Then, as above, there is  $2 \leq i \leq l + 1$  such that  $[x', y'] = \langle [s_j, r_j] \rangle_{\mathbb{Z}}$ , hence in particular we have  $\|[s_j, r_j]\| < \sqrt{b}$ . Then Hadamard’s inequality, see [26, Theorem 16.6], implies that

$$\det \left( \begin{bmatrix} s_i & r_i \\ s_j & r_j \end{bmatrix} \right) \leq \|[s_i, r_i]\| \cdot \|[s_j, r_j]\| < b.$$

Since  $|\det(L_{a,b})| = b$  divides  $\det \left( \begin{bmatrix} s_i & r_i \\ s_j & r_j \end{bmatrix} \right)$  this entails  $\langle [s_i, r_i] \rangle_{\mathbb{Q}} = \langle [s_j, r_j] \rangle_{\mathbb{Q}}$ , and hence  $i = j$  by Lemma 3.1. □

A comparison of the above treatment with the references already mentioned seems to be in order: The statement of Proposition 3.2(a) is roughly equivalent to [3, Theorem] and [21, Theorem 1], respectively. Alone, the proof given in [3] appears to be too concise, and provides a slightly worse bound for  $b$  to be large enough. And [21, Theorem 1] is attributed in turn to [2], while for a proof the reader is referred to [26]. Unfortunately, [26, Theorem 5.26] is not immediately conclusive for the statements under consideration here.

The main difference between the above-mentioned approaches and ours is the break condition used to actually determine the index  $i$  referred to in Proposition 3.2(a): In [2, 3, 26] a bound on the residues  $r_i$  is used, while in [21, Section 3] the quotients  $q_i$  are considered instead (yielding a randomized algorithm). In contrast, in our decisive Proposition 3.2(b) we are using the minimum of the lattice  $L_{a,b}$ , which hence treats both the  $r_i$  and  $s_i$  (in other words the unknown numbers  $y$  and  $x$ ) on a “symmetric” footing. To our knowledge, this point of view is new, its algorithmic relevance being explained below.

### 3.2 Recovering Rational Numbers

We are now prepared to describe our first computational task, which will appear both in computations over  $\mathbb{Z}$  in Sect. 3.3, and over the polynomial ring  $\mathbb{Q}[X]$  in Sect. 4.2:

Let  $x \in \mathbb{N}$  and  $0 \neq y \in \mathbb{Z}$  such that  $\gcd(x, y) = 1$ . Assume we are given  $a, b \in \mathbb{N}$  such that  $\gcd(x, b) = 1$  and  $y \equiv ax \pmod{b}$ ; note that since  $x$  is invertible modulo  $b$  we may write  $\frac{y}{x} \equiv a \pmod{b}$  instead, which we will feel free to do if convenient. Now, if  $b$  is large enough compared to  $x$  and  $|y|$ , the task is to recover  $\frac{y}{x} \in \mathbb{Q}$  from its congruence class  $a \pmod{b}$ .



In view of Proposition 3.2(b), this is straightforward: Assuming that  $x^2 + y^2 < b$ , the  $\mathbb{Z}$ -lattice  $L_{a,b} = \langle [1, a], [0, b] \rangle_{\mathbb{Z}} \subseteq \mathbb{Z}^2$  has precisely two shortest non-zero elements, namely the primitive elements  $\pm[x, y]$ . In other words, the rational number  $\frac{y}{x} \in \mathbb{Q}$  can be found by computing a shortest non-zero element of  $L_{a,b}$ . This in turn can be done algorithmically by the Gauß reduction algorithm for  $\mathbb{Z}$ -lattices of rank 2, see [1, Algorithm 1.3.14]. Moreover, compared to the general case, for the particular lattice  $L_{a,b}$  we have a better break condition: We may stop early as soon as we have found an element  $[x, y] \in L_{a,b}$  such that  $x^2 + y^2 < b$ . If then  $[x, y]$  is primitive, the rational number  $\frac{y}{x}$  fulfills all assumptions made, where of course its correctness has to be verified independently. Otherwise, if  $[x, y]$  is not primitive, or the shortest element  $[x', y'] \in L_{a,b}$  found fulfills  $x'^2 + y'^2 \geq b$ , then we report failure. Thus, in practice, we choose  $b$  small, and rerun the above algorithm with  $b$  increasing, until we find a valid candidate passing independent verification.

At this stage, we should point out the algorithmic advantage of our approach, compared to the other ones mentioned: The latter refer to the convergents of continued fraction expansions, and thus to the full sequence of non-negative residues of the extended Euclidean algorithm. In contrast, the Gauß reduction algorithm to find a lattice minimum proceeds by iterated pair reduction, starting with the pair  $[0, b]$  and  $[1, a]$ . Although this is essentially equivalent to running the extended Euclidean algorithm on  $a$  and  $b$ , here we are allowed to use best approximation. This amounts to using numerically smallest residues, instead of non-negative ones as was necessary in the context of continued fraction expansions. Although we have not carried out a detailed comparison, it is well-known that this saves a non-negligible amount of quotient and remainder steps.

### 3.3 Finding Linear Combinations

We are now going to describe *the* basic task we are faced with in order to be able to do computational linear algebra over  $\mathbb{Z}$ . To do so, we of course avoid the Gauß algorithm over  $\mathbb{Q}$ , but we also do not refer to pure “lattice algorithms”, as they are called in [1, Section 2.1], for example those to compute Hermite normal forms or reduced lattice bases described in [1, Sections 2.4–2.7]. Instead, we use a modular technique, which is a keystone to make use of the ideas of the **MeatAxe** in the framework of the **IntegralMeatAxe**. To our knowledge, this has only been discussed very briefly in the literature, for example in [3, 24]. Moreover, our approach differs from those cited, at least in detail; in particular, [3] only allows for regular square matrices.

To describe the computational task, we again need some preparations first: Given a (rectangular) matrix  $A \in \mathbb{Z}^{m \times n}$ , with  $\mathbb{Q}$ -linearly independent rows  $w_1, \dots, w_m \in \mathbb{Z}^n$ , where  $m, n \in \mathbb{N}$ , let

$$L := \langle w_1, \dots, w_m \rangle_{\mathbb{Z}} \leq \mathbb{Z}^n$$

be the  $\mathbb{Z}$ -lattice spanned by the rows of  $A$ , and let  $L \leq \widehat{L} \leq \mathbb{Z}^n$  be its *pure closure* in  $\mathbb{Z}^n$ , that is the smallest pure  $\mathbb{Z}$ -sublattice of  $\mathbb{Z}^n$  containing  $L$ . Then the index  $\det(L) := [\widehat{L}:L]$  is finite; of course, if  $m = n$  then we have  $\det(L) = |\det(A)|$ . Thus for any vector  $v \in \mathbb{Z}^n$ , we have  $v \in \widehat{L}$  if and only if there is  $a \in \mathbb{N}$  such that  $av \in L$ ; in this case, if  $a$  is chosen minimal then it divides  $\det(L)$ .

Now, given  $v \in \mathbb{Z}^n$ , the task is to decide whether or not  $v \in \widehat{L}$ , and if this is the case to compute  $a_1, \dots, a_m \in \mathbb{Z}$  and  $a \in \mathbb{N}$  such that  $\gcd(a, a_1, \dots, a_m) = 1$  and

$$v = \frac{1}{a} \cdot \sum_{j=1}^m a_j w_j = \frac{1}{a} \cdot [a_1, \dots, a_m] \cdot A;$$

in this case  $a$  and the  $a_i$  are uniquely determined.

**The  $p$ -Adic Decomposition Algorithm** To do so, we choose a (large) prime  $p$ . Then reduction modulo  $p$  yields the matrix  $\bar{A} \in \mathbb{F}_p^{m \times n}$  over the prime field  $\mathbb{F}_p$ . We assume that the rows  $\bar{w}_1, \dots, \bar{w}_m \in \mathbb{F}_p^n$  of  $\bar{A}$  are  $\mathbb{F}_p$ -linearly independent as well; otherwise we choose another prime  $p$ . By the structure theory of finitely generated modules over principal ideal domains, this condition is equivalent to saying  $\widehat{\bar{L}} = \bar{L}$ , which in turn is equivalent to  $p$  not dividing  $\det(L)$ . In particular, the independence condition on  $\bar{w}_1, \dots, \bar{w}_m \in \mathbb{F}_p^n$  is fulfilled for all but finitely many primes  $p$ .

Thus we have  $v \in \widehat{L}$  if and only if  $\bar{v} \in \bar{L} = \langle \bar{w}_1, \dots, \bar{w}_m \rangle_{\mathbb{F}_p}$ , solving the decision problem. Furthermore, if  $v \in \widehat{L}$  then set  $v_0 := v$ , and for  $d \in \mathbb{N}_0$  proceed successively as follows: Since  $v_d \in \widehat{L}$ , there are  $[a_{d,1}, \dots, a_{d,m}] \in \mathbb{Z}^m$  such that  $-\frac{p}{2} < a_{d,j} \leq \frac{p}{2}$  for all  $1 \leq j \leq m$ , and

$$\bar{v}_d = \sum_{j=1}^m \bar{a}_{d,1} \bar{w}_j = [\bar{a}_{d,1}, \dots, \bar{a}_{d,m}] \cdot \bar{A} \in \mathbb{F}_p^n.$$

Then we let

$$v_{d+1} := \frac{1}{p} \cdot \left( v_d - [a_{d,1}, \dots, a_{d,m}] \cdot A \right) \in \mathbb{Z}^n.$$

Hence we have  $v_{d+1} \in \widehat{L}$  as well, and we may recurse. This yields

$$v \equiv \left( \sum_{i=0}^d p^i \cdot [a_{i,1}, \dots, a_{i,m}] \right) \cdot A \equiv \left[ \sum_{i=0}^d p^i a_{i,1}, \dots, \sum_{i=0}^d p^i a_{i,m} \right] \cdot A \pmod{p^{d+1} \mathbb{Z}^n},$$

or equivalently

$$\frac{a_j}{a} \equiv \sum_{i=0}^d p^i a_{i,j} \pmod{p^{d+1}}, \quad \text{for all } 1 \leq j \leq m.$$

Thus, if  $v \in L$ , or equivalently  $a = 1$ , then since  $-\frac{v^{d+1}}{2} < \sum_{i=0}^d p^i a_{i,j} \leq \frac{v^{d+1}}{2}$  there is some  $d \in \mathbb{N}_0$  such that  $v_{d+1} = 0$ , implying that  $a_j = \sum_{i=0}^d p^i a_{i,j}$ , for all  $1 \leq j \leq m$ , without further independent verification necessary. Otherwise, if  $v \in \widehat{L} \setminus L$ , then applying rational number recovery for some  $d \in \mathbb{N}_0$  large enough, see Sect. 3.2, reveals the vector  $\frac{1}{a} \cdot [a_1, \dots, a_m] \in \mathbb{Q}^m$ ; note that under the assumptions made  $p$  does not divide  $a$ . In the latter case correctness is independently verified by computing  $[a_1, \dots, a_m] \cdot A \in \mathbb{Z}^n$  and checking whether it equals  $av \in \mathbb{Z}^n$ .

**Modular Computations** In practice, to check  $\overline{w}_1, \dots, \overline{w}_m \in \mathbb{F}_p^n$  for  $\mathbb{F}_p$ -linear independence, and to compute the vectors  $[\overline{a}_{d,1}, \dots, \overline{a}_{d,m}] \in \mathbb{F}_p^m$  we use ideas taken from the `MeatAxe`. In particular, in order to keep the depth  $d$  needed smallish, but still to be able to make efficient use of fast arithmetic over small finite prime fields, we choose the prime  $p$  amongst the largest primes smaller than  $2^8 = 256$ . (In our application we for example use  $p = 251$  as the default prime.)

### 3.4 Nullspace

In the framework of the `IntegralMeatAxe` there is a general method to compute a  $\mathbb{Z}$ -basis of the row kernel of a matrix with entries in  $\mathbb{Z}$ , see [24]. But in view of the application to row kernels of matrices over  $\mathbb{Q}[X]$  in Sect. 5.1, here we only deal with the following restricted nullspace problem:

Given a matrix  $A \in \mathbb{Q}^{m \times n}$ , where  $m, n \in \mathbb{N}$ , such that  $\dim_{\mathbb{Q}}(\ker(A)) = 1$ , where  $\ker(A)$  denotes the row kernel of  $A$ , compute a primitive vector  $v \in \mathbb{Z}^m$  such that  $\ker(A) = \langle v \rangle_{\mathbb{Q}}$ ; then  $v$  is unique up to sign.

To do so, by going over to a suitable  $\mathbb{Q}$ -multiple we may assume that  $A \in \mathbb{Z}^{m \times n}$ . Let  $w_1, \dots, w_m \in \mathbb{Z}^n$  be the rows of  $A$ . We may assume that  $w_1 \neq 0$ , since otherwise we trivially set  $v := [1, 0, \dots, 0] \in \mathbb{Z}^m$ . Then for  $2 \leq i \leq m$  we successively check, using the  $p$ -adic decomposition algorithm in Sect. 3.3, whether or not  $w_i \in \langle w_1, \dots, w_{i-1} \rangle_{\mathbb{Q}}$ . If this is not the case, that is  $\{w_1, \dots, w_i\}$  is  $\mathbb{Q}$ -linearly independent, then if  $\overline{w}_1, \dots, \overline{w}_i \in \mathbb{F}_p^n$  turns out to be  $\mathbb{F}_p$ -linearly independent we increment  $i$ , while otherwise we return failure in order to choose another prime  $p$ . If  $\{w_1, \dots, w_i\}$  is  $\mathbb{Q}$ -linearly dependent, then the  $p$ -adic decomposition algorithm returns  $a_1, \dots, a_{i-1} \in \mathbb{Z}$  and  $a \in \mathbb{N}$  such that  $\gcd(a, a_1, \dots, a_{i-1}) = 1$  and  $w_i = \frac{1}{a} \cdot \sum_{j=1}^{i-1} a_j w_j$ . Thus  $v := [a_1, \dots, a_{i-1}, -a, 0, \dots, 0] \in \ker(A) \leq \mathbb{Z}^m$  is primitive.

### 3.5 Inverse

Matrix inversion over  $\mathbb{Q}$ , from the point of view of reducing to computations over  $\mathbb{Z}$  as much as possible, can be formulated as the following task:

Given a matrix  $A \in \mathbb{Q}^{n \times n}$ , where  $n \in \mathbb{N}$ , such that  $\det(A) \neq 0$ , compute  $B \in \mathbb{Z}^{n \times n}$  and  $c \in \mathbb{N}$ , such that  $A^{-1} = \frac{1}{c} \cdot B \in \mathbb{Q}^{n \times n}$  and the overall greatest common divisor  $\gcd(B, c)$  of the entries of  $B$  and  $c$  equals  $\gcd(B, c) = 1$ ; then  $(B, c)$  is unique.

To do so, by going over to a suitable  $\mathbb{Q}$ -multiple we may assume that  $A \in \mathbb{Z}^{n \times n}$ . Then the equation  $BA = c \cdot E_n$ , where  $E_n$  denotes the identity matrix, implies that  $\gcd(B)$  divides  $c$ , and hence  $B$  is necessarily primitive. Solving the equations  $\mathcal{X}A = E_n$ , for the unknown matrix  $\mathcal{X} \in \mathbb{Q}^{n \times n}$ , amounts to writing the rows of the identity matrix as  $\mathbb{Q}$ -linear combinations of the rows of  $A$ , which is done using the  $p$ -adic decomposition algorithm in Sect. 3.3; recall that the rows of  $A$  indeed are assumed to be  $\mathbb{Q}$ -linearly independent.

### 3.6 The Exponent of a Matrix

Given a square matrix  $A \in \mathbb{Z}^{n \times n}$  such that  $\det(A) \neq 0$  as above, the number  $c \in \mathbb{N}$  found in the expression  $A^{-1} = \frac{1}{c} \cdot B$ , where  $B \in \mathbb{Z}^{n \times n}$  is chosen to be primitive, turns out to have another interpretation:

Let  $\text{im}(A) \leq \mathbb{Z}^n$  be the  $\mathbb{Z}$ -span of the rows of  $A$ . By the structure theory of finitely generated modules over principal ideal domains, the annihilator of the  $\mathbb{Z}$ -module  $\mathbb{Z}^n / \text{im}(A)$  is a non-zero ideal of  $\mathbb{Z}$ , the positive generator  $\exp(A)$  of which is called the *exponent* of  $A$ . Moreover,  $\exp(A)$  divides  $\det(A)$ , which in turn divides some power of  $\exp(A)$ . Thus the prime divisors of  $\exp(A)$  are precisely the primes  $p \in \mathbb{Z}$  such that  $\bar{A} \in \mathbb{F}_p^{n \times n}$  is not invertible.

Now, actually  $\exp(A)$  and  $c$  coincide: From  $BA = c \cdot E_n$  we conclude that  $(c\mathbb{Z})^n \leq \text{im}(A)$ , hence  $\exp(A)$  divides  $c$ ; conversely, since  $(\exp(A) \cdot \mathbb{Z})^n \leq \text{im}(A)$  there is  $B' \in \mathbb{Z}^{n \times n}$  such that  $B'A = \exp(A) \cdot E_n$ , implying that  $\exp(A) \cdot B = c \cdot B'$ , which by the primitivity of  $B$  shows that  $c$  divides  $\exp(A)$ . In other words, computing the inverse of  $A$  as described in Sect. 3.5 also yields a method to compute  $\exp(A)$ .

## 4 Computing with Polynomials

Having the necessary pieces of linear algebra over the integers in place, in this section we describe computational aspects of single polynomials, before we turn to linear algebra over polynomials rings in Sect. 5.

## 4.1 Polynomial Arithmetic

As our general strategy is to use linear algebra over  $\mathbb{Z}$  or  $\mathbb{Q}$  to do linear algebra over  $\mathbb{Z}[X]$  or  $\mathbb{Q}[X]$ , for all arithmetically heavy computations we recurse to  $\mathbb{Z}$  or  $\mathbb{Q}$ . Consequently, for the remaining pieces of explicit computation in  $\mathbb{Z}[X]$  or  $\mathbb{Q}[X]$  we may use a simple straightforward approach:

We use our own standard arithmetic for polynomials over  $\mathbb{Z}$  or  $\mathbb{Q}$ , where a polynomial  $0 \neq f = \sum_{i=0}^d z_i X^i \in \mathbb{Q}[X]$  is just represented by its coefficient list  $[z_0, \dots, z_d] \in \mathbb{Q}^{d+1}$  of length  $d + 1$ , where  $d = \deg(f)$ . Thus we avoid structural overhead as much as possible, and may use directly the facilities to handle row vectors provided by GAP. But we would like to stress that this is just tailored for our aim of doing linear algebra over polynomial rings, and not intended to become a new general-purpose polynomial arithmetic. For example, we are not providing asymptotically fast multiplication, as is for example described in [26, Section 8.3].

In particular, we only rarely need to compute polynomial greatest common divisors. Hence we avoid sophisticated (modular) techniques, as are for example described and compared in [26, Chapter 6], but we are content with a simple variant of the Euclidean algorithm: Assuming that the operands have integral coefficients, by going over to  $\mathbb{Q}$ -multiples if necessary, in order to avoid coefficient explosion we just use denominator-free pseudo-division as described in [1, Algorithm 3.1.2], and Collins's sub-resultant algorithm given in [1, Algorithm 3.3.1], albeit the latter without intermediate primitivisation.

On the other hand, we very often have to evaluate polynomials at various places, where our strategy is to use as few of these specializations as possible, so that evaluation at distinct places is done step by step. Thus we are not in a position to use multi-point evaluation techniques, as are for example described in [26, Section 10.1]. Hence we are just using the Horner scheme, which under these circumstances is well-known to need the optimal number of multiplications.

We now describe the special tasks needed to be solved in our approach:

## 4.2 Recovering Polynomials

The aim is to recover a polynomial with rational coefficients, which we are able to evaluate at arbitrary integral places, from as few such evaluations (at "small" places) as possible. More precisely:

Let  $0 \neq f := \sum_{i=0}^d z_i X^i \in \mathbb{Q}[X]$  be a polynomial of degree  $d = \deg(f) \in \mathbb{N}_0$ , having coefficients  $z_i = \frac{y_i}{x_i} \in \mathbb{Q}$ , where  $x_i \in \mathbb{N}$  and  $y_i \in \mathbb{Z}$  such that  $\gcd(x_i, y_i) = 1$ . Then the task is to find pairwise coprime places  $b_1, \dots, b_k \in \mathbb{Z} \setminus \{0, \pm 1\}$ , for some (small)  $k \in \mathbb{N}$ , such that the degree  $d$  and the coefficients  $z_0, \dots, z_d$  of  $f$  can be computed from the values  $f(b_1), \dots, f(b_k) \in \mathbb{Q}$  alone. Note that, in particular, we do not assume that  $k > d$ , so that polynomial interpolation is not applicable. (Actually,

in our application we often enough have  $k \ll d$ , where for example  $k \sim 5$ , but  $d \lesssim 200$ .)

To this end, let  $b := \prod_{j=1}^k |b_j| \in \mathbb{N}$ , and assume that we have  $\gcd(x_i, b) = 1$  and  $x_i^2 + y_i^2 < b$  for all  $0 \leq i \leq d$ . Hence the congruence classes  $z_i \equiv \frac{y_i}{x_i} \pmod{b_j}$  and  $f(b_j) \pmod{b_j}$  are well-defined, and for the constant coefficient of  $f$  we get

$$z_0 \equiv \sum_{i=0}^d z_i b_j^i \equiv f(b_j) \pmod{b_j}, \quad \text{for } 1 \leq j \leq k.$$

Thus by the Chinese Remainder Theorem, see for example [1, Theorem 1.3.9], there is a unique congruence class  $a \pmod{b}$ , where  $a \in \mathbb{Z}$ , such that  $a \equiv z_0 \pmod{b}$ . To compute  $a \in \mathbb{Z}$ , we let  $a_j \in \mathbb{Z}$  such that

$$f(b_j) \equiv a_j \pmod{b_j}, \quad \text{for } 1 \leq j \leq k.$$

An application of Chinese remainder lifting in  $\mathbb{Z}$  to the congruence classes  $a_1 \pmod{b_1}, \dots, a_k \pmod{b_k}$  yields the congruence class  $a \pmod{b}$ , and by our choice of  $b$  applying rational number recovery as described in Sect. 3.2 reveals  $z_0 \in \mathbb{Q}$ . Now we recurse to  $\tilde{f} := \frac{f - z_0}{X} \in \mathbb{Q}[X]$ , whose value at the place  $b_j$  can of course be determined directly from  $f(b_j)$  as  $\tilde{f}(b_j) = \frac{f(b_j) - z_0}{b_j} \in \mathbb{Q}$ .

**Chinese Remainder Lifting** Hence, apart from rational number recovery, the key computational task to be solved is to perform Chinese remainder lifting in  $\mathbb{Z}$ :

We are using the straightforward approach based on the extended Euclidean algorithm, as is described in [1, Section 1.3.3]. Since we are computing many lifts with respect to the same places  $b_1, \dots, b_k$ , we make use of a precomputation step, as in [1, Algorithm 1.3.11]. But, since again for reasons of time and memory efficiency we are choosing small places  $b_j$ , the specially tailored approach in [1, Algorithm 1.3.11] to keep the intermediate numbers occurring small, at the expense of needing more multiplications, does not pay off as experiments show. Moreover, as we are computing the values  $f(b_j)$  for  $1 \leq j \leq k$  step by step, where even the number  $k$  of places is not determined in advance, we cannot take advantage of fast Chinese remainder lifting techniques, as are described for example in [26, Section 10.3], either.

Our strategy is to rerun the above algorithm with  $k$  increasing, choosing small integral  $2 \leq b_1 < b_2 < \dots < b_k$ , and to discard quickly erroneous guesses by an independent verification, until the correct answer passing the verification is found. By the above discussion, this happens after finitely many iterations. Before that, if  $b = |\prod_{j=1}^k b_j|$  is too small, or not coprime to all the denominators  $x_i$ , the Chinese remainder lifting process does not terminate, or it terminates with a wrong guess. To catch the first case, we impose a degree bound, and stop the lifting process with a failure message if it is exceeded, in order to increment  $k$ . (In our application, 200 turned out to be a suitable degree bound in all cases.)

To catch the second case, we only allow for denominators  $x_i$  dividing an imposed bound. This is justified, since rational number recovery as described in Sect. 3.2 is a trade-off between finding the numerator  $y$  and the denominator  $x$  of the rational number  $\frac{y}{x}$  to be reconstructed: In practice, we typically encounter small denominators  $x$  and large numerators  $y$ , which escape the Gauß reduction algorithm if  $b$  is chosen too small, since then the latter tends to return a larger denominator  $x' > x$  and a smaller numerator  $|y'| < |y|$ . (In our application, denominator bounds such as small 2-powers, or 12, or 20 turned out to be sufficient in all cases.)

### 4.3 Degree Detection

We keep the setting of Sect. 4.2. The technique to be described now has arisen out of an attempt to determine the degree of  $f$  without determining its coefficients. Actually, it deals with the following more general situation (whose relevance for our computations will be explained in Sect. 4.5 below):

Assume that instead of the values  $f(b_1), \dots, f(b_k)$  we are only able to compute “rescaled values”  $a_1 f(b_1), \dots, a_k f(b_k) \in \mathbb{Q}$ , with scalar factors  $a_j \in \mathbb{Q}$  such that  $a_j > 0$ , which are only known to come from a finite pool  $\mathcal{R}$  of positive rational numbers associated with  $f$ . Thus the task now becomes to find  $k \in \mathbb{N}$  and coprime places  $b_1, \dots, b_k \in \mathbb{Z} \setminus \{0, \pm 1\}$  as above, allowing to determine  $f$  up to some positive rational scalar multiple, that is to find  $af \in \mathbb{Q}[X]$ , for some  $a \in \mathbb{Q}$  such that  $a > 0$ ; note that this also determines all the quotients  $\frac{a_j}{a}$ .

To this end, we let  $\alpha_1, \dots, \alpha_d \in \mathbb{C}$  be the complex roots of  $f$ , and set  $\mu := \max\{0, |\alpha_1|, \dots, |\alpha_d|\}$ . Moreover, since  $\mathcal{R}$  is a finite set, we have

$$\delta := \min\{|\ln(a') - \ln(a)| \in \mathbb{R}; a, a' \in \mathcal{R}, a \neq a'\} > 0.$$

Now, let  $k \geq 2$ , and for the places  $b_1, \dots, b_k$  we additionally assume that

$$(1 + 2d) \cdot \mu < b_1 < \dots < b_k \quad \text{and} \quad \ln(b_k) - \ln(b_1) < \delta;$$

hence, in particular, the  $f(b_j)$  are non-zero and have the same sign. The necessity of these choices will become clear below. But this forces us to show that for all  $k \geq 2$  and all  $x > 0$  and  $\delta > 0$  there actually exist pairwise coprime integers  $b_1 < \dots < b_k$  such that  $x < b_1$  and  $\ln\left(\frac{b_k}{b_1}\right) < \delta$ . Indeed, we are going to show that the latter can always be chosen to be primes (where the mere existence proof to follow is impractical, but in practice considering small primes works well, see Example 4.4):

Let  $p_0 < p_1 < \dots$  be the sequence of all primes exceeding  $x$ , and assume to the contrary that for all  $k$ -subsets thereof,  $q_1 < \dots < q_k$  say, we have  $\ln\left(\frac{q_k}{q_1}\right) \geq \delta$ . Then we have  $p_{k-1} \geq e^\delta \cdot p_0$ , and thus  $p_{j(k-1)} \geq e^{j\delta} \cdot p_0$ , for all  $j \in \mathbb{N}$ . Using the prime

number function  $\pi(x) := |\{p \in \mathbb{N}; p \text{ prime}, p \leq x\}|$  this implies

$$\pi(e^{j\delta} \cdot p_0) \leq \pi(p_0) + j(k - 1).$$

From this we get

$$\lim_{j \rightarrow \infty} \frac{\pi(e^{j\delta} \cdot p_0) \cdot \ln(e^{j\delta} \cdot p_0)}{e^{j\delta} \cdot p_0} \leq \lim_{j \rightarrow \infty} \frac{(\pi(p_0) + j(k - 1)) \cdot (j\delta + \ln(p_0))}{e^{j\delta} \cdot p_0} = 0,$$

contradicting the Prime Number Theorem, see [13, Section 1.8, Theorem 6], saying that  $\lim_{x \rightarrow \infty} \frac{\pi(x) \cdot \ln(x)}{x} = 1$ .

**Growth Behavior of Polynomials** We now consider the growth behavior of the polynomial  $f$ . For  $x > \mu$  we have

$$\frac{\partial}{\partial x}(f(x)) = z_d \cdot \frac{\partial}{\partial x} \left( \prod_{r=1}^d (x - \alpha_r) \right) = f(x) \cdot \sum_{r=1}^d \frac{1}{x - \alpha_r},$$

implying

$$\frac{\partial}{\partial x}(\ln(f(x))) = \frac{\partial}{\partial x}(f(x)) \cdot \frac{1}{f(x)} = \sum_{r=1}^d \frac{1}{x - \alpha_r}.$$

Thus, for  $1 \leq i < j \leq k$ , by the mean value theorem for derivatives there is  $b_i < \beta < b_j$  such that

$$\frac{\ln(f(b_j)) - \ln(f(b_i))}{\ln(b_j) - \ln(b_i)} = \sum_{r=1}^d \frac{\beta}{\beta - \alpha_r}.$$

Since by assumption  $b_i > (1 + 2d) \cdot \mu \geq (1 + 2d) \cdot |\alpha_r|$ , we have

$$\left| \frac{\beta}{\beta - \alpha_r} - 1 \right| = \left| \frac{\alpha_r}{\beta - \alpha_r} \right| \leq \frac{|\alpha_r|}{\beta - |\alpha_r|} < \frac{|\alpha_r|}{(1 + 2d) \cdot |\alpha_r| - |\alpha_r|} \leq \frac{1}{2d}$$

for all  $1 \leq r \leq d$ . All differences  $\beta - \alpha_r \in \mathbb{C}$  having positive real parts, we get

$$d < \frac{\ln(f(b_j)) - \ln(f(b_i))}{\ln(b_j) - \ln(b_i)} < d + \frac{1}{2}.$$

Moreover, by assumption we have  $0 < \ln(b_j) - \ln(b_i) < \delta \leq |\ln(a_j) - \ln(a_i)|$ , hence

$$\left| \frac{\ln(a_j) - \ln(a_i)}{\ln(b_j) - \ln(b_i)} \right| > 1.$$



Now, letting  $\lceil x \rceil := \lfloor x + \frac{1}{2} \rfloor \in \mathbb{Z}$  denote the integer nearest to  $x \in \mathbb{R}$ , we set

$$d_{ij} := \left\lceil \frac{\ln(a_j f(b_j)) - \ln(a_i f(b_i))}{\ln(b_j) - \ln(b_i)} \right\rceil = \left\lfloor \frac{\ln(f(b_j)) - \ln(f(b_i))}{\ln(b_j) - \ln(b_i)} + \frac{\ln(a_j) - \ln(a_i)}{\ln(b_j) - \ln(b_i)} \right\rfloor$$

for all  $1 \leq i, j \leq k$  such that  $i \neq j$ ; note that  $d_{ij} = d_{ji}$ . Hence from the above estimates we infer that  $d_{ij} = d$  if and only if  $a_i = a_j$ . In particular, all these numbers  $d_{ij}$  coincide if and only if  $a_1 = \dots = a_k$ , hence in this case immediately determining  $d$ .

**Combinatorial Translation** Thus our task can now be rephrased in combinatorial terms as follows: For  $c \in \mathbb{Z}$  let  $\Gamma_{d+c}$  be the undirected graph on the vertex set  $\{1, \dots, k\}$ , whose edges are the 2-subsets  $\{i, j\} \subseteq \{1, \dots, k\}$  such that  $d_{ij} = d + c$ .

Then by the above discussion the connected components of  $\Gamma_d$  are complete graphs, whose vertex sets coincide with the sets of  $j \in \{1, \dots, k\}$  such that the associated scalars  $a_j$  assume one and the same value. On the other hand, if  $\Gamma_{d+c}$ , for some  $c \neq 0$ , has a complete connected component with  $r \geq 2$  vertices  $b_{j_1} < \dots < b_{j_r}$ , then for all  $i, j \in \{j_1, \dots, j_r\}$  such that  $i < j$  we have

$$c - 1 < \left| \frac{\ln(a_j) - \ln(a_i)}{\ln(b_j) - \ln(b_i)} \right| < c + \frac{1}{2}.$$

Thus we infer that the sequence  $a_{j_1}, \dots, a_{j_r}$  is strictly increasing if  $c > 0$ , and strictly decreasing if  $c < 0$ . In particular this implies that  $r \leq |\mathcal{R}|$ . In other words, as soon as we find a complete connected component of a graph  $\Gamma_{d+c}$  having more than  $|\mathcal{R}|$  elements, then we may conclude that  $c = 0$ , and we have determined  $d$ . Moreover, if  $k > |\mathcal{R}|^2$  than this case actually happens.

Our algorithm to determine the degree  $d$  of  $f$ , and  $af$  for some  $a > 0$ , is now straightforward: Again our strategy is to increase  $k$  step by step, and to choose places  $2 \leq b_1 < b_2 < \dots < b_k$  such that  $b_1$  is growing and  $\ln(b_k) - \ln(b_1)$  tends to zero. Having made a choice, we compute the numbers  $d_{ij} \in \mathbb{Z}$  for all  $1 \leq i < j \leq k$ ; note that here we do not see a way to avoid using non-exact floating point arithmetic (to evaluate logarithms), while everywhere else we are computing exactly. For all numbers  $d' \in \mathbb{Z}$  thus occurring we then determine the graph  $\Gamma_{d'}$ . Amongst all the graphs found we choose one, again  $\Gamma_{d'}$  say, having a complete connected component of maximal cardinality, with vertex set  $\mathcal{J} \subseteq \{1, \dots, k\}$  say. Then we run polynomial recovery, see Sect. 4.2, using the places  $\{b_j; j \in \mathcal{J}\}$  and the values  $\{a_j f(b_j); j \in \mathcal{J}\}$ , with degree bound  $d'$ .

### 4.4 An Example

Here is an example to illustrate the above process. (It is a modified version of an example which actually occurred in our application.) Assume as places  $b_j$ , for  $1 \leq$

**Table 3** An example for degree detection

j	$b_j$	$a_j f(b_j)$	$a_j$
1	29	471132000262895400	$\frac{1}{25}$
2	31	5556161802048405504	$\frac{1}{5}$
3	37	271378870503231142344	1
4	41	203982274364082601464	$\frac{1}{5}$
5	43	1885780898401789278912	1
6	47	5946135224244400779264	1
7	53	28077873950889396256392	1
8	59	4493456499569142283200	$\frac{1}{25}$
9	61	34577756822169042208584	$\frac{1}{5}$
10	67	581970465933078043504704	1
11	71	246522309921169431519744	$\frac{1}{5}$
12	73	1766015503219395154436952	1
13	79	196427398952317706342400	$\frac{1}{25}$

$j \leq k = 13$ , we have chosen the rational primes between 29 and 79, and evaluating the unknown polynomial  $f$  has resulted in the list of values  $a_j f(b_j)$  given in Table 3; the scalars  $a_j$  are of course not known either.

Then it turns out that the numbers  $d' \in \mathbb{Z}$ , where  $1 \leq i < j \leq 13$ , come from an 34-element subset of  $\{-27, \dots, 71\}$ . For seven of them the associated graph  $\Gamma_{d'}$  has a connected component with at least three vertices, but only for two of them we find a complete connected component amongst them: The graph  $\Gamma_7$  has a complete connected component consisting of the vertices  $\mathcal{B}_0 := \{47, 61, 79\}$ , while the graph  $\Gamma_{13}$  consists of three connected components, which all are complete, having the vertices

$$\mathcal{B}_1 := \{37, 43, 47, 53, 67, 73\}, \quad \mathcal{B}_2 := \{31, 41, 61, 71\}, \quad \mathcal{B}_3 := \{29, 59, 79\}.$$

Running polynomial recovery, see Sect. 4.2, using the places  $\mathcal{B}_0$  fails by exceeding the degree bound. But running it using  $\mathcal{B}_1$  yields  $af = \sum_{i=0}^{13} z_i X^i \in \mathbb{Z}[X]$ , where

$$[z_0, \dots, z_{13}] = [1, 4, 8, 11, 12, 12, 12, 12, 12, 12, 11, 8, 4, 1],$$

while running it using  $\mathcal{B}_2$  and  $\mathcal{B}_3$  yields  $\frac{1}{5} \cdot af \in \mathbb{Q}[X]$  and  $\frac{1}{25} \cdot af \in \mathbb{Q}[X]$ , respectively. Thus we indeed have  $d = \deg(f) = 13$ , and assuming that  $a = 1$  we have determined the scalars  $a_j$ , for  $1 \leq j \leq 13$ , as well. Note that the bounds assumed in Sect. 4.2 are fulfilled; and the roots of  $f$  turning out to be complex roots of unity, implying  $\mu = 1$ , the bounds assumed in Sect. 4.3 are fulfilled as well.

It should be noted that for the preceding discussion we have chosen  $k$  large enough to exhibit the occurrence of the erroneous set  $\mathcal{B}_0$ , for which we indeed observe that the associated scalars  $a_j$  are pairwise distinct. But this also reveals another practical observation, at least for polynomials occurring in the applications

in Sect. 5: The scalars  $a_j$ , here coming from the three-element set  $\mathcal{R} = \{1, \frac{1}{5}, \frac{1}{25}\}$ , typically are not uniformly distributed throughout  $\mathcal{R}$ , but the scalar  $a_j = 1$  occurs much more frequently than the other ones.

As was already mentioned, in practice we instead increase  $k$  step by step. Then for the smallest  $k \geq 3$  such that the graph  $\Gamma_{13}$  has a complete connected component with at least three vertices, that is for  $k = 6$ , we find the set  $\mathcal{B} := \{37, 43, 47\}$  of places, indeed being associated to the case  $a_j = 1$ . Now polynomial recovery using  $\mathcal{B}$  readily returns  $f$ ; note that the bounds assumed in Sect. 4.2 are still fulfilled.

## 4.5 Catching Projectivities

We now have to explain where the conditions imposed in Sect. 4.3 come from: Typically, for example for the tasks described in Sects. 5.1 and 5.2, our aim is to determine a matrix over  $\mathbb{Z}[X]$  or  $\mathbb{Q}[X]$  by computing various specializations first, that is evaluating at certain places  $b_1, \dots, b_k$ , performing some linear algebra over  $\mathbb{Z}$  or  $\mathbb{Q}$ , as described in Sect. 3, for each of the specializations, and then lifting back to polynomials as explained in Sect. 4.2. But the linear algebra step in between might only be unique up to a scalar in  $\mathbb{Q}$ , which additionally depends on the particular specialization considered. On the other hand, the matrix we are looking for might also only be unique up to a scalar in  $\mathbb{Q}(X)$ .

Let us now, again, agree on the following convention: Given  $f, g \in \mathbb{Z}[X]$ , not both zero, let  $\gcd(f, g) \in \mathbb{Z}[X]$  denote the polynomial greatest common divisor of  $f$  and  $g$  with positive leading coefficient. A vector  $0 \neq v \in \mathbb{Q}[X]^m$ , where  $m \in \mathbb{N}$ , is called *primitive*, if actually  $v \in \mathbb{Z}[X]^m$ , and for the greatest common divisor  $\gcd(v)$  of its entries we have  $\gcd(v) = 1$ . Clearly greatest common divisor computations in  $\mathbb{Z}$  and in  $\mathbb{Z}[X]$  yield a  $\mathbb{Q}(X)$ -multiple of  $v$  which is primitive. Similarly, a matrix  $A \in \mathbb{Q}[X]^{m \times n}$ , where  $m, n \in \mathbb{N}$ , is called *primitive*, if actually  $A \in \mathbb{Z}[X]^{m \times n}$ , and for the greatest common divisor  $\gcd(A)$  of its entries we have  $\gcd(A) = 1$ .

**Specializing Primitive Vectors** Hence, in the above context the task is to recover a primitive vector  $[f_1, \dots, f_m] \in \mathbb{Z}[X]^m$  not from specializations  $[f_1(b_j), \dots, f_m(b_j)] \in \mathbb{Z}^m$ , for  $1 \leq j \leq k$ , but from “rescaled” versions  $[a_j f_1(b_j), \dots, a_j f_m(b_j)] \in \mathbb{Q}^m$  instead. This places us in the setting of Sect. 4.3, but it remains to justify the assumption that the scalars  $a_j \in \mathbb{Q}$  involved indeed come from a finite pool:

**Proposition 4.1** *Let  $f_1, \dots, f_m \in \mathbb{Z}[X]$ , where  $m \in \mathbb{N}$ , such that  $\gcd(f_1, \dots, f_m) = 1 \in \mathbb{Z}[X]$ . Then there is a finite set  $\mathcal{P} \subseteq \mathbb{N}$  such that for all  $b \in \mathbb{Z}$  we have*

$$\gcd(f_1(b), \dots, f_m(b)) \in \mathcal{P}.$$

*Proof* Note first that by assumption  $f_1, \dots, f_m$  do not have any common zeroes, so that  $\gcd(f_1(b), \dots, f_m(b)) \in \mathbb{N}$  is well-defined for any  $b \in \mathbb{Z}$ . We proceed by induction on  $m \in \mathbb{N}$ . For  $m = 1$  we have  $f_1 = \pm 1$ , and we may let  $\mathcal{P} := \{\pm 1\}$ .

Hence let  $m \geq 2$ , where we may assume that all the  $f_i$ , for  $1 \leq i \leq m$ , are non-constant. Letting  $g := \gcd(f_1, \dots, f_{m-1}) \in \mathbb{Z}[X]$  we have  $\gcd(g, f_m) = 1$ . Letting  $g_i := f_i/g \in \mathbb{Z}[X]$  for  $1 \leq i \leq m-1$ , we have  $\gcd(g_1, \dots, g_{m-1}) = 1$ , thus by induction let  $\mathcal{Q} \subseteq \mathbb{N}$  be a set as asserted associated with  $g_1, \dots, g_{m-1}$ . Now, given  $b \in \mathbb{Z}$ , we may write

$$x := \gcd(f_1(b), \dots, f_m(b)) = \gcd(g(b)g_1(b), \dots, g(b)g_{m-1}(b), f_m(b))$$

as  $x = yz$ , where  $y = \gcd(g(b), f_m(b))$ , and  $z$  divides  $\gcd(g_1(b), \dots, g_{m-1}(b), f_m(b))$ . Hence  $z$  divides  $\gcd(g_1(b), \dots, g_{m-1}(b))$ , and thus divides an element of  $\mathcal{Q}$ . Moreover, from  $\gcd(g, f_m) = 1$  we infer that the resultant  $\rho := \text{res}(g, f_m) \in \mathbb{Z}$  is different from zero, see [26, Corollary 6.20], which by von zur Gathen and Gerhard [26, Corollary 6.21] implies that  $y = \gcd(g(b), f_m(b))$  divides  $\rho$ . Thus the set  $\mathcal{P}$  of all positive divisors of the elements of  $\rho\mathcal{Q} := \{\rho r \in \mathbb{N}; r \in \mathcal{Q}\}$  is as desired.  $\square$

## 5 Linear Algebra Over Polynomial Rings

As was already mentioned, our general strategy to determine matrices over  $\mathbb{Z}[X]$  or  $\mathbb{Q}[X]$  is to specialize first at integral places, to apply linear algebra techniques as described in Sect. 3 to the matrices over  $\mathbb{Z}$  or  $\mathbb{Q}$  thus obtained, and subsequently to recover the polynomial entries in question by the Chinese remainder lifting technique described in Sect. 4.2, applying degree detection as described in Sect. 4.3 if necessary. In this section we describe how we can do linear algebra over  $\mathbb{Z}[X]$  or  $\mathbb{Q}[X]$  using this approach.

Since we are faced with both sparse and dense matrices, we keep two corresponding formats for matrices over polynomial rings. (In our application, representing matrices for  $W$ -graph representations, see Definition 2.4, are extremely sparse, while Gram matrices for them, see Remark 2.2, typically are dense; see also Example 9.2.) We have conversion and multiplication routines between them, but whenever it comes to linear algebra computations we always use the dense matrix format. From the arithmetical side, we are only using standard matrix multiplication, but no asymptotically faster methods, as are for example indicated in [26, Section 12.1].

### 5.1 Nullspace

We have developed a solution to the following restricted nullspace problem only (which is sufficient for our application):

Given a matrix  $A \in \mathbb{Q}[X]^{m \times n}$ , where  $m, n \in \mathbb{N}$ , such that  $\text{rk}_{\mathbb{Q}[X]}(\ker(A)) = 1$ , the task is to determine a primitive vector  $v \in \mathbb{Z}[X]^m$  such that  $\ker(A) = \langle v \rangle_{\mathbb{Q}[X]}$ ; then the vector  $v$  is unique up to sign.

To do so, by going over to a suitable  $\mathbb{Q}(X)$ -multiple we may assume that  $A \in \mathbb{Z}[X]^{m \times n}$  is primitive. Then we specialize the matrix  $A$  successively at integral places  $b_1, \dots, b_k$ , yielding matrices  $A(b_j) \in \mathbb{Z}^{m \times n}$ . Since the rank condition on  $A$  is equivalent to saying that  $\det(A') = 0$  for all  $(m \times m)$ -submatrices  $A'$  of  $A$ , while there is an  $((m - 1) \times (m - 1))$ -submatrix  $A''$  of  $A$  such that  $\det(A'') \neq 0$ , we have  $\text{rk}_{\mathbb{Z}}(\ker(A(b))) \geq 1$  for any  $b \in \mathbb{Z}$ , and for all but finitely many such  $b$  we indeed have  $\text{rk}_{\mathbb{Z}}(\ker(A(b))) = 1$ . Thus we may assume that all the chosen specializations  $A(b_j)$  also fulfill  $\text{rk}_{\mathbb{Z}}(\ker(A(b_j))) = 1$ . Note that this provides an implicit check whether the rank condition on  $A$  indeed holds.

Hence we are in a position to compute the row kernels  $\ker(A(b_j)) = \langle v_j \rangle_{\mathbb{Z}} \leq \mathbb{Z}^m$  as described in Sect. 3.4, where the  $v_j \in \mathbb{Z}^m$  are primitive, for all  $1 \leq j \leq k$ . Thus the latter are of the form  $v_j = \frac{1}{a_j} \cdot v(b_j)$ , where  $a_j = \text{gcd}(v(b_j)) \in \mathbb{N}$ , and  $v \in \mathbb{Z}[X]^m$  is the desired primitive solution vector from above. By Proposition 4.1 we conclude that the scalars  $a_j$  involved indeed come from a finite pool only depending on  $v$ .

Now applying degree detection, see Sect. 4.3, and polynomial recovery, see Sect. 4.2, yields candidate vectors  $0 \neq \tilde{v} \in \mathbb{Q}[X]^m$ , which by going over to a suitable  $\mathbb{Q}$ -multiple can be assumed to be primitive. Then the correctness of  $\tilde{v}$  can be independently verified by explicitly computing  $\tilde{v}A$  and checking whether this is zero.

## 5.2 Inverse

Given a matrix  $A \in \mathbb{Q}[X]^{n \times n}$ , where  $n \in \mathbb{N}$ , such that  $\det(A) \neq 0$ , the task is to find  $B \in \mathbb{Z}[X]^{n \times n}$  and  $c \in \mathbb{Z}[X]$ , such that  $A^{-1} = \frac{1}{c} \cdot B \in \mathbb{Q}(X)^{n \times n}$  and the overall greatest common divisor  $\text{gcd}(B, c) \in \mathbb{Z}[X]$  of the entries of  $B$  and  $c$  equals  $\text{gcd}(B, c) = 1$ ; then the pair  $(B, c)$  is unique up to sign.

To do so, by going over to a suitable  $\mathbb{Q}$ -multiple we may assume that  $A \in \mathbb{Z}[X]^{n \times n}$ . Thus the equation  $BA = c \cdot E_n$  implies that  $\text{gcd}(B)$  divides  $c$ , and hence  $B$  is primitive. Then we specialize the matrix  $A$  successively at integral places  $b_1, \dots, b_k$ , yielding matrices  $A(b_j) \in \mathbb{Z}^{n \times n}$ . Since for all but finitely many  $b \in \mathbb{Z}$  we have  $\det(A(b)) \neq 0$ , we may assume that all the chosen specializations  $A(b_j)$  indeed also fulfill  $\det(A(b_j)) \neq 0$ . Note that this provides an implicit check whether the invertibility condition on  $A$  indeed holds.

Hence we are in a position to compute the inverses  $A(b_j)^{-1} \in \mathbb{Q}^{n \times n}$  as described in Sect. 3.5, yielding  $B_j \in \mathbb{Z}^{n \times n}$  and  $c_j \in \mathbb{Z}$ , such that  $B_j$  is primitive and  $A(b_j)^{-1} = \frac{1}{c_j} \cdot B_j$ , for all  $1 \leq j \leq k$ . Thus, if  $B \in \mathbb{Z}[X]^{n \times n}$  and  $c \in \mathbb{Z}[X]$  are the desired solutions from above, we infer

$$B_j = \frac{1}{a_j} \cdot B(b_j) \quad \text{and} \quad c_j = \frac{1}{a_j} \cdot c(b_j), \quad \text{where} \quad a_j := \text{gcd}(B(b_j), c(b_j)) \in \mathbb{N}.$$

By Proposition 4.1 we conclude that the scalars  $a_j$  involved indeed come from a finite pool only depending on  $B$  and  $c$ .

Now applying degree detection, see Sect. 4.3, and polynomial recovery, see Sect. 4.2, yields candidate solutions  $\widetilde{B} \in \mathbb{Q}[X]^{n \times n}$  and  $\widetilde{c} \in \mathbb{Q}[X]^n$ , for which by going over to a suitable  $\mathbb{Q}$ -multiple we may assume that  $\widetilde{c} \in \mathbb{Z}[X]^n$  and  $\widetilde{B} \in \mathbb{Z}[X]^{n \times n}$  is primitive. Then the correctness of  $(\widetilde{B}, \widetilde{c})$  can be independently verified by explicitly computing  $A\widetilde{B}$  and checking whether it equals  $\widetilde{c} \cdot E_n$ .

### 5.3 The Exponent of a Matrix

In view of the discussion in Sect. 3.6, and noting that  $\mathbb{Q}[X]$  is a principal ideal domain as well, we pursue the analogy between matrix inverses over  $\mathbb{Z}$  and over  $\mathbb{Q}[X]$  still a little further. Indeed, given a square matrix  $A \in \mathbb{Z}[X]^{n \times n}$  such that  $\det(A) \neq 0$  as above, the polynomial  $c \in \mathbb{Z}[X]$  in the expression  $A^{-1} = \frac{1}{c} \cdot B$ , where  $B \in \mathbb{Z}[X]^{n \times n}$  is chosen primitive, again has another interpretation:

Let the *exponent*  $\exp(A) \in \mathbb{Z}[X]$  of  $A$  be a primitive generator of the annihilator of the  $\mathbb{Q}[X]$ -module  $\mathbb{Q}[X]^n/\text{im}(A)$ , where  $\text{im}(A) \leq \mathbb{Q}[X]^n$  is the  $\mathbb{Q}[X]$ -span of the rows of  $A$ ; then  $\exp(A)$  is unique up to sign. Then, similar to Sect. 3.6, we conclude that  $\exp(A)$  and  $c$  are associated in  $\mathbb{Q}[X]$ , and thus the primitivity of  $\exp(A)$  yields

$$c = \text{gcd}(c) \cdot \exp(A) \in \mathbb{Z}[X].$$

In other words, computing the inverse of  $A$  as described in Sect. 5.2 also yields a method to compute the exponent of  $A$  as  $\exp(A) = \frac{1}{\text{gcd}(c)} \cdot c$ . Moreover,  $c$  governs modular invertibility of  $A$  as follows:

**Proposition 5.1** *We keep the notation of Sect. 5.3. Let  $\{0\} \neq \mathfrak{p} \triangleleft \mathbb{Z}[X]$  be a prime ideal, let  $\mathcal{Q}_{\mathfrak{p}} := \text{Quot}(\mathbb{Z}[X]/\mathfrak{p})$  be the field of fractions of the integral domain  $\mathbb{Z}[X]/\mathfrak{p}$ , and let  $A_{\mathfrak{p}} \in (\mathbb{Z}[X]/\mathfrak{p})^{n \times n}$  be the matrix obtained from  $A$  by reduction modulo  $\mathfrak{p}$ . Then  $A_{\mathfrak{p}}$  is invertible in  $\mathcal{Q}_{\mathfrak{p}}^{n \times n}$  if and only if  $c \notin \mathfrak{p}$ .*

*Proof* The prime ideals of  $\mathbb{Z}[X]$  being well-understood, we are in precisely one of the following cases: (i) We have  $\mathfrak{p} = (p)$ , where  $p \in \mathbb{Z}$  is a prime; then we have  $\mathcal{Q}_{\mathfrak{p}} \cong \text{Quot}(\mathbb{F}_p[X]) = \mathbb{F}_p(X)$ , a rational function field; (ii) we have  $\mathfrak{p} = (f)$ , where  $f \in \mathbb{Z}[X]$  is non-constant and irreducible, hence in particular is primitive; then we have  $\mathcal{Q}_{\mathfrak{p}} \cong \mathbb{Q}[X]/(f)$ , an algebraic number field; (iii) we have  $\mathfrak{p} = (p, f)$ , where  $p$  and  $f$  are as above; then we have  $\mathcal{Q}_{\mathfrak{p}} = \mathbb{Z}[X]/\mathfrak{p} \cong \mathbb{F}_p[X]/(\overline{f})$ , a finite field.

Now  $A_{\mathfrak{p}}$  is non-invertible in  $\mathcal{Q}_{\mathfrak{p}}^{n \times n}$  if and only if  $\det(A) \in \mathfrak{p}$ , which holds if and only if there is an irreducible divisor of  $\det(A)$  being contained in  $\mathfrak{p}$ . Thus it suffices to determine (i) the primes  $p \in \mathbb{Z}$ , and (ii) the non-constant irreducible polynomials  $f \in \mathbb{Z}[X]$  dividing  $\det(A)$  in  $\mathbb{Z}[X]$ .

- (i) From  $A^{-1} = \frac{1}{\det(A)} \cdot \text{adj}(A) \in \mathbb{Q}(X)^{n \times n}$ , where  $\text{adj}(A) \in \mathbb{Z}[X]^{n \times n}$  is the adjoint matrix of  $A$ , we infer that  $c$  divides  $\det(A)$  in  $\mathbb{Z}[X]$ . Hence any prime  $p \in \mathbb{Z}$  dividing  $\text{gcd}(c)$  also divides  $\det(A)$  in  $\mathbb{Z}[X]$ . Conversely, if  $p$  does not divide  $\text{gcd}(c)$ , then  $p$ -modular reduction yields  $\overline{A\widetilde{B}} = \overline{cE_n} \neq 0 \in \mathbb{F}_p[X]^{n \times n}$ , hence

$\det(\bar{A}) \neq 0 \in \mathbb{F}_p[X]$ . Hence the primes  $p \in \mathbb{Z}$  we are looking for are precisely the prime divisors of  $\gcd(c)$ .

- (ii) This is equivalent to finding the irreducible polynomials in  $\mathbb{Q}[X]$  dividing  $\det(A)$  in  $\mathbb{Q}[X]$ . Again similar to Sect. 3.6 we conclude that the latter are precisely the irreducible polynomials dividing  $\exp(A)$ . Hence the polynomials  $f \in \mathbb{Z}[X]$  we are looking for are precisely the non-constant irreducible divisors of  $\frac{1}{\gcd(c)} \cdot c$ .  $\square$

## 5.4 Product

Given matrices  $A \in \mathbb{Q}[X]^{l \times m}$  and  $B \in \mathbb{Q}[X]^{m \times n}$ , where  $l, m, n \in \mathbb{N}$ , the task is to compute their product  $AB \in \mathbb{Q}[X]^{l \times n}$ .

This is straightforwardly done: Again, by going over to suitable  $\mathbb{Q}$ -multiples we may assume that  $A \in \mathbb{Z}[X]^{l \times m}$  and  $B \in \mathbb{Z}[X]^{m \times n}$ . Then we specialize the matrices  $A$  and  $B$  successively at integral places  $b_1, \dots, b_k$ , yielding matrices  $A(b_j) \in \mathbb{Z}^{l \times m}$  and  $B(b_j) \in \mathbb{Z}^{m \times n}$ , whose products  $A(b_j)B(b_j) \in \mathbb{Z}^{l \times n}$  we compute. Now applying polynomial recovery, see Sect. 4.2, yields candidate solutions  $\tilde{C} \in \mathbb{Q}[X]^{l \times n}$ . (Note that since no “rescaling” takes place here it is not necessary to apply degree detection.)

As for correctness, there are a few necessary conditions which can be used as break conditions in polynomial recovery: All entries of  $\tilde{C}$  must be polynomials with integer coefficients, and the degrees of the entries of the input matrices yield bounds on the degrees of those of  $\tilde{C}$ . But these conditions are far from being sufficient, so that, in contrast to the tasks in Sects. 5.1 and 5.2, here we do not have a general way of independently verifying correctness. (In our application, as a very efficient break condition we have used the fact that the entries of  $\tilde{C}$  have to be of a particular form, see Sect. 8.3.)

## 5.5 An Alternative Approach

The idea of our approach is, essentially, to reduce computations over  $\mathbb{Q}[X]$  to computations over  $\mathbb{Z}$ , where lifting back to polynomials is done in one step by combining specialization and Chinese remainder lifting. In consequence, we almost entirely use arithmetic in characteristic zero (except the use of a large prime field in the  $p$ -adic decomposition algorithm in Sect. 3.3). But it seems to be worth-while to say a few more words on the following “two-step” approach, which was already mentioned briefly in Sects. 1 and 2.5:

Assume our aim is to determine a matrix  $0 \neq A \in \mathbb{Q}[X]^{m \times n}$ , where  $m, n \in \mathbb{N}$ . To this end, we choose pairwise distinct places  $b_1, \dots, b_k \in \mathbb{Z}$ , for some  $k \in \mathbb{N}$  such that  $k > d$ , where  $d \in \mathbb{N}_0$  is the maximum of the degrees of the non-zero entries of

A. Thus, if we are able to compute the specializations  $A(b_j) \in \mathbb{Q}^{n \times n}$ , for  $1 \leq j \leq k$ , we may recover the entries of  $A$  by polynomial interpolation, as for example is described in [26, Section 10.2]. In turn, to find the specializations  $A(b_j)$  we choose pairwise distinct primes  $p_1, \dots, p_l \in \mathbb{N}$ , for some  $l \in \mathbb{N}$ , such that the denominators of all the entries of  $A(b_j)$  are coprime to  $p_i$ , for all  $1 \leq j \leq k$  and  $1 \leq i \leq l$ . Then reduction modulo the chosen primes yields matrices  $A_{p_i}(b_j) \in \mathbb{F}_{p_i}^{m \times n}$ . Hence, if  $\prod_{i=1}^l p_i$  is large enough, and we are able to compute the modular reductions  $A_{p_i}(b_j)$ , for  $1 \leq i \leq l$ , then rational number recovery, see Sect. 3.2, reveals the entries of  $A(b_j)$ . Hence this reduces finding the matrix  $A$  to finding the matrices  $A_{p_i}(b_j)$  over prime fields, for which we in turn may use techniques of the **MeatAxe**.

Thus here specialization and Chinese remainder lifting are done in two separate steps, aiming at taking advantage of the efficiency of computations in prime characteristic. But the “two-step” approach has severe disadvantages: The number  $k$  of places to specialize at is at least as large as the degree of the polynomials in question, hence many more and larger  $b_j$  than in our approach are needed, increasing time and memory requirements, presumably drastically. (In our application this means  $k \lesssim 200$ .) Moreover, in order to use rational number recovery, the number  $l$  of primes used for modular reduction must not be too small, at the expense of possibly losing the very fast arithmetic over small finite fields, which otherwise is a major advantage of the **MeatAxe**.

Actually, apart from our own experiences, this kind of approach is pursued in [19], and the figures on timings and memory consumption given there seem to support the above comments. But it should be stressed that the emphasis of [19] is on parallelizing this kind of computations, which we here do not consider at all.

## 6 Computing with Representations

As was already mentioned in Sect. 1, in our application we will make use of a suitable variant of the “standard basis algorithm”, which was originally used in [23] for computations over finite fields. In this section we present the necessary ideas from computational representation theory, which can be formulated in terms of the following general setting:

### 6.1 Standard Bases

Let  $\mathcal{A}$  be a  $K$ -algebra, where  $K$  is a field, being generated by the (ordered) set  $A_1, \dots, A_r$ , where  $r \in \mathbb{N}_0$ . Moreover, let  $\mathfrak{X}: \mathcal{A} \rightarrow K^{n \times n}$  be an absolutely irreducible matrix representation of  $\mathcal{A}$ , where  $n \in \mathbb{N}$ . Then the task is to find a “canonical”  $K$ -basis of the row space  $K^n$  with respect to the representation  $\mathfrak{X}$ , where we consider right actions, as is common in the computational world.



To this end, let  $A_0 \in \mathcal{A}$  such that  $\dim_K(\ker(\mathfrak{X}(A_0))) = 1$ ; note that whenever  $\mathfrak{X}$  is irreducible such an element  $A_0$  exists if and only if  $\mathfrak{X}$  is absolutely irreducible. This leads to the following breadth-first search algorithm; see also [23]: Choose a seed vector  $0 \neq u \in \ker(\mathfrak{X}(A_0))$ , let  $\mathfrak{B} := [u]$  and  $\mathfrak{T} := [[0, 0]]$ , and set  $i := 1$ . As long as  $i$  does not exceed the cardinality of  $\mathfrak{B}$ , let  $v$  be the  $i$ -th element of  $\mathfrak{B}$ . Then for  $1 \leq j \leq r$  let successively  $w := v \cdot \mathfrak{X}(A_j)$ , and check whether or not  $w \in \langle \mathfrak{B} \rangle_K$ . If so, then discard  $w$ ; if not, then append  $w$  to  $\mathfrak{B}$ , and append  $[i, j]$  to  $\mathfrak{T}$ . Having done this for all  $j$ , increment  $i$  and recurse.

Since the growing set  $\mathfrak{B}$  is  $K$ -linearly independent throughout, this algorithm terminates after at most  $n$  loops. After termination,  $\langle \mathfrak{B} \rangle_K$  is a non-zero submodule of the irreducible  $\mathcal{A}$ -module  $K^n$ , and thus  $\mathfrak{B}$  indeed is a  $K$ -basis. (Of course, we may terminate early, without any further checking, as soon as the cardinality of  $\mathfrak{B}$  equals  $n$ , since from this point on  $\mathfrak{B}$  would not change anymore anyway.) The (ordered) set  $\mathfrak{B}$  is called a *standard basis* of  $K^n$  with respect to the representation  $\mathfrak{X}$ , the generators  $A_1, \dots, A_r$ , and the distinguished element  $A_0$ , and the “bookkeeping list”  $\mathfrak{T}$  is called the associated *Schreier tree*.

Strictly speaking,  $\mathfrak{B}$  also depends on the chosen seed vector, but it is essentially unique in the following sense: If  $0 \neq \tilde{u} \in \ker(\mathfrak{X}(A_0))$  gives rise to the standard basis  $\tilde{\mathfrak{B}}$  with Schreier tree  $\tilde{\mathfrak{T}}$ , then we have  $\tilde{u} = c \cdot u$ , for some  $0 \neq c \in K$ , and thus  $\tilde{\mathfrak{B}} = c \cdot \mathfrak{B}$  and  $\tilde{\mathfrak{T}} = \mathfrak{T}$ . Moreover, using the Schreier tree  $\mathfrak{T} = [[i_1, j_1], \dots, [i_n, j_n]]$ , we may recover  $\mathfrak{B} = [u_1, \dots, u_n]$ , up to a scalar, without any searching as follows: Choose  $0 \neq u_1 \in \ker(\mathfrak{X}(A_0))$ , and for  $2 \leq k \leq n$  let successively  $u_k := u_{i_k} \cdot \mathfrak{X}(A_{j_k})$ .

**In Practice** We are able to run the above standard basis algorithm in the following particular cases: If  $K$  is a (small) finite field, then this can of course be done using ideas from the **MeatAxe**, as is already described in [23].

More important from our point of view is the case  $K = \mathbb{Q}$ . Then we may assume that  $u \in \mathbb{Z}^n$ , and if additionally  $\mathfrak{X}(A_i) \in \mathbb{Z}^{n \times n}$ , for all  $1 \leq i \leq r$ , then we have  $\mathfrak{B} \subseteq \mathbb{Z}^n$ , hence the key step in the above algorithm, to decide whether or not  $w \in \langle \mathfrak{B} \rangle_{\mathbb{Q}}$ , can be done using the  $p$ -adic decomposition algorithm in Sect. 3.3, where whenever  $\mathfrak{B}$  is enlarged we also check whether its  $p$ -modular reduction  $\overline{\mathfrak{B}} \subseteq \mathbb{F}_p^n$  is  $\mathbb{F}_p$ -linearly independent; if not, then we return failure in order to choose another prime  $p$ . (Note that this is reminiscent of the strategy in Sect. 3.4.)

## 6.2 Computing Homomorphisms

We return to the general setting in Sect. 6.1, and let  $\mathfrak{X}': \mathcal{A} \rightarrow K^{n \times n}$  be a matrix representation of  $\mathcal{A}$ , which is equivalent to  $\mathfrak{X}$ . Then a standard basis  $\mathfrak{B}' = [v'_1, \dots, v'_n]$  of  $K^n$  with respect to the representation  $\mathfrak{X}'$  is found by choosing  $0 \neq v'_1 \in \ker(\mathfrak{X}'(A_0))$  and just applying the Schreier tree  $\mathfrak{T} = [[i_1, j_1], \dots, [i_n, j_n]]$  already known from the standard basis computation for  $\mathfrak{X}$  by letting successively  $v'_k := v'_{i_k} \cdot \mathfrak{X}'(A_{j_k})$ , for  $2 \leq k \leq n$ ; note that by assumption we indeed have  $\dim_K(\ker(\mathfrak{X}'(A_0))) = 1$ .

Now let  $0 \neq C \in K^{n \times n}$  be an  $\mathcal{A}$ -homomorphism from  $\mathfrak{X}$  to  $\mathfrak{X}'$ , that is we have

$$\mathfrak{X}(A) \cdot C = C \cdot \mathfrak{X}'(A) \quad \text{for all } A \in \mathcal{A};$$

of course, it suffices to require this condition for the generators  $A_1, \dots, A_r$  only. Since  $\mathfrak{X}$  is absolutely irreducible, it follows that  $C \in \text{GL}_n(K)$  and is unique up to a scalar. Moreover, we have  $\ker(\mathfrak{X}(A_0)) \cdot C = \ker(\mathfrak{X}'(A_0))$ , and thus going over from the standard bases  $\mathfrak{B}$  and  $\mathfrak{B}'$  with respect to  $\mathfrak{X}$  and  $\mathfrak{X}'$ , respectively, to the associated invertible matrices  $B$  and  $B'$  with rows  $v_1, \dots, v_n \in K^n$  and  $v'_1, \dots, v'_n \in K^n$ , respectively, we get  $B \cdot C = B'$ , or equivalently

$$C = B^{-1} \cdot B' \in \text{GL}_n(K).$$

Thus to determine  $C$  we have to perform the following steps: find  $A_0 \in \mathcal{A}$  such that  $\dim_K(\ker(\mathfrak{X}(A_0))) = 1$ ; compute  $\ker(\mathfrak{X}(A_0)) \leq K^n$  and  $\ker(\mathfrak{X}'(A_0)) \leq K^n$ ; compute a Schreier tree  $\mathfrak{T}$  with respect to  $\mathfrak{X} \cong \mathfrak{X}'$  and  $A_0$ ; apply the Schreier tree  $\mathfrak{T}$  in order to compute standard bases  $\mathfrak{B}$  and  $\mathfrak{B}'$  of  $K^n$  with respect to  $\mathfrak{X}$  and  $\mathfrak{X}'$ , respectively; going over to matrices, compute the inverse  $B^{-1} \in \text{GL}_n(K)$ ; and compute the product  $C = B^{-1} \cdot B' \in \text{GL}_n(K)$ .

**In Practice** If  $K = \mathbb{Q}(X)$ , the nullspaces required can be found as described in Sect. 5.1, where we may assume that  $v_1$  and  $v'_1$  are primitive. Moreover, computing matrix inverses and matrix products can be done as described in Sects. 5.2 and 5.4, respectively; by multiplying with a suitable element of  $K$  we may assume that  $C$  is primitive as well, then  $C$  is unique up to sign. Hence for our application it remains to describe how a distinguished element and a Schreier tree can be found, and we have to give an efficient break condition for the algorithm in Sect. 5.4.

## 7 Finding Standard Bases for $W$ -Graph Representations

We have now described the necessary infrastructure from linear algebra over integral domains, and some relevant general ideas how to compute with representations, to proceed to the explicit determination of Gram matrices of invariant bilinear forms for balanced representations of Iwahori–Hecke algebras. We recall the setting of Sect. 2.5, which we keep from now on:

Let  $(W, S)$  be a finite Coxeter group, and let  $\mathcal{H}_A \subseteq \mathcal{H}_K$  be the associated generic Iwahori–Hecke algebras with equal parameters over the ring  $A = \mathbb{Z}[v, v^{-1}]$  and the field  $K = \mathbb{Q}(v)$ , respectively, being generated by  $\{T_s; s \in S\}$ . Moreover, let  $\mathfrak{X}^\lambda: \mathcal{H}_K \rightarrow K^{n \times n}$ , where  $n = d_\lambda$ , be a  $W$ -graph representation associated with  $\lambda \in \Lambda$ , and let

$$(\mathfrak{X}^\lambda)': \mathcal{H}_K \rightarrow K^{n \times n}: T_w \mapsto \mathfrak{X}^\lambda(T_{w^{-1}})^{\text{tr}} \quad \text{for all } w \in W.$$

As far as computer implementations are concerned, it is more convenient and more efficient to work with row vectors instead of column vectors. Therefore, we will now work throughout with right actions rather than left actions as in Sect. 2. Our aim is to find a primitive Gram matrix  $P \in \mathbb{Z}[v]^{n \times n}$  for  $\mathfrak{X}^\lambda$ , that is, using the language of right actions, a primitive matrix such that

$$\mathfrak{X}^\lambda(T_w) \cdot P = P \cdot (\mathfrak{X}^\lambda)'(T_w) \quad \text{for all } w \in W.$$

Thus the task is to find a non-zero  $\mathcal{H}_K$ -homomorphism from  $\mathfrak{X}^\lambda$  to  $(\mathfrak{X}^\lambda)'$ . In order to use the approach described in Sect. 6.2, we proceed as follows, where the basic idea of this strategy has already been indicated in [9, Section 4.3]:

### 7.1 Finding Seed Vectors

To find a suitable seed vector  $u_1 \in K^n$  for the standard basis algorithm with respect to  $\mathfrak{X}^\lambda$ , we proceed as follows:

Specializing  $v \mapsto 1$  we from  $\mathcal{H}_A$  recover the group algebra  $\mathbb{Q}[W]$ , and  $\mathfrak{X}^\lambda$  corresponds to an irreducible representation  $\mathfrak{Y}^\lambda: \mathbb{Q}[W] \rightarrow \mathbb{Q}^{n \times n}$ . In particular, the index and sign representations of  $\mathcal{H}_K$ , given by  $\text{ind}_{\mathcal{H}}: T_s \mapsto v$  and  $\text{sgn}_{\mathcal{H}}: T_s \mapsto -v^{-1}$ , respectively, for all  $s \in S$ , correspond to the trivial and sign representations of  $\mathbb{Q}[W]$ , given by  $1_W: s \mapsto 1$  and  $\text{sgn}_W: s \mapsto -1$ , respectively.

As was observed by Benson and Curtis (see [10, Section 6.3] and the references there), there is a subset  $J \subseteq S$  (depending on  $\lambda$ , and in general not being unique), such that the restriction of  $\mathfrak{Y}^\lambda$  to the parabolic subgroup  $\widetilde{W} := W_J \leq W$  associated with  $J$  fulfills

$$\dim_{\mathbb{Q}} (\text{Hom}_{\mathbb{Q}[\widetilde{W}]}(\text{sgn}_{\widetilde{W}}, \mathfrak{Y}^\lambda)) = 1.$$

Note that  $J = \emptyset$  and  $J = S$  if and only if  $\mathfrak{Y}^\lambda$  equals  $1_W$  and  $\text{sgn}_W$ , respectively. Letting  $\widetilde{\mathcal{H}}_K \subseteq \mathcal{H}_K$  be the parabolic subalgebra associated with  $J$ , this implies

$$\dim_K (\text{Hom}_{\widetilde{\mathcal{H}}_K}(\text{sgn}_{\widetilde{\mathcal{H}}}, \mathfrak{X}^\lambda)) = 1.$$

In other words, we equivalently have

$$\dim_K \left( \bigcap_{s \in J} \ker (\mathfrak{X}^\lambda(T_s + v^{-1})) \right) = 1.$$

Now we are going to use the fact that  $\mathfrak{X}^\lambda$  is a  $W$ -graph representation: Using the  $I$ -sets associated with  $\mathfrak{X}^\lambda$ , see Definition 2.4, we conclude that  $\ker(\mathfrak{X}^\lambda(T_s + v^{-1})) =$

$\langle e_i; s \in I_i \rangle_K$  for all  $s \in S$ , where  $e_i \in K^n$  denotes the  $i$ -th “unit” vector. This implies

$$\bigcap_{s \in J} \ker(\mathfrak{X}^\lambda(T_s + v^{-1})) = \langle e_i; J \subseteq I_i \rangle_K.$$

Hence we may let  $u_1 := e_i$ , where  $1 \leq i \leq n$  is the unique index such that  $J \subseteq I_i$ .

Note that this conversely also yields a way to find all subsets of  $S$  fulfilling the Benson–Curtis condition: We run through all subsets  $J \subseteq S$ , and just check whether there is precisely one index  $1 \leq i \leq n$  such that  $J \subseteq I_i$ .

### 7.2 Finding a Distinguished Element

The above immediate approach strongly uses the fact that  $\mathfrak{X}^\lambda$  is a  $W$ -graph representation. Thus, in order to find a suitable seed vector  $u'_1 \in K^n$  for the standard basis algorithm with respect to  $(\mathfrak{X}^\lambda)'$  we specify a distinguished element  $T^\lambda \in \mathcal{H}_K$  such that  $\dim_K(\ker(\mathfrak{X}^\lambda(T^\lambda))) = 1$ . Let

$$T^\lambda := \left( \sum_{s \in J} T_s \right) + v^{-1} \cdot |J| \in \mathcal{H}_A \subseteq \mathcal{H}_K.$$

Hence we have  $\bigcap_{s \in J} \ker(\mathfrak{X}^\lambda(T_s + v^{-1})) \leq \ker(\mathfrak{X}^\lambda(T^\lambda))$ , and it remains to be shown that  $\dim_K(\ker(\mathfrak{X}^\lambda(T^\lambda))) = 1$ :

Assume to the contrary that  $\dim_K(\ker(\mathfrak{X}^\lambda(T^\lambda))) \geq 2$ . Then letting

$$\sigma_J := \frac{1}{|J|} \cdot \sum_{s \in J} s \in \mathbb{Q}[\tilde{W}],$$

specializing  $v \mapsto 1$  shows that  $\dim_{\mathbb{Q}}(\ker(\mathfrak{Y}^\lambda(1 + \sigma_J))) \geq 2$  as well. Since for any vector  $u \in \ker(\mathfrak{Y}^\lambda(1 + \sigma_J))$  we have  $u \cdot \mathfrak{Y}^\lambda(\sigma_J^k) = (-1)^k \cdot u$ , for all  $k \in \mathbb{N}_0$ , Lemma 7.1 proven below implies that  $\langle u \rangle_{\mathbb{Q}} \leq K^n$  is  $\mathbb{Q}[\tilde{W}]$ -invariant and carries the sign representation. Thus we have  $\dim_{\mathbb{Q}}(\text{Hom}_{\mathbb{Q}[\tilde{W}]}(\text{sgn}_{\tilde{W}}, \mathfrak{Y}^\lambda)) \geq 2$ , a contradiction.

**Lemma 7.1** For  $\epsilon \in \{0, 1\}$  let  $W_\epsilon := \{w \in W; \text{sgn}(w) = (-1)^\epsilon\}$ . Moreover, let

$$\sigma_S := \frac{1}{|S|} \cdot \sum_{s \in S} s \in \mathbb{Q}[W].$$

Then, with respect to the natural topology on  $\mathbb{Q}[W] \cong \mathbb{Q}^{|W|}$ , we have

$$\lim_{k \rightarrow \infty} \sigma_S^{2k+\epsilon} = \frac{1}{|W_\epsilon|} \cdot \sum_{w \in W_\epsilon} w \in \mathbb{Q}[W].$$

*Proof* We consider the Markov chain with (finite) state space  $W = W_0 \dot{\cup} W_1$ , and transition matrix  $M = \text{reg}_W(\sigma_S) \in \mathbb{Q}^{|W| \times |W|}$ , where  $\text{reg}_W: \mathbb{Q}[W] \rightarrow \mathbb{Q}^{|W| \times |W|}$  denotes the regular matrix representation of  $\mathbb{Q}[W]$ . In other words, the matrix entry  $M_{w,w'}$ , where  $w, w' \in W$ , is given as

$$M_{w,w'} := \begin{cases} \frac{1}{|S|}, & \text{if } w' = ws \text{ for some } s \in S, \\ 0, & \text{otherwise.} \end{cases}$$

Now, since  $\text{sgn}(ws) = -\text{sgn}(w)$  for all  $w \in W$  and  $s \in S$ , we conclude that  $M^2 = \text{reg}_W(\sigma_S^2)$  induces Markov chains on both  $W_0$  and  $W_1$ . Moreover, since any element of  $W$  can be written as a word of length at most  $l(w_0)$  in the generators  $S$ , we infer that  $M^{2l(w_0)}$  has positive entries in both the block submatrices belonging to  $W_0$  and  $W_1$ , respectively. Hence the induced Markov chains are both irreducible and aperiodic. They thus converge towards stationary distributions, which since  $M$  is doubly-stochastic are both equal to the respective uniform distributions. Thus, in particular, the initial state  $\sigma_S^\epsilon \in \langle W_\epsilon \rangle_{\mathbb{Q}}$  yields

$$\lim_{k \rightarrow \infty} \sigma_S^{2k+\epsilon} = \sigma_S^\epsilon \cdot \left( \lim_{k \rightarrow \infty} (M^2)^k \right) = \frac{1}{|W_\epsilon|} \cdot \sum_{w \in W_\epsilon} w.$$

□

### 7.3 Finding Standard Bases

The distinguished element  $T^\lambda$  can now be used to find a primitive vector  $u'_1 \in \ker((\mathfrak{X}^\lambda)'(T^\lambda))$ . Next, having both seed vectors  $u_1$  and  $u'_1$  in place, we aim at computing the associated standard bases  $\mathfrak{B}$  with respect to  $\mathfrak{X}^\lambda$ , and  $\mathfrak{B}'$  with respect to  $(\mathfrak{X}^\lambda)'$ , for the  $A$ -algebra generated by  $\{vT_s; s \in S\}$ . But since we do not have a standard basis algorithm available for representations over the field  $K$ , we again use suitable specializations:

Given a place  $0 \neq b \in \mathbb{Z}$ , let  $\mathfrak{Y}_b^\lambda: \mathcal{H}_{\mathbb{Q}} \rightarrow \mathbb{Q}^{n \times n}$  be the representation of  $\mathcal{H}_{\mathbb{Q}}$  obtained by specializing  $v \mapsto b$ , that is, considering  $\mathcal{H}_{\mathbb{Q}}$  as the  $\mathbb{Q}$ -algebra generated by  $\{bT_s; s \in S\}$  we have

$$\mathfrak{Y}_b^\lambda: bT_s \mapsto (\mathfrak{X}^\lambda(vT_s))(b) := \mathfrak{X}^\lambda(vT_s)|_{v \mapsto b} \in \mathbb{Z}^{n \times n};$$

thus in particular for  $b = 1$ , identifying  $\mathcal{H}_{\mathbb{Q}}$  with  $\mathbb{Q}[W]$ , we recover  $\mathfrak{Y}_1^\lambda = \mathfrak{Y}^\lambda$ .

Now we compare a putative run of the standard basis algorithm, as described in Sect. 6.1, with respect to the seed vector  $u_1 \in \mathbb{Z}[v]^n$  and the generators  $\{\mathfrak{X}^\lambda(vT_s) \in \mathbb{Z}[v]^{n \times n}; s \in S\}$ , with a run with respect to the specialized seed vector  $u_1(b) \in \mathbb{Z}^n$  and the generators  $\{\mathfrak{Y}_b^\lambda(bT_s) \in \mathbb{Z}^{n \times n}; s \in S\}$ . These successively produce standard bases  $\mathfrak{B} \subseteq \mathbb{Z}[v]^n$  and  $\mathfrak{C} \subseteq \mathbb{Z}^n$ , respectively. We show by induction on the cardinality

$0 \leq m \leq n$  of the intermediate sets  $\mathfrak{B}$ , that for all but finitely many  $b$  the set  $\mathfrak{C}$  is obtained by specializing  $\mathfrak{B}$ , and that the Schreier trees found in both runs coincide:

Indeed, the key steps are to decide for some  $w := u \cdot \mathfrak{X}^\lambda(vT_s) \in \mathbb{Z}[v]^n$  whether or not  $w \in \langle \mathfrak{B} \rangle_K$ , and similarly for its specialization  $w(b) := u(b) \cdot \mathfrak{Y}_b^\lambda(bT_s) \in \mathbb{Z}^n$  whether or not  $w(b) \in \langle \mathfrak{C} \rangle_{\mathbb{Q}}$ . Identifying  $\mathfrak{B}$  and  $\mathfrak{C}$  with matrices  $B \in \mathbb{Z}[v]^{m \times n}$  and  $C \in \mathbb{Z}^{m \times n}$ , respectively, we have  $C = B(b)$ . Considering the matrix  $B_w \in \mathbb{Z}[v]^{(m+1) \times n}$  obtained by concatenating  $B$  and  $w$ , we have  $w \notin \langle \mathfrak{B} \rangle_K$  if and only if there is an  $((m + 1) \times (m + 1))$ -submatrix  $B'$  of  $B_w$  such that  $\det(B'_w) \neq 0$ . Similarly, we have  $w(b) \notin \langle \mathfrak{C} \rangle_{\mathbb{Q}}$  if and only if there is an  $((m + 1) \times (m + 1))$ -submatrix  $C'$  of  $C_{w(b)} = B_w(b) \in \mathbb{Z}^{(m+1) \times n}$  such that  $\det(C') \neq 0$ . Hence, whenever  $w(b) \notin \langle \mathfrak{C} \rangle_{\mathbb{Q}}$  we also have  $w \notin \langle \mathfrak{B} \rangle_K$ , and conversely for all but finitely many  $b$  from  $w \notin \langle \mathfrak{B} \rangle_K$  we may conclude that  $w(b) \notin \langle \mathfrak{C} \rangle_{\mathbb{Q}}$ . (We have used a similar argument in Sect. 5.1.)

Thus assuming that  $0 \neq b \in \mathbb{Z}$  is suitably chosen, we may just run the standard basis algorithm for the seed vector  $u_1(b) = u_1 = e_i \in \mathbb{Z}^n$ , the  $i$ -th “unit” vector, and the generators  $\mathfrak{Y}_b^\lambda(bT_s) \in \mathbb{Z}^{n \times n}$ , as described in Sect. 6.1, yielding a Schreier tree  $\mathfrak{T}$ . Letting  $w_1 := 1 \in W$ , and  $w_i := w_j \cdot s \in W$ , if  $[j, s]$  is the  $i$ -th entry in  $\mathfrak{T}$ , for  $2 \leq i \leq n$ , we thus obtain reduced expressions of the elements  $w_i \in W$ , and hence the number of steps needed to find the  $i$ -th element of  $\mathfrak{C}$  equals the length  $l(w_i) \in \mathbb{N}_0$ . (In practice, it turns out that choosing either  $b = 1$  or  $b = 2$  is sufficient, where actually almost always  $b = 1$  works.)

Applying the Schreier tree  $\mathfrak{T}$  to  $u_1$  and  $\{\mathfrak{X}^\lambda(vT_s); s \in S\}$  this yields a standard basis  $\mathfrak{B} \subseteq \mathbb{Z}[v]^n$  of  $K^n$ . Similarly, applying  $\mathfrak{T}$  to  $u'_1 \in \mathbb{Z}[v]^n$  and  $\{(\mathfrak{X}^\lambda)'(vT_s) \in \mathbb{Z}[v]^{n \times n}; s \in S\}$  we get a standard basis  $\mathfrak{B}' \subseteq \mathbb{Z}[v]^n$  of  $K^n$ . But note that this does *not* ensure that the  $A$ -lattices  $\langle \mathfrak{B} \rangle_A$  and  $\langle \mathfrak{B}' \rangle_A$  are invariant under the  $A$ -algebras generated by  $\{\mathfrak{X}^\lambda(vT_s); s \in S\}$  and  $\{(\mathfrak{X}^\lambda)'(vT_s); s \in S\}$ , respectively. (In practice they are not, typically.)

## 8 Finding Gram Matrices for $W$ -Graph Representations

We keep the setting of Sect. 7; in particular  $\mathfrak{X}^\lambda$  still is a  $W$ -graph representation. Having found standard bases  $\mathfrak{B}$  and  $\mathfrak{B}'$  for  $\mathfrak{X}^\lambda$  and  $(\mathfrak{X}^\lambda)'$ , respectively, we proceed by writing them as matrices  $B \in \mathbb{Z}[v]^{n \times n}$  and  $B' \in \mathbb{Z}[v]^{n \times n}$ , respectively, where by construction both  $B$  and  $B'$  are primitive. In order to complete the final task of computing the product  $B^{-1} \cdot B' \in \mathbb{Z}[v]^{n \times n}$  efficiently, we need a few preparations.

### 8.1 Palindromicity

Let  $*$ :  $K \rightarrow K$  be the involutory field automorphism given by  $*$ :  $v \mapsto v^{-1}$ . Hence  $A$  is  $*$ -invariant, and by entry-wise application we get involutory module automorphisms on  $K^n$  and  $A^n$ , and algebra automorphisms on  $K^{n \times n}$  and  $A^{n \times n}$ , all of which will also be denoted by  $*$ .

A polynomial  $0 \neq f \in \mathbb{Z}[v]$  is called  $(k)$ -palindromic, for some  $k \in \mathbb{N}_0$ , if  $v^k \cdot f^* = f \in A$ , and  $f$  is called  $(k)$ -skew-palindromic if  $v^k \cdot f^* = -f \in A$ . In these cases, letting  $\delta(f) \in \mathbb{N}_0$  be the maximum power of  $v$  dividing  $f$  in  $\mathbb{Z}[v]$ , we have  $k = \delta(f) + \deg(f)$ . Hence  $f$  is palindromic or skew-palindromic if and only if  $f \in \mathbb{Z}[v]$  and  $f^* \in \mathbb{Z}[v^{-1}]$  are associated in  $A$ . Moreover, if  $f$  is  $k$ -skew-palindromic, then specializing  $v \mapsto 1$  we get  $f(1) = -f(1)$ , implying that  $v - 1$  divides  $f$  in  $\mathbb{Z}[v]$ ; similarly, if  $f$  is  $k$ -palindromic, then specializing  $v \mapsto -1$  we get  $(-1)^k \cdot f(-1) = f(-1)$ , implying that  $k$  is even, or  $v + 1$  divides  $f$  in  $\mathbb{Z}[v]$ .

**Proposition 8.1**

- (a) Let  $P \in \mathbb{Z}[v]^{n \times n}$  be a primitive Gram matrix for  $\mathfrak{X}^\lambda$ . Then we have  $v^m \cdot P^* = P$ , where  $m = m_P \in \mathbb{N}$  is even and coincides with the maximum of the degrees of the non-zero entries of  $P$ .
- (b) For the primitive seed vector  $u'_1 \in \mathbb{Z}[v]^n$  we have  $v^m \cdot (u'_1)^* = u'_1$ , where  $m = m_{u'_1} \in \mathbb{N}_0$  is even and coincides with the maximum of the degrees of the non-zero entries of  $u'_1$ . (Trivially, the analogous statement holds for  $u_1 \in \mathbb{Z}[v]^n$  with  $m_{u_1} = 0$ .)

*Proof* Letting  $E_n \in A^{n \times n}$  be the identity matrix, by Definition 2.4 for  $s \in S$  we have

$$\mathfrak{X}^\lambda(T_s)^* = \mathfrak{X}^\lambda(T_s) - (v - v^{-1}) \cdot E_n = \mathfrak{X}^\lambda(T_s - (v - v^{-1})).$$

In particular, this yields

$$\mathfrak{X}^\lambda(T_s + v^{-1})^* = \mathfrak{X}^\lambda(T_s)^* + v \cdot E_n = \mathfrak{X}^\lambda(T_s - (v - v^{-1})) + v \cdot E_n = \mathfrak{X}^\lambda(T_s + v^{-1}).$$

- (a) We consider the matrix  $P^* \in \mathbb{Z}[v^{-1}]^{n \times n}$ : For all  $s \in S$  we have

$$\begin{aligned} \mathfrak{X}^\lambda(T_s) \cdot P^* &= \left( \mathfrak{X}^\lambda(T_s - (v - v^{-1})) \cdot P \right)^* = \left( P \cdot \mathfrak{X}^\lambda(T_s - (v - v^{-1}))^{\text{tr}} \right)^* \\ &= \left( P \cdot \mathfrak{X}^\lambda(T_s)^{* \text{tr}} \right)^* = \left( P \cdot \mathfrak{X}^\lambda(T_s)^{\text{tr} *} \right)^* = P^* \cdot \mathfrak{X}^\lambda(T_s)^{\text{tr}}. \end{aligned}$$

Now  $m = m_P \in \mathbb{N}$  as above is minimal such that  $v^m P^* \in \mathbb{Z}[v]^{n \times n}$ , hence we infer that  $v^m P^*$  is a primitive Gram matrix for  $\mathfrak{X}^\lambda$  as well, and thus we have  $v^m P^* = P$  or  $v^m P^* = -P$ . Assume the latter case holds, then all non-zero entries of  $P$  are  $m$ -skew-palindromic, implying that  $v - 1$  divides  $\gcd(P)$ , contradicting the primitivity of  $P$ . Hence we have  $v^m P^* = P$ , that is all non-zero entries of  $P$  are  $m$ -palindromic. Assume that  $m$  is odd, then we infer that  $v + 1$  divides  $\gcd(P)$ , again contradicting the primitivity of  $P$ . Hence  $m$  is even.

(b) We consider the vector  $(u'_1)^* \in \mathbb{Z}[v^{-1}]^n$ : We have

$$\begin{aligned} (u'_1)^* \cdot (\mathfrak{X}^\lambda)'(T^\lambda) &= \left( u'_1 \cdot (\mathfrak{X}^\lambda)'(T^\lambda)^* \right)^* = \left( u'_1 \cdot \left( \sum_{s \in J} \mathfrak{X}^\lambda(T_s + v^{-1}) \right)^{\text{tr}*} \right)^* \\ &= \left( u'_1 \cdot \left( \sum_{s \in J} \mathfrak{X}^\lambda(T_s + v^{-1}) \right)^{\text{tr}} \right)^* = \left( u'_1 \cdot (\mathfrak{X}^\lambda)'(T^\lambda) \right)^* = 0. \end{aligned}$$

Now  $m = m_{u'_1} \in \mathbb{N}_0$  as above is minimal such that  $v^m \cdot (u'_1)^* \in \mathbb{Z}[v]^n$ , hence we infer that  $v^m \cdot (u'_1)^*$  is primitive. Thus from  $\dim_K(\ker((\mathfrak{X}^\lambda)'(T^\lambda))) = 1$  we conclude that  $v^m \cdot (u'_1)^* = u'_1$  or  $v^m \cdot (u'_1)^* = -u'_1$ . Now we argue as above. □

### 8.2 Properties of the Standard Bases

We have a closer look at the standard bases  $\mathfrak{B}$  and  $\mathfrak{B}'$ , and the associated matrices  $B$  and  $B'$ , where we assume  $\mathfrak{B}$  to be chosen according to Sect. 7.3. The facts collected are largely due to experimental observation, and will be helpful in the final computational steps in Sect. 8.3. Still, these properties seem to be stronger than expected from general principles, and it should be worth-while to try and prove the particular observations specified below. (In particular, we have checked the standard bases associated with *all* subsets  $J \subseteq S$  fulfilling the Benson–Curtis condition, see Sect. 7.1, for the types  $E_6, E_7$  and  $E_8$ .)

Recall that for all  $s \in S$  we have

$$(vT_s)^{-1} = v^{-1} \cdot (T_s - (v - v^{-1})) = v^{-2} \cdot (vT_s - (v^2 - 1)),$$

hence by the proof of Proposition 8.1 we get

$$\mathfrak{X}^\lambda(vT_s)^* = v^{-1} \cdot \mathfrak{X}^\lambda(T_s - (v - v^{-1})) = v^{-2} \cdot \mathfrak{X}^\lambda(vT_s - (v^2 - 1)) = \mathfrak{X}^\lambda((vT_s)^{-1}).$$

**The Elements of  $\mathfrak{B}$**  For any  $u_i \in \mathfrak{B}$ , where  $2 \leq i \leq n$ , we have  $u_i = u_j \cdot \mathfrak{X}^\lambda(vT_s)$ , for some  $1 \leq j < i$  and  $s \in S$ . This yields

$$v^2 \cdot u_j = v^2 \cdot u_i \cdot \mathfrak{X}^\lambda((vT_s)^{-1}) = u_i \cdot \mathfrak{X}^\lambda(vT_s - (v^2 - 1)).$$

We conclude that  $\gcd(u_i) \in \mathbb{Z}[v]$  and  $\gcd(u_j) \in \mathbb{Z}[v]$  are associated in  $A$ . Hence by recursion, since  $u_1$  is primitive, we infer that  $\gcd(u_i) = v^{d_i} \in \mathbb{Z}[v]$  for some  $d_i \in \mathbb{N}_0$ .

Moreover, we have  $d_j \leq d_i \leq d_j + 2$ . Since  $d_1 = 0 = l(w_1)$ , this implies  $d_i \leq 2l(w_i)$  for all  $1 \leq i \leq n$ , where  $w_i \in W$  is as in Sect. 7.3. (Experiments show that all three cases  $d_i \in \{d_j, d_j + 1, d_j + 2\}$  actually occur.) But the growth behavior



of the  $d_i$  seems to be more restricted than given by these bounds: Considering the case  $l(w_i) = 1$ , we have  $w_i = s$  for some  $s \in S$  such that the “unit” vector  $u_1$  is not an eigenvector of  $T_s$ , hence using the shape of  $\mathfrak{X}^\lambda(vT_s)$  we conclude that  $d_i = 1 = l(w_i)$ .

Now, experimentally, we have made the following

**Observation 8.1** *We have  $d_i \leq l(w_i) + 1$ , for all  $1 \leq i \leq n$ .*

(Actually, almost always we have got  $d_i \leq l(w_i)$ , for all  $1 \leq i \leq n$ , where often we have even seen equality throughout; the only cases found where actually  $d_i = l(w_i) + 1$ , for some  $i$ , are for type  $E_8$ , the representation labeled by 3200<sub>x</sub>, and two out of the twelve Benson–Curtis subsets of generators.)

**The Matrix  $B$**  Letting  $1 \leq j < i \leq n$  and  $s \in S$  be as above, we get

$$v^2 \cdot u_i^* = v^2 \cdot u_j^* \cdot \mathfrak{X}^\lambda(vT_s)^* = u_j^* \cdot \mathfrak{X}^\lambda(vT_s - (v^2 - 1)).$$

Since the standard basis algorithm is a breadth-first search, from  $u_1^* = u_1$  we conclude that there is lower untriangular matrix  $U \in K^{n \times n}$  and a diagonal matrix  $D = \text{diag}[v^{2l(w_1)}, \dots, v^{2l(w_n)}] \in \mathbb{Z}[v]^{n \times n}$ , such that

$$D \cdot B^* = U \cdot B.$$

(Note that if the  $A$ -lattice  $\langle \mathfrak{B} \rangle_A$  was invariant under the  $A$ -algebra generated by  $\{\mathfrak{X}^\lambda(vT_s); s \in S\}$ , then we even had  $U \in A^{n \times n}$ .)

In particular, letting  $l := \sum_{i=1}^n l(w_i) \in \mathbb{N}_0$ , we infer that

$$\det(B) = v^{2l} \cdot \det(B^*),$$

hence  $\det(B) \in \mathbb{Z}[v]$  is palindromic. Letting  $\text{exp}(B) \in \mathbb{Z}[v]$  denote the exponent of  $B$  in the sense of Sect. 5.3, it follows from Proposition 5.1 that the non-constant irreducible polynomials dividing  $\det(B)$  are precisely those dividing  $\text{exp}(B)$ . Now, experimentally, we have made the following

**Observation 8.2** *Any irreducible divisor of  $\text{exp}(B)$  in  $\mathbb{Z}[v]$  is monic and palindromic.*

(Actually, in general the entries of the matrix  $B$  are neither palindromic nor skew-palindromic; moreover, quite often  $\text{exp}(B)$  is a product of cyclotomic polynomials, but this does not always happen.)

In particular, if  $\widehat{u}_k^{\text{tr}} \in \mathbb{Z}[v]^{1 \times n}$  denotes the  $k$ -th column of  $B$ , for  $1 \leq k \leq n$ , then  $\text{gcd}(\widehat{u}_k) \in \mathbb{Z}[v]$  divides  $\det(B)$ , hence  $\text{gcd}(\widehat{u}_k)$  is palindromic as well. (Actually, contrary to  $\text{gcd}(u_k) = v^{d_k}$ , in general the  $\text{gcd}(\widehat{u}_k)$  are not just powers of  $v$ .)

**The Elements of  $\mathfrak{B}'$**  The recursion used in the standard basis algorithm only depends on the Schreier tree  $\mathfrak{T}$ , but is independent of the representation considered. Hence for  $u'_i \in \mathfrak{B}'$ , where  $1 \leq i \leq n$ , and  $u'_1$  is primitive, we get  $\text{gcd}(u'_i) = v^{d'_i} \in \mathbb{Z}[v]$  for some  $d'_i \in \mathbb{N}_0$ . Moreover, if  $1 \leq j < i \leq n$  and  $s \in S$  are as above, we get

$d'_j \leq d'_i \leq d'_j + 2$  and  $d'_i \leq 2l(w_i)$ . Actually, the  $d'_i$  seem to be closely related to the  $d_i$  from above, inasmuch experimentally we have made the following

**Observation 8.3** *We have  $d'_i = d_i$ , for all  $1 \leq i \leq n$ .*

**The Matrix  $B'$**  Again by the fact that the recursion used in the standard basis algorithm only depends on  $\mathfrak{T}$ , and using  $v^m \cdot (u'_1)^* = u'_1$ , where  $m = m_{u'_1} \in \mathbb{N}_0$  is as in Proposition 8.1, we get

$$v^m \cdot D \cdot (B')^* = U \cdot B',$$

for the same matrices  $U$  and  $D$ . In particular, it follows that  $\det(B')$  is palindromic. (In general neither  $\det(B')$  and  $\det(B)$ , nor  $\exp(B')$  and  $\exp(B)$  are associated in  $A$ , so that  $\langle \mathfrak{B} \rangle_A$  and  $\langle \mathfrak{B}' \rangle_A$  are inequivalent  $A$ -sublattices of  $A^n$ , which typically are not included in each other.) Again, experimentally we have made the following

**Observation 8.4** *Any irreducible divisor of  $\exp(B')$  in  $\mathbb{Z}[v]$  is monic and palindromic.*

In particular, similarly, if  $\widehat{u}'_k \text{tr} \in \mathbb{Z}[v]^{1 \times n}$  denotes the  $k$ -th column of  $B'$ , for  $1 \leq k \leq n$ , then  $\gcd(\widehat{u}'_k) \in \mathbb{Z}[v]$  is palindromic.

**The Product  $B^{-1} \cdot B'$**  In combination the above yields

$$v^m \cdot (B^{-1} \cdot B')^* = v^m \cdot (B^*)^{-1} \cdot (B')^* = (D^{-1} \cdot U \cdot B)^{-1} \cdot (D^{-1} \cdot U \cdot B') = B^{-1} \cdot B'.$$

Hence the non-zero entries of  $B^{-1} \cdot B'$  are palindromic.

Letting  $0 \neq b \in \mathbb{Z}$  and  $\widehat{B} \in \mathbb{Z}[v]^{n \times n}$  primitive such that  $B^{-1} = \frac{1}{b \cdot \exp(B)} \cdot \widehat{B}$ , we get

$$b \cdot \exp(B) \cdot B^{-1} \cdot B' = \widehat{B} \cdot B' = c \cdot P,$$

where  $P \in \mathbb{Z}[v]^{n \times n}$  is a primitive Gram matrix, and  $0 \neq c \in \mathbb{Z}[v]$ . In particular, since by Observation 8.2 the exponent  $\exp(B)$  is palindromic, we conclude that the non-zero entries of  $\widehat{B} \cdot B'$  are palindromic as well.

Moreover, letting  $\widetilde{m} = m_{\exp(B)} \in \mathbb{N}_0$  such that  $v^{\widetilde{m}} \cdot \exp(B)^* = \exp(B)$ , we get

$$v^{m+\widetilde{m}} \cdot (b \cdot \exp(B) \cdot B^{-1} \cdot B')^* = b \cdot \exp(B) \cdot B^{-1} \cdot B' \in \mathbb{Z}[v]^{n \times n}.$$

Hence from  $v^{m_P} \cdot P^* = P$ , where  $m_P \in \mathbb{N}_0$  is as in Proposition 8.1, we get

$$m_P \leq m + \widetilde{m} = m_{u'_1} + m_{\exp(B)},$$

providing an upper bound on the degrees of the non-zero entries of  $P$ .

### 8.3 The Final Product

We are now prepared to do the last computational steps. To do so, we could quite straightforwardly compute first the inverse  $B^{-1}$ , that is essentially  $\widehat{B}$ , and then the product  $\widehat{B} \cdot B'$ . But it will substantially add to the efficiency if we keep the degrees of the non-zero entries of the matrices involved as small as possible. Now we have already observed above that the rows of  $B$  and  $B'$  are far from being primitive, and it turns out in practice that this also holds for their columns. We take advantage of this as follows:

Keeping the notation of Sect. 8.2, let  $R := \text{diag}[v^{d_1}, \dots, v^{d_n}] \in \mathbb{Z}[v]^{n \times n}$ . Then the rows of  $R^{-1} \cdot B \in \mathbb{Z}[v]^{n \times n}$  are primitive. As for its columns, letting  $\widetilde{u}_k^{\text{tr}} \in \mathbb{Z}[v]^{1 \times n}$  denote the  $k$ -th column of  $R^{-1} \cdot B$ , for  $1 \leq k \leq n$ , let

$$C := \text{diag}[\text{gcd}(\widetilde{u}_1), \dots, \text{gcd}(\widetilde{u}_n)] \in \mathbb{Z}[v]^{n \times n}.$$

Since by Observation 8.2 the polynomial  $\text{gcd}(\widehat{u}_k)$  is palindromic, using the particular form of  $R$ , we conclude that the  $\text{gcd}(\widetilde{u}_k)$  are palindromic as well. We let  $0 \neq \widehat{c} \in \mathbb{Z}[v]$  and  $\widehat{C} \in \mathbb{Z}[v]^{n \times n}$  be primitive such that  $C^{-1} = \frac{1}{\widehat{c}} \cdot \widehat{C}$ . The latter are of course straightforwardly computed, where both  $\widehat{c}$  and the diagonal entries of  $\widehat{C}$  are palindromic.

Then we get  $\widetilde{B} \in \mathbb{Z}[v]^{n \times n}$  such that  $B = R \cdot \widetilde{B} \cdot C$ , where now all the rows and all the columns of  $\widetilde{B}$  are primitive. We use the algorithm in Sect. 5.2 to compute  $0 \neq \widehat{b} \in \mathbb{Z}[v]$  and  $\widehat{B} \in \mathbb{Z}[v]^{n \times n}$  primitive such that  $\widetilde{B}^{-1} = \frac{1}{\widehat{b}} \cdot \widehat{B}$ . Since by Observation 8.2 the exponent  $\text{exp}(B)$  is palindromic, using the particular form of  $R$  and  $C$ , we conclude that  $\widehat{b}$  is palindromic as well. Thus altogether we have

$$B^{-1} = \frac{1}{\widehat{b} \cdot \widehat{c}} \cdot \widehat{C} \cdot \widehat{B} \cdot R^{-1}.$$

Similarly, let  $R' := \text{diag}[v^{d'_1}, \dots, v^{d'_n}] \in \mathbb{Z}[v]^{n \times n}$  and

$$C' := \text{diag}[\text{gcd}(\widetilde{u}'_1), \dots, \text{gcd}(\widetilde{u}'_n)] \in \mathbb{Z}[v]^{n \times n},$$

where  $\widetilde{u}'_k^{\text{tr}} \in \mathbb{Z}[v]^{1 \times n}$  denotes the  $k$ -th column of  $(R')^{-1} \cdot B'$ , for  $1 \leq k \leq n$ . As above, using Observation 8.4 implying the palindromicity of  $\text{gcd}(\widehat{u}'_k)$ , we conclude that the diagonal entries of  $C'$  are palindromic as well, and thus those of  $(C')^{-1}$  are too. Then we get  $\widetilde{B}' \in \mathbb{Z}[v]^{n \times n}$  such that  $B' = R' \cdot \widetilde{B}' \cdot C'$ , where now all the rows and all the columns of  $\widetilde{B}'$  are primitive.

In combination this yields

$$Q := \widehat{b} \cdot \widehat{c} \cdot B^{-1} \cdot B' = \widehat{C} \cdot \widehat{B} \cdot R^{-1} \cdot R' \cdot \widetilde{B}' \cdot C'.$$

By the above considerations we conclude that the non-zero entries of  $Q$  are palindromic, which entails that those of  $\widehat{B} \cdot R^{-1} \cdot R' \cdot \widetilde{B}'$  are as well. Now by

Observation 8.3 we have  $R' = R$ , hence this simplifies to

$$Q = \widehat{C} \cdot (\widehat{B} \cdot \widetilde{B}') \cdot C' \in \mathbb{Z}[v]^{n \times n},$$

where the non-zero entries of  $\widehat{B} \cdot \widetilde{B}' \in \mathbb{Z}[v]^{n \times n}$  are palindromic.

**In Practice** To find  $Q$ , finally, we apply the matrix multiplication algorithm in Sect. 5.4 to compute the product  $\widehat{B} \cdot \widetilde{B}'$ . As was already mentioned, in order to apply it efficiently we need good break conditions to discard erroneous guesses quickly: Apart from requiring that rational number recovery, see Sect. 3.2, returns only integral coefficients but not rational ones, it turns out that checking for palindromicity is highly effective in this respect.

Having found a good candidate for  $\widehat{B} \cdot \widetilde{B}' \in \mathbb{Z}[v]^{n \times n}$ , multiplying with the diagonal matrices  $\widehat{C} \in \mathbb{Z}[v]^{n \times n}$  and  $C' \in \mathbb{Z}[v]^{n \times n}$  is straightforward. Note that, since the result is expected to be a symmetric matrix, it is sufficient to compute only the lower triangular half of the product. Thus we get a candidate for a primitive Gram matrix  $P$  from  $Q = \gcd(Q) \cdot P \in \mathbb{Z}[v]^{n \times n}$ . (In many cases  $Q$  already is primitive, but this does not happen always, in which cases  $\gcd(Q)$  typically has a smallish degree.)

As independent verification we of course just explicitly check whether the candidate  $P$  fulfills the condition

$$\mathfrak{X}^\lambda(vT_s) \cdot P = P \cdot \mathfrak{X}^\lambda(vT_s)^{\text{tr}} \in \mathbb{Z}[v]^{n \times n} \quad \text{for all } s \in S.$$

## 9 Timings

We conclude by providing running times and workspace requirements for our computations in types  $E_7$  and  $E_8$ , and by presenting an explicit example for type  $E_6$ .

### 9.1 Timings

In Table 4, we give the running time (on a single processor running at a clock speed of 3.5 GHz) and GAP workspace requirements needed to compute primitive Gram matrices for types  $E_7$  and  $E_8$ , and the irreducible  $W$ -graph representations of  $\mathcal{H}_K$  given in [14, 15]. The figures for  $E_7$  should be compared with those given in Sect. 2.5 for the approach used there. Recalling that in [9, Remark 4.10] degree 2500 was the limit of feasibility, in Table 5 we present the resources now needed for the individual representations of degree at least 2500, where for comparison we repeat the first three columns of the relevant part of Table 2.

**Table 4** Time and space consumption

	Degree	No. repr.	Time	Workspace
$E_7$	All	60	4 min	0.2 GB
$E_8$	$\leq 1000$	50	30 min	0.7 GB
	1000–2000	20	137 min	2.2 GB
	2000–2500	10	329 min	4.3 GB
	2500–3000	5	350 min	5.9 GB
	3000–4000	7	874 min	11.6 GB
	4000–5000	13	3175 min	16.3 GB
	5000–7000	6	2784 min	23.2 GB
	$\geq 7000$	1	1183 min	31.5 GB

**Table 5** Time and space consumption for degree  $\geq 2500$ 

$E_8$	$m_p$	Abs.val.	Time	Workspace
$2688_y$	24	169,180	39 min	3.9 GB
$2800_z$	20	38,038	61 min	3.7 GB
$2800'_z$	30	882,222	116 min	5.9 GB
$2835_x$	24	1,344,484	52 min	3.1 GB
$2835'_x$	32	5,391,418	82 min	5.3 GB
$3150_y$	26	6,166,994	72 min	5.8 GB
$3200_x$	24	266,284	79 min	4.9 GB
$3200'_x$	30	587,345	104 min	6.1 GB
$3240_z$	16	25,586	60 min	4.0 GB
$3240'_z$	48	33,653,538	326 min	11.6 GB
$3360_z$	20	29,722	74 min	5.1 GB
$3360'_z$	32	775,084	159 min	8.1 GB
$4096_x$	22	531,634	156 min	8.0 GB
$4096'_x$	44	234,956,568	392 min	16.0 GB
$4096_z$	22	531,634	143 min	8.1 GB
$4096'_z$	44	234,956,568	428 min	16.1 GB
$4200_y$	28	58,249,760	171 min	10.1 GB
$4200_x$	24	5,413,484	171 min	9.8 GB
$4200'_x$	36	129,331,224	277 min	13.3 GB
$4200_z$	26	728,053	183 min	10.4 GB
$4200'_z$	28	1,298,612	199 min	10.3 GB
$4480_y$	32	85,556,320,920	239 min	13.9 GB
$4536_y$	28	3,887,856	180 min	11.7 GB
$4536_z$	24	2,728,756	217 min	11.4 GB
$4536'_z$	38	50,779,421	419 min	16.3 GB
$5600_w$	26	372,230	331 min	16.6 GB
$5600_z$	26	3,115,126	335 min	15.4 GB
$5600'_z$	30	3,848,044	473 min	17.5 GB
$5670_y$	30	10,762,741	351 min	21.7 GB
$6075_x$	26	894,864	542 min	19.5 GB
$6075'_x$	34	10,488,013	752 min	23.2 GB
$7168_w$	32	1,190,470,476	1183 min	31.5 GB

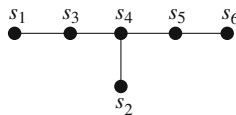
**Table 6** Time and space consumption for  $7168_w$

$7168_w$	Time	Workspace	Space	Disc
$\mathfrak{T}$	9 min	0.6 GB		
$u'_1$	5 min	1.3 GB		
$\widetilde{B}$	925 min	7.6 GB	1.7 GB	0.3 GB
$\widetilde{B}'$	29 min	17.5 GB	12.6 GB	4.7 GB
$\widetilde{B} \cdot \widetilde{B}'$	207 min	31.5 GB	5.8 GB	2.4 GB
P	8 min	7.9 GB	5.8 GB	2.5 GB

Finally, in Table 6 we give some details about the various steps in the computation for the unique representation of largest degree, which is labeled by  $7168_w$ . In the two last columns we indicate the actual size of the object under consideration in the GAP workspace, and the disc space needed to store it (as an uncompressed text file), respectively; the difference is accounted for by the space consumption of the data structure we are using within GAP, where matrices with polynomial entries are kept as lists of lists of (short) lists of (small long) integers. In particular, in the workspace needed to compute the product, next to the matrices  $\widetilde{B}$  and  $\widetilde{B}'$  and (the lower triangular half of) the product  $\widetilde{B} \cdot \widetilde{B}'$ , we also keep various specializations of the right hand factor  $\widetilde{B}'$ , which have a cumulative size of 7.1 GB. Hence to compute a primitive Gram matrix for the representation labeled by  $7168_w$  we need a running time of 1183 min  $\sim$  20h and a workspace of size 31.5 GB.

### 9.2 An Explicit Example

We conclude by revisiting the (tiny) example already presented in [9, Example 4.9] (which of course in practice runs in a fraction of a second): Let  $W$  be of type  $E_6$  with Dynkin diagram

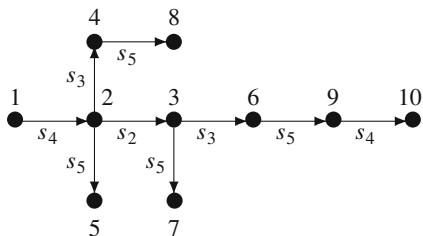


We consider the irreducible  $W$ -graph representation of  $\mathcal{H}_K$ , see [22], labeled by the representation  $10_s$  of  $\mathbb{Q}[W]$ , which is the unique one of degree 10, see Table 1. The  $W$ -graph in question is depicted in [9, Example 4.9], hence we do not repeat it here. But to illustrate the shape, and in particular the sparseness of the representing

matrices for the generators  $vT_{s_1}, \dots, vT_{s_6}$  we present a few of them:

$$vT_1 \mapsto \begin{bmatrix} v^2 & . & . & v & . & . & . & . & . & v & . \\ . & v^2 & . & v & . & . & . & . & . & . & v \\ . & . & -1 & . & . & . & . & . & . & . & . \\ . & . & . & -1 & . & . & . & . & . & . & . \\ . & . & . & . & -1 & . & . & . & . & . & . \\ . & . & . & v & . & . & v^2 & . & . & . & . \\ . & . & . & . & . & . & . & v^2 & . & v & v \\ . & . & . & . & . & . & . & . & v^2 & . & v \\ . & . & . & . & . & . & . & . & . & v^2 & . \\ . & . & . & . & . & . & . & . & . & . & -1 \\ . & . & . & . & . & . & . & . & . & . & . & -1 \end{bmatrix} \quad vT_6 \mapsto \begin{bmatrix} v^2 & . & . & . & . & . & . & . & . & v & v & . \\ . & v^2 & . & . & . & . & . & . & . & v & . & v \\ . & . & v^2 & . & . & . & . & . & . & . & v & v \\ . & . & . & v^2 & v & . & . & . & . & . & . & . \\ . & . & . & . & -1 & . & . & . & . & . & . & . \\ . & . & . & . & . & v^2 & v & . & . & . & . & . \\ . & . & . & . & . & . & . & -1 & . & . & . & . \\ . & . & . & . & . & . & . & . & -1 & . & . & . \\ . & . & . & . & . & . & . & . & . & -1 & . & . \\ . & . & . & . & . & . & . & . & . & . & -1 & . \\ . & . & . & . & . & . & . & . & . & . & . & -1 \end{bmatrix}$$

As it turns out, there are 22 possible choices of a distinguished subset  $J \subseteq S$ . We choose  $J := \{s_1, s_2, s_3, s_5, s_6\}$ , in accordance with [10, Table C.4]. Then associated primitive seed vectors  $u_1$  and  $u'_1$  are as given below, in the first row of the matrices  $B$  and  $\widetilde{B}'$ , respectively. Running the standard basis algorithm on the specialization of the above  $W$ -graph representation with respect to  $v \mapsto 1$  yields the following Schreier tree  $\mathfrak{T}$ , which we depict as an oriented graph, whose vertices  $1, \dots, 10$  correspond to the vectors in the (ordered) standard bases, and where an arrow from vertex  $j$  to vertex  $i$  with label  $s_k$  says that  $[j, s_k]$  is the  $i$ -th entry of  $\mathfrak{T}$ :



We find the standard basis  $\mathfrak{B}$  with associated matrix  $B$  as shown below. (It is not always the case that the entries of  $B$  are only monomials.) Hence we have  $R = \text{diag}[v^{d_1}, \dots, v^{d_{10}}]$ , where  $[d_1, \dots, d_{10}] = [0, 1, 2, 2, 2, 3, 3, 3, 4, 5] = [l(w_1), \dots, l(w_{10})]$ , and  $C$  is the identity matrix. Thus we get the matrix  $\widetilde{B}$ , and from that  $\widehat{b} = 1$  and the matrix  $\widehat{B}$  as also shown below. Note that the entries of  $\widehat{B}$  are not necessarily palindromic or skew-palindromic, and that the maximum degree of the

non-zero entries of  $B, \widetilde{B}$  and  $\widehat{B}$  equals 8, 3 and 5, respectively:

$$B = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & v & \cdot & \cdot & \cdot & v^2 \\ \cdot & \cdot & \cdot & \cdot & v^3 & \cdot & \cdot & \cdot & v^2 \\ \cdot & \cdot & \cdot & \cdot & v^3 & \cdot & v^2 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & v^2 v^3 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & v^5 & v^3 v^4 & v^4 v^4 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & v^3 v^4 v^5 & \cdot & \cdot & v^4 v^4 & \cdot \\ \cdot & v^3 & \cdot & v^4 v^5 & \cdot & \cdot & v^4 & v^4 & \cdot \\ \cdot & v^5 & v^5 & v^6 v^7 & v^4 v^5 & v^6 v^6 & v^6 v^6 & \cdot & \cdot \\ v^5 & v^7 & v^7 & v^6 & \cdot & v^6 v^7 & v^6 v^6 & v^8 & \cdot \end{bmatrix}$$

$$\widetilde{B} = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & v \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & v \\ \cdot & \cdot & \cdot & \cdot & v & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & v & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & v & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & v & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & v^2 & \cdot & 1 & v & v & v \\ \cdot & \cdot & \cdot & \cdot & v^2 & \cdot & 1 & v & v & v \\ \cdot & 1 & v & v^2 & \cdot & \cdot & v & v & \cdot & \cdot \\ \cdot & 1 & v & v^2 & \cdot & \cdot & v & \cdot & v & \cdot \\ \cdot & v & v & v^2 v^3 & 1 & v & v^2 & v^2 & v^2 & \cdot \\ 1 & v^2 & v^2 & v & \cdot & v & v^2 & v & v & v^3 \end{bmatrix}$$

$$\widehat{B} = \begin{bmatrix} 2v^5 - 3v^3 - 2v^4 + 3v^2 & v^3 - v & v^3 - v & v^3 - v & \cdot & \cdot & \cdot & \cdot & -v & 1 \\ -v^3 - v & v^2 & \cdot & -v & -v & \cdot & \cdot & \cdot & 1 & \cdot \\ -v^3 - v & v^2 & -v & \cdot & -v & \cdot & 1 & \cdot & \cdot & \cdot \\ v^2 & -v & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ -v & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ v^4 + 2v^2 & -v^3 & v^2 & v^2 & v^2 & v^2 & -v & -v & -v & 1 \\ -v^3 - v & v^2 & -v & -v & \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ v^2 & -v & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ v^2 & -v & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

Similarly, we find the standard basis  $\mathfrak{B}'$  with associated matrix  $B'$ . As it turns out we indeed have  $R' = R$ , and  $C'$  is the identity matrix. This yields the matrix  $\widetilde{B}'$  as shown below. Note that the entries of  $\widetilde{B}'$  are not necessarily palindromic or skew-palindromic, and that the maximum degree of the non-zero entries of  $\widetilde{B}'$  is 9:

$$\widetilde{B}' = \begin{bmatrix} 2v^3 & v^5 + 2v^3 + v & v^5 + 2v^3 + v & -v^4 - v^2 & v^5 + 2v^3 + v \\ -2v^2 & v^6 - v^2 & v^6 - v^2 & v^3 + v & -v^4 - 2v^2 - 1 \\ -v^5 + v & -2v^5 & -v^5 + v & v^6 + v^4 & -v^7 - 2v^5 - v^3 \\ -v^5 + v & -v^5 + v & -2v^5 & v^6 + v^4 & -v^7 - 2v^5 - v^3 \\ -v^5 + v & -v^5 + v & -v^5 + v & -v^2 - 1 & -v^7 - 2v^5 - v^3 \\ 2v^4 & 2v^4 & 2v^4 & v^7 - v^5 & -v^8 + v^4 \\ 2v^4 & 2v^4 & v^4 - 1 & -v^5 - v^3 & -v^8 + v^4 \\ 2v^4 & v^4 - 1 & 2v^4 & -v^5 - v^3 & -v^8 + v^4 \\ v^7 + v^5 - v^3 + v & -2v^3 & -2v^3 & -v^6 + v^4 & -v^9 + v^7 - v^5 - v^3 \\ -v^6 - v^4 + v^2 - 1 & -2v^6 & -2v^6 & v^5 - v^3 & v^8 - v^6 + v^4 + v^2 \end{bmatrix}$$



$$\left[ \begin{array}{ccccc} -v^4 - v^2 & v^5 + 2v^3 + v & -v^4 - v^2 & -v^4 - v^2 & -v^6 - 2v^4 - 2v^2 - 1 \\ -v^5 + v^3 & v^6 - v^2 & v^3 + v & v^3 + v & -v^7 - v^5 \\ v^4 - v^2 & -v^5 + v & v^6 + v^4 & -v^2 - 1 & v^6 + v^4 \\ v^4 - v^2 & -v^5 + v & -v^2 - 1 & v^6 + v^4 & v^6 + v^4 \\ v^4 - v^2 & -2v^5 & v^6 + v^4 & v^6 + v^4 & v^6 + v^4 \\ -v^3 + v & v^4 - 1 & -v^5 - v^3 & -v^5 - v^3 & -v^5 - v^3 \\ -v^3 + v & 2v^4 & v^7 - v^5 & -v^5 - v^3 & -v^5 - v^3 \\ -v^3 + v & 2v^4 & -v^5 - v^3 & v^7 - v^5 & -v^5 - v^3 \\ v^2 - 1 & -2v^3 & -v^6 + v^4 & -v^6 + v^4 & v^4 + v^2 \\ v^7 + v^5 & -2v^6 & v^5 - v^3 & v^5 - v^3 & -v^9 + v^7 \end{array} \right]$$

From this we get  $Q = \widehat{B} \cdot \widetilde{B}'$ . As it turns out we already have  $\text{gcd}(Q) = 1$ , thus we may let  $P = -Q$  be as shown below. Indeed, independent verification shows that  $P$  is a primitive Gram matrix as desired, coinciding with the one already given in [9, Example 4.9]. Note that indeed  $P$  is a completely dense matrix, all of whose entries are 6-palindromic, where the maximum degree occurring is 6, and that in accordance with Table 1 the largest coefficient occurring has absolute value 3, and that the specialization  $v \mapsto 0$  yields the identity matrix:

$$\left[ \begin{array}{ccccc} v^6 + 3v^4 + 3v^2 + 1 & 2v^4 + 2v^2 & 2v^4 + 2v^2 & -v^5 - 2v^3 - v & 2v^4 + 2v^2 \\ 2v^4 + 2v^2 & v^6 + 3v^4 + 3v^2 + 1 & 2v^4 + 2v^2 & -v^5 - 2v^3 - v & 2v^4 + 2v^2 \\ 2v^4 + 2v^2 & 2v^4 + 2v^2 & v^6 + 3v^4 + 3v^2 + 1 & -v^5 - 2v^3 - v & 2v^4 + 2v^2 \\ -v^5 - 2v^3 - v & -v^5 - 2v^3 - v & -v^5 - 2v^3 - v & v^6 + 2v^4 + 2v^2 + 1 & -v^5 - 2v^3 - v \\ 2v^4 + 2v^2 & 2v^4 + 2v^2 & 2v^4 + 2v^2 & -v^5 - 2v^3 - v & v^6 + 3v^4 + 3v^2 + 1 \\ -v^5 - 2v^3 - v & -v^5 - 2v^3 - v & -v^5 - 2v^3 - v & v^4 + v^2 & -2v^3 \\ 2v^4 + 2v^2 & 2v^4 + 2v^2 & 2v^4 + 2v^2 & -2v^3 & 2v^4 + 2v^2 \\ -v^5 - 2v^3 - v & -v^5 - 2v^3 - v & -2v^3 & v^4 + v^2 & -v^5 - 2v^3 - v \\ -v^5 - 2v^3 - v & -2v^3 & -v^5 - 2v^3 - v & v^4 + v^2 & -v^5 - 2v^3 - v \\ -2v^3 & -v^5 - 2v^3 - v & -v^5 - 2v^3 - v & v^4 + v^2 & -v^5 - 2v^3 - v \\ \\ -v^5 - 2v^3 - v & 2v^4 + 2v^2 & -v^5 - 2v^3 - v & -v^5 - 2v^3 - v & -2v^3 \\ -v^5 - 2v^3 - v & 2v^4 + 2v^2 & -v^5 - 2v^3 - v & -2v^3 & -v^5 - 2v^3 - v \\ -v^5 - 2v^3 - v & 2v^4 + 2v^2 & -2v^3 & -v^5 - 2v^3 - v & -v^5 - 2v^3 - v \\ v^4 + v^2 & -2v^3 & v^4 + v^2 & v^4 + v^2 & v^4 + v^2 \\ -2v^3 & 2v^4 + 2v^2 & -v^5 - 2v^3 - v & -v^5 - 2v^3 - v & -v^5 - 2v^3 - v \\ v^6 + 2v^4 + 2v^2 + 1 & -v^5 - 2v^3 - v & v^4 + v^2 & v^4 + v^2 & v^4 + v^2 \\ -v^5 - 2v^3 - v & v^6 + 3v^4 + 3v^2 + 1 & -v^5 - 2v^3 - v & -v^5 - 2v^3 - v & -v^5 - 2v^3 - v \\ v^4 + v^2 & -v^5 - 2v^3 - v & v^6 + 2v^4 + 2v^2 + 1 & v^4 + v^2 & v^4 + v^2 \\ v^4 + v^2 & -v^5 - 2v^3 - v & v^4 + v^2 & v^6 + 2v^4 + 2v^2 + 1 & v^4 + v^2 \\ v^4 + v^2 & -v^5 - 2v^3 - v & v^4 + v^2 & v^4 + v^2 & v^6 + 2v^4 + 2v^2 + 1 \end{array} \right]$$

## References

1. H. Cohen, A course in computational algebraic number theory, in *Graduate Texts in Mathematics*, vol. 138 (Springer, New York, 1993)
2. J. Davenport, M. Guy, P. Wang,  $p$ -adic reconstruction of rational numbers. *SIGSAM Bull.* 16(2), 2–33 (1982)
3. J. Dixon, Exact solution of linear equations using  $p$ -adic expansions. *Numer. Math.* 40, 137–141 (1982)
4. The GAP Group: GAP — Groups, Algorithms, Programming — A System for Computational Discrete Algebra. Version 4.8.5 (2016). <http://www.gap-system.org>
5. S. Garibaldi,  $E_8$ , the most exceptional group. *Bull. Am. Math. Soc.* 53, 643–671 (2016)
6. M. Geck, Leading coefficients and cellular bases of Hecke algebras. *Proc. Edinb. Math. Soc.* 52, 653–677 (2009)
7. M. Geck, A. Halls, On the Kazhdan–Lusztig cells in type  $E_8$ . *Math. Comp.* 84, 3029–3049 (2015)
8. M. Geck, N. Jacon, Representations of Hecke algebras at roots of unity, in *Algebra and Applications*, vol. 15 (Springer, New York, 2011)
9. M. Geck, J. Müller, James’ conjecture for Hecke algebras of exceptional type I. *J. Algebra* 321, 3274–3298 (2009)
10. M. Geck, G. Pfeiffer, *Characters of Finite Coxeter Groups and Iwahori–Hecke Algebras*. London Mathematical Society Monographs, New Series, vol. 21 (Oxford University Press, Oxford, 2000)
11. M. Geck, G. Hiß, F. Lübeck, G. Malle, G. Pfeiffer, CHEVIE—a system for computing and processing generic character tables for finite groups of Lie type, Weyl groups and Hecke algebras. *Appl. Algebra Eng. Commun. Comput.* 7, 175–210 (1996)
12. The GMP Development Team: GMP — The GNU Multiple Precision Arithmetic Library. Version 6.1.1 (2016). <http://www.gmp-lib.org>
13. G. Hardy, E. Wright, *An Introduction to the Theory of Numbers*, 6th edn. (Oxford University Press, Oxford, 2008)
14. R.B. Howlett,  $W$ -graphs for the irreducible representations of the Hecke algebra of type  $E_8$ . Private communication with J. Michel (December 2003)
15. R.B. Howlett, Y. Yin, Computational construction of irreducible  $W$ -graphs for types  $E_6$  and  $E_7$ . *J. Algebra* 321, 2055–2067 (2009)
16. D.A. Kazhdan, G. Lusztig, Representations of Coxeter groups and Hecke algebras. *Invent. Math.* 53, 165–184 (1979)
17. G. Lusztig, *Hecke Algebras with Unequal Parameters*. CRM Monographs Series, vol. 18 (American Mathematical Society, American Mathematical Society, 2003). Enlarged and updated version at [arXiv:0208154v2](https://arxiv.org/abs/0208154v2)
18. G. Lusztig, Algebraic and geometric methods in representation theory (September 2014). [arxiv:1409.8003](https://arxiv.org/abs/1409.8003)
19. P. Maier, D. Livesey, H.-W. Loidl, P. Trinder, High-performance computer algebra: a Hecke algebra case study, ed. by F. Silva, I. Dutra, V. Santos Costa, in *Euro-Par 2014 Parallel Processing: 20th International Conference*, Porto (August 2014). *Lecture Notes in Computer Science*, vol. 8632, pp. 415–426 (2014)
20. J. Michel, The development version of the CHEVIE package of GAP3. *J. Algebra* 435, 308–336 (2015); see also <http://www.math.rwth-aachen.de/~CHEVIE/>
21. M. Monagan, Maximal quotient rational reconstruction: an almost optimal algorithm for rational reconstruction, in *Proceedings of ISSAC '04* (ACM Press, New York, 2004), pp. 243–249
22. H. Naruse,  $W$ -graphs for the irreducible representations of the Iwahori–Hecke algebras of type  $F_4$  and  $E_6$ . Private communication with M. Geck (January and July, 1998)

23. R.A. Parker, The computer calculation of modular characters (the Meat-Axe), in *Computational Group Theory, Durham (1982)*, ed. by M.D. Atkinson (Academic Press, Cambridge, 1984), pp. 267–274.
24. R.A. Parker, An integral meataxe, in *The Atlas of Finite Groups: Ten Years On, Birmingham (1995)*. London Mathematical Society. Lecture Note Series, vol. 249. (Cambridge University Press, Cambridge, 1998), pp. 215–228
25. D.A. Vogan, The character table of  $E_8$ . *Not. Am. Math. Soc.* **9**, 1022–1034 (2007); see also [http://atlas.math.umd.edu/AIM\\_E8/technicaldetails.html](http://atlas.math.umd.edu/AIM_E8/technicaldetails.html)
26. J. von zur Gathen, J. Gerhard, *Modern Computer Algebra*, 3rd edn. (Cambridge University Press, Cambridge, 2013)

# Tropical Computations in `polymake`



Simon Hampe and Michael Joswig

**Abstract** We give an overview of recently implemented `polymake` features for computations in tropical geometry. The main focus is on explicit examples rather than technical explanations. Our computations employ tropical hypersurfaces, moduli of tropical plane curves, tropical linear spaces and Grassmannians, lines on tropical cubic surfaces as well as intersection rings of matroids.

**Keywords** Mathematical software • Tropical hypersurfaces • Tropical linear spaces

**Subject Classifications** 14-04 (14T05, 14Q99, 52-04)

## 1 Introduction

Many avenues lead to tropical geometry as we know it today. One motivation comes from studying algebraic varieties (over some field with a non-Archimedean valuation) via their piecewise-linear images (under the valuation map). This is useful since many interesting properties are preserved, and they often become algorithmically accessible via tools from polyhedral geometry [39].

Many of these methods are actually implemented, and we start out with giving a brief overview. The standard software to compute tropical varieties with constant coefficients is Jensen's `Gfan` [30]. Its main function is to traverse the dual graph of the Gröbner fan of an ideal and to construct the associated tropical variety as a subfan. The `Singular` [12] library `tropical.lib` [32] by Jensen, Markwig, Markwig and Ren interfaces to `Gfan` and implements extra functionality on top. The most recent version also covers Ren's implementation of tropical varieties

---

S. Hampe • M. Joswig (✉)

Institut für Mathematik, Technische Universität Berlin, Sekretariat MA 6-2, Straß des 17. Juni 136, 10623 Berlin, Germany

e-mail: [hampe@math.tu-berlin.de](mailto:hampe@math.tu-berlin.de); [joswig@math.tu-berlin.de](mailto:joswig@math.tu-berlin.de)

with arbitrary coefficients [43]. Rincón’s program `TropLi` computes tropical linear spaces from matrix input [45], whereas the `Tropical Polyhedra Library` by Allamigeon allows to manipulate tropical polyhedra [1]. The `polymake` system is a comprehensive system for polyhedral geometry and adjacent areas of discrete mathematics. Basic support for computations with tropical hypersurfaces and tropical polytopes goes back as far as version 2.0 from 2004. A much more substantial contribution to `polymake` was the extension `a-tint` for tropical intersection theory [26].

In this paper we report on the recent rewrite of all functionality related to tropical geometry in `polymake`. This largely builds on `a-tint`, which is now a bundled extension and which itself has undergone a massive refactoring. Here we refer to the current version 3.0 of `polymake` from 2016.

Our paper is organized as follows. We start out with the basics of tropical arithmetic and tropical matrix operations. This topic connects tropical geometry to combinatorial optimization [46]. The most basic geometric objects in our investigation are tropical hypersurfaces. These are the vanishing loci of tropical polynomials. Since the latter are equivalent to finite point sets in  $\mathbb{Z}^d$ , equipped with real-valued lifting functions, their study is closely related to regular subdivisions [11]. An interesting new vein in tropical geometry are applications to economics. As an example, we look at arrangements of tropical hypersurfaces as they occur in the product-mix auctions of Baldwin and Klemperer [4, 36, 37]. For a regular subdivision  $\Sigma$  of a point configuration  $P$ , the secondary cone comprises all lifting functions, which induce  $\Sigma$ . That cone forms a stratum in the moduli space of tropical hypersurfaces with support set  $P$ . We exhibit an example computation concerning moduli of tropical plane curves of genus three [8].

Going from hypersurfaces to more general tropical varieties is a major step. Historically, the first explicit computations dealt with the tropicalization of the Grassmannians [49]. The classical Grassmannians are the moduli spaces of linear subspaces in a complex vector space. Their tropical analogues parametrize those tropical linear spaces which arise as tropicalizations. We explore the combinatorics of one tropical linear space. Employing the `polymake` interface to `Singular`, we verify that it is realizable. Tropical linear spaces are interesting in their own right for their connection with matroid theory [35, 48]. We briefly compare several polyhedral structures on the Bergman fan of a matroid.

A famous classical result of Cayley and Salmon states that every smooth cubic surface in  $\mathbb{P}^3$  over an algebraically closed field contains exactly 27 lines. Vigeland studied the question, whether a similar result holds in the tropical setting [53]. A counterexample was given by Maclagan and Sturmfels in [39, Theorem 4.5.8]. Based on a `polymake` computation with `a-tint`’s specialized algorithms for tropical intersection theory, we exhibit a generic tropical cubic surface,  $V$ , which does not match any of the types listed by Vigeland. The surface  $V$  contains 26 isolated lines and three infinite families. Cohomological methods are an indispensable tool in modern algebraic geometry. Tropical intersection theory is a first step towards a similar approach to tropical varieties [2, 40]. Interestingly, tropical intersection theory is also useful in combinatorics, if applied to tropical linear spaces associated

with matroids. Our final example computation shows how the Tutte polynomial of a matroid can be computed from the nested components [27].

In addition to the features presented here, `polymake` also provides functions for computing with Puiseux fractions [34], tropical polytopes [13, 33], general tropical cycles, tropical morphisms and rational functions [26].

## 2 Arithmetic and Linear Algebra

The *tropical semiring*  $\mathbb{T}$  is the set  $\mathbb{R} \cup \{\infty\}$  equipped with  $\oplus := \min$  as the *tropical addition* and  $\odot := +$  as the *tropical multiplication*. Clearly one could also use  $\max$  instead of  $\min$ , but here we will stick to  $\oplus = \min$ . In `polymake` there is a corresponding data type, which allows to compute in  $\mathbb{Q} \cup \{\infty\}$ , e.g., the following.

```
polytope > application "tropical";
tropical > $a = new TropicalNumber<Min>(3);
tropical > $b = new TropicalNumber<Min>(5);
tropical > $c = new TropicalNumber<Min>(8);
tropical > print $a*$c, ", ", $b*$c;
11, 13
tropical > print (($a + $b) * $c);
11
tropical > print $a * (new TropicalNumber<Min>("inf"));
inf
```

**Listing 1** Adding and multiplying tropically

Note that `polymake` is organized into several *applications*, which serve to separate the various functionalities. Most of our computations take place in the application `tropical`, but we will occasionally make use of other types of objects, such as matroids, fans and ideals. One can either switch to an application as shown in Listing 1 or prefix the corresponding types and commands with the name of the application and two colons, such as `matroid::`. We will see examples below. The software system `polymake` is a hybrid design, written in C++ and Perl. In the `polymake` shell the user's commands are interpreted in an enriched dialect of Perl. Note that here the usual operators “+” and “\*” are overloaded, i.e., they are interpreted as tropical matrix addition and tropical matrix multiplication, respectively. It is always necessary to explicitly specify the tropical addition via the template parameter `Min` or `Max`. Mixing expressions with `Min` and `Max` is not defined and results in an error. Templates are not part of standard Perl but rather part of `polymake`'s Perl enrichment.

The type `TropicalNumber` may be used for coefficients of vectors, matrices and polynomials. Matrix addition and multiplication are defined—and interpreted tropically. Here is a basic application of tropical matrix computations: Let  $A = (a_{ij}) \in \mathbb{R}^{d \times d}$  be a square matrix encoding edge lengths on the complete directed graph  $K_d$ . If there are no directed cycles of negative length then there is a well defined shortest path between any two nodes, which may be of infinite length. These

shortest path lengths are given by the so-called *Kleene star*

$$A^* := I \oplus A \oplus (A \odot A) \oplus (A \odot A \odot A) \oplus \dots, \quad (1)$$

where  $I$  is the tropical identity matrix, which has zeros on the diagonal and infinity as a coefficient otherwise. The assumption that there are no directed cycles of negative length makes the above tropical sum of tropical matrix powers stabilize after finitely many steps. The direct evaluation of (1) is precisely the Floyd–Warshall-Algorithm, for computing all shortest paths, known from combinatorial optimization [46, §8.4]. In Listing 2 below we compute the Kleene star of a  $3 \times 3$ -matrix, called  $A$ . Here we verify that  $I \oplus A = I \oplus A \oplus (A \odot A)$ , which implies  $A^* = I \oplus A$ , i.e., all shortest paths are direct.

```
tropical > $A = new Matrix<TropicalNumber<Min>>(
  [[1,2,3],[1,2,4],[1,0,1]]);
tropical > $I = new Matrix<TropicalNumber<Min>>(
  [[0,"inf","inf"],["inf",0,"inf"],["inf","inf",0]]);
tropical > print $I + $A;
0 2 3
1 0 4
1 0 0
tropical > print $I + $A + $A*$A;
0 2 3
1 0 4
1 0 0
```

**Listing 2** Adding and multiplying matrices tropically to obtain the Kleene star  $A^*$

The *tropical determinant* of  $A$  is defined as

$$\begin{aligned} \text{tdet } A &:= \bigoplus_{\sigma \in \text{Sym}(d)} a_{1,\sigma(1)} \odot \dots \odot a_{d,\sigma(d)} \\ &= \min\{a_{1,\sigma(1)} + \dots + a_{d,\sigma(d)} \mid \sigma \in \text{Sym}(d)\}, \end{aligned}$$

where  $\text{Sym}(d)$  denotes the symmetric group of degree  $d$ . This arises from tropicalizing Leibniz' formula for the classical determinant. Notice that evaluating the tropical determinant is tantamount to solving a linear assignment problem from combinatorial optimization. Via the Hungarian method this can be performed in  $O(d^3)$  time; see [46, §17.3]. This is implemented in `polymake` and can be used as shown in Listing 3.

```
tropical > print tdet($A);
4
tropical > print tdet_and_perm($A);
4 <0 1 2>
tropical > print $A->elem(0,0) * $A->elem(1,1) * $A->elem(2,2);
4
```

**Listing 3** Computing a tropical determinant

The user can choose to only compute the value of `tdetA` or also one optimal permutation. In the example from Listing 3 that would be the identity permutation.

### 3 Hypersurfaces

A polyhedral complex is *weighted*, if it is equipped with a function  $\omega$ , assigning integers to its maximal cells. A *tropical cycle* is a weighted pure rational polyhedral complex  $C$ , such that each cell  $C$  of codimension one satisfies a certain *balancing condition*. The interested reader is referred to [39, 41]. Note that the latter only considers *varieties*, which are cycles with strictly positive weights. Tropical hypersurfaces and linear spaces are special cases of tropical varieties and our examples involve only these. We mention tropical cycles since the `polymake` implementation is based on this concept. Moreover, the tropical intersection theory, which we consider in Sect. 6, makes more sense in this general setting.

#### 3.1 Tropical Hypersurfaces and Dual Subdivisions

Let

$$F := \bigoplus_{a \in A} c_a \odot x^{\odot a} \in \mathbb{T}[x_0^\pm, \dots, x_n^\pm] ,$$

be a *tropical (Laurent) polynomial* with *support*  $A \subset \mathbb{Z}^{n+1}$ , i.e., the coefficients  $c_a$  are real numbers and  $A$  is finite. The *tropical hypersurface* of  $F$  is the set

$$T(F) := \{p \in \mathbb{R}^{n+1} \mid \text{the minimum in } F(p) \text{ is attained at least twice}\} .$$

Often we will assume that, for some  $\delta \in \mathbb{N}$ , we have  $a_0 + a_1 + \dots + a_n = \delta$  for all  $a \in A$ . This means that the tropical polynomial  $F$  is *homogeneous* (of degree  $\delta$ ). In this case for each point  $p \in T(F)$  we have  $p + \mathbb{R}\mathbf{1} \subseteq T(F)$ . Thus we usually consider the tropical hypersurface of a homogeneous polynomial as a subset of the quotient  $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$ , which is called the *tropical projective  $n$ -torus*. Note that one could also consider hypersurfaces in *tropical projective space*  $(\mathbb{T}^n \setminus \{(\infty)^n\})/\mathbb{R}\mathbf{1}$ . However, from a computational point of view this incurs several challenges. In `polymake`'s implementation all tropical cycles live in the tropical projective torus and we will thus also adopt this viewpoint mathematically.

The *dual subdivision*  $\Delta(F)$  induced by  $F$  is the collection of sets

$$\Delta_p := \{\arg \min_{a \in A} \{c_a \odot p^{\odot a}\}\} \subseteq A ,$$



where  $p$  ranges over all points in  $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$ . Instead of  $\Delta(F)$  we also write  $T(F)^*$ . Notice that, by definition  $\Delta(F)$  is a set of subsets of  $A$  whose union is  $A$ . More precisely,  $\Delta(F)$  is the combinatorial description of the regular subdivision of the support of  $F$  induced by the coefficients. We say that  $T(F)$  is *smooth* if the dual subdivision  $\Delta(F)$  is unimodular, i.e., every maximal  $\Delta_p$  is the vertex set of a unimodular simplex.

The tropical hypersurface of  $F$  is the codimension one skeleton of the following subdivision of  $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$ : We define two elements  $p, p' \in \mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$  to be equivalent, if  $\Delta_p = \Delta_{p'}$ . The equivalence classes are open polyhedral cones and their closures form a complete polyhedral complex, the *normal complex*  $\mathcal{D}(F)$ .

As an example we consider the cubic polynomial

$$\begin{aligned}
 F := & 12x_0^{\odot 3} \oplus (-131)x_0^{\odot 2}x_1 \oplus (-67)x_0^{\odot 2}x_2 \oplus (-9)x_0^{\odot 2}x_3 \oplus (-131)x_0x_1^{\odot 2} \\
 & \oplus (-129)x_0x_1x_2 \oplus (-131)x_0x_1x_3 \oplus (-116)x_0x_2^{\odot 2} \oplus (-76)x_0x_2x_3 \\
 & \oplus (-24)x_0x_3^{\odot 2} \oplus (-95)x_1^{\odot 3} \oplus (-108)x_1^{\odot 2}x_2 \oplus (-92)x_1^{\odot 2}x_3 \\
 & \oplus (-115)x_1x_2^{\odot 2} \oplus (-117)x_1x_2x_3 \oplus (-83)x_1x_3^{\odot 2} \oplus (-119)x_2^{\odot 3} \\
 & \oplus (-119)x_2^{\odot 2}x_3 \oplus (-82)x_2x_3^{\odot 2} \oplus (-36)x_3^{\odot 3}
 \end{aligned} \tag{2}$$

in four variables, i.e., the tropical hypersurface  $V := T(F)$  is a cubic surface in  $\mathbb{R}^4/\mathbb{R}\mathbf{1}$ . This is constructed in Listing 4 along with the dual subdivision  $V^* = \Delta(F)$ .

```

tropical > $F = toTropicalPolynomial("min(12+3*x0, -131+2*x0+x1,
-67+2*x0+x2, -9+2*x0+x3, -131+x0+2*x1, -129+x0+x1+x2,
-131+x0+x1+x3, -116+x0+2*x2, -76+x0+x2+x3, -24+x0+2*x3, -95+3*x1,
-108+2*x1+x2, -92+2*x1+x3, -115+x1+2*x2, -117+x1+x2+x3,
-83+x1+2*x3, -119+3*x2, -119+2*x2+x3, -82+x2+2*x3, -36+3*x3)");
tropical > $V = new Hypersurface<Min>(POLYNOMIAL=>$F);
tropical > print $V->DEGREE;
3
tropical > print $V->dual_subdivision()->N_MAXIMAL_CELLS;
27

```

**Listing 4** Computing a tropical cubic surface

The computation shows that  $V$  is smooth. Indeed, the support of  $F$  are the lattice points of the scaled 3-dimensional simplex  $3\Delta_3$ , whose normalized volume equals 27. Since there are exactly 27 maximal cells, every single one must have volume 1 (and thus has to be a simplex as well). We will come back to this example later to see that  $V$  has some interesting enumerative properties.

While the entire design of the `polymake` system follows the paradigm of object orientation there is a fundamental difference between an object of type `Matrix`, as in Listing 2, and an object of type `Hypersurface`, as in Listing 4. Matrices form an example of a *small object* class, while tropical hypersurfaces are *big objects*. To understand the difference it is important to know that, by design, `polymake` employs both `Perl` and `C++` as main programming languages. Essentially, the `C++`

code deals with the computations for which speed matters. The small objects belong to container classes which entirely live in the C++ world. On the Perl side this occurs as a mere reference, which is opaque. Calling a member function on a small object from within the `polymake` shell is always deferred to a corresponding C++ function. If new template instantiations occur for the first time this triggers just-in-time compilation. The user experiences this as an occasional short time lag. The newly compiled instantiation is kept in the `polymake` folder in the user's home directory, such that it does not need to be compiled again.

Big objects are very different. Technically, they entirely live in the Perl world. More importantly, the user should think of them as technical realizations of actual mathematical objects, such as a tropical hypersurface. For each big object class there is a certain number of *properties* of which some subset is known at any given point in time. In Listing 4 the variable `$V` is initialized as an object of type `Hypersurface<Min>` with the single property `POLYNOMIAL`, which is clearly enough to define a unique hypersurface. The subsequent command prints a new property, the `DEGREE`, which is automatically derived from the input. The essential idea is that this (and other properties computed on the way) will be kept and stored with the big object. In this way it is avoided to repeat costly computations. More details on `polymake`'s big object concept are found in [21].

### 3.2 *Product-Mix Auctions*

A fascinating connection between tropical geometry and economics was discovered by Baldwin and Klemperer [4]. They showed that the mechanics of many *product-mix auctions* [36, 37] can be modeled using tropical hypersurfaces. Further analysis was given by Tran and Yu [52], whose notation we mostly adopt.

In a product-mix auction, several bidders (“agents”) compete for combinations of several goods. For example, one of the original motivations for this approach was, when the Bank of England wanted to auction off loans of funds during the financial crisis in 2007. These loans could be secured—in various combinations—against either weak or strong collateral. These two types of loans would be the goods in this case.

For our example, we will assume that there are two types of goods, sold in discrete quantities, and only two agents. Every agent now provides a *valuation*  $u^j : A^j \rightarrow \mathbb{R}$ ,  $j = 1, 2$ , where  $A^j \subseteq \mathbb{N}^2$  is the set of bundles of goods the agent is interested in. Negative quantities could also be allowed, thus expressing an interest in selling the corresponding quantity. The valuation measures how valuable a bundle is to the agent. Now, if the auctioneer fixes a price  $p = (p_1, p_2)$ , the agent will naturally be interested in the bundles which maximize her profit. These bundles form the *demand set*

$$D_{u^j}(p) := \arg \max_{a \in A^j} \{u^j(a) - p \cdot a\} ,$$

which depends not only on the price, but also on the choice of the valuation. The *aggregate demand* for the combined valuations  $U = (u^1, u^2)$  is

$$D_U(p) := \{a^1 + a^2 \mid a^j \in D_{u^j}(p)\} \subseteq A^1 + A^2 .$$

Given an actual supply of  $a \in \mathbb{N}^2$ , the auctioneer will be interested in whether there exists a price such that all of the supply can be split between the agents such that every agent obtains a bundle which maximizes their profit, i.e., if there is a  $p$  such that  $a \in D_U(p)$ . In this case, we say that *competitive equilibrium* exists at  $a$ .

In tropical language, every agent defines a hypersurface, corresponding to (the homogenization of) the tropical polynomial

$$F_j := \bigoplus_{a \in A^j} (-u^j(a)) \odot x^{\odot a} .$$

In this formulation, we see that  $-F_j(p)$  is the maximal profit of agent  $j$  at price  $p$ . The tropical hypersurface  $T(F_j)$  is the set of prices where the agent is indifferent between at least two bundles.

Let  $f := f_1 \odot f_2$  be the product of the two polynomials and  $A := A^1 + A^2$  its support. Now, competitive equilibrium exists at a point  $a \in A$ , if and only if for some price vector  $p$  the point  $a$  is contained in the cell  $\Delta_p$  of the dual subdivision  $\Delta(F)$ .

To illustrate, we compute Example 2 from [52]. In Listing 5 we define one hypersurface for each agent and a third one,  $H$ , which is the union.

```
tropical > $H1 = new Hypersurface<Min>(
  MONOMIALS=>[[3,0,0],[2,0,1],[1,0,2],[0,1,2]],
  COEFFICIENTS=>[0,-3,-5,-9]);
tropical > $H2 = new Hypersurface<Min>(
  MONOMIALS=>[[1,0,0],[0,1,0],[0,0,1]],
  COEFFICIENTS=>[0,-1,-1]);
tropical > $H = new Hypersurface<Min>(POLYNOMIAL=>
  $H1->POLYNOMIAL * $H2->POLYNOMIAL);
```

**Listing 5** Constructing a tropical hypersurface from a product of polynomials

We need to homogenize the polynomial, so every bundle of goods has an additional coordinate in front. Our goal is to determine the competitive equilibria. Note that, since  $H$  is a union of two tropical hypersurfaces, the dual subdivision  $H^*$  is the common refinement of the dual subdivisions of the factors.

```
tropical > $ds = $H->dual_subdivision();
tropical > $dehomog = $ds->POINTS->minor(All,~[0,1]);
tropical > $cells = transpose($ds->MAXIMAL_CELLS);
```

**Listing 6** The dual subdivision

The monomials of the tropical polynomial defining  $H$  arise as the vertices of the cells of  $H^*$ . To reinterpret them as bundles we dehomogenize, i.e., we strip the first two coordinates; the first one equals one (since ordinary points are homogenized),

while the second one equals zero (due to the tropical homogenization). For each bundle or monomial we can now print the number of cells containing it.

```
tropical > for (my $i=0; $i<$ds->N_POINTS; ++$i) {
  print $dehomog->row($i), " : ", $cells->row($i)->size(), "\n" }
0 0 : 2
1 0 : 1
1 3 : 2
0 1 : 2
1 1 : 0
0 2 : 2
1 2 : 5
2 2 : 2
0 3 : 1
```

**Listing 7** Checking all bundles

Indeed, we see that every bundle, except for (1, 1), is in at least one cell of the dual subdivision. That is, competitive equilibrium exists precisely at the nine remaining bundles.

## 4 Moduli of Tropical Plane Curves

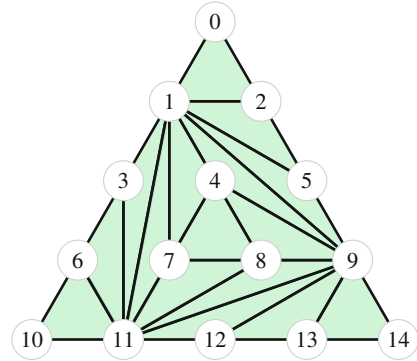
So far we investigated individual examples of tropical hypersurfaces. Now we will look into families which are obtained by varying the coefficients. To this end we start out with a point configuration and a given subdivision  $\Sigma$ . The goal is to determine all possible tropical polynomials  $F$  with  $\Delta(F) = \Sigma$ . Our example computation will deal with a planar point configuration, and hence the tropical hypersurfaces  $T(F)$  will be tropical plane curves. These objects stood at the cradle of tropical geometry; see, in particular, Mikhalkin [40]. Yet the study of their moduli spaces is more recent [8], and this is the direction where we are heading here; see also [5].

The Listing 8 shows `polymake` code to visualize a triangulation of 15 points in the affine hyperplane  $\sum x_i = 4$  in  $\mathbb{R}^3$ ; see Fig. 1. For technical reasons the set of points is converted into a matrix with leading ones. We start out by switching the application.

```
tropical > application "fan";
fan > $points = [ [0,0,4], [1,0,3], [0,1,3], [2,0,2], [1,1,2],
  [0,2,2], [3,0,1], [2,1,1], [1,2,1], [0,3,1], [4,0,0], [3,1,0],
  [2,2,0], [1,3,0], [0,4,0] ];
fan > $triangulation = [[0,1,2], [9,11,12], [9,12,13],
  [9,13,14], [1,2,5], [6,10,11], [3,6,11], [1,5,9], [1,3,11],
  [8,9,11], [1,4,9], [1,7,11], [7,8,11], [4,8,9], [1,4,7], [4,7,8]];
fan > $pointMatrix =
  (ones_vector<Rational>(15)) | (new Matrix<Rational>($points));
fan > $Sigma = new SubdivisionOfPoints(
  POINTS=>$pointMatrix, MAXIMAL_CELLS=>$triangulation);
fan > $Sigma->VISUAL;
```

**Listing 8** Constructing and visualizing a triangulation

**Fig. 1** Unimodular triangulation of  $4\Delta_2$

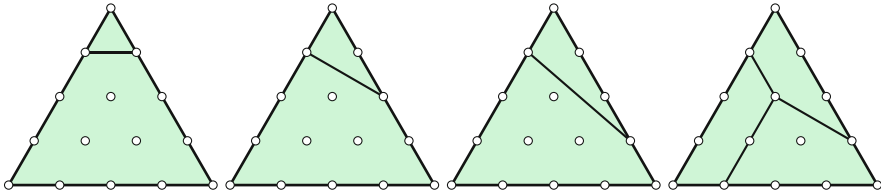


Let  $P$  be a finite set of points in  $\mathbb{R}^d$ , and let  $\Sigma$  be a (*polytopal*) *subdivision* of  $P$ , i.e.,  $\Sigma$  is a polytopal complex whose vertices form a subset of  $P$  and which covers the convex hull of  $P$ . The *secondary cone*  $\text{seccone } \Sigma$  is the topological closure of the set of lifting functions on the set  $P$  which induce  $\Sigma$ . Scaling any lifting function in  $\text{seccone } \Sigma$  by a positive number does not lead outside, and neither does adding two such lifting functions. We infer that  $\text{seccone } \Sigma$ , indeed, is a cone. Each cell of codimension one in  $\Sigma$  yields one linear inequality, and these give an exterior description of the secondary cone. This means that the secondary cone is polyhedral. It is of interest to determine the rays of  $\text{seccone } \Sigma$ . For our example this is accomplished in Listing 9.

```
fan > $sc=$Sigma->secondary_cone();
fan > print $sc->N_RAYS;
12
fan > for (my $i=0; $i<$sc->N_RAYS; ++$i) {
    my $c=new SubdivisionOfPoints(
        POINTS=>$pointMatrix,WEIGHTS=>$sc->RAYS->[$i]);
    print $i, ":", $c->N_MAXIMAL_CELLS, " ";
}
0:2 1:2 2:2 3:2 4:2 5:2 6:2 7:2 8:2 9:3 10:3 11:3
```

**Listing 9** Analyzing the secondary cone

The rays of  $\text{seccone } \Sigma$  induce those *coarsest subdivisions* of the point set  $P$  from which  $\Sigma$  arises as their common refinement. In our example the secondary cone has 12 rays, which come in no particular order. In Listing 9 we list the index of each ray (from 0 to 11) with the number of maximal cells in the corresponding coarsest subdivision. Throughout these numbers are either two or three, from which one can tell right away that the former are 2-splits, while the latter are 3-splits; see Herrmann [28]. The 12 rays come in four orbits, with respect to the symmetry group of  $P$  which fixes  $\Sigma$ . The order of that group is three. There are three orbits of 2-splits, represented by 0, 2 and 8, and one orbit of 3-splits, represented by 11. These are shown in Fig. 2.



**Fig. 2** The coarsest subdivisions of the rays labeled 0, 2, 8 and 11

One specific lifting function on the 15 points which yields our example triangulation,  $\Sigma$ , is shown in Listing 10. We use it as the coefficient vector of a tropical polynomial, and this defines a tropical hypersurface, which we call  $C$ . To verify that this vector, indeed, lies in the relative interior of the secondary cone of  $\Sigma$  we can compute the scalar products with all facet normal vectors.

```
fan > application "tropical";
tropical > $C = new Hypersurface<Min>(MONOMIALS=>$points,
  COEFFICIENTS=>[6, 0,3, 1,-1/3,1, 3,-1/3,-1/3,0, 6,0,1,3,6]);
tropical > $ratCoeff = new Vector<Rational>($C->COEFFICIENTS);
tropical > print $sc->FACETS * $ratCoeff;
4 4 8/3 4 4 4 4 8/3 8/3 4/3 4/3 4/3
tropical > print $sc->LINEAR_SPAN * $ratCoeff;
```

**Listing 10** A tropical plane curve

The fact that all these numbers are positive serves as a certificate for strict containment; the actual values do not matter. For a general subdivision, which is not a triangulation, there are additional linear equations to be checked which describe the linear span. There are no such equations in this case, which is why the last command has no output. Note that prior to computing the scalar products it is necessary to explicitly convert the lifting function into a vector with rational coefficients. This is unavoidable since we want to use the ordinary scalar multiplication here, not the tropical one.

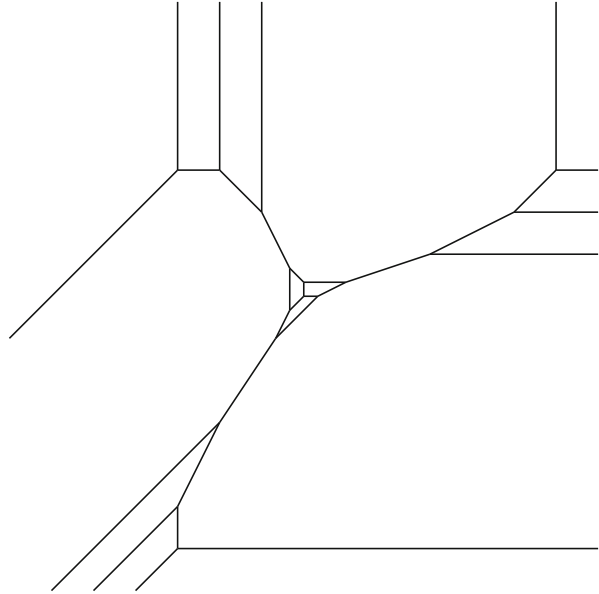
The curve  $C$  defined in Listing 10 and shown in Fig. 3 is a tropical plane quartic of genus three: the Newton polytope is the dilated standard simplex  $4\Delta_2$ , and this has exactly three interior lattice points, all of which are used in the triangulation  $\Sigma$ .

```
tropical > print $C->DEGREE;
4
tropical > print $C->GENUS;
3
```

**Listing 11** Degree and genus of a tropical plane curve

In the sequel we want to locate  $C$  in the moduli space of tropical plane curves of genus three. This will go hand in hand with our initial goal to determine all such curves which fit the initial triangulation  $\Sigma$  of  $4\Delta_2$ . To this end we determine the (lattice) length of each edge of  $C$ , considered as a one-dimensional ordinary polytopal complex. The integer lattice  $\mathbb{Z}^3$  induces sublattices on the line spanned

**Fig. 3** Tropical plane quartic of genus three



by any edge of  $\Sigma$ . Since each edge of  $C$  is dual to an edge of  $\Sigma$  we can measure its length with respect to that sublattice. The result is shown in Listing 12. The 30 edges are labeled from 0 through 29; again they do not come in any particular order.

```
tropical > print labeled($C->CURVE_EDGE_LENGTHS);
0:inf 1:inf 2:1 3:inf 4:1 5:1 6:inf 7:inf 8:1/3 9:1/3 10:1/3
11:1/3 12:2/3 13:1/3 14:inf 15:1 16:inf 17:2/3 18:1 19:1 20:1
21:1/3 22:2/3 23:inf 24:1 25:inf 26:1 27:inf 28:inf 29:inf
tropical > $C->VISUAL(LengthLabels=>"show");
```

**Listing 12** Moduli of a tropical plane curve

As a generic curve of degree four, the curve  $C$  has four edges of infinite length in each of the three coordinate directions. Contracting these infinite edges including the edges of finite length which “lead” to them yields the “essential part” of  $C$ . A picture of this with the remaining edge lengths is given in Fig. 4 (left), the remaining vertices are labeled with the corresponding triangles of  $\Sigma$ . In the essential part of  $C$  we have vertices of degree two or three. Joining edges by omitting those of degree two gives the *combinatorial skeleton* of  $C$ , which is a planar graph with  $2g - 2$  vertices and  $3g - 3$  edges, where  $g$  is the genus, i.e.,  $g = 3$  in our case. The name comes about from its (loose) connection to the Berkovich skeleton of the analytification of a smooth complete curve; see [25]. The joined edges receive the sum of the lengths of the original edges of the curve, and this way we arrive at a metric graph. The lengths of the skeleton edges are the *moduli* of  $C$ . In this case the skeleton is the *honeycomb graph* of genus three, denoted as “(000)” in [8]. Figure 4 (right) shows the skeleton

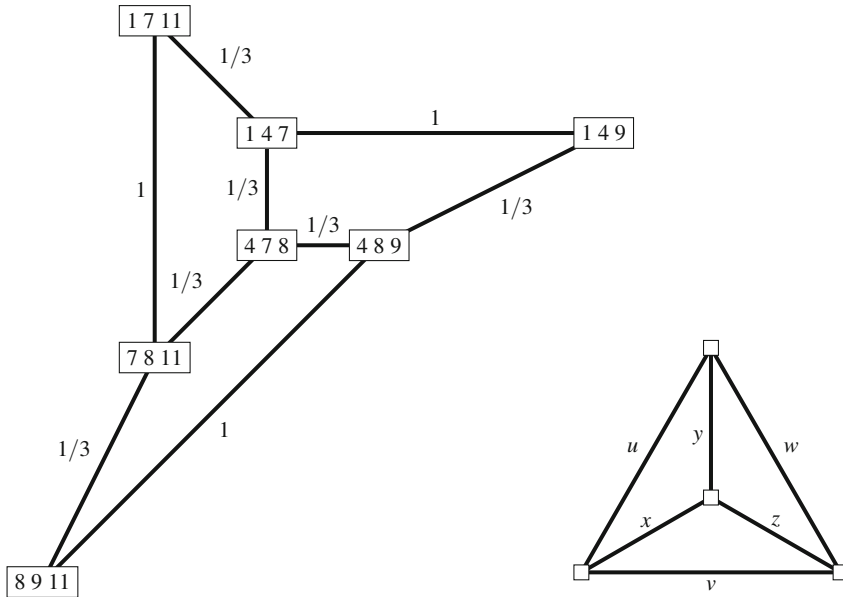


Fig. 4 Essential part of a tropical plane quartic (with edge lengths) and skeleton (with edge labels)

of  $C$  with the edge labels as in [8, Fig. 4]; the moduli are

$$u = v = w = 1 + \frac{1}{3} = \frac{4}{3} \quad \text{and} \quad x = y = z = \frac{1}{3} ,$$

and this agrees with [8, Thm. 5.1].

Without going through all the remaining computations explicitly we now want to sketch how the secondary cone of  $\Sigma$  contributes to the moduli space of tropical plane curves of genus three. Via measuring lattice lengths of edges and properly attributing their contributions to the  $3g-3$  edges of a fixed skeleton curve, seccone  $\Sigma$  is linearly mapped to a cone in  $\mathbb{R}^{3g-3}$ ; see [8, pp. 3ff] for the details. We apply this procedure to the 12 rays computed in Listing 9 and visualized in Fig. 2. Comparing the dual pair of pictures in Figs. 1 and 3 we can see that the three orbits of 2-splits do not contribute anything to essential part of the curve. This means they are mapped to zero in the moduli cone. On the other hand, e.g., the 3-split corresponding to ray 11 corresponds to a curve with moduli  $v = 1, x = z = \frac{1}{3}$  and  $u = w = y = 0$ . The three rays in this orbit (with labels 9, 10 and 11) span a 3-dimensional cone in the moduli space, and this is the image of seccone  $\Sigma$  under the linear map described above. The coefficient vector which defines the curve is, up to scaling, the sum of those three rays. That is, it sits right in the center of that moduli cone.

The moduli space of tropical plane curves of genus 3 is generated from the secondary cones of all 1278 unimodular triangulations (up to symmetry) of the



dilated triangle  $4\Delta_2$ , plus an additional contribution from hyperelliptic curves. The dimension of the entire moduli space is six. Each of the 1278 triangulations contributes a moduli cone whose dimension lies between three and six; see [8, Table 1]. For higher genus it is necessary to consider several polygons; see [10]. Since that moduli space describes *isomorphism classes* of tropical curves it is necessary to take symmetries (of the triangulations and the skeleta) into account. This entails that the global structure is *not* a polyhedral fan but rather a quotient structure called *stacky fan*; see [7].

## 5 Grassmannians, Linear Spaces and Matroids

There are many equivalent ways to define matroids, and we recommend that the interested reader look at [42, 54]. For our purposes the following criterion [15, 22] is the most convenient: Let  $M \subseteq \binom{[n]}{k}$  be a set of subsets of  $[n] = \{1, 2, \dots, n\}$  of size  $k$ . To this collection we can associate a polytope

$$P_M := \text{conv}\{v_B := \sum_{i \in B} e_i \mid B \in M\} ,$$

where  $e_i$  is a standard basis vector. If  $M = \binom{[n]}{k}$  consists of all  $k$ -sets, we call  $P_M =: \Delta(k, n)$  a *hypersimplex*. We say that  $M$  is a *matroid* of rank  $k$  on  $[n]$ , if every edge of  $P_M$  is parallel to  $e_i - e_j$  for some  $i \neq j$ . We call  $P_M$  the *matroid (basis) polytope* of  $M$  and the elements of  $M$  its *bases*. We say that  $M$  is *loopfree*, if every element of  $[n]$  is contained in some basis.

### 5.1 The Tropical Grassmannian

The *tropical Grassmannian*  $\text{TGr}(k, n)$  was first studied in detail by Speyer and Sturmfels [49]. It is the tropicalization of the (complex) Grassmannian  $\text{Gr}(k, n)$ , intersected with the torus. It also parametrizes tropicalizations of uniform linear spaces. This can be seen in the following fashion: Every element of  $\text{TGr}(k, n)$  is a *tropical Plücker vector*  $p \in \mathbb{R}^{\binom{[n]}{k}}$ . Equivalently, we can view it as a height function on the hypersimplex  $\Delta(k, n) \subseteq \mathbb{R}^n$ . The set of all tropical Plücker vectors is the *Dressian*  $\text{Dr}(k, n)$ . In general,  $\text{TGr}(k, n)$  is a proper subset of  $\text{Dr}(k, n)$ . Throughout the following we assume that  $k \leq n$ .

This height function thus induces a regular subdivision of  $\Delta(k, n)$ . The fact that  $p$  is a Plücker vector implies (but is generally not equivalent) that this subdivision is matroidal, i.e., every cell is again a matroid basis polytope. That is, if the set  $\{v_{B_1}, \dots, v_{B_k}\}$  comprises the vertices of a cell, then  $\{B_1, \dots, B_k\}$  is the set of bases of a matroid.

The combinatorics of  $\text{TGr}(3, 6)$  were studied in detail in [29]. The authors compute that there are seven combinatorial types of generic uniform linear spaces, basically encoded in their bounded complexes. As an example, we want to consider a particular vector  $p \in \mathbb{R}^{\binom{6}{3}}$ . We start out with analyzing the combinatorics, and we turn to algebraic computations later.

```
tropical > $Delta=polytope::hypersimplex(3,6);
tropical > $p=new Vector<Int>(
  [0,0,3,1,2,1,0,1,0,2,2,0,3,0,4,1,2,2,0,0]);
tropical > $tlinear=new fan::SubdivisionOfPoints(
  POINTS=>$Delta->VERTICES, WEIGHTS=>$p);
tropical > print $tlinear->TIGHT_SPAN->MAXIMAL_POLYTOPES;
{0 4}
{1 5}
{1 2 3 4}
tropical > $bases = new IncidenceMatrix(
  matroid::uniform_matroid(3,6)->BASES);
tropical > @subdiv_bases = map {
  new Array<Set>(rows($bases->minor($_,All)))
} @{$tlinear->MAXIMAL_CELLS};
tropical > print join(",", map {
  matroid::check_basis_exchange_axiom($_) } @subdiv_bases );
1,1,1,1,1,1
```

**Listing 13** Computing the tight span of a tropical Plücker vector

Now we want to verify that the vector  $p$ , indeed, lies in  $\text{TGr}(3, 6)$ . It is known that for  $(k, n) = (3, 6)$  the tropical Grassmannian and the Dressian agree as sets. However, in general, this does not hold. We employ `polymake`'s interface to `Singular`, and we switch to the application ideal.

```
tropical > application "ideal";
ideal > $I=pluecker_ideal(3,6);
ideal > $pp=new Vector<Int>(5*ones_vector(20)-$p);
ideal > $J=new Ideal(GENERATORS=>
  $I->GROEBNER(ORDER_VECTOR=>$pp)->INITIAL_FORMS);
ideal > print $J->contains_monomial();
0
```

**Listing 14** Computing a generalized initial ideal of the Plücker ideal

A few explanations are in order. The *Plücker ideal*  $I(k, n)$  describes the algebraic relations among the  $k \times k$ -minors of a general  $k \times n$ -matrix. This is an ideal in the polynomial ring over the integers with  $\binom{n}{k}$  indeterminates, each of which encodes a choice of  $k$  columns to specify one such minor. There is a purely combinatorial description of the reverse lex Gröbner basis of  $I(k, n)$ , and this is what is computed by `polymake` directly; see [50, Chapter 3]. This function is also implemented in `Macaulay2` [24]. The tropical variety  $\text{TGr}(k, n)$  arises as a subfan of the Gröbner fan of  $I(k, n)$ . Yet it is common that the interpretation of the vectors in the Gröbner fan refer to maximization, while our choice for regular subdivisions relies on minimization. This entails that we need to swap from the tropical Plücker vector

$p$  to its negative. For technical reasons `Singular` requires such weight vectors to be positive, which is why we consider  $p' = 5 \cdot \mathbf{1} - p$ ; notice that  $\binom{6}{3} = 20$ . The condition for  $p'$  to lie in  $\text{TGr}(3, 6)$  is that the ideal which is generated by the leading forms of  $I(3, 6)$  with respect to  $p'$  does not contain any monomial.

*Remark 5.1* Our choice for  $p$  or rather  $p'$  corresponds to a generic tropical 2-plane in 5-space of type EEFG in the notation of [49].

## 5.2 Tropical Linear Spaces

So far, we have only considered special Plücker vectors—in the sense that the underlying matroid is a *uniform* matroid. There is a general theory of valuated matroids, originally developed by Dress and Wenzel [14].

**Definition 5.2** A *valuated matroid*  $(M, v)$  is a matroid  $M$  together with a function  $v$  from the set of its bases to the real numbers such that the induced regular subdivision on the matroid basis polytope of  $M$  is matroidal, i.e., every cell is a matroid polytope again.

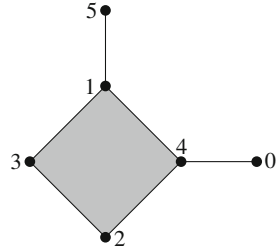
We denote the regular subdivision by  $\Delta(M, v)$ . It has a normal complex  $\mathcal{D}(M, v)$ ; see Sect. 3.1. The *tropical linear space*  $B(M, v)$  associated to  $(M, v)$  is the subcomplex of  $\mathcal{D}(M, v)$  consisting of all faces whose corresponding dual cell if the basis polytope of a loopfree matroid.

If the matroid is the uniform matroid  $U_{k,n}$ , then  $v$  is an element in the Dressian  $\text{Dr}(k, n)$ . If, moreover, it is realizable, then  $v$  lies in the tropical Grassmannian  $\text{TGr}(k, n)$ . Tropical linear spaces play an important role in tropical geometry. They are exactly the tropical varieties of degree 1 [16]. If the valuation  $v \equiv 0$  is trivial, the tropical linear space is a fan, also called the *Bergman fan*  $B(M)$ . These fans are the basic building blocks for *smooth* tropical varieties; see the discussion in Sect. 6. Also, the tropical homology of a Bergman fan encodes the Orlik-Solomon algebra of the matroid [55].

The definition itself already suggests an algorithm to compute a tropical linear space: Find all cells of the matroidal subdivision corresponding to a loopfree matroid and compute its cell in the normal complex. For this, `polymake` uses a variant of Ganter’s algorithm [19, 20], which computes, in fact, the full face lattice of the tropical linear space.

As an example, we want to compute the tropical linear space associated to the (uniform) Plücker vector we considered in Listing 13. We then compute the complex of its bounded faces to confirm that its combinatorics matches the one in Fig. 5. In this example, the bounded faces are identified as those, which do not contain any rays. The latter are stored in the property `FAR_VERTICES`.

**Fig. 5** The combinatorial type of the tropical linear space computed in Listing 15



```
ideal > application "tropical";
tropical > $mat = new matroid::ValuatedMatroid<Min>(
  BASES=>\@{\rows($bases)}, VALUATION_ON_BASES=>$tlinear->WEIGHTS,
  N_ELEMENTS=>6);
tropical > $tl= linear_space($mat);
tropical > @bounded = map {
  grep { ($_ * $tl->FAR_VERTICES)->size == 0 } @{\rows($_)}
} @{$tl->CONES};
tropical > print join(", ",@bounded);
{0},{1},{2},{3},{4},{5},{0 4},{1 4},{2 4},{1 3},{2 3},{1 5},
{1 2 3 4}
```

**Listing 15** Computing a tropical linear space

In the case of Bergman fans there are two more methods to compute the space:

### 5.2.1 Cyclic Fan Structure

Rincón studies the *cyclic Bergman fan* of a matroid [45], which is a refinement of the polyhedral structure we have considered so far. It relies heavily on computing *fundamental circuits* and is particularly fast in the case of matrix matroids, where these computations can be carried out by standard linear algebra methods.

### 5.2.2 Order Complex of the Lattice of Flats

Ardila and Klivans [3] proved that the order complex of the lattice of flats of a matroid  $M$  can be realized as a fan which is supported on  $B(M)$ . That is, we obtain a refinement of  $B(M)$ , such that every ray corresponds to a flat and every cone to a chain of flats. Due to the potentially large number of flats, this is naturally not a very efficient method to compute a tropical linear space. Nevertheless, this is still feasible for small matroids and this particular subdivision is often useful.

In Listing 16, we compute both these fans for the Fano matroid, plus the linear space of its trivial valuation. We see that in this particular case, all these fans are actually the same.

```
tropical > $fano = matroid::fano_matroid();
tropical > $cyclic = matroid_fan<Min>($fano);
tropical > $linear = linear_space(
  matroid::trivial_valuation<Min>($fano));
tropical > $order = matroid_fan_from_flats<Min>($fano);
tropical > print join(",", map {
  $_->N_MAXIMAL_POLYTOPES } ($cyclic,$linear,$order));
21,21,21
```

**Listing 16** Computing a Bergman fan in various ways

## 6 Intersection Theory

The basics for a tropical intersection theory were already laid out in [40]. They are closely related to the *fan displacement rule* by Fulton and Sturmfels [18]. The upshot is that, to intersect two tropical cycles in  $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$ , one of them is shifted in a generic direction until they intersect transversely. The actual intersection product is the limit of the transversal product when shifting back.

This definition is of course hardly suitable for computations. Instead, `a-tint` uses the criterion by Jensen and Yu [31] for computations. This is a local criterion, which states that a point in the set-theoretic intersection of two cycles  $X$  and  $Y$  is also in the intersection product, if and only if the Minkowski sum of the local fans at this point is full-dimensional (a more detailed account of various definitions of the intersection product in the tropical torus can be found in [26]).

The situation becomes more difficult when the ambient variety is not the whole projective torus. As in the algebraic case, even the theory is only fully understood in the case of *smooth varieties*. This is a tropical cycle with positive weights, which is everywhere locally isomorphic to the Bergman fan of a matroid (where an isomorphism of fans is a map in  $\mathrm{GL}_n(\mathbb{Z})$  respecting weights). Every tropical linear space is smooth by definition. For hypersurfaces, this definition of smoothness coincides with the one given in Sect. 3.1.

Since intersection products should be local, it is enough to understand how to compute them in Bergman fans of matroids. In this case, there are two equivalent definitions by Shaw [47] and François and Rau [17]. In general, both are not very computation-friendly. The first is a recursive procedure using projections and pull-backs. The second employs the idea of “cutting out the diagonal”, i.e., writing down rational functions whose consecutive application gives the diagonal of  $B(M) \times B(M)$ . Due to the large dimension and the quadratic increase in the number of cones, this method does not perform well.

However, Shaw’s method does simplify nicely in the case of surfaces (i.e., matroids of rank 3), where the intersection product can be computed in a nice combinatorial manner [47, Section 4]. Using a smoothness detection algorithm, which was implemented in `a-tint` by Dennis Diefenbach, this enabled us to write a procedure which computes intersection products of cycles in smooth surfaces. The

upshot of the detection algorithm is that in the case of surfaces, one can basically do a brute force search over all possible ways to assign flats to the rays of the fan.

As a demonstration of both intersection in the tropical torus and in a smooth surface, we will consider the hypersurface defined by the polynomial (2) in Sect. 3.1. The reader will have to believe (or verify) that this hypersurface contains the standard tropical line with apex  $(0, 0, 0, 0)$  or, in other words, the Bergman fan  $B$  of  $U_{2,4}$ . We want to compute the self-intersection of  $B$  in  $V$ . Also, we will calculate the threefold intersection of  $V$  in the torus. We already know from the Tropical Bernstein Theorem [39, Theorem 4.6.8] that this is the lattice volume of  $3\Delta_3$ .

```
tropical > print intersect(intersect($V,$V),$V)->DEGREE;
27
tropical > $B = matroid_fan<Min>(matroid::uniform_matroid(2,4));
tropical > print intersect_in_smooth_surface($V,$B,$B)->DEGREE;
-1
```

**Listing 17** Self-intersection in a smooth surface

This seems to tie in nicely with the classic fact that a line in a smooth cubic has self-intersection  $-1$ . Hence we want to take a closer look at that situation.

## 6.1 Lines in Tropical Cubics

In algebraic geometry it is a well-known fact that any smooth cubic surface in  $\mathbb{P}^3$  contains exactly 27 lines. It is known that the incidence structure arising from the 27 lines and their 45 points of intersection is the unique generalized quadrangle of order  $(4, 2)$ ; see [51, §3] for details and related constructions. For instance, it is known that any line intersects exactly ten other lines, and for any two disjoint lines there are five lines that intersect both of them. Furthermore, as mentioned before, they all have self-intersection  $-1$ .

In tropical geometry, the situation is much more complicated—or, possibly, much more interesting, depending on your point of view. First of all, we need to establish what a tropical line is:

**Definition 6.1** A tropical line in  $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$  is the tropical linear space of a valuation on  $U_{2,n+1}$ .

Now the peculiarities begin with the fact that a smooth tropical cubic surface may actually contain *families* of tropical lines. A first systematic study of this problem was undertaken by Vigeland in [53]. He provided an example of a secondary cone of  $3\Delta_3$ , such that any general element of that cone defines a tropical cubic which contains exactly 27 lines. He also gave a list of possible *combinatorial types*, which describe how a line can lie in a tropical surface. Our example, which does not occur in Vigeland’s list, was found via a systematic search through the secondary fan of  $3\Delta_3$ . Here we just show our example, while its complete analysis is beyond the scope of the present paper.

Other approaches focused on counting the lines “in the correct manner”, e.g., by showing that certain lines could not be *realized* as tropicalizations of lines in a cubic surface [6, 9]. The paper [44] finds the 27 lines as trees in the boundary of the tropicalization. The interest in this particular problem stems from the fact that it provides a nontrivial, yet computationally feasible testing ground for studying the problem of (relative) *realizability*. It may also serve as an indicator of what possible additional structure should be associated to a tropical variety, i.e., in bold terms, what a *tropical scheme theory* could be; see for example [23, 38].

## 6.2 Vigeland’s Missing Type

We will reconsider the polynomial (2) from Sect. 3.1. We compute the list of lines (and families thereof) in the tropical hypersurface corresponding to  $f$ :

```
tropical > $L = lines_in_cubic($F);
[Output omitted]
tropical > print $L->N_ISOLATED, ", ", $L->N_FAMILIES, "\n";
26, 3
```

**Listing 18** Computing lines in a tropical cubic surface

This demonstrates that there are in fact 26 isolated lines and three different families of lines in the tropical hypersurface  $V$  defined by  $f$ . It can be shown that small random changes to the coefficients do not affect the combinatorics of the lines in the corresponding cubic. In particular, this contradicts Conjecture 1 in [53] that a general cubic contains exactly 27 lines. In Vigeland’s terminology, a general cubic with a fixed dual subdivision corresponds to a dense open subset in the euclidean topology in the secondary cone. A formal proof is beyond the scope of this paper; a different counterexample can be found in [39, Theorem 4.5.8].

Now we want to take a closer look at the families. One can ask `polymake` for a picture of the families using

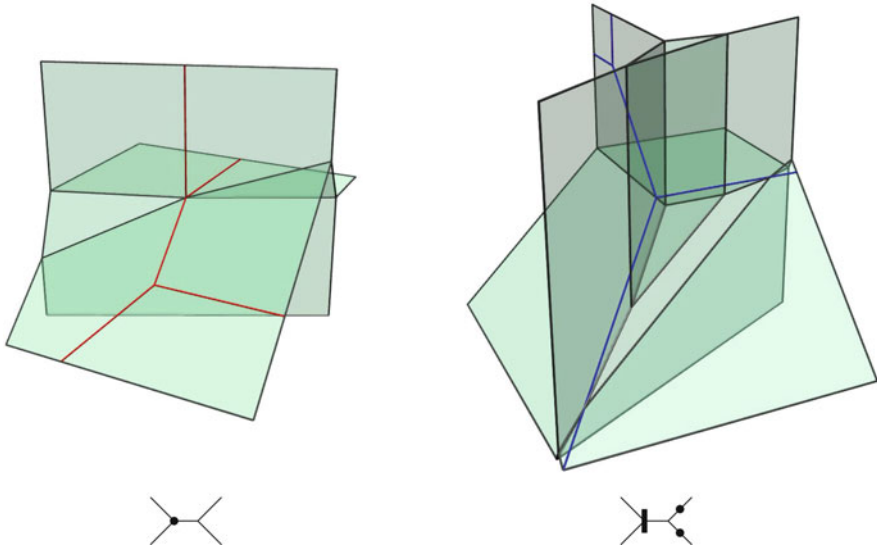
```
tropical > @rp = map { $_->representative() } $L->all_families;
tropical > visualize_in_surface($V, @rp);
```

**Listing 19** Visualizing families in the cubic

This produces a (more or less generic) representative of each family and visualizes them together in the hypersurface.

We see that the first two families have the same combinatorial type: One of the vertices lies on a vertex of the surface, while the other one is allowed to move on a halfline; see Fig. 6(left). These lines can in fact be explained away using an argument from [9]: They are not relatively realizable.

The third and last family, however, is somewhat baffling: One of the vertices lies on an *edge* of the surface, while the other one can move (see Fig. 6, right hand side). In fact, this combinatorial type is missing from the table provided by Vigeland in



**Fig. 6** Local pictures of representatives of families in the cubic and their combinatorial type in Vigeland's notation

[53, Table 2]. There is currently no known obstruction (i.e., non-realizability result) for any of the lines in this family. In fact, we will shortly see that even in terms of intersection combinatorics, any one of them—even the degenerate one—would fit the bill. Note that the vertex on the edge has coordinates  $(0, 0, 0, 0)$ , so if we shrink the bounded edge to length 0, the line is actually the Bergman fan of  $U_{2,4}$ ; see also Listing 17.

### 6.3 Intersection Products in Smooth Surfaces

We want to apply the algorithm we saw in Listing 17 to the case of lines in a cubic surface and check that the 26 lines, together with a representative of the “odd” family we found above, fulfill the following criteria (we denote by  $\cdot_V$  the intersection product in the surface  $V$ ):

1. For any line  $L$ , there are exactly 10 other lines  $L'$ , such that  $L \cdot_V L' = 1$ . Also,  $L \cdot_V L'' = 0$  for all other lines  $L''$ .
2.  $L \cdot_V L = -1$  for all lines  $L$ .

In fact, to save space we will only verify this here for the representative of the last family and leave it to the interested reader to complete the computation for all lines.



```

tropical > $family_line = $rp[2];
tropical > @products = map {
  intersect_in_smooth_surface ($V,$family_line, $_)->DEGREE }
  $L->all_isolated();
tropical > print join(", ",@products);
0,1,0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0,0,1,1,1,1,0,1,1,0
tropical > print intersect_in_smooth_surface (
  $V,$family_line,$family_line)->DEGREE;
-1

```

**Listing 20** Intersecting lines in a smooth surface

One can also check that any two disjoint lines (by which we mean lines with intersection product 0) intersect five other lines and that changing the representative of the family does not affect any of the intersection multiplicities. In fact, we already saw in Listing 17 that the degenerate representative of the last family has self-intersection -1.

## 6.4 Rings of Matroids

Arbitrary intersections of tropical cycles are generally costly to compute, since they involve numerous convex hull computations. It is therefore desirable to make use of additional information whenever possible. One such case is the stable intersection of two tropical linear spaces. Speyer described this in purely combinatorial terms using the underlying valuated matroids [48]. In the case of trivial valuation, i.e., when intersecting two Bergman fans, this corresponds to the operation of *matroid intersection*. In [27], the tropical cycle ring  $\mathbb{M}_n$  generated by Bergman fans of loopfree matroids on  $n$  elements is studied in detail. It is shown that *nested matroids* form a basis for this space and that it is, in fact, the cohomology ring of a toric variety.

**Theorem 6.4** *The ring  $\mathbb{M}_n$  is isomorphic to the cohomology ring  $A^*(X(\text{Perm}_n))$  of the toric variety corresponding to the normal fan of the permutahedron of order  $n$ .*

Using the explicit representation in terms of nested matroids, one can much more easily compute sums and products of cycles lying in this ring. Various properties can also be read off of this data—such as the degree, which is simply the sum of the coefficients. It was also shown in [27], that various matroid invariants, such as the Tutte polynomial, are linear maps on  $\mathbb{M}_n$ .

In the example in Listing 21, we consider the direct sum of two uniform matroids  $U_{1,2}$ . In  $\mathbb{M}_4$ , it is the sum of three nested matroids and we compute both its Tutte polynomial and the corresponding linear combination of the Tutte polynomials of the nested matroids to see that they are equal.

```

tropical > $u = matroid::uniform_matroid(1,2);
tropical > $m = matroid::direct_sum($u,$u);
tropical > print $m->TUTTE_POLYNOMIAL;
x^2 + 2*x*y + y^2
tropical > $r = matroid_ring_cycle<Min>($m);
tropical > print $r->NESTED_COEFFICIENTS;
-1 1 1
tropical > @n = $r->nested_matroids();
tropical > print - $n[0]->TUTTE_POLYNOMIAL
+ $n[1]->TUTTE_POLYNOMIAL + $n[2]->TUTTE_POLYNOMIAL ;
x^2 + 2*x*y + y^2

```

**Listing 21** Computing the Tutte polynomial of a direct sum of matroids

**Acknowledgements** We would like to thank Elizabeth Baldwin and Diane Maclagan for many helpful suggestions on improving this paper.

## References

1. X. Allamigeon, TPLib (Tropical Polyhedra Library) (2013), <http://www.cmap.polytechnique.fr/~allamigeon/software/>
2. L. Allermann, J. Rau, First steps in tropical intersection theory. *Math. Z.* **264**(3), 633–670 (2010)
3. F. Ardila, C.J. Klivans, The Bergman complex of a matroid and phylogenetic trees. *J. Combin. Theory Ser. B* **96**(1), 38–49 (2006)
4. E. Baldwin, P. Klempner, Understanding preferences: “Demand Types”, and the existence of equilibrium with indivisibilities, Nuffield College, Working Paper (2015)
5. A.L. Birkmeyer, A. Gathmann, Realizability of tropical curves in a plane in the non-constant coefficient case (2014, Preprint), arXiv:1412.3035
6. T. Bogart, E. Katz, Obstructions to lifting tropical curves in surfaces in 3-space. *SIAM J. Discret. Math.* **26**(3), 1050–1067 (2012)
7. S. Brannetti, M. Melo, F. Viviani, On the tropical Torelli map. *Adv. Math.* **226**(3), 2546–2586 (2011)
8. S. Brodsky, M. Joswig, R. Morrison, B. Sturmfels, Moduli of tropical plane curves. *Res. Math. Sci.* **2**(4), 1–31 (2015)
9. E. Brugallé, K. Shaw, Obstructions to approximating tropical curves in surfaces via intersection theory. *Canad. J. Math.* **67**(3), 527–572 (2015)
10. W. Castryck, J. Voight, On nondegeneracy of curves. *Algebra Number Theory* **3**(3), 255–281 (2009)
11. J.A. De Loera, J. Rambau, F. Santos, *Triangulations. Structures for Algorithms and Applications*. Algorithms and Computation in Mathematics, vol. 25 (Springer, Berlin, 2010)
12. W. Decker, G.-M. Greuel, G. Pfister, H. Schönemann, Singular 4-1-0 – A computer algebra system for polynomial computations (2016), <http://www.singular.uni-kl.de>
13. M. Develin, B. Sturmfels, Tropical convexity. *Doc. Math.* **9**, 1–27 (electronic) (2004). Correction: *ibid.*, pp. 205–206
14. A.W.M. Dress, W. Wenzel, Valuated matroids. *Adv. Math.* **93**(2), 214–250 (1992)
15. J. Edmonds, Submodular functions, matroids, and certain polyhedra, in *Combinatorial Structures and Their Applications (Proc. Calgary Internat. Conf., Calgary, Alta., 1969)* (Gordon and Breach, New York, 1970), pp. 69–87

16. A. Fink, Tropical cycles and chow polytopes. *Beiträge zur Algebra und Geometrie/Contributions to Algebra and Geometry* **54**(1), 13–40 (2013)
17. G. François, J. Rau, The diagonal of tropical matroid varieties and cycle intersections. *Collect. Math.* **64**(2), 185–210 (2013)
18. W. Fulton, B. Sturmfels, Intersection theory on toric varieties. *Topology* **36**(2), 335–353 (1997)
19. B. Ganter, *Algorithmen zur formalen Begriffsanalyse*, ed. by B. Ganter, R. Wille, K.E. Wolff. *Beiträge zur Begriffsanalyse* (Bibliographisches Institut, Mannheim, 1987), pp. 241–254
20. B. Ganter, K. Reuter, Finding all closed sets: a general approach. *Order* **8**(3), 283–290 (1991)
21. E. Gawrilow, M. Joswig, Flexible object hierarchies in `polymake`, in *Proceedings of the 2nd International Congress of Mathematical Software*, Castro Urdiales, Spanien, 1–3 Sept 2006, ed. by A. Igelesias, N. Takayama (2006), pp. 219–221
22. I.M. Gel'fand, M. Goresky, R.D. MacPherson, V.V. Serganova, Combinatorial geometries, convex polyhedra, and Schubert cells. *Adv. Math.* **63**(3), 301–316 (1987)
23. J. Giansiracusa, N. Giansiracusa, Equations of tropical varieties. *Duke Math. J.* **165**(18), 3379–3433 (2016)
24. D.R. Grayson, M.E. Stillman, *Macaulay2*, a software system for research in algebraic geometry (2017), available at <http://www.math.uiuc.edu/Macaulay2/>
25. W. Gubler, J. Rabinoff, A. Werner, Tropical skeletons (2015, preprint), arXiv:1508.01179
26. S. Hampe, a-tint: a polymake extension for algorithmic tropical intersection theory. *Eur. J. Combin.* **36**, 579–607 (2014)
27. S. Hampe, The intersection ring of matroids. *J. Comb. Theory Ser. B* **122**, 578–614 (2017)
28. S. Herrmann, On the facets of the secondary polytope. *J. Comb. Theory Ser. A* **118**(2), 425–447 (2011)
29. S. Herrmann, A. Jensen, M. Joswig, B. Sturmfels, How to draw tropical planes. *Electron. J. Comb.* **16**(2), Special volume in honor of Anders Björner, Research Paper 6, 26 (2009)
30. A.N. Jensen, Gfan, a software system for Gröbner fans and tropical varieties, version 0.6, available at <http://home.imf.au.dk/jensen/software/gfan/gfan.html> (2017)
31. A. Jensen, J. Yu, Stable intersections of tropical varieties. *J. Algebraic Comb.* **43**(1), 101–128 (2016)
32. A.N. Jensen, H. Markwig, T. Markwig, Y. Ren, *tropical.lib*. a Singular 4-1-0 library for computations in tropical geometry, Tech. report (2016)
33. M. Joswig, Tropical convex hull computations, in *Tropical and Idempotent Mathematics*, ed. by G.L. Litvinov, S.N. Sergeev. *Contemporary Mathematics*, vol. 495 (American Mathematical Society, Providence, RI, 2009)
34. M. Joswig, G. Loho, B. Lorenz, B. Schröter, Linear programs and convex hulls over fields of Puiseux fractions, in *Proceedings of MACIS 2015, LNCS 9582*, Berlin, 11–13 Nov 2015 (2016), pp. 429–445
35. M.M. Kapranov, *Chow Quotients of Grassmannians. I*, ed. by I.M. Gel'fand Seminar. *Advances in Soviet Mathematics*, vol. 16 (American Mathematical Society, Providence, RI, 1993), pp. 29–110
36. P. Klemperer, A new auction for substitutes: central bank liquidity auctions, the U.S. TARP, and variable product-mix auctions, University of Oxford, Working Paper (2008)
37. P. Klemperer, The product-mix auction: a new auction design for differentiated goods. *J. Eur. Econ. Assoc.* **8**, 526–536 (2010)
38. D. Maclagan, F. Rincón, Tropical schemes, tropical cycles, and valuated matroids (2014, preprint), arXiv:1401.4654
39. D. Maclagan, B. Sturmfels, *Introduction to Tropical Geometry*. *Graduate Studies in Mathematics*, vol. 161 (American Mathematical Society, Providence, RI, 2015)
40. G. Mikhalkin, Enumerative tropical algebraic geometry in  $\mathbb{R}^2$ . *J. Am. Math. Soc.* **18**(2), 313–377 (2005)
41. G. Mikhalkin, J. Rau, Tropical geometry, work in progress, available at <https://www.math.uni-tuebingen.de/user/jora/downloads/main.pdf> (2015)
42. J. Oxley, *Matroid Theory*. *Oxford Graduate Texts in Mathematics*, vol. 21, 2nd edn. (Oxford University Press, Oxford, 2011)

43. Y. Ren, Computing tropical varieties over fields with valuation using classical standard basis techniques. *ACM Commun. Comput. Algebra* **49**(4), 127–129 (2015)
44. Q. Ren, K. Shaw, B. Sturmfels, Tropicalization of del Pezzo surfaces. *Adv. Math.* **300**, 156–189 (2016)
45. F. Rincón, Computing tropical linear spaces. *J. Symb. Comput.* **51**, 86–98 (2013)
46. A. Schrijver, *Combinatorial Optimization. Polyhedra and Efficiency. Vol. A. Algorithms and Combinatorics*, vol. 24 (Springer, Berlin, 2003). Paths, flows, matchings, Chapters 1–38
47. K.M. Shaw, A tropical intersection product in matroidal fans. *SIAM J. Discret. Math.* **27**(1), 459–491 (2013)
48. D.E. Speyer, Tropical linear spaces. *SIAM J. Discret. Math.* **22**(4), 1527–1558 (2008)
49. D. Speyer, B. Sturmfels, The tropical Grassmannian. *Adv. Geom.* **4**(3), 389–411 (2004)
50. B. Sturmfels, *Algorithms in Invariant Theory*. Texts and Monographs in Symbolic Computation, 2nd edn. (Springer, New York, 2008)
51. J.A. Thas, H. Van Maldeghem, Embeddings of small generalized polygons. *Finite Fields Appl.* **12**(4), 565–594 (2006)
52. N.M. Tran, J. Yu, Product-mix auctions and tropical geometry (2015, preprint), arXiv:1505.05737
53. M.D. Vigeland, Tropical lines on smooth tropical surfaces (2007, preprint), arXiv:0708.3847
54. N. White (ed.), *Theory of Matroids*. Encyclopedia of Mathematics and Its Applications, vol. 26 (Cambridge University Press, Cambridge, 1986)
55. I. Zharkov, The Orlik-Solomon algebra and the Bergman fan of a matroid. *J. Gökova Geom. Topol.* **GGT 7**, 25–31 (2013)

# Focal Schemes to Families of Secant Spaces to Canonical Curves



Michael Hoff

**Abstract** For a general canonically embedded curve  $C$  of genus  $g \geq 5$ , let  $d \leq g - 1$  be an integer such that the Brill–Noether number  $\rho(g, d, 1) = g - 2(g - d + 1) \geq 1$ . We study the family of  $d$ -secant  $\mathbf{P}^{d-2}$ 's to  $C$  induced by the smooth locus of the Brill–Noether locus  $W_d^1(C)$ . Using the theory of foci and a structure theorem for the rank one locus of special 1-generic matrices by Eisenbud and Harris, we prove a Torelli-type theorem for general curves by reconstructing the curve from its Brill–Noether loci  $W_d^1(C)$  of dimension at least 1.

**Keywords** Focal scheme • Brill–Noether locus • Torelli-type theorem

**Subject Classifications** 14H51, 14M12, 14C34

## 1 Introduction and Motivation

For a general canonically embedded curve  $C$  of genus  $g \geq 5$  over  $\mathbf{C}$ , we study the local structure of the Brill–Noether locus  $W_d^1(C)$  for an integer  $\lceil \frac{g+3}{2} \rceil \leq d \leq g - 1$ . Our main object of interest is the focal scheme associated to the family of  $d$ -secant  $\mathbf{P}^{d-2}$ 's to  $C$ . The focal scheme arises in a natural way as the degeneracy locus of a map of locally free sheaves associated to a family of secant spaces to a curve. In other words, the focal scheme (or the scheme of first-order foci) consists of all points where a secant intersects its infinitesimal first-order deformation.

In [5] and [6], Ciliberto and Sernesi studied the geometry of the focal scheme associated to the family of  $(g - 1)$ -secant  $\mathbf{P}^{g-3}$ 's induced by the singular locus  $W_{g-1}^1(C)$  of the theta divisor, and they gave a conceptual new proof of Torelli's theorem. Using higher-order focal schemes for general canonical curves of genus  $g = 2m + 1$ , they showed in [7] that the family of  $(m + 2)$ -secants induced by

---

M. Hoff (✉)

Universität des Saarlandes, Campus E2 4, 66123 Saarbrücken, Germany  
e-mail: [hahn@math.uni-sb.de](mailto:hahn@math.uni-sb.de)

$W_{m+2}^1(C)$  also determines the curve. These are the extremal cases, that is, the degree  $d$  is maximal or minimal with respect to the genus  $g$  (in symbols  $d = g - 1$  or  $d = \frac{g+3}{2}$  and  $g$  odd). The article [2] of Bajravani can be seen as a first extension of the previous results to another Brill–Noether locus ( $g = 8$  and  $d = 6 = \lceil \frac{g+3}{2} \rceil$ ).

Combining methods of [6, 7] and [8], we will give a unified proof which shows that the canonical curve is contained in the focal schemes parametrised by the smooth locus of any  $W_d^1(C)$  if  $d \leq g - 1$  and  $\rho(g, d, 1) = g - 2(g - d + 1) \geq 1$ . Moreover, we have the following Torelli-type theorem (see also Corollary 3.13).

**Theorem 1.1** *A general canonically embedded curve of genus  $g$  can be reconstructed from its Brill–Noether locus  $W_d^1(C)$  if  $\lceil \frac{g+3}{2} \rceil \leq d \leq g - 1$ .*

In [12], Pirola and Teixidor i Bigas proved a generic Torelli-type theorem for  $W_d^r(C)$  if  $\rho(g, d, r) \geq 2$ , or  $\rho(g, d, r) = 1$  and  $r = 1$ . Whereas they used the global geometry of the Brill–Noether locus to recover the curve, our theorem is based on the local structure around a smooth point of  $W_d^1(C) \subset W_d(C)$ . Only first-order deformations are needed.

Our proof follows [7]. We show that the first-order focal map is in general 1-generic and apply a result of Eisenbud and Harris [10] in order to describe the rank one locus of the focal matrix. Two cases are possible. The rank one locus of the focal matrix consists either of the support of a divisor  $D$  of degree  $d$  corresponding to a line bundle  $\mathcal{O}_C(D) \in W_d^1(C)$  or of a rational normal curve. Even if we are not able to decide which case should occur on a general curve (see Sect. 4 for a discussion), we finish our proof by studying focal schemes to a family of rational normal curves induced by the first-order focal map.

In Sect. 2, we recall the definition of focal schemes as well as general facts and known results about focal schemes. Section 3 is devoted to prove the generalisation of the main theorem of [7] to arbitrary positive dimensional Brill–Noether loci.

## 2 The Theory of Foci

We recall the definition as well as the construction of the family of  $d$ -secant  $\mathbf{P}^{d-2}$ 's induced by an open dense subset of  $C_d^1$ . Afterwards we introduce the characteristic or focal map and define the scheme of first-order foci of rank  $k$  associated to the above family. We give a slightly generalised definition of the scheme of first- and second-order foci compared to [6]. In Sect. 2.2, we recall the basic properties of the scheme of first-order foci. Our approach follows [8].

### 2.1 Definition of the Scheme of First-Order Foci

Let  $C$  be a Brill–Noether general canonically embedded curve of genus  $g \geq 5$ , and let  $d \leq g - 1$  be an integer such that the Brill–Noether number  $\rho := \rho(g, d, 1) =$

$g - 2(g - d + 1) \geq 1$ . Let  $C_d^1$  be the variety parametrising effective divisors of degree  $d$  on  $C$  moving in a linear system of dimension at least 1 (see [1, IV, §1]). Let  $\Sigma \subset W_d^1(C)$  be the smooth locus of  $W_d^1(C)$ . Furthermore, let  $\alpha_d : C_d^1 \rightarrow W_d^1(C)$  be the Abel-Jacobi map (see [1, I, §3]) and let  $S = \alpha_d^{-1}(\Sigma)$ . Then  $\alpha : S \rightarrow \Sigma$  is a  $\mathbf{P}^1$ -bundle, and in particular  $S$  is smooth of pure dimension  $\rho + 1$ . For every  $s \in S$ , we denote by  $D_s$  the divisor of degree  $d$  on  $C$  defined by  $s$  and  $\Lambda_s = \overline{D_s} \subset \mathbf{P}^{g-1}$  its linear span, which is a  $d$ -secant  $\mathbf{P}^{d-2}$  to  $C$ . We get a  $(\rho + 1)$ -dimensional family of  $d$ -secant  $\mathbf{P}^{d-2}$ 's parametrised by  $S$ :

$$\begin{array}{ccc} \underline{\Lambda} \subset S \times \mathbf{P}^{g-1} & \xrightarrow{q} & \mathbf{P}^{g-1} \\ \downarrow p & & \\ S & & \end{array}$$

We denote by  $f : \underline{\Lambda} \rightarrow \mathbf{P}^{g-1}$  the induced map.

**Construction 1 (Of the Family  $\underline{\Lambda}$ )** Let  $\mathbf{D}_d \subset C_d \times C$  be the universal divisor of degree  $d$  and let  $\mathbf{D}_S \subset S \times C$  be its restriction to  $S \times C$ . We denote by  $\pi : S \times C \rightarrow S$  the projection. We consider the short exact sequence

$$0 \rightarrow \mathcal{O}_{S \times C} \rightarrow \mathcal{O}_{S \times C}(\mathbf{D}_S) \rightarrow \mathcal{O}_{\mathbf{D}_S}(\mathbf{D}_S) \rightarrow 0.$$

By Grauert's Theorem, the higher direct image  $R^1 \pi_*(\mathcal{O}_{\mathbf{D}_S}(\mathbf{D}_S)) = 0$  vanishes and we get a map of locally free sheaves on  $S$

$$R^1 \pi_*(\mathcal{O}_{S \times C}) \rightarrow R^1 \pi_*(\mathcal{O}_{S \times C}(\mathbf{D}_S)) \rightarrow 0$$

whose kernel is a locally free sheaf  $\mathcal{F} \subset R^1 \pi_*(\mathcal{O}_{S \times C}) \cong \mathcal{O}_S \otimes H^1(C, \mathcal{O}_C)$  of rank  $d - 1 = g - (g - d + 1)$ . The family  $\underline{\Lambda}$  is the associated projective bundle

$$\underline{\Lambda} = \mathbf{P}(\mathcal{F}) \subset S \times \mathbf{P}^{g-1}.$$

*Remark 2.1* We can also construct the family  $\underline{\Lambda}$  from the Brill-Noether locus  $W_d(C)$  and its singular locus  $W_d^1(C)$ . At a singular point  $L \in W_d^1(C) \setminus W_d^2(C)$ , the projectivised tangent cone to  $W_d(C)$  at  $L$  in the canonical space  $\mathbf{P}^{g-1}$  coincides with the scroll

$$X_L = \bigcup_{D \in |L|} \overline{D}$$

swept out by the pencil  $g_d^1 = |L|$ . Hence, the ruling of  $X_L$  is the one-dimensional family of secants induced by  $|L|$ . Varying the point  $L$  yields the family  $\underline{\Lambda}$ . See also [6, Theorem 1.2]. We conclude that the family  $\underline{\Lambda}$  is determined by  $W_d(C)$  and its singular locus  $W_d^1(C)$ .

In order to define the first-order focal map of the family  $\underline{A}$ , we make a short digression. We consider a flat family  $F$  of closed subschemes of a projective scheme  $X$  over a base  $B$ , that is,

$$\begin{array}{ccc} F \subset B \times X & \xrightarrow{\pi_2} & X \\ \downarrow \pi_1 & & \\ B & & \end{array}$$

Let  $T(\pi_1)|_F := \pi_1^*(T_B)|_F$  be the tangent sheaf along the fibers of  $\pi_2$  restricted to the family  $F$  and let  $\mathcal{N}_{F/B \times X}$  be the normal sheaf of  $F \subset B \times X$ . There is a map

$$\psi : T(\pi_1)|_F \rightarrow \mathcal{N}_{F/B \times X}$$

called the *global characteristic map* of the family  $F$  which is defined by the following exact and commutative diagram:

$$\begin{array}{ccccccc} & & & 0 & & & \\ & & & \downarrow & & & \\ & & & T(\pi_1)|_F & \xrightarrow{\psi} & \mathcal{N}_{F/B \times X} & \\ & & & \downarrow & & \parallel & \\ 0 & \longrightarrow & T_F & \longrightarrow & T_{B \times X}|_F & \longrightarrow & \mathcal{N}_{F/B \times X} \longrightarrow 0 \\ & & \downarrow d(\pi_2|_F) & & \downarrow & & \\ & & (\pi_2|_F)^*(T_X) & \xlongequal{\quad} & \pi_2^*(T_X)|_F & & \end{array}$$

For every  $b \in B$  the homomorphism  $\psi$  induces a homomorphism

$$\psi_b : T_{B,b} \otimes \mathcal{O}_{\pi_1^{-1}(b)} \rightarrow \mathcal{N}_{\pi_1^{-1}(b)/X}$$

called the *(local) characteristic map* of the family  $F$  at a point  $b$ . Since  $F$  is a flat family, we get a classifying morphism

$$\varphi : B \rightarrow \text{Hilb}_Y$$

by the universal property of the Hilbert scheme  $\text{Hilb}_Y$ . The linear map induced by the characteristic map

$$H^0(\psi_b) : T_{B,b} \rightarrow H^0(\mathcal{N}_{\pi_1^{-1}(b)/X})$$



is the differential  $d\varphi_b$  at the point  $b$  (see also [11, p. 198 f]). Assuming that  $B, Y$  and the family  $F$  are smooth, all sheaves in the above diagram are locally free and by diagram-chasing, it follows that

$$\ker(d(\pi_2|_F)) = \ker(\psi) \quad \text{and} \quad \dim(\pi_2(F)) = \dim(F) - \text{rk}(\ker(\psi)).$$

We come back to the smooth family  $\underline{\Delta}$  and fix some notation for the rest of the article. Let  $\mathcal{N} := \mathcal{N}_{\underline{\Delta}/S \times \mathbf{P}^{g-1}}$  be the normal bundle of  $\underline{\Delta}$  in  $S \times \mathbf{P}^{g-1}$  and let  $T(p)|_{\underline{\Delta}} := p^*(T_S)|_{\underline{\Delta}}$  be the restriction of the tangent bundle along the fibers of  $q$  to  $\underline{\Delta}$ . Let

$$\chi : T(p)|_{\underline{\Delta}} \rightarrow \mathcal{N}$$

be the global characteristic map defined as above. For every  $s \in S$  the homomorphism  $\chi$  induces a homomorphism

$$\chi_s : T_{S,s} \otimes \mathcal{O}_{\Lambda_s} \rightarrow \mathcal{N}_{\Lambda_s/\mathbf{P}^{g-1}}$$

also called the *characteristic map* or *first-order focal map* of the family  $\underline{\Delta}$  at a point  $s$ .

*Remark 2.2* Fix an  $s \in S$ . We have  $\Lambda_s = \mathbf{P}(U)$ , where  $U \subset V = H^1(C, \mathcal{O}_C)$  is a vector subspace of dimension  $d - 1$ . The normal bundle of  $\Lambda_s$  in  $\mathbf{P}^{g-1}$  is given by

$$\mathcal{N}_{\Lambda_s/\mathbf{P}^{g-1}} = V/U \otimes \mathcal{O}_{\Lambda_s}(1)$$

and

$$H^0(\Lambda_s, \mathcal{N}_{\Lambda_s/\mathbf{P}^{g-1}}) = \text{Hom}(U, V/U).$$

The characteristic map is of the form

$$\chi_s : T_{S,s} \otimes \mathcal{O}_{\Lambda_s} \rightarrow V/U \otimes \mathcal{O}_{\Lambda_s}(1).$$

Hence, it is given by a matrix of linear form on  $\Lambda_s$ .

We define the first- and the second-order foci (of rank  $k$ ) of a family  $\underline{\Delta}$ .

**Definition 2.3**

(a) Let  $V(\chi)_k$  be the closed subscheme of  $\underline{\Delta}$  defined by

$$V(\chi)_k = \{p \in \underline{\Delta} \mid \text{rk}(\chi(p)) \leq k\}.$$

Then,  $V(\chi)_k$  is the *scheme of first-order foci of rank  $k$*  and the fiber of  $V(\chi)_k$  over a point  $s \in S$

$$(V(\chi)_k)_s = V(\chi_s)_k \subset \Lambda_s$$

is the *scheme of first-order foci of rank  $k$  at  $s$* .

(b) Assume that  $V(\chi)_k$  induces a family of rational normal curves  $\underline{\Gamma}$ , that is, for a general  $s \in S$  the fiber  $\Gamma_s = V(\chi_s)_k$  is a rational normal curve. Let  $\psi$  be the global characteristic map of  $\underline{\Gamma}$ . We call the first-order foci of rank  $k$  of the family  $\underline{\Gamma}$ , that is,

$$V(\psi)_k = \{p \in \underline{\Gamma} \mid \text{rk}(\psi(p)) \leq k\}$$

the *second-order foci of rank  $k$*  of the family  $\underline{\Delta}$ .

*Remark 2.4* Our definition of scheme of first-order foci is a slight generalisation of the definition given in [5, 6]. Note that if

$$k = \min\{\text{rk}(T(p)|_{\underline{\Delta}}), \text{rk}(\mathcal{N})\} - 1 = \min\{\text{codim}S, \text{codim}_{S \times \mathbf{P}^{g-1}}(\underline{\Delta})\} - 1,$$

we get the classical definition of first-order foci. Furthermore, our definition of the second-order foci of rank  $k$  is inspired by the definition of higher-order foci of [8].

*Remark 2.5*

- (a) The equality  $V(\chi)_s = V(\chi_s)$  is shown in [4, Proposition 14].
- (b) If  $\chi$  has maximal rank, that is, if  $\chi$  is either injective or has torsion cokernel, then  $V(\chi)_k$  is a proper closed subscheme of  $\underline{\Delta}$  for  $k \leq \min\{\text{rk}(T(p)|_{\underline{\Delta}}), \text{rk}(\mathcal{N})\} - 1$ .
- (c) In Sect. 3 we study the scheme of first-order foci of rank 1 of the family  $\underline{\Delta}$ .

**2.2 Properties of the Scheme of First-Order Foci**

We assume in this section that  $C$  is a Brill–Noether general curve. The following proposition is proven in [6] which can be easily generalised to the case of divisors of degree  $d < g - 1$ .

**Proposition 2.6** *For  $s \in S$ , we have*

$$D_s \subset V(\chi_s)_1.$$

*In particular, the canonical curve  $C$  is contained in the scheme of first-order foci.*

*Proof* Let  $p \in \text{Supp}(D_s)$ . Then there exists a codimension 1 family of effective divisors and hence  $d$ -secants containing the point  $p$ . Therefore, there is a codimension 1 subspace  $T \subset T_{S,s}$  such that the map  $\chi_s(p)|_T$  is zero. We conclude that the focal map  $\chi_s$  has rank at most 1 in points of  $\text{Supp}(D_s)$ .  $\square$

An important step in the proof of our main theorem is to show that the first-order focal map  $\chi_s$  is 1-generic. The general definition of 1-genericity can be found in [9]. In our case, a reformulation of the definition is the following.

**Proposition 2.7** *The matrix  $\chi_s$  is 1-generic if and only if for each nonzero element  $v \in T_{S,s}$ , the homomorphism*

$$H^0(\chi_s)(v) \in \text{Hom}(U, V/U)$$

is surjective.

We recall what is known about the 1-genericity of the matrix  $\chi_s$ .

**Proposition 2.8** ([6, Theorem 2.5], [7, Theorem 2], [2]) *Let  $s \in S$  be a general point.*

- (a) *If  $D_s$  is a divisor of degree  $g - 1$  cut on  $C$  by  $\Lambda_s$ , then the matrix  $\chi_s$  is 1-generic (equivalently,  $V(\chi_s)_1$  is a rational normal curve) if and only if the pencil  $|D_s|$  is base point free.*
- (b) *If  $\rho = \rho(g, d, 1) = 1$ , then the matrix  $\chi_s$  is 1-generic (equivalently,  $V(\chi_s)_1$  is a rational normal curve).*
- (c) *If  $g = 8$  and  $d = 6$ , then the matrix  $\chi_s$  is 1-generic.*

*Remark 2.9* ([13, p. 253]) Another fact related to the 1-genericity of  $\chi_s$  is the following: Let

$$\begin{array}{c} \underline{\Delta}_\varepsilon \subset \text{Spec}(\mathbb{C}[\varepsilon]) \times \mathbb{P}^{g-1} \\ \downarrow \\ \text{Spec}(\mathbb{C}[\varepsilon]) \end{array}$$

be the first order deformation of  $\Lambda_s$  defined by  $H^0(\chi_s)(v)$  for a vector  $v \in T_{S,s}$ . Then,  $H^0(\chi_s)(v)$  is surjective if and only if  $q(\underline{\Delta}_\varepsilon) \subset \mathbb{P}^{g-1}$  is not contained in a hyperplane. Furthermore, the definition of the first-order foci at a point  $s \in S$  depends only on the geometry of the family  $\underline{\Delta}$  in a neighbourhood of  $s$ . A point in  $V(\chi_s)_k$  is a point where the fiber  $\Lambda_s$  intersects a codimension  $k$  family of its infinitesimally near ones.

### 3 Proof of the Main Theorem

The strategy of the proof is the same as in [7]. We assume that the canonically embedded curve  $C$  is a Brill–Noether general curve. Recall that  $g$  and  $d$  are chosen

such that the Brill–Noether number  $\rho := \rho(g, d, 1) \geq 1$ . We begin by showing some standard properties of a line bundle over a Brill–Noether general curve which we will use later on. Then we prove that the matrix  $\chi_s$  is 1-generic for general  $s \in S$  and study the rank one locus of  $\chi_s$  which will be the divisor  $D_s$  or a rational normal curve. In the second case, we study the second-order focal locus. In both cases we can recover the canonical curve.

**Lemma 3.1** *Let  $C$  be a Brill–Noether general curve and let  $L \in W_d^1(C)$  be a smooth point. Then  $|L|$  is base point free,  $H^1(C, L^2) = 0$  and  $g_{2d}^{\rho+2} = |L^2|$  maps  $C$  birational to its image (it is not composed with an involution).*

*Proof* All of our claims follow directly from the generality assumption. We just mention that the map induced by  $|L^2|$  cannot be composed with an irrational involution. Hence, if the map is not birational, it is composed with a  $g_{d'}^1$  for  $d' \leq \frac{2d}{\rho+2}$  which is impossible for a Brill–Noether general curve.  $\square$

**Corollary 3.2** *Let  $C$  be a Brill–Noether general curve and let  $L \in W_d^1(C)$  be a smooth point. For  $i \geq 1$  and  $p_1, \dots, p_i \in \text{Supp}(D)$  for  $D \in |L|$  general, we have*

$$h^0(C, L^2(-p_1 - \dots - p_i)) = 2d - i + 1 - g.$$

*In particular,  $H^0(C, L^2(-p_1 - \dots - p_{\rho+1})) = H^0(C, L)$  and  $H^1(C, L^2(-p_1 - \dots - p_i)) = 0$  for  $i = 1, \dots, \rho + 1$ .*

*Proof*  $H^0(C, L^2(-p_1 - \dots - p_i)) = H^0(C, L^2(-p_1 - \dots - p_{i+1}))$  if the images under  $|L^2|$  of the two points  $p_i$  and  $p_{i+1}$  are the same point. Since  $|L^2|$  maps  $C$  birational to its image, this does not happen for a general choice.  $\square$

Using Lemma 3.1 and Corollary 3.2, our proof of the following lemma is identical to [7, Theorem 2]. We clarify and generalise the arguments given in [7, Theorem 2].

**Lemma 3.3** *With the assumptions of Lemma 3.1, the focal matrix  $\chi_s : T_{S,s} \otimes \mathcal{O}_{\Lambda_s} \rightarrow \mathcal{N}_{\Lambda_s/\mathbf{P}^{g-1}}$  is 1-generic for a sufficiently general  $s \in S$ .*

*Proof* By Proposition 2.7, the matrix  $\chi_s$  is 1-generic if and only if for each nonzero element  $v \in T_{S,s}$ , the homomorphism  $H^0(\chi_s)(v) \in \text{Hom}(U, V/U)$  is surjective.

We consider the first order deformation  $\underline{\Delta}_\varepsilon \subset \text{Spec}(k[\varepsilon]) \times \mathbf{P}^{g-1}$  defined by  $H^0(\chi_s)(\theta)$  for a nonzero vector  $\theta \in T_{S,s}$ . Note that  $H^0(\chi_s)(\theta)$  is surjective if and only if the image  $q(\underline{\Delta}_\varepsilon) \subset \mathbf{P}^{g-1}$  is not contained in a hyperplane. Let  $D_\varepsilon \subset \text{Spec}(k[\varepsilon]) \times \mathbf{P}^{g-1}$  be the first order deformation of the divisor  $D_s$  defined by  $\theta \in T_{S,s}$ . Then

$$q(\underline{\Delta}_\varepsilon) \supset q(D_\varepsilon)$$

and the curvilinear scheme  $q(D_\varepsilon)$  corresponding to a divisor on  $C$  satisfies

$$D_s \leq q(D_\varepsilon) \leq 2D_s.$$

We show for all possible cases that  $q(D_\varepsilon)$  is not contained in a hyperplane.

*Case 1:* The vector  $\theta$  is tangent to  $\alpha_d^{-1}(L)$ , equivalently the family  $D_\varepsilon$  deforms the divisor  $D_s$  in the linear pencil  $|L|$ . Let  $\varphi_L$  be the morphism defined by the pencil. Then we get

$$q(D_\varepsilon) = \varphi_L^*(\theta),$$

where we identify  $\theta$  with a curvilinear scheme of  $\mathbf{P}^1$  supported at the point  $s \in \mathbf{P}^1$ . Since  $|L|$  is base point free, we have  $q(D_\varepsilon) = 2D_s$ . Therefore, the curvilinear scheme  $q(D_\varepsilon)$  is not contained in a hyperplane since  $H^0(C, K_C - 2D_s)^* = H^1(C, 2D_s) = H^1(C, L^2) = 0$ . We are done in this case.

*Case 2:* We assume that  $\theta \in T_{s,s} \setminus \{0\}$  is not tangent to  $\alpha_d^{-1}(L)$  at  $s$ . Let

$$q(D_\varepsilon) = p_1 + \dots + p_k + 2(p_{k+1} + \dots + p_d)$$

where  $D_s = p_1 + \dots + p_d$  and  $k \geq 0$ .

*Case 2 (a):* We assume  $k \leq \rho$ . We have

$$\begin{aligned} H^0(C, K_C - q(D_\varepsilon))^* &= H^1(C, p_1 + \dots + p_k + 2(p_{k+1} + \dots + p_d)) \\ &= H^1(C, 2D_s - p_1 - \dots - p_k) \\ &= H^1(C, L^2(-p_1 - \dots - p_k)) = 0 \end{aligned}$$

by Corollary 3.2. Hence, the curvilinear scheme  $q(D_\varepsilon)$  is not contained in a hyperplane and  $H^0(\chi_s)(\theta)$  is surjective.

*Case 2 (b):* We assume  $k \geq \rho + 1$ . In the following, we will show that this case cannot occur. The vector  $\theta$  is also tangent to  $p_1 + \dots + p_k + C_{d-k}$ . We denote by  $E_s$  the divisor  $E_s = p_{k+1} + \dots + p_d$ . Then the tangent space to  $p_1 + \dots + p_k + C_{d-k}$  is given by  $H^0(E_s, \mathcal{O}_{E_s}(D_s))$  which is a subspace of  $H^0(D_s, \mathcal{O}_{D_s}(D_s))$ . The short exact sequence

$$0 \rightarrow \mathcal{O}_C \rightarrow L \rightarrow \mathcal{O}_{D_s}(D_s) \rightarrow 0$$

induces a linear map

$$H^0(D_s, \mathcal{O}_{D_s}(D_s)) \xrightarrow{\delta} H^1(C, \mathcal{O}_C)$$

which we identify with the differential of  $\alpha_d$  at  $s$  (see [1, IV, §2, Lemma 2.3]). The image of  $\theta \in H^0(E_s, \mathcal{O}_{E_s}(D_s))$  is therefore contained in the linear span of  $E_s$ . After projectivising, we get

$$[\delta(\theta)] \in \overline{E_s} = \overline{p_{k+1} + \dots + p_d} \subset A_s \subset \mathbf{P}^{g-1}.$$

Since  $\theta$  is not tangent to  $\alpha_d^{-1}(L)$ , the vector  $\theta$  is also tangent to  $W_d^1(C)$  and therefore the image point  $[\delta(\theta)]$  is contained in the vertex  $V = T_L(W_d^1(C))$  of

$X_L$ , the scroll swept out by the linear pencil  $|L|$ . Hence, for every sufficiently general  $D \in |L|$ , there is an effective divisor  $E$  of degree  $d - \rho - 1$  such that  $D = E + p_1 + \dots + p_{\rho+1}$  and  $V \cap \overline{E} \neq \emptyset$ . Hence,  $\dim(\overline{D_s + E}) \leq d - 2 + d - \rho - 1$  and equivalently,

$$h^0(C, D_s + E) = \deg(D_s + E) + \dim(\overline{D_s + E}) + 1 \geq 3.$$

But by Corollary 3.2,  $H^0(C, L^2(-p_1 - \dots - p_{\rho+1})) = H^0(C, L)$ , a contradiction. □

Note that

$$\chi_s : T_{S,s} \otimes \mathcal{O}_{\Lambda_s} \rightarrow \mathcal{N}_{\Lambda_s/\mathbf{P}^{g-1}}$$

is a map between rank  $\rho + 1$  and  $n = h^1(C, L)$  vector bundles of linear forms in  $\mathbf{P}^{d-2} = \Lambda_s$ . Since  $d = \rho + 1 + n$  and  $\chi_s$  is 1-generic by Lemma 3.3, we may apply the following theorem due to Eisenbud and Harris.

**Theorem 3.4 ([10, Proposition 5.1])** *Let  $M$  be an  $(a + 1) \times (b + 1)$  1-generic matrix of linear forms on  $\mathbf{P}^{a+b}$ . If  $D_1(M) = \{x \in \mathbf{P}^{a+b} \mid \text{rk}(M(x)) \leq 1\}$  contains a finite scheme  $\Gamma$  of length  $\geq a + b + 3$ , then  $D_1(M)$  is the unique rational normal curve through  $\Gamma$  and  $M$  is equivalent to the catalecticant matrix.*

We get the following corollary.

**Corollary 3.5** *For  $s \in S$  sufficiently general, the rank one locus  $V(\chi_s)_1$  is either  $D_s$  or a rational normal curve through  $D_s$ .*

*Proof* By Lemma 3.3, we may apply Theorem 3.4. Note that  $D_s \subset V(\chi_s)_1$  (there exists a codimension 1 family in  $S$  of  $\Lambda_s$  containing a point of the support of  $D_s$ ). □

*Remark 3.6*

- (a) The scheme of first-order foci at  $s \in S$  of the family  $\underline{\Delta}$  is a secant variety to  $V(\chi_s)_1$ .
- (b) If  $d = g - 1$  or  $\rho = 1$ , the focal matrix  $\chi_s$  is a  $2 \times (g - 3)$  or  $n \times 2$ -matrix, respectively. Hence, the rank one locus is the scheme of first-order foci, which is a rational normal curve in  $\Lambda_s$ . We recover the cases of [6] and [7].

**Corollary 3.7** *Let  $C$  be a Brill–Noether general canonically embedded curve. If  $V(\chi_s)_1 = D_s$  for sufficiently general  $s \in S$ , the family  $\underline{\Delta}$  determines the canonical curve  $C$ .*

**For the rest of this section, we assume that  $\Gamma_s = V(\chi_s)_1$  is a rational normal curve for  $s \in S$  sufficiently general.**

Let  $\Sigma$  be the smooth locus of  $W_d^1(C)$  and  $L \in \Sigma$ . Let  $U \subset \alpha_d^{-1}(L)$  be a Zariski open dense set such that  $\Gamma_s = V(\chi_s)_1$  for all  $s \in U$ . We define the surface

$$\Gamma_L = \bigcup_{s \in U} \overline{\Gamma_s}$$

and

$$\Gamma_{\mathbf{P}^{g-1}} = \overline{\bigcup_{L \in \Sigma} F_L}.$$

Let

$$\begin{array}{ccc} \underline{\Gamma} \subset S' \times \mathbf{P}^{g-1} & \xrightarrow{q} & \mathbf{P}^{g-1} \\ \downarrow p & & \\ S' & & \end{array}$$

be the family induced by all rational normal curves, that is, for  $s \in S' \subset S$ ,  $\Gamma_s = V(\chi_s)_1$  is a rational normal curve. The family  $\underline{\Gamma}$  is the rank one locus of the global characteristic map  $\chi$  and the variety  $\Gamma_{\mathbf{P}^{g-1}}$  is the image of the family  $\underline{\Gamma}$  under the second projection  $q$ .

*Remark 3.8* In the cases  $d = g - 1$  or  $\rho = 1$  the rational surface  $\Gamma_L$  is birational to  $\mathbf{P}^1 \times \mathbf{P}^1 \subset \mathbf{P}^3$  or a quadric cone in  $\mathbf{P}^3$ , respectively. This can be explained in terms of the curve  $C$  and the line bundle  $L$ :

For  $d = g - 1$  we consider the birational image  $C'$  of  $C \xrightarrow{|L| \times |\omega_C \otimes L^{-1}|} \mathbf{P}^1 \times \mathbf{P}^1$  given by the line bundle  $L$  and its Serre dual  $\omega_C \otimes L^{-1}$ . Then the rational surface  $\Gamma_L$  is the image of the blow up of  $\mathbf{P}^1 \times \mathbf{P}^1$  along the singular points of  $C'$  under the adjoint morphism.

For  $\rho = 1$  we consider the birational image  $C'$  of the curve  $C$  in the quadric cone  $Q$  in  $\mathbf{P}^3$  induced by the line bundle  $L^2$ . Note that  $H^0(C, L^2)$  is four-dimensional and the multiplication map  $H^0(C, L) \otimes H^0(C, L) \rightarrow H^0(C, L^2)$  has a one-dimensional kernel. Then the rational surface  $\Gamma_L$  is again the image of the blow up of  $Q$  along the singular points of  $C'$  under the adjoint morphism.

We have not found a similar geometrical meaning of the surface  $\Gamma_L$  in the other cases (see also Question 4.1).

**Lemma 3.9** *The variety  $\Gamma_{\mathbf{P}^{g-1}}$  has dimension at least 3.*

*Proof* Note that there is a map  $\Gamma_L \rightarrow \mathbf{P}^1 = \alpha_d^{-1}(L)$  such that the general fiber is a rational curve. Hence, the surface  $\Gamma_L$  is rational. Assume that  $\Gamma_L = \Gamma_{L'}$  for all  $L' \in \Sigma$ . Since the scrolls  $X_{L'}$  are algebraically equivalent to each other, the rulings on them cut out a  $(\rho + 1)$ -dimensional family of algebraically equivalent rational curves on  $\Gamma_L$ , the focal curves. (We can also argue that all  $d$ -secant to  $C$  are algebraically equivalent, thus the intersection with  $\Gamma_L$  yields a  $(\rho + 1)$ -dimensional family of algebraically equivalent focal curves.) On the desingularization of  $\Gamma_L$ , all of them are linear equivalent since  $\Gamma_L$  is regular ( $H^1(\Gamma_L, \mathcal{O}_{\Gamma_L}) = 0$ ). This implies that all  $g_d^1$ 's on  $C$  are linear equivalent, hence  $C$  has a  $g_d^{\rho+1}$ . A contradiction to the generality assumption on  $C$ . □

For the convenience of the reader, we recall the definition of the second-order foci of the family  $\underline{\Delta}$  (see also Definition 2.3). We apply the theory of foci to the

family  $\underline{\Gamma} \subset S' \times \mathbf{P}^{g-1}$  and get the characteristic map

$$\psi : T(p)|_{\underline{\Gamma}} \rightarrow \mathcal{N}_{\underline{\Gamma}/S' \times \mathbf{P}^{g-1}}$$

of vector bundles of rank  $\rho + 1$  and  $g - 2$ , respectively. For  $s \in S'$ , we call the closed subscheme of  $\Gamma_s$  defined by  $\text{rank}(\psi_s) \leq k$  the *scheme of second-order foci of rank  $k$  at  $s$*  (of the family  $\underline{\Delta}$ ).

We will show that the scheme of second-order foci of rank 1 at  $s \in S'$  of the family  $\underline{\Delta}$  is a finite scheme containing the divisor  $D_s$  and compute its degree.

**Lemma 3.10** *Let  $\psi_s : T_{S',s} \otimes \mathcal{O}_{\Gamma_s} \rightarrow \mathcal{N}_{\Gamma_s/\mathbf{P}^{g-1}}$  be the characteristic map for general  $s \in S'$ . Then the rank of  $\psi_s$  at a general point of  $\Gamma_s$  is at least 2.*

*Proof* We recall the connection of the rank and the dimension of  $\Gamma_{\mathbf{P}^{g-1}}$  as in [8, page 6]. Since  $\dim(\Gamma_{\mathbf{P}^{g-1}}) = \rho + 2 - \text{rank}(\ker(\psi))$ , the rank of  $\psi_s$  at the general point  $p \in \Gamma_s$  is

$$\begin{aligned} \text{rank}(\psi_s(p)) &= \dim(T(p)|_{\underline{\Gamma}}) - \text{rank}(\ker(\psi)) \\ &= \rho + 1 - \text{rank} \ker(\psi) \\ &= \dim(\Gamma_{\mathbf{P}^{g-1}}) - 1. \end{aligned}$$

The lemma follows from Lemma 3.9. This fact is also shown in [4, page 98]. □

We now consider for a general  $s \in S'$  the rank one locus of  $\psi_s$  which is a proper subset of  $\Gamma_s$  by Proposition 3.10.

**Lemma 3.11** *The degree of  $V(\psi_s)_1 \subset \Gamma_s = V(\chi_s)_1$  is at most  $d + \rho$ .*

*Proof* We imitate the proof of [7, Theorem 3]. Let  $s \in S' \subset S$  be a general point and let  $\Gamma_s \subset \mathbf{P}^{d-2} = \Lambda_s$  be the rank 1 locus of the map

$$\chi_s : T_{S',s} \otimes \mathcal{O}_{\Lambda_s} \rightarrow \mathcal{N}_{\Lambda_s/\mathbf{P}^{g-1}}.$$

Note that the normal bundle of  $\Gamma_s$  splits

$$\mathcal{N}_{\Gamma_s/\mathbf{P}^{g-1}} = (\mathcal{N}_{\Lambda_s/\mathbf{P}^{g-1}} \otimes \mathcal{O}_{\Gamma_s}) \oplus \mathcal{N}_{\Gamma_s/\Lambda_s} = \mathcal{O}_{\Gamma_s}(d - 2)^{\oplus n} \oplus \mathcal{O}_{\Gamma_s}(d)^{\oplus d-3}.$$

Hence, the map  $\psi_s$  is given by a matrix

$$\psi_s = \begin{pmatrix} A \\ B \end{pmatrix}$$

where  $A$  is a  $n \times (\rho + 1)$ -matrix and  $B$  is a  $(d - 3) \times (\rho + 1)$ -matrix. The matrix  $A$  represents the map  $(\chi_s : T_{S,s} \otimes \mathcal{O}_{\Lambda_s} \rightarrow \mathcal{N}_{\Gamma_s/\mathbf{P}^{g-1}})|_{\Gamma_s}$  and therefore has rank 1 and is equivalent to a catalecticant matrix. Let  $\{s, t\}$  be a basis of  $H^0(\Gamma_s, \mathcal{O}_{\Gamma_s}(1))$ . In an



appropriate basis, the matrix  $A$  is of the following form

$$A = \begin{pmatrix} t^{d-2} & t^{d-3}s & \dots & t^{d-2-\rho}s^\rho \\ t^{d-3}s & \ddots & & t^{d-2-\rho-1}s^{\rho+1} \\ \vdots & & & \vdots \\ t^{d-2-n+1}s^{n-1} & t^{d-2-n}s^n & \dots & s^{d-2} \end{pmatrix} = \begin{matrix} t^{n-1} \cdot \\ t^{n-1}s \cdot \\ \vdots \\ s^{n-1} \cdot \end{matrix} \begin{pmatrix} t^\rho & t^{\rho-1}s & \dots & s^\rho \\ t^\rho & \ddots & & s^\rho \\ \vdots & & & \vdots \\ t^\rho & t^{\rho-1}s & \dots & s^\rho \end{pmatrix}.$$

We see that the rank 1 locus of  $\psi_s$  is the rank 1 locus of the following matrix

$$N = \begin{pmatrix} t^\rho & t^{\rho-1}s & \dots & s^\rho \\ & B & & \end{pmatrix}.$$

Since  $V(\psi_s)_1 \neq \Gamma_s$  by Lemma 3.10, we have

$$\deg(V(\psi_s)_1) = \deg(D_1(N)) \leq \min\{\text{degree of elements of } I_{2 \times 2}(N)\} \leq \rho + d.$$

□

**Proposition 3.12** *Let  $s \in L$  be a sufficiently general point. Then,  $V(\psi_s)_1$  is the union of  $D_s$  and  $\rho$  points which are the intersection of  $\Gamma_s = V(\chi_s)_1$ , and the vertex  $V$  of the scroll  $X_L$  swept out by the pencil  $|L|$ .*

*Proof* As in the proof of Proposition 2.6, one can show that the points in the support of  $D_s$  are contained in  $V(\psi_s)_1$ .

Next, we show that the vertex in  $\Lambda_s$  is given by a column of the matrix  $\chi_s$ . Again, we imitate the proof of [6, Proposition 4.2]. Each column of the  $n \times (\rho + 1)$ -matrix  $\chi_s$  is a section of the rank  $n$  vector bundle  $V/U \otimes \mathcal{O}_{\Lambda_s}(1)$  (where  $U \subset V$  is the affine subspace representing  $\Lambda_s$ ) corresponding to an infinitesimal deformation of  $\Lambda_s$ . Each section vanishes in a  $\rho - 1 = (d - 2 - n)$ -subspace of  $\Lambda_s$  which is a  $\rho$ -secant of  $\Gamma_s$ . Since  $\chi_s$  is 1-generic, we get a  $(\rho + 1)$ -dimensional family of infinitesimal deformations of  $\Lambda_s$  induced by all columns. Hence, one column corresponds to the deformation in the scroll  $X_L$ . The corresponding section vanishes at the vertex. □

As in the case  $V(\chi_s)_1 = D_s$ , we get the following Torelli-type theorem using Remark 2.1.

**Corollary 3.13** *A Brill–Noether general canonically embedded curve  $C$  is uniquely determined by the family  $\underline{\Lambda}$ . More precise, the canonical curve  $C$  is a component of the scheme of first- or second-order foci of the family  $\underline{\Lambda}$  induced by the Brill–Noether locus  $W_d(C)$  and (the smooth locus of) its singular locus  $W_d^1(C)$  of dimension at least one (equivalently  $\lceil \frac{g+3}{2} \rceil \leq d \leq g - 1$ ).*

### 4 The First-Order Focal Map

For a general curve  $C$  and a sufficiently general point  $s \in S$ , the rank one locus of the focal map  $\chi_s$  at  $s$  is either  $d$  points or a rational normal curve. In the second case, the focal matrix at  $s$  is catalecticant (see Corollary 3.5).

As mentioned above, the articles [6] and [7] of Ciliberto and Sernesi are the extremal cases ( $d = g - 1$  and  $\rho = 1$ , respectively), where the rank one locus is always a rational normal curve. We propose the following question.

*Question 4.1* When is the focal matrix  $\chi_s$  catalecticant for a general curve  $C$  and a sufficiently general point  $s \in S$ ?

We conjecture that only in the extremal cases  $d = g - 1$  and  $\rho = 1$  the rank one locus of  $\chi_s$  is a rational normal curve for a general curve  $C$  and a general point  $s \in S$ . For the rest of this section we explain the reason for our conjecture.

Let  $C \subset \mathbf{P}^{g-1}$  be a canonically embedded curve of genus  $g$  and let  $L \in W_d^1(C)$  be a smooth point such that the rank one locus of the focal matrix  $\chi_s : T_{S,s} \otimes \mathcal{O}_{\Lambda_s} \rightarrow \mathcal{N}_{\Lambda_s/\mathbf{P}^{g-1}}$  is a rational normal curve  $\Gamma_s$  in  $\mathbf{P}^{d-2}$  for  $s \in |L|$  sufficiently general. Let  $X_L = \bigcup_{s \in |L|} \overline{D}_s$  be the scroll swept out by the pencil  $|L|$ . We get a rational surface

$$\Gamma_L = \overline{\bigcup_{s \in |L| \text{ gen}} \Gamma_s} \subset X_L$$

defined as in the previous section. The rational normal curve  $\Gamma_s$  intersects the vertex  $V$  of  $X_L$  in  $\rho = \rho(g, d, 1)$  points by Proposition 3.12. Note that the scroll  $X_L$  is a cone over  $\mathbf{P}^1 \times \mathbf{P}^{h^1(C,L)-1}$  with vertex  $V$ . Hence, projection from the vertex  $V$  yields a rational surface in  $\mathbf{P}^1 \times \mathbf{P}^{h^1(C,L)-1}$  whose general fiber in  $\mathbf{P}^{h^1(C,L)-1}$  is again a rational normal curve. We have shown the following proposition.

**Proposition 4.2** *Let  $C \subset \mathbf{P}^{g-1}$  be a canonically embedded curve of genus  $g$  and let  $L \in W_d^1(C)$  be a smooth point such that the rank one locus of the focal matrix  $\chi_s$  is a rational normal for  $s \in |L|$  sufficiently general. Then, the image of  $C$  in  $\mathbf{P}^1 \times \mathbf{P}^{h^1(C,L)-1}$  given by  $|L| \times |\omega_C \otimes L^{-1}|$  lies on a rational surface of bidegree  $(d', h^1(C, L) - 1)$  for some  $d'$ .*

*Proof* The proposition follows from the preceding discussion. We only note that the map given by  $|L| \times |\omega_C \otimes L^{-1}|$  is the same as the projection of  $\mathbf{P}^{g-1}$  along the vertex  $V$  of the canonically embedded  $C$ . □

*Example 4.3* We explain the above circumstance for a curve  $C$  of genus 8 with a line bundle  $L \in W_6^1(C)$ . The residual line bundle  $\omega_C \otimes L^{-1}$  has degree 8 and  $H^0(C, \omega_C \otimes L^{-1})$  is three-dimensional. Let  $C'$  be the image of  $C$  in  $\mathbf{P}^1 \times \mathbf{P}^2$  given by  $|L| \times |\omega_C \otimes L^{-1}|$ . We think of  $C' \rightarrow \mathbf{P}^1$  as a one-dimensional family of six points in the plane. If our assumption of Proposition 4.2 is true, the six points lie on a conic in every fiber over  $\mathbf{P}^1$ . Computing a curve of genus 8 with a  $g_6^1$  in Macaulay2 shows that these conics do not exist. Hence, our assumption of Proposition 4.2, that

is, the rank one locus of the focal matrix  $\chi_s$  is a rational normal curve for  $s \in |L|$  sufficiently general, does not hold for a general curve.

If  $\rho(g, d, 1) = 2d - g - 2 \geq 2$  and  $d < g - 1$ , we do not expect the existence of such a rational surface for a curve of genus  $g$  and a line bundle of degree  $d$  as above. Indeed,  $m$  general points in  $\mathbf{P}^r$  do not lie on a rational normal curve if  $m > r + 3$ . But the inequality  $\rho(g, d, 1) = 2d - g - 2 \geq 2$  implies  $d > (h^1(C, L) - 1) + 3$ . Using our `Macaulay2` package (see [3]), we could show in several examples  $((g, d) = (8, 6), (9, 7), (10, 8), (9, 6))$  that the rational surface of bidegree  $(d', h^1(C, L) - 1)$  of Proposition 4.2 does not exist. This confirms our conjectural behaviour of the first-order focal map.

**Acknowledgements** I would like to thank Edoardo Sernesi for sharing his knowledge about focal schemes and for valuable and enjoyable discussions while my visit at the university Roma Tre. The author was partially supported by the DFG-Grant SPP 1489 Schr. 307/5-2. I also thank George H. Hitching for useful questions regarding the geometry of focal schemes.

## References

1. E. Arbarello, M. Cornalba, P.A. Griffiths, J. Harris, *Geometry of Algebraic Curves. Vol. I.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 267 (Springer, New York, 1985)
2. A. Bajravani, Focal varieties of curves of genus 6 and 8. *Rend. Circ. Mat. Palermo* (2) **59**(1), 127–135 (2010)
3. C. Bopp, M. Hoff, `RelativeCanonicalResolution.m2` – construction of relative canonical resolutions and Eagon-Northcott type complexes, a `Macaulay2` package (2015). Available at <http://www.math.uni-sb.de/ag-schreyer/index.php/people/researchers/75-christian-bopp>
4. L. Chiantini, C. Ciliberto, A few remarks on the lifting problem. *Astérisque* **218**, 95–109 (1993). *Journées de Géométrie Algébrique d’Orsay* (Orsay, 1992)
5. C. Ciliberto, E. Sernesi, Singularities of the theta divisor and congruences of planes. *J. Algebr. Geom.* **1**(2), 231–250 (1992)
6. C. Ciliberto, E. Sernesi, Singularities of the theta divisor and families of secant spaces to a canonical curve. *J. Algebra* **171**(3), 867–893 (1995)
7. C. Ciliberto, E. Sernesi, On the geometry of canonical curves of odd genus. *Commun. Algebra* **28**(12), 5993–6001 (2000). Special issue in honor of Robin Hartshorne
8. C. Ciliberto, E. Sernesi, Projective geometry related to the singularities of theta divisors of Jacobians. *Boll. Unione Mat. Ital.* (9) **3**(1), 93–109 (2010)
9. D. Eisenbud, Linear sections of determinantal varieties. *Am. J. Math.* **110**(3), 541–575 (1988)
10. D. Eisenbud, J. Harris, An intersection bound for rank 1 loci, with applications to Castelnuovo and Clifford theory. *J. Algebr. Geom.* **1**(1), 31–59 (1992)
11. W. Fulton, *Intersection Theory.* *Ergebnisse der Mathematik und ihrer Grenzgebiete* (3) [Results in Mathematics and Related Areas (3)], vol. 2 (Springer, Berlin, 1984)
12. G.P. Pirola, M. Teixidor i Bigas, Generic Torelli for  $W_d^r$ . *Math. Z.* **209**(1), 53–54 (1992)
13. E. Sernesi, *Deformations of Algebraic Schemes.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 334 (Springer, Berlin, 2006)

# Inductive and Recursive Freeness of Localizations of Multiarrangements



Torsten Hoge, Gerhard Röhrle, and Anne Schauenburg

**Abstract** The class of free multiarrangements is known to be closed under taking localizations. We extend this result to the stronger notions of inductive and recursive freeness. As an application, we prove that recursively free (multi)arrangements are compatible with the product construction for (multi)arrangements. In addition, we show how our results can be used to derive that some canonical classes of free multiarrangements are not inductively free.

**Keywords** Multiarrangement • Free arrangement • Inductively free arrangement • Recursively free arrangement • Localization of an arrangement

**Subject Classifications** Primary 52C35, 14N20; Secondary 51D20

## 1 Introduction

The class of free arrangements plays a pivotal role in the study of hyperplane arrangements. While an arbitrary subarrangement of a free arrangement need not be free, freeness is retained by special types of subarrangements, so called localizations, [14], [10, Thm. 4.37]. It is natural to investigate this property for other classes of free arrangements.

For that purpose, let  $\mathcal{F}$ ,  $\mathcal{IF}$ ,  $\mathcal{RF}$  and  $\mathcal{HIF}$  denote the classes of free, inductively free, recursively free and hereditarily inductively free hyperplane arrangements, respectively (see [10, Defs. 4.53, 4.60]). Note that we have proper

---

T. Hoge

Institut für Algebra, Zahlentheorie und Diskrete Mathematik, Fakultät für Mathematik und Physik, Leibniz Universität Hannover, Welfengarten 1, 30167 Hannover, Germany  
e-mail: [hoge@math.uni-hannover.de](mailto:hoge@math.uni-hannover.de)

G. Röhrle (✉) • A. Schauenburg

Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany  
e-mail: [gerhard.roehrle@rub.de](mailto:gerhard.roehrle@rub.de); [anne.schauenburg@rub.de](mailto:anne.schauenburg@rub.de)

inclusions throughout  $\mathcal{HIF} \subsetneq \mathcal{IF} \subsetneq \mathcal{RF} \subsetneq \mathcal{F}$ , see [7, Ex. 2.16], [10, Ex. 4.56], and [6, Rem. 3.7], respectively. Our first result shows that localization preserves each of these stronger notions of freeness.

**Theorem 1.1** *Each of the classes  $\mathcal{IF}$ ,  $\mathcal{RF}$  and  $\mathcal{HIF}$  is closed under taking localizations.*

Moreover, freeness is compatible with the product construction for arrangements [10, Prop. 4.28]. It was shown in [7, Prop. 2.10, Cor. 2.12] that this property also holds for both  $\mathcal{IF}$  and  $\mathcal{HIF}$ . Our second main result extends this property to the class  $\mathcal{RF}$ .

**Theorem 1.2** *A product of arrangements belongs to  $\mathcal{RF}$  if and only if each factor belongs to  $\mathcal{RF}$ .*

It can be a rather complicated affair to show that a given arrangement is inductively free, see for instance [5, §5.2], [4, Lem. 4.2], and [7, Lem. 3.5]. In principle, one might have to search through all possible chains of free subarrangements. The notion of recursive freeness is even more elusive. In that sense, Theorem 1.1 serves to be a very useful tool in deciding that a given arrangement is not inductively or recursively free by exhibiting a localization which is known to lack this property.

In his seminal work [18], Ziegler introduced the notion of multiarrangements and initiated the study of their freeness. In their ground breaking work [2, Thm. 0.8], Abe, Terao and Wakefield proved the Addition Deletion Theorem for multiarrangements.

The class of free multiarrangements is known to be closed under taking localizations, see Theorem 2.7. Our third main result shows that localization also preserves the notions of inductive and recursive freeness in the setting of multiarrangements. For this purpose, let  $\mathcal{IF.M}$  and  $\mathcal{RF.M}$  denote the classes of inductively free and recursively free multiarrangements, see Definitions 2.18 and 2.21.

**Theorem 1.3** *The classes  $\mathcal{IF.M}$  and  $\mathcal{RF.M}$  are closed under taking localizations.*

Theorem 1.1 follows for  $\mathcal{IF}$  and  $\mathcal{RF}$  from Theorem 1.3 as a special case, cf. Remark 2.19.

It follows from [2, Lem. 1.3] that a product of multiarrangements is free if and only if each factor is free. Armed with Theorem 1.3, we can readily extend this further to the classes  $\mathcal{IF.M}$  and  $\mathcal{RF.M}$ .

**Theorem 1.4** *A product of multiarrangements belongs to  $\mathcal{IF.M}$  (resp.  $\mathcal{RF.M}$ ) if and only if each factor belongs to  $\mathcal{IF.M}$  (resp.  $\mathcal{RF.M}$ ).*

Theorem 1.2 follows from Theorem 1.4 for  $\mathcal{RF.M}$  as a special case.

In Sect. 4 we further demonstrate the versatility of Theorem 1.3 by showing that certain multiarrangements stemming from complex reflection groups are not inductively free. Among them are multiarrangements of a restricted arrangement equipped with Ziegler's natural multiplicity on the restriction to a hyperplane, see Definition 2.9.

For applications of Theorems 1.1 and 1.2 in the context of the classification of recursively free reflection arrangements, see [8, §3, Lem. 3.2].

## 2 Recollections and Preliminaries

### 2.1 Hyperplane Arrangements

Let  $V = \mathbb{K}^\ell$  be an  $\ell$ -dimensional  $\mathbb{K}$ -vector space. A *hyperplane arrangement* is a pair  $(\mathcal{A}, V)$ , where  $\mathcal{A}$  is a finite collection of hyperplanes in  $V$ . Usually, we simply write  $\mathcal{A}$  in place of  $(\mathcal{A}, V)$ . Sometimes we also call  $\mathcal{A}$  an  $\ell$ -*arrangement* if we want to emphasize the dimension of the ambient vector space. We write  $|\mathcal{A}|$  for the number of hyperplanes in  $\mathcal{A}$ . The empty arrangement in  $V$  is denoted by  $\Phi_\ell$ .

The *lattice*  $L(\mathcal{A})$  of  $\mathcal{A}$  is the set of subspaces of  $V$  of the form  $H_1 \cap \dots \cap H_i$  where  $\{H_1, \dots, H_i\}$  is a subset of  $\mathcal{A}$ . For  $X \in L(\mathcal{A})$ , we have two associated arrangements, firstly  $\mathcal{A}_X := \{H \in \mathcal{A} \mid X \subseteq H\} \subseteq \mathcal{A}$ , the *localization of  $\mathcal{A}$  at  $X$* , and secondly, the *restriction of  $\mathcal{A}$  to  $X$* ,  $(\mathcal{A}^X, X)$ , where  $\mathcal{A}^X := \{X \cap H \mid H \in \mathcal{A} \setminus \mathcal{A}_X\}$ . Note that  $V$  belongs to  $L(\mathcal{A})$  as the intersection of the empty collection of hyperplanes and  $\mathcal{A}^V = \mathcal{A}$ . The lattice  $L(\mathcal{A})$  is a partially ordered set by reverse inclusion:  $X \leq Y$  provided  $Y \subseteq X$  for  $X, Y \in L(\mathcal{A})$ .

If  $0 \in H$  for each  $H$  in  $\mathcal{A}$ , then  $\mathcal{A}$  is called *central*. If  $\mathcal{A}$  is central, then the *center*  $T_{\mathcal{A}} := \bigcap_{H \in \mathcal{A}} H$  of  $\mathcal{A}$  is the unique maximal element in  $L(\mathcal{A})$  with respect to the partial order. We have a *rank function* on  $L(\mathcal{A})$ :  $r(X) := \text{codim}_V(X)$ . The *rank*  $r := r(\mathcal{A})$  of  $\mathcal{A}$  is the rank of a maximal element in  $L(\mathcal{A})$ . The  $\ell$ -arrangement  $\mathcal{A}$  is *essential* provided  $r(\mathcal{A}) = \ell$ . If  $\mathcal{A}$  is central and essential, then  $T_{\mathcal{A}} = \{0\}$ . Throughout this article, we only consider central arrangements.

More generally, for  $U$  an arbitrary subspace of  $V$ , define the *localization of  $\mathcal{A}$  at  $U$*  by  $\mathcal{A}_U := \{H \in \mathcal{A} \mid U \subseteq H\} \subseteq \mathcal{A}$ , and  $\mathcal{A}^U := \{U \cap H \mid H \in \mathcal{A} \setminus \mathcal{A}_U\}$ , a subarrangement in  $U$ . The following observations are immediate from the definitions, cf. [10, §2].

**Lemma 2.1** *Let  $\mathcal{B} \subseteq \mathcal{A}$  be a subarrangement and  $Y \leq X$  in  $L(\mathcal{A})$ . Then we have*

- (i)  $\mathcal{B} \cap \mathcal{A}_X = \mathcal{B}_X$ ; and
- (ii)  $(\mathcal{B}_X)^Y = (\mathcal{B}^Y)_X$ .

*Note that  $X$  and  $Y$  need not be members of  $L(\mathcal{B})$ .*

### 2.2 Free Hyperplane Arrangements

Let  $S = S(V^*)$  be the symmetric algebra of the dual space  $V^*$  of  $V$ . If  $x_1, \dots, x_\ell$  is a basis of  $V^*$ , then we identify  $S$  with the polynomial ring  $\mathbb{K}[x_1, \dots, x_\ell]$ . Letting  $S_p$

denote the  $\mathbb{K}$ -subspace of  $S$  consisting of the homogeneous polynomials of degree  $p$  (along with 0),  $S$  is naturally  $\mathbb{Z}$ -graded:  $S = \bigoplus_{p \in \mathbb{Z}} S_p$ , where  $S_p = 0$  in case  $p < 0$ .

Let  $\text{Der}(S)$  be the  $S$ -module of algebraic  $\mathbb{K}$ -derivations of  $S$ . Using the  $\mathbb{Z}$ -grading on  $S$ ,  $\text{Der}(S)$  becomes a graded  $S$ -module. For  $i = 1, \dots, \ell$ , let  $D_i := \partial/\partial x_i$  be the usual derivation of  $S$ . Then  $D_1, \dots, D_\ell$  is an  $S$ -basis of  $\text{Der}(S)$ . We say that  $\theta \in \text{Der}(S)$  is *homogeneous of polynomial degree  $p$*  provided  $\theta = \sum_{i=1}^\ell f_i D_i$ , where  $f_i$  is either 0 or homogeneous of degree  $p$  for each  $1 \leq i \leq \ell$ . In this case we write  $\text{pdeg } \theta = p$ .

Let  $\mathcal{A}$  be a central arrangement in  $V$ . Then for  $H \in \mathcal{A}$  we fix  $\alpha_H \in V^*$  with  $H = \ker(\alpha_H)$ . The *defining polynomial*  $Q(\mathcal{A})$  of  $\mathcal{A}$  is given by  $Q(\mathcal{A}) := \prod_{H \in \mathcal{A}} \alpha_H \in S$ .

The *module of  $\mathcal{A}$ -derivations* of  $\mathcal{A}$  is defined by

$$D(\mathcal{A}) := \{ \theta \in \text{Der}(S) \mid \theta(\alpha_H) \in \alpha_H S \ \forall H \in \mathcal{A} \}.$$

We say that  $\mathcal{A}$  is *free* if the module of  $\mathcal{A}$ -derivations  $D(\mathcal{A})$  is a free  $S$ -module.

With the  $\mathbb{Z}$ -grading of  $\text{Der}(S)$ , also  $D(\mathcal{A})$  becomes a graded  $S$ -module, [10, Prop. 4.10]. If  $\mathcal{A}$  is a free arrangement, then the  $S$ -module  $D(\mathcal{A})$  admits a basis of  $\ell$  homogeneous derivations, say  $\theta_1, \dots, \theta_\ell$ , [10, Prop. 4.18]. While the  $\theta_i$ 's are not unique, their polynomial degrees  $\text{pdeg } \theta_i$  are unique (up to ordering). This multiset is the set of *exponents* of the free arrangement  $\mathcal{A}$  and is denoted by  $\text{exp } \mathcal{A}$ .

The fundamental *Addition Deletion Theorem* due to Terao [12] plays a pivotal role in the study of free arrangements, [10, Thm. 4.51].

**Theorem 2.2** *Suppose  $\mathcal{A} \neq \Phi_\ell$ . Let  $H_0 \in \mathcal{A}$ . Set  $\mathcal{A}' = \mathcal{A} \setminus \{H_0\}$  and  $\mathcal{A}'' = \mathcal{A}^{H_0}$ . Then any two of the following statements imply the third:*

- (i)  $\mathcal{A}$  is free with  $\text{exp } \mathcal{A} = \{b_1, \dots, b_{\ell-1}, b_\ell\}$ ;
- (ii)  $\mathcal{A}'$  is free with  $\text{exp } \mathcal{A}' = \{b_1, \dots, b_{\ell-1}, b_\ell - 1\}$ ;
- (iii)  $\mathcal{A}''$  is free with  $\text{exp } \mathcal{A}'' = \{b_1, \dots, b_{\ell-1}\}$ .

Theorem 2.2 motivates the notion of *inductive freeness*, cf. [10, Def. 4.53]:

**Definition 2.3** The class  $\mathcal{IF}$  of *inductively free* arrangements is the smallest class of arrangements subject to

- (i)  $\Phi_\ell \in \mathcal{IF}$  for each  $\ell \geq 0$ ;
- (ii) if there exists a hyperplane  $H_0 \in \mathcal{A}$  such that both  $\mathcal{A}'$  and  $\mathcal{A}''$  belong to  $\mathcal{IF}$ , and  $\text{exp } \mathcal{A}'' \subseteq \text{exp } \mathcal{A}'$ , then  $\mathcal{A}$  also belongs to  $\mathcal{IF}$ .

There is an even stronger notion of freeness, cf. [10, §6.4, p. 253].

**Definition 2.4** The arrangement  $\mathcal{A}$  is called *hereditarily inductively free* provided  $\mathcal{A}^X$  is inductively free for each  $X \in L(\mathcal{A})$ . We abbreviate this class by  $\mathcal{HIF}$ .

As  $V \in L(\mathcal{A})$  and  $\mathcal{A}^V = \mathcal{A}$ ,  $\mathcal{A}$  is inductively free, if it is hereditarily inductively free. Also,  $\mathcal{HIF}$  is a proper subclass of  $\mathcal{IF}$ , see [7, Ex. 2.16].

Let  $U \subseteq V$  be a subspace of  $V$ . Thanks to work of Terao, [13, Prop. 5.5], [14, Prop. 2],  $\mathcal{A}_U$  is free whenever  $\mathcal{A}$  is, cf. [19, Thm. 1.7(i)], [10, Thm. 4.37], or [17, Prop. 1.15].

### 2.3 Multiarrangements

A *multiarrangement* is a pair  $(\mathcal{A}, \nu)$  consisting of a hyperplane arrangement  $\mathcal{A}$  and a *multiplicity function*  $\nu : \mathcal{A} \rightarrow \mathbb{Z}_{\geq 0}$  associating to each hyperplane  $H$  in  $\mathcal{A}$  a non-negative integer  $\nu(H)$ . Alternately, the multiarrangement  $(\mathcal{A}, \nu)$  can also be thought of as the multiset of hyperplanes

$$(\mathcal{A}, \nu) = \{H^{\nu(H)} \mid H \in \mathcal{A}\}.$$

The *order* of the multiarrangement  $(\mathcal{A}, \nu)$  is the cardinality of the multiset  $(\mathcal{A}, \nu)$ ; we write  $|\nu| := |(\mathcal{A}, \nu)| = \sum_{H \in \mathcal{A}} \nu(H)$ . For a multiarrangement  $(\mathcal{A}, \nu)$ , the underlying arrangement  $\mathcal{A}$  is sometimes called the associated *simple arrangement*, and so  $(\mathcal{A}, \nu)$  itself is simple if and only if  $\nu(H) = 1$  for each  $H \in \mathcal{A}$ .

**Definition 2.5** Let  $\nu_i$  be a multiplicity of  $\mathcal{A}_i$  for  $i = 1, 2$ . When viewed as multisets, suppose that  $(\mathcal{A}_1, \nu_1)$  is a subset of  $(\mathcal{A}_2, \nu_2)$ . Then we say that  $(\mathcal{A}_1, \nu_1)$  is a *submultiarrangement* of  $(\mathcal{A}_2, \nu_2)$  and write  $(\mathcal{A}_1, \nu_1) \subseteq (\mathcal{A}_2, \nu_2)$ , i.e. we have  $\nu_1(H) \leq \nu_2(H)$  for each  $H \in \mathcal{A}_1$ .

**Definition 2.6** Let  $(\mathcal{A}, \nu)$  be a multiarrangement in  $V$  and let  $U \subseteq V$  be a subspace of  $V$ . The *localization* of  $(\mathcal{A}, \nu)$  at  $U$  is  $(\mathcal{A}_U, \nu_U)$ , where  $\nu_U = \nu|_{\mathcal{A}_U}$ . Note that for  $X = \bigcap_{H \in \mathcal{A}_U} H$ , we have  $\mathcal{A}_X = \mathcal{A}_U$  and  $X$  belongs to the intersection lattice of  $\mathcal{A}$ .

### 2.4 Freeness of Multiarrangements

Following Ziegler [18], we extend the notion of freeness to multiarrangements as follows. The *defining polynomial* of the multiarrangement  $(\mathcal{A}, \nu)$  is given by

$$Q(\mathcal{A}, \nu) := \prod_{H \in \mathcal{A}} \alpha_H^{\nu(H)},$$

a polynomial of degree  $|\nu|$  in  $S$ .

The *module of  $\mathcal{A}$ -derivations* of  $(\mathcal{A}, \nu)$  is defined by

$$D(\mathcal{A}, \nu) := \{\theta \in \text{Der}(S) \mid \theta(\alpha_H) \in \alpha_H^{\nu(H)} S \ \forall H \in \mathcal{A}\}.$$

We say that  $(\mathcal{A}, \nu)$  is *free* if  $D(\mathcal{A}, \nu)$  is a free  $S$ -module, [18, Def. 6].

As in the case of simple arrangements,  $D(\mathcal{A}, \nu)$  is a  $\mathbb{Z}$ -graded  $S$ -module and thus, if  $(\mathcal{A}, \nu)$  is free, there is a homogeneous basis  $\theta_1, \dots, \theta_\ell$  of  $D(\mathcal{A}, \nu)$ . The multiset of the unique polynomial degrees  $\text{pdeg } \theta_i$  forms the set of *exponents* of the free multiarrangement  $(\mathcal{A}, \nu)$  and is denoted by  $\text{exp}(\mathcal{A}, \nu)$ . It follows from Ziegler’s



analogue of Saito’s criterion [18, Thm. 8] that

$$\sum \text{pdeg } \theta_i = \text{deg } Q(\mathcal{A}, \nu) = |\nu|.$$

Freeness for multiarrangements is preserved under localizations. The argument in the proof of [10, Thm. 4.37] readily extends to this more general setting.

**Theorem 2.7 ([1, Prop. 1.7])** *For  $U \subseteq V$  a subspace, the localization  $(\mathcal{A}_U, \nu_U)$  of  $(\mathcal{A}, \nu)$  at  $U$  is free provided  $(\mathcal{A}, \nu)$  is free.*

Though constructive, the proof of Theorem 2.7 does not shed any light on the exponents of  $(\mathcal{A}_U, \nu_U)$  in relation to the exponents of  $(\mathcal{A}, \nu)$ . We do however have the following elementary observation.

*Remark 2.8* Let  $(\mathcal{A}_1, \nu_1) \subseteq (\mathcal{A}_2, \nu_2)$  be free multiarrangements with ordered sets of exponents  $\text{exp}(\mathcal{A}_i, \nu_i) = \{a_{i,1} \leq \dots \leq a_{i,\ell}\}$  for  $i = 1, 2$ . Then  $a_{1,j} \leq a_{2,j}$  for each  $1 \leq j \leq \ell$ . For, let  $\{\theta_{i,1}, \dots, \theta_{i,\ell}\}$  be a homogeneous  $S$ -basis of the free  $S$ -module  $D(\mathcal{A}_i, \nu_i)$  for  $i = 1, 2$ . For a contradiction, suppose that  $k$  is the smallest index such that  $a_{1,k} > a_{2,k}$ . Then the grading of both  $S$ -modules and the fact that  $D(\mathcal{A}_2, \nu_2) \subseteq D(\mathcal{A}_1, \nu_1)$  imply that  $\theta_{2,1}, \dots, \theta_{2,k} \in S\theta_{1,1} + \dots + S\theta_{1,k-1}$ . But this shows that  $\{\theta_{2,1}, \dots, \theta_{2,\ell}\}$  is not algebraically independent over  $S$ , a contradiction.

We recall a fundamental construction due to Ziegler, [18, Ex. 2].

**Definition 2.9 (Ziegler Restriction)** Let  $\mathcal{A}$  be a simple arrangement. Fix  $H_0 \in \mathcal{A}$  and consider the restriction  $\mathcal{A}''$  with respect to  $H_0$ . Define the *canonical multiplicity*  $\kappa$  on  $\mathcal{A}''$  as follows. For  $Y \in \mathcal{A}''$  set

$$\kappa(Y) := |\mathcal{A}_Y| - 1,$$

i.e.,  $\kappa(Y)$  is the number of hyperplanes in  $\mathcal{A} \setminus \{H_0\}$  lying above  $Y$ . Ziegler showed that freeness of  $\mathcal{A}$  implies freeness of  $(\mathcal{A}'', \kappa)$  as follows.

**Theorem 2.10 ([18, Thm. 11])** *Let  $\mathcal{A}$  be a free arrangement with exponents  $\text{exp } \mathcal{A} = \{1, e_2, \dots, e_\ell\}$ . Let  $H_0 \in \mathcal{A}$  and consider the restriction  $\mathcal{A}''$  with respect to  $H_0$ . Then the multiarrangement  $(\mathcal{A}'', \kappa)$  is free with exponents  $\text{exp}(\mathcal{A}'', \kappa) = \{e_2, \dots, e_\ell\}$ .*

Note that the converse of Theorem 2.10 is false. For example, let  $\mathcal{A}$  be a non-free 3-arrangement, cf. [10, Ex. 4.34]. Since  $\mathcal{A}''$  is of rank 2,  $(\mathcal{A}'', \kappa)$  is free, [18, Cor. 7]. Nevertheless, Ziegler’s construction and in particular the question of a converse of Theorem 2.10 under suitable additional hypotheses play an important role in the study of free simple arrangements, e.g. see [15, Thm. 2.1, Thm. 2.2], [16], [3, Cor. 4.2], [11, Thm. 2] and [17, Cor. 1.35].

### 2.5 The Addition Deletion Theorem for Multiarrangements

We recall the construction from [2].

**Definition 2.11** Let  $(\mathcal{A}, \nu) \neq \Phi_\ell$  be a multiarrangement. Fix  $H_0$  in  $\mathcal{A}$ . We define the *deletion*  $(\mathcal{A}', \nu')$  and *restriction*  $(\mathcal{A}'', \nu^*)$  of  $(\mathcal{A}, \nu)$  with respect to  $H_0$  as follows. If  $\nu(H_0) = 1$ , then set  $\mathcal{A}' = \mathcal{A} \setminus \{H_0\}$  and define  $\nu'(H) = \nu(H)$  for all  $H \in \mathcal{A}'$ . If  $\nu(H_0) > 1$ , then set  $\mathcal{A}' = \mathcal{A}$  and define  $\nu'(H_0) = \nu(H_0) - 1$  and  $\nu'(H) = \nu(H)$  for all  $H \neq H_0$ .

Let  $\mathcal{A}'' = \{H \cap H_0 \mid H \in \mathcal{A} \setminus \{H_0\}\}$ . The Euler multiplicity  $\nu^*$  of  $\mathcal{A}''$  is defined as follows. Let  $Y \in \mathcal{A}''$ . Since the localization  $\mathcal{A}_Y$  is of rank 2, the multiarrangement  $(\mathcal{A}_Y, \nu_Y)$  is free, [18, Cor. 7]. According to [2, Prop. 2.1], the module of derivations  $D(\mathcal{A}_Y, \nu_Y)$  admits a particular homogeneous basis  $\{\theta_Y, \psi_Y, D_3, \dots, D_\ell\}$ , where  $\theta_Y$  is identified by the property that  $\theta_Y \notin \alpha_0 \text{Der}(S)$  and  $\psi_Y$  by the property that  $\psi_Y \in \alpha_0 \text{Der}(S)$ , where  $H_0 = \ker \alpha_0$ . Then the Euler multiplicity  $\nu^*$  is defined on  $Y$  as  $\nu^*(Y) = \text{pdeg } \theta_Y$ . Crucial for our purpose is the fact that the value  $\nu^*(Y)$  only depends on the  $S$ -module  $D(\mathcal{A}_Y, \nu_Y)$ .

Frequently,  $(\mathcal{A}, \nu)$ ,  $(\mathcal{A}', \nu')$  and  $(\mathcal{A}'', \nu^*)$  is referred to as the *triple* of multiarrangements with respect to  $H_0$ .

**Theorem 2.12 ([2, Thm. 0.8] Addition Deletion Theorem for Multiarrangements)** *Suppose that  $(\mathcal{A}, \nu) \neq \Phi_\ell$ . Fix  $H_0$  in  $\mathcal{A}$  and let  $(\mathcal{A}, \nu)$ ,  $(\mathcal{A}', \nu')$  and  $(\mathcal{A}'', \nu^*)$  be the triple with respect to  $H_0$ . Then any two of the following statements imply the third:*

- (i)  $(\mathcal{A}, \nu)$  is free with  $\text{exp}(\mathcal{A}, \nu) = \{b_1, \dots, b_{\ell-1}, b_\ell\}$ ;
- (ii)  $(\mathcal{A}', \nu')$  is free with  $\text{exp}(\mathcal{A}', \nu') = \{b_1, \dots, b_{\ell-1}, b_\ell - 1\}$ ;
- (iii)  $(\mathcal{A}'', \nu^*)$  is free with  $\text{exp}(\mathcal{A}'', \nu^*) = \{b_1, \dots, b_{\ell-1}\}$ .

*Remark 2.13* We require a slightly stronger version of the restriction part of Theorem 2.12, where we do not prescribe the exponents a priori. Let  $(\mathcal{A}, \nu)$ ,  $(\mathcal{A}', \nu')$  and  $(\mathcal{A}'', \nu^*)$  be the triple with respect to a fixed hyperplane. It follows from [2, Thm. 0.4] that if both  $(\mathcal{A}, \nu)$  and  $(\mathcal{A}', \nu')$  are free, then their exponents are as given by parts (i) and (ii) in Theorem 2.12 (i.e., the exponents differing by 1 in one term is automatic, cf. [10, Thm. 4.46]). It then follows from the restriction part of Theorem 2.12 that  $(\mathcal{A}'', \nu^*)$  is also free with exponents as in part (iii).

Next we observe that localization is compatible with both deletion and restriction for multiarrangements.

**Lemma 2.14** *Let  $(\mathcal{A}, \nu)$  be a multiarrangement,  $X \in L(\mathcal{A})$ , and  $H \in \mathcal{A}_X$ . Let  $(\mathcal{A}, \nu)$ ,  $(\mathcal{A}', \nu')$  and  $(\mathcal{A}'', \nu^*)$  be the triple with respect to  $H$ . Then we have*

- (i)  $((\mathcal{A}_X)', (\nu_X)') = ((\mathcal{A}')_X, (\nu')_X)$ ; and
- (ii)  $((\mathcal{A}_X)'', (\nu_X)^*) = ((\mathcal{A}_X)^H, (\nu_X)^*) = ((\mathcal{A}^H)_X, (\nu^*)_X) = ((\mathcal{A}'')_X, (\nu^*)_X)$ .

*Proof* (i) The proof follows easily from Definitions 2.6 and 2.11.

(ii) Thanks to Lemma 2.1(ii), we have  $(\mathcal{A}_X)^H = (\mathcal{A}^H)_X$ . By definition,  $(v^*)_X = v^*$  on  $(\mathcal{A}^H)_X$ . So it suffices to show that  $(v_X)^* = v^*$  on  $(\mathcal{A}_X)^H$ . Let  $Y \in (\mathcal{A}_X)^H$ . Then  $Y = H \cap H'$  for some  $H' \in \mathcal{A}_X \setminus \{H\}$ . Consequently,  $(\mathcal{A}_X)_Y = \mathcal{A}_Y$  and  $(v_X)_Y = v_Y$ . Therefore,  $D((\mathcal{A}_X)_Y, (v_X)_Y) = D(\mathcal{A}_Y, v_Y)$  and so by definition of the Euler multiplicity  $(v_X)^* = v^*$ , as desired.

We recast Lemma 2.14 in terms of triples as follows.

**Corollary 2.15** *Let  $(\mathcal{A}, v)$  be a multiarrangement,  $X \in L(\mathcal{A})$ , and  $H \in \mathcal{A}_X$ . Let  $(\mathcal{A}, v)$ ,  $(\mathcal{A}', v')$  and  $(\mathcal{A}'', v^*)$  be the triple of  $(\mathcal{A}, v)$  with respect to  $H$ . Then  $(\mathcal{A}_X, v_X)$ ,  $((\mathcal{A}')_X, (v')_X)$  and  $((\mathcal{A}'')_X, (v^*)_X)$  is the triple of  $(\mathcal{A}_X, v_X)$  with respect to  $H$ .*

In general, for  $\mathcal{A}$  a free hyperplane arrangement,  $(\mathcal{A}, v)$  does not need to be free for an arbitrary multiplicity  $v$ , e.g. see [18, Ex. 14]. However, for the following special class of multiarrangements this is always the case, [2, Prop. 5.2].

**Definition 2.16** Let  $\mathcal{A}$  be a simple arrangement. Fix  $H_0 \in \mathcal{A}$  and  $m_0 \in \mathbb{Z}_{>1}$  and define the *multiplicity  $\delta$  concentrated at  $H_0$*  by

$$\delta(H) := \delta_{H_0, m_0}(H) := \begin{cases} m_0 & \text{if } H = H_0, \\ 1 & \text{else.} \end{cases}$$

The following combines [2, Prop. 5.2], parts of its proof and Theorem 2.10. Recall the definition of Ziegler’s multiplicity  $\kappa$  from Definition 2.9. The proof of Proposition 2.17(i) given in [2] depends on Theorem 2.12. We present an elementary explicit construction for a homogeneous  $S$ -basis of  $D(\mathcal{A}, \delta)$ .

**Proposition 2.17** *Let  $\mathcal{A}$  be a free simple arrangement with  $\exp \mathcal{A} = \{1, e_2, \dots, e_\ell\}$ . Fix  $H_0 \in \mathcal{A}$ ,  $m_0 \in \mathbb{Z}_{>1}$  and let  $\delta = \delta_{H_0, m_0}$  be as in Definition 2.16. Let  $(\mathcal{A}'', \delta^*)$  be the restriction of  $(\mathcal{A}, \delta)$  with respect to  $H_0$ . Then we have*

- (i)  $(\mathcal{A}, \delta)$  is free with exponents  $\exp(\mathcal{A}, \delta) = \{m_0, e_2, \dots, e_\ell\}$ ;
- (ii)  $(\mathcal{A}'', \delta^*) = (\mathcal{A}'', \kappa)$  is free with exponents  $\exp(\mathcal{A}'', \kappa) = \{e_2, \dots, e_\ell\}$ .

*Proof* (i) We utilize the construction from the proof of [10, Prop. 4.27]. Let  $\alpha_0 \in V^*$  with  $H_0 = \ker \alpha_0$  and let  $\text{Ann}(H_0) = \{\theta \in D(\mathcal{A}) \mid \theta(\alpha_0) = 0\}$  be the annihilator of  $H_0$  in  $D(\mathcal{A})$ . Let  $\theta_E$  be the Euler derivation in  $\text{Der}(S)$  [10, Def. 4.7]. Then

$$D(\mathcal{A}) = S\theta_E \oplus \text{Ann}(H_0)$$

is a direct sum of  $S$ -modules. Let  $\{\theta_2, \dots, \theta_\ell\}$  be a homogeneous  $S$ -basis of  $\text{Ann}(H_0)$ . Then  $\{\theta_E, \theta_2, \dots, \theta_\ell\}$  is a homogeneous  $S$ -basis of  $D(\mathcal{A})$ . We thus conclude that  $\{\alpha_0^{m_0-1}\theta_E, \theta_2, \dots, \theta_\ell\}$  is a homogeneous  $S$ -basis of  $D(\mathcal{A}, \delta)$ .

(ii) The equality  $(\mathcal{A}'', \delta^*) = (\mathcal{A}'', \kappa)$  is derived as in the proof of [2, Prop. 5.2]. The remaining statements then follow from Theorem 2.10.

## 2.6 Inductive and Recursive Freeness for Multiarrangements

As in the simple case, Theorem 2.12 motivates the notion of inductive freeness.

**Definition 2.18** ([2, Def. 0.9]) The class  $\mathcal{IFM}$  of *inductively free* multiarrangements is the smallest class of arrangements subject to

- (i)  $\Phi_\ell \in \mathcal{IFM}$  for each  $\ell \geq 0$ ;
- (ii) for a multiarrangement  $(\mathcal{A}, \nu)$ , if there exists a hyperplane  $H_0 \in \mathcal{A}$  such that both  $(\mathcal{A}', \nu')$  and  $(\mathcal{A}'', \nu^*)$  belong to  $\mathcal{IFM}$ , and  $\exp(\mathcal{A}'', \nu^*) \subseteq \exp(\mathcal{A}', \nu')$ , then  $(\mathcal{A}, \nu)$  also belongs to  $\mathcal{IFM}$ .

*Remark 2.19* ([2, Rem. 0.10]) The intersection of  $\mathcal{IFM}$  with the class of simple arrangements is  $\mathcal{IF}$ .

As for simple arrangements, multiarrangements of rank at most 2 are inductively free. For, it follows from [18, Cor. 7] that every such multiarrangement is free. We can pick any chain of free subarrangements for such a multiarrangement starting with  $\Phi_2$ . By Theorem 2.12, for any member  $(\mathcal{A}, \nu)$  of such a chain, we have  $\exp(\mathcal{A}'', \nu^*) \subseteq \exp(\mathcal{A}', \nu')$ , so that Definition 2.18(ii) is satisfied and so the claim follows by induction on  $|\nu|$ .

*Remark 2.20* Suppose that  $(\mathcal{A}, \nu) \in \mathcal{IFM}$ . Then by Definition 2.18 there exists a chain of inductively free submultiarrangements, starting with the empty arrangement

$$\Phi_\ell \subseteq (\mathcal{A}_1, \nu_1) \subseteq (\mathcal{A}_2, \nu_2) \subseteq \dots \subseteq (\mathcal{A}_n, \nu_n) = (\mathcal{A}, \nu)$$

such that each consecutive pair obeys Definition 2.18(ii). In particular,  $|\nu_i| = i$  for each  $1 \leq i \leq n$  and so  $|\nu| = n$ . Letting  $H_i$  be the hyperplane in the  $i$ th inductive step, we have  $(\mathcal{A}_i'', \nu_i^*) = (\mathcal{A}_i^{H_i}, \nu_i^*)$ . In particular,  $(\mathcal{A}, \nu) = \{H_1, \dots, H_n\}$  as a multiset. So a fixed hyperplane may occur as one of the  $H_i$  for different indices  $i$ . We frequently refer to a sequence as above as an *inductive chain* of  $(\mathcal{A}, \nu)$ .

As in the simple case, Theorem 2.12 also motivates the notion of recursive freeness for multiarrangements, cf. [10, Def. 4.60].

**Definition 2.21** The class  $\mathcal{RFM}$  of *recursively free* multiarrangements is the smallest class of arrangements subject to

- (i)  $\Phi_\ell \in \mathcal{RFM}$  for each  $\ell \geq 0$ ;
- (ii) for a multiarrangement  $(\mathcal{A}, \nu)$ , if there exists a hyperplane  $H_0 \in \mathcal{A}$  such that both  $(\mathcal{A}', \nu')$  and  $(\mathcal{A}'', \nu^*)$  belong to  $\mathcal{RFM}$ , and  $\exp(\mathcal{A}'', \nu^*) \subseteq \exp(\mathcal{A}', \nu')$ , then  $(\mathcal{A}, \nu)$  also belongs to  $\mathcal{RFM}$ ;
- (iii) for a multiarrangement  $(\mathcal{A}, \nu)$ , if there exists a hyperplane  $H_0 \in \mathcal{A}$  such that both  $(\mathcal{A}, \nu)$  and  $(\mathcal{A}'', \nu^*)$  belong to  $\mathcal{RFM}$ , and  $\exp(\mathcal{A}'', \nu^*) \subseteq \exp(\mathcal{A}, \nu)$ , then  $(\mathcal{A}', \nu')$  also belongs to  $\mathcal{RFM}$ .

By Definitions 2.18 and 2.21,  $\mathcal{IFM} \subseteq \mathcal{RFM}$ .

*Remark 2.22* Suppose that  $(\mathcal{A}, \nu) \in \mathcal{RFM}$ . It follows from Definition 2.21 that there exists a chain of recursively free submultiarrangements, starting with the empty arrangement

$$\Phi_\ell \subseteq (\mathcal{A}_1, \nu_1) \subseteq (\mathcal{A}_2, \nu_2) \dots (\mathcal{A}_n, \nu_n) = (\mathcal{A}, \nu)$$

such that each consecutive pair obeys Definition 2.21. In particular,  $|\nu_i| = |\nu_{i-1}| \pm 1$  for each  $1 \leq i \leq n$  and  $|\nu| = n$ . We also refer to a sequence as above as a *recursive chain* of  $(\mathcal{A}, \nu)$ .

### 2.7 Hereditary Inductive Freeness for Multiarrangements

It is tempting to define the notion of a hereditarily inductively free multiarrangement simply by iterating the construction of the Euler multiplicity from Definition 2.11. However, the following two examples demonstrate that the resulting multiplicity on the restriction depends on the order in which the iteration is taking place. The first is an instance of a constant multiplicity while the second is an example of a multiplicity concentrated at a single hyperplane, cf. Definition 2.16. Thus, such a notion is only well-defined with respect to a fixed total order on  $\mathcal{A}$ . We introduce such a notion in Definition 2.25 below without further pursuing it seriously, because of its lack of uniqueness.

*Example 2.23* Define the rank 3 multiarrangement  $(\mathcal{A}, \nu)$  by

$$Q((\mathcal{A}, \nu)) = x^2y^2(x+y)^2(x+z)^2(y+z)^2.$$

Let  $H_1 := \ker x$ ,  $H_2 := \ker(y+z)$  and  $Y := H_1 \cap H_2$ . Then we have  $(\mathcal{A}^{H_1}, \nu^{*1}) = (\mathcal{A}^{H_1}, (3, 2, 2))$  and  $(\mathcal{A}^{H_2}, \nu^{*2}) = (\mathcal{A}^{H_2}, (2, 2, 2, 2))$ , by [2, Prop. 4.1(6)]. Moreover, we have

$$\left( (\mathcal{A}^{H_1})^{H_1 \cap H_2}, (\nu^{*1})^{*12} \right) = (\mathcal{A}^Y, (3)) \neq (\mathcal{A}^Y, (4)) = \left( (\mathcal{A}^{H_2})^{H_1 \cap H_2}, (\nu^{*2})^{*12} \right),$$

according to [2, Prop. 4.1(7), (6)].

*Example 2.24* Let  $\mathcal{A} = \mathcal{A}(G(3, 3, 3))$  be the reflection arrangement of the unitary reflection group  $G(3, 3, 3)$  with defining polynomial

$$Q(\mathcal{A}(G(3, 3, 3))) = (x^3 - y^3)(x^3 - z^3)(y^3 - z^3).$$

Fix  $H_1 := \ker(x - y) \in \mathcal{A}$ ,  $m_1 \in \mathbb{Z}_{>1}$  and let  $\delta = \delta_{H_1, m_1}$  be as in Definition 2.16. Let  $H_2 := \ker(x - z)$  and  $Y := H_1 \cap H_2$ . Then  $(\mathcal{A}^{H_1}, \delta^{*1}) = (\mathcal{A}^{H_1}, (2, 2, 2, 2))$ , owing to [2, Prop. 4.1(2)], and  $((\mathcal{A}^{H_1})^{H_1 \cap H_2}, (\delta^{*1})^{*12}) = (\mathcal{A}^Y, (4))$ , thanks to [2, Prop. 4.1(6)]. On the other hand we have  $(\mathcal{A}^{H_2}, \delta^{*2}) = (\mathcal{A}^{H_2}, (m_1, 1, 1, 1))$ ,

by [2, Prop. 4.1(3)] and  $((\mathcal{A}^{H_2})^{H_1 \cap H_2}, (\delta^{*2})^{*12}) = (\mathcal{A}^Y, (3))$  for  $m_1 \geq 3$ , by [2, Prop. 4.1(2)], and for  $m_1 = 2$  by [2, Prop. 4.1(4)]. Therefore,

$$\left( (\mathcal{A}^{H_1})^{H_1 \cap H_2}, (\delta^{*1})^{*12} \right) = (\mathcal{A}^Y, (4)) \neq (\mathcal{A}^Y, (3)) = \left( (\mathcal{A}^{H_2})^{H_1 \cap H_2}, (\delta^{*2})^{*12} \right).$$

In view of these examples, we extend the construction of a restriction of a multiarrangement to a hyperplane from Definition 2.11 to restrictions of arbitrary members of  $L(\mathcal{A})$  as follows.

**Definition 2.25** Fix a total order  $<$  on  $\mathcal{A}$ . Let  $Y \in L(\mathcal{A})$  be of rank  $m$ . Then  $<$  descends to give a total order on  $\mathcal{A}_Y$ . Then pick  $H_1 < \dots < H_m$  in  $\mathcal{A}_Y$  minimally with respect to  $<$  such that  $Y = H_1 \cap \dots \cap H_m$ . One readily checks that

$$\mathcal{A}^Y = \left( \dots \left( (\mathcal{A}^{H_1})^{H_1 \cap H_2} \right) \dots \right)^{H_1 \cap \dots \cap H_m}. \tag{1}$$

Note that this is independent of the chosen order on  $\{H_1, \dots, H_m\}$ . Because of (1), if  $\nu$  is a multiplicity on  $\mathcal{A}$ , we can iterate the Euler multiplicity on consecutive restrictions in (1) to obtain the *restricted multiarrangement* on  $\mathcal{A}^Y$  with corresponding *Euler multiplicity* which we denote again just by  $\nu^*$  for simplicity

$$(\mathcal{A}^Y, \nu^*) := \left( \left( \dots \left( (\mathcal{A}^{H_1})^{H_1 \cap H_2} \right) \dots \right)^{H_1 \cap \dots \cap H_m}, \left( \dots (\nu^*)^* \dots \right)^* \right).$$

As demonstrated in Examples 2.23 and 2.24, the construction of  $(\mathcal{A}^Y, \nu^*)$  in Definition 2.25 depends on the chosen order of the iterated Euler multiplicities. Nevertheless, using Lemma 2.14(ii) repeatedly, we get compatibility of restricted multiarrangements with taking localizations.

**Corollary 2.26** Fix an order on  $\mathcal{A}$ . Let  $(\mathcal{A}, \nu)$  be a multiarrangement,  $X \in L(\mathcal{A})$  and  $Y \in L(\mathcal{A}_X)$ . Then with the notation as in Definition 2.25, we have

$$((\mathcal{A}_X)^Y, (\nu_X)^*) = ((\mathcal{A}^Y)_X, (\nu^*)_X).$$

**Definition 2.27** Fix an order on  $\mathcal{A}$ . The multiarrangement  $(\mathcal{A}, \nu)$  is called *hereditarily inductively free* (with respect to the order on  $\mathcal{A}$ ) provided  $(\mathcal{A}^Y, \nu^*)$  is inductively free for every  $Y \in L(\mathcal{A})$ . We abbreviate this class by  $\mathcal{HIFM}$ .

Clearly,  $\mathcal{HIFM} \subseteq \mathcal{IFM}$ . With the aid of Corollary 2.26 and Theorem 1.3 for  $\mathcal{IFM}$ , one can extend the latter to the class  $\mathcal{HIFM}$ . Also, using Theorem 1.4, one readily obtains the compatibility of  $\mathcal{HIFM}$  with the product construction for multiarrangements. We leave the details to the interested reader.

### 3 Proofs of Theorems 1.3 and 1.4

#### 3.1 Inductive and Recursive Freeness of Localizations of Multiarrangements

The following is a reformulation of Theorem 1.3.

**Theorem 3.1** *Let  $U \subseteq V$  be a subspace and let  $(\mathcal{A}, \nu)$  be a multiarrangement in  $V$ .*

- (i) *If  $(\mathcal{A}, \nu)$  is inductively free, then so is the localization  $(\mathcal{A}_U, \nu_U)$ .*
- (ii) *If  $(\mathcal{A}, \nu)$  is recursively free, then so is the localization  $(\mathcal{A}_U, \nu_U)$ .*

*Proof* We readily reduce to the case where we localize with respect to a space  $X$  belonging to the intersection lattice of  $\mathcal{A}$ . For, letting  $X = \bigcap_{H \in \mathcal{A}_U} H \in L(\mathcal{A})$ , we have  $\mathcal{A}_X = \mathcal{A}_U$ .

(i) We argue by induction on the rank  $r(\mathcal{A})$ . If  $r(\mathcal{A}) \leq 3$ , then  $r(\mathcal{A}_X) \leq 2$  for  $X \neq T_{\mathcal{A}}$ , so the result follows thanks to [18, Cor. 7].

So suppose  $(\mathcal{A}, \nu)$  is inductively free of rank  $r > 3$  and that the statement holds for all inductively free multiarrangements of rank less than  $r$ .

Since  $(\mathcal{A}, \nu)$  is inductively free, there is an inductive chain  $(\mathcal{A}_i, \nu_i)$  of  $(\mathcal{A}, \nu)$ , where  $|\nu_i| = i$ , for  $i = 1, \dots, n = |\nu|$ , see Remark 2.20. Then thanks to Lemma 2.1(i), we have

$$\mathcal{A}_X \cap \mathcal{A}_i = (\mathcal{A}_i)_X. \tag{2}$$

For  $H \in \mathcal{A}_X \cap \mathcal{A}_i$ , we have  $H \leq X$ , and so by (2) and Lemma 2.1(ii),

$$(\mathcal{A}_X \cap \mathcal{A}_i)^H = ((\mathcal{A}_i)_X)^H = (\mathcal{A}_i^H)_X. \tag{3}$$

Consequently, localizing each member of the sequence  $(\mathcal{A}_i, \nu_i)$  at  $X$ , removing redundant terms if necessary and reindexing the resulting distinct multiarrangements, we obtain the following sequence of submultiarrangements of  $(\mathcal{A}_X, \nu_X)$ ,

$$(\mathcal{A}_{1,X}, \nu_{1,X}) \subsetneq (\mathcal{A}_{2,X}, \nu_{2,X}) \subsetneq \dots \subsetneq (\mathcal{A}_{m,X}, \nu_{m,X}) = (\mathcal{A}_X, \nu_X), \tag{4}$$

where  $\mathcal{A}_{i,X}$  is short for  $(\mathcal{A}_i)_X$  and  $\nu_{i,X}$  for  $\nu_i|_{(\mathcal{A}_i)_X}$ . In particular,  $|\nu_{i,X}| = i$  and  $m = |\nu_X|$ . We claim that (4) is an inductive chain of  $(\mathcal{A}_X, \nu_X)$ .

Now let  $H_i \in \mathcal{A}_X \cap \mathcal{A}_i = \mathcal{A}_{i,X}$  be the relevant hyperplane in the  $i$ th step in the sequence (4). Let  $(\mathcal{A}_{i,X}, \nu_{i,X})$ ,  $(\mathcal{A}'_{i,X}, \nu'_{i,X})$  and  $(\mathcal{A}''_{i,X}, \nu^*_{i,X})$  be the triple with respect to  $H_i$ .

Note that, since  $(\mathcal{A}_{i-1,X}, \nu_{i-1,X}) \subsetneq (\mathcal{A}_{i,X}, \nu_{i,X})$ , it follows from Definitions 2.5 and 2.11 that  $(\mathcal{A}_i, \nu_i), (\mathcal{A}'_i, \nu'_i) = (\mathcal{A}_{i-1}, \nu_{i-1})$  and  $(\mathcal{A}''_i, \nu^*_i)$  is the triple with respect

to  $H_i$ . Therefore, by the construction of the chain in (4) and Lemma 2.14(i), we have

$$((\mathcal{A}_{i,X})', v'_{i,X}) = ((\mathcal{A}'_i)_X, (v'_i)_X) = (\mathcal{A}_{i-1,X}, v_{i-1,X}). \tag{5}$$

Since  $(\mathcal{A}_i, v_i)$  is free by assumption, it follows from Theorem 2.7 that  $(\mathcal{A}_{i,X}, v_{i,X})$  is free for each  $i$ . Consequently, it follows from Remark 2.13 and (5) that also each restriction  $((\mathcal{A}_{i,X})^{H_i}, v_{i,X}^*) = (\mathcal{A}''_{i,X}, v_{i,X}^*)$  is free with exponents given by Theorem 2.12(iii).

Since  $(\mathcal{A}_i^{H_i}, v_i^*) = (\mathcal{A}''_i, v_i^*)$  is inductively free by assumption and  $r(\mathcal{A}''_i) < r$ , it follows from our induction hypothesis that the localization  $((\mathcal{A}''_i)_X, (v_i^*)_X)$  is also inductively free for each  $i$ . Thus, thanks to (3) and Lemma 2.14(ii),

$$(\mathcal{A}''_{i,X}, v_{i,X}^*) = ((\mathcal{A}_{i,X})'', (v_{i,X})^*) = ((\mathcal{A}''_i)_X, (v_i^*)_X)$$

is inductively free for each  $i$ .

Since the rank of  $\mathcal{A}_{1,X}$  is 1,  $(\mathcal{A}_{1,X}, v_{1,X})$  is inductively free. Together with the fact that each of the restrictions  $(\mathcal{A}''_{i,X}, v_{i,X}^*)$  is also inductively free for each  $i$ , a repeated application of the addition part of Theorem 2.12 then shows that the sequence (4) is an inductive chain of  $(\mathcal{A}_X, v_X)$ , satisfying Definition 2.18, as claimed.

(ii) The argument is very similar to the one above. We argue again by induction on the rank  $r(\mathcal{A})$ . If  $r(\mathcal{A}) \leq 3$ , then  $r(\mathcal{A}_X) \leq 2$  for  $X \neq T_{\mathcal{A}}$ , so the result follows by [18, Cor. 7].

So suppose  $(\mathcal{A}, v)$  is recursively free of rank  $r > 3$  and that the statement holds for all recursively free multiarrangements of rank less than  $r$ .

Since  $(\mathcal{A}, v)$  is recursively free, there is a recursive chain  $(\mathcal{A}_i, v_i)$  of  $(\mathcal{A}, v)$ , where  $|v_i| = |v_{i-1}| \pm 1$  for  $i = 1, \dots, n$ , and  $(\mathcal{A}_n, v_n) = (\mathcal{A}, v)$ , see Remark 2.22.

Since  $X$  is a subspace in  $V$ , as above, we can consider the localization  $(\mathcal{A}_{i,X}, v_{i,X})$  of each member of the recursive chain, where again  $\mathcal{A}_{i,X}$  is short for  $(\mathcal{A}_i)_X$  and  $v_{i,X}$  for  $v_i|_{(\mathcal{A}_i)_X}$ , cf. (2).

Then removing redundant terms and reindexing the resulting distinct multiarrangements if needed, we obtain a sequence of multiarrangements starting with the empty arrangement

$$\Phi_\ell \neq (\mathcal{A}_{1,X}, v_{1,X}) \subsetneq (\mathcal{A}_{2,X}, v_{2,X}) \subsetneq \dots (\mathcal{A}_{m,X}, v_{m,X}) = (\mathcal{A}_X, v_X), \tag{6}$$

where by construction, at each stage we either increase or decrease the multiplicity of a single hyperplane by 1.

Since  $(\mathcal{A}_i, v_i)$  is free by assumption, it follows from Theorem 2.7 that  $(\mathcal{A}_{i,X}, v_{i,X})$  is free for each  $i$ , and so (6) is a chain of free submultiarrangements of  $(\mathcal{A}_X, v_X)$ .

Now fix  $i$  and let  $H$  be the relevant hyperplane in the  $i$ th step in the sequence (6) above, i.e., the multiplicity of  $H$  is either increased or decreased in this step. In the first instance, letting  $(\mathcal{A}_{i,X}, v_{i,X}), (\mathcal{A}'_{i,X}, v'_{i,X}) = (\mathcal{A}_{i-1,X}, v_{i-1,X})$ , and  $(\mathcal{A}''_{i,X}, v_{i,X}^*)$  be the triple with respect to  $H$ , we are in the situation of (5) above. On the other hand, if the multiplicity of  $H$  is decreased in this step, then let  $(\mathcal{A}_{i-1,X}, v_{i-1,X}),$



$(\mathcal{A}'_{i-1,X}, v'_{i-1,X}) = (\mathcal{A}_{i,X}, v_{i,X})$ , and  $(\mathcal{A}''_{i-1,X}, v^*_{i-1,X})$  be the triple with respect to  $H$ . In the first instance we argue as in (i) above to see that  $((\mathcal{A}_{i,X})^H, v^*_{i,X}) = (\mathcal{A}''_{i,X}, v^*_{i,X})$  is free with exponents given by Theorem 2.12(iii). In the second case we argue in just the same way to get that  $((\mathcal{A}_{i-1,X})^H, v^*_{i-1,X}) = (\mathcal{A}''_{i-1,X}, v^*_{i-1,X})$  is free with exponents given by Theorem 2.12(iii).

If  $(\mathcal{A}_X, v_X)$  is inductively free, then it is recursively free and we are done. So we may assume that  $(\mathcal{A}_X, v_X)$  is not inductively free. Then in particular, the sequence (6) is not an inductive chain. We claim that (6) is a recursive chain of  $(\mathcal{A}_X, v_X)$ . Clearly, the initial part of this sequence is necessarily a chain of inductively free arrangements (one needs to add hyperplanes first before one can start removing them again). Let  $k$  be maximal so that

$$\Phi_\ell \neq (\mathcal{A}_{1,X}, v_{1,X}) \subsetneq (\mathcal{A}_{2,X}, v_{2,X}) \subsetneq \dots \subsetneq (\mathcal{A}_{k,X}, v_{k,X}) \tag{7}$$

is a sequence of inductively free terms in the chain (6). Then in particular,  $(\mathcal{A}_{k,X}, v_{k,X})$  is inductively free, hence recursively free.

Since  $(\mathcal{A}''_i, v^*_i)$  is recursively free by assumption and  $r(\mathcal{A}''_i) < r$ , it follows from our induction hypothesis that the localization  $((\mathcal{A}''_i)_X, (v^*_i)_X)$  is also recursively free for each  $i$ . Thus, thanks to Lemma 2.14(ii),  $((\mathcal{A}_i)_X)^H, v^*_{i,X}) = ((\mathcal{A}''_i)_X, (v^*_i)_X)$  is recursively free for each  $i$ .

In particular, returning to the sequence (7) and the  $(k + 1)$ -st step, where we reduce a multiplicity for the first time in the chain in (6), it follows from the argument above that

$$\exp((\mathcal{A}_{k,X})^{H_{k+1}}, v^*_{k,X}) \subseteq \exp(\mathcal{A}_{k,X}, v_{k,X}).$$

Therefore, applying the deletion part of Theorem 2.12 and using Lemma 2.14(i), it follows that

$$(((\mathcal{A}_{k,X})', (v_{k,X})') = ((\mathcal{A}'_k)_X, (v'_k)_X) = (\mathcal{A}_{k+1,X}, v_{k+1,X})$$

is recursively free, where the deletion is with respect to  $H_{k+1}$ . Now iterate this process.

The special case when  $v \equiv 1$  in Theorem 3.1 gives Theorem 1.1 for the classes  $\mathcal{IF}$  and  $\mathcal{RF}$ . Armed with Theorem 1.1 for  $\mathcal{IF}$ , we obtain the statement of Theorem 1.1 for the class  $\mathcal{HSF}$ .

**Corollary 3.2** *Let  $U \subseteq V$  be a subspace and let  $\mathcal{A}$  be an arrangement in  $V$ . If  $\mathcal{A}$  is hereditarily inductively free, then so is the localization  $\mathcal{A}_U$ .*

*Proof* As before, for  $X = \bigcap_{H \in \mathcal{A}_U} H$ , we have  $\mathcal{A}_X = \mathcal{A}_U$  and  $X \in L(\mathcal{A})$ . Let  $Y \in L(\mathcal{A}_X)$ . Then  $Y \leq X$  in  $L(\mathcal{A})$ . Since  $\mathcal{A}$  is hereditarily inductively free,  $\mathcal{A}^Y$  is inductively free. So by Theorem 1.1 and Lemma 2.1(ii), we get that  $(\mathcal{A}_X)^Y = (\mathcal{A}^Y)_X$  is inductively free.

Theorem 1.1 thus follows from Theorem 3.1 and Corollary 3.2.

*Remark 3.3* It is worth noting that the proof of Theorem 3.1 shows that any given inductive (resp. recursive) chain of the ambient multiarrangement descends to give an inductive (resp. recursive) chain of any localization.

### 3.2 Products of Inductively Free and Recursively Free Arrangements

Thanks to [10, Prop. 4.28], the product of two arrangements is free if and only if each factor is free. In [7, Prop. 2.10], the first two authors showed that this factorization property descends to the class of inductively free arrangements.

Let  $(\mathcal{A}_i, \nu_i)$  be a multiarrangement in  $V_i$  for  $i = 1, 2$ . We consider the product  $(\mathcal{A} := \mathcal{A}_1 \times \mathcal{A}_2, \nu)$  which is a multiarrangement in  $V = V_1 \oplus V_2$  with multiplicity  $\nu := \nu_1 \times \nu_2$ , see [2].

The following is just a reformulation of Theorem 1.4.

**Theorem 3.4** *Let  $(\mathcal{A}_i, \nu_i)$  be a multiarrangement in  $V_i$  for  $i = 1, 2$ . Then the product  $(\mathcal{A}, \nu)$  is inductively free (resp. recursively free) if and only if each factor  $(\mathcal{A}_i, \nu_i)$  is inductively free (resp. recursively free).*

*Proof* We just give the argument for the case of recursive freeness, the argument for inductive freeness is identical.

The reverse implication is straightforward, cf. [7, Prop. 2.10]. For the forward implication, assume that  $(\mathcal{A}, \nu)$  is recursively free. Set  $X_1 := T_{\mathcal{A}_1} \oplus V_2$  and  $X_2 := V_1 \oplus T_{\mathcal{A}_2}$ . Then both  $X_1$  and  $X_2$  belong to the intersection lattice of  $\mathcal{A}$ , [10, Prop. 2.14]. Note that  $\mathcal{A}_{X_1} = \{H_1 \oplus V_2 \mid H_1 \in \mathcal{A}_1\} \cong \mathcal{A}_1$  and  $\mathcal{A}_{X_2} = \{V_1 \oplus H_2 \mid H_2 \in \mathcal{A}_2\} \cong \mathcal{A}_2$ . It thus follows from Theorem 1.3 that both  $(\mathcal{A}_{X_1}, \nu_{X_1}) = (\mathcal{A}_1, \nu_1)$  and  $(\mathcal{A}_{X_2}, \nu_{X_2}) = (\mathcal{A}_2, \nu_2)$  are recursively free.

The special case of Theorem 3.4 for  $\mathcal{RFM}$  when  $\nu_i \equiv 1$  gives Theorem 1.2.

## 4 Applications to Reflection Arrangements

### 4.1 Inductive Freeness of Reflection Arrangements

In this section we demonstrate how Theorem 1.1 can be used to show that certain arrangements are not inductively free.

Let  $W$  be one of the exceptional complex reflection groups  $G_{29}$ ,  $G_{33}$ , or  $G_{34}$ . Then by [10, Tables C.10, C.14, C.15],  $W$  admits a parabolic subgroup  $W_X$  of type  $G(4, 4, 3)$ ,  $G(3, 3, 4)$ , or  $G(3, 3, 5)$ , respectively. Thanks to [7, Prop. 3.2], the reflection arrangement of  $G(r, r, \ell)$  is not inductively free for  $r, \ell \geq 3$ . By [10, Cor. 6.28], we have  $\mathcal{A}(W_X) = \mathcal{A}(W)_X$  and so it follows from Theorem 1.1, that

$\mathcal{A}(W)$  is not inductively free in each of the three instances. This was proved in [7, §3.1.4] by different means.

### 4.2 Inductive Freeness of Ziegler’s Canonical Multiplicity for Monomial Groups

Theorem 1.3 is very useful in showing that a given multiarrangement is not inductively free by exhibiting a suitable localization which is known to not be inductively free. We demonstrate this in the following results.

Let  $\mathcal{A} = \mathcal{A}(W)$  be the reflection arrangement of the complex reflection group  $W := G(r, r, \ell)$  for  $r, \ell \geq 3$ . Let  $H_{ij}(\zeta) := \ker(x_i - \zeta x_j) \in \mathcal{A}$ , where  $1 \leq i < j \leq \ell$  and  $\zeta$  is an  $r$ th root of unity and let  $H_i := \ker x_i$  be the  $i$ th coordinate hyperplane for  $1 \leq i \leq \ell$ , [10, §6.4].

Using results from [7] and Theorem 1.3, we show that  $(\mathcal{A}'', \kappa)$  fails to be inductively free for  $\ell \geq 5$ , where  $\kappa$  is Ziegler’s canonical multiplicity from Definition 2.9.

In view of Proposition 2.17, we also consider a concentrated multiplicity on  $\mathcal{A}$ . Fix  $H_0 \in \mathcal{A}$ ,  $m_0 \in \mathbb{Z}_{>1}$  and let  $\delta = \delta_{H_0, m_0}$  be as in Definition 2.16. Then, since  $\mathcal{A}$  is free (cf. [10, §6.3]) so is  $(\mathcal{A}, \delta)$ , by Proposition 2.17(i). Let  $\mathcal{A}'''$  be the restriction of  $\mathcal{A}$  with respect to  $H_0$ .

**Proposition 4.1** *Let  $\mathcal{A} = \mathcal{A}(W)$  be the reflection arrangement of  $W = G(r, r, \ell)$  for  $r \geq 3, \ell \geq 5$ . Then both  $(\mathcal{A}'', \kappa)$  and  $(\mathcal{A}, \delta)$  are not inductively free.*

*Proof* Since  $W$  is transitive on  $\mathcal{A}$ , without loss, we may choose  $H_0 := H_{1,2}(1) = \ker(x_1 - x_2)$ . Define

$$X := \bigcap_{3 \leq i < j \leq \ell} H_{ij}(\zeta) = \bigcap_{3 \leq i \leq \ell} H_i.$$

Then  $X$  is of rank  $\ell - 2$  in  $L(\mathcal{A})$ .

Set  $Y_{ij}(\zeta) := H_0 \cap H_{ij}(\zeta) \in \mathcal{A}'''$ . Then one readily checks that for  $Y \in \mathcal{A}'''$ , we have

$$\kappa(Y) = \begin{cases} r - 1 & \text{for } Y = Y_{1,2}(\zeta), \\ 2 & \text{for } Y = Y_{1,i}(\zeta), Y_{2,i}(\zeta) \text{ and } 3 \leq i \leq \ell, \\ 1 & \text{for } Y = Y_{i,j}(\zeta) \text{ and } 3 \leq i < j \leq \ell. \end{cases} \tag{8}$$

According to (8), the multiplicity  $\kappa_X$  of the localization  $((\mathcal{A}''')_X, \kappa_X)$  satisfies  $\kappa_X \equiv 1$ . Thus,  $((\mathcal{A}''')_X, \kappa_X)$  is isomorphic to the simple reflection arrangement  $\mathcal{A}(G(r, r, \ell - 2))$ . According to [7, Prop. 3.2], the latter is not inductively free, as  $\ell \geq 5$ . Therefore,  $(\mathcal{A}'', \kappa)$ , is not inductively free either, thanks to Theorem 1.3.

By definition of  $(\mathcal{A}, \delta)$  and [10, Cor. 6.28], we have  $\mathcal{A}(W_X) = \mathcal{A}(W)_X \cong \mathcal{A}(G(r, r, \ell - 2))$  and  $\delta_X \equiv 1$ . Consequently,  $(\mathcal{A}_X, \delta_X)$  is also isomorphic to the simple reflection arrangement  $\mathcal{A}(G(r, r, \ell - 2))$ . Again, thanks to [7, Prop. 3.2], the latter is not inductively free, as  $\ell \geq 5$ . Therefore,  $(\mathcal{A}, \delta)$  is not inductively free, owing to Theorem 1.3.

Proposition 4.1 generalizes to a larger class of multiarrangements stemming from complex reflection groups. Orlik and Solomon defined complex  $\ell$ -arrangements  $\mathcal{A}_\ell^k(r)$  in [9, §2] (cf. [10, §6.4]) which interpolate between the reflection arrangements of the complex reflection groups  $G(r, r, \ell)$  and  $G(r, 1, \ell)$ . For  $r, \ell \geq 3$  and  $0 \leq k \leq \ell$  the defining polynomial of  $\mathcal{A}_\ell^k(r)$  is given by

$$Q(\mathcal{A}_\ell^k(r)) = x_1 \cdots x_k \prod_{1 \leq i < j \leq \ell} (x_i^r - x_j^r),$$

so that  $\mathcal{A}_\ell^k(r) = \mathcal{A}(G(r, 1, \ell))$  and  $\mathcal{A}_\ell^0(r) = \mathcal{A}(G(r, r, \ell))$ .

Again fix  $H_0 \in \mathcal{A}$ ,  $m_0 \in \mathbb{Z}_{>1}$  and let  $\delta = \delta_{H_0, m_0}$  be as in Definition 2.16. Then, since  $\mathcal{A}$  is free (cf. [10, Prop. 6.85]), so is  $(\mathcal{A}, \delta)$ , by Proposition 2.17(i). Let  $\mathcal{A}''$  be the restriction of  $\mathcal{A}$  with respect to  $H_0$ . Then  $(\mathcal{A}'', \delta^*) = (\mathcal{A}'', \kappa)$  is free, thanks to Proposition 2.17(ii).

Combining results from [7] with Theorem 1.3, we show that also for these more general arrangements  $(\mathcal{A}'', \kappa)$  fails to be inductively free provided  $\ell \geq 5$ ,  $0 \leq k \leq \ell - 3$  and  $H_0$  is of the form  $H_{i,j}(\zeta)$ .

**Proposition 4.2** *Let  $\mathcal{A} = \mathcal{A}_\ell^k(r)$  for  $r \geq 3, \ell \geq 5$  and  $0 \leq k \leq \ell - 3$ . Fix  $H_0 = H_{i,j}(\zeta) \in \mathcal{A}$ . Then both  $(\mathcal{A}'', \kappa)$  and  $(\mathcal{A}, \delta)$  are not inductively free.*

*Proof* For  $k = 0$ , this is just Proposition 4.1. So we may assume that  $1 \leq k \leq \ell - 3$ .

We may suppose without loss that  $H_0 := H_{1,2}(1) = \ker(x_1 - x_2)$ . Set  $Y_{i,j}(\zeta) := H_0 \cap H_{i,j}(\zeta)$  and  $Y_i := H_0 \cap H_i$  in  $\mathcal{A}''$ . Then one readily checks that for  $Y \in \mathcal{A}''$ , we have

$$\kappa(Y) = \begin{cases} r + 1 \text{ (resp. } r) & \text{for } Y = Y_{1,2}(\zeta) \text{ and } k \geq 2 \text{ (resp. } k = 1), \\ 2 & \text{for } Y = Y_{1,i}(\zeta), Y_{2,i}(\zeta) \text{ and } 3 \leq i \leq \ell, \\ 1 & \text{for } Y = Y_{i,j}(\zeta) \text{ and } 3 \leq i < j \leq \ell, \\ 1 & \text{for } Y = Y_i \text{ and } 3 \leq i \leq k, \end{cases} \tag{9}$$

where the value of  $\kappa(Y)$  in the first case depends on  $k$  and the last instance only occurs if  $k \geq 3$ .

Define

$$Z := \bigcap_{\ell-2 \leq i < j \leq \ell} H_{i,j}(\zeta) = \bigcap_{\ell-2 \leq i \leq \ell} H_i,$$

which is of rank 3 in  $L(\mathcal{A})$  and

$$X := H_0 \cap Z = H_0 \cap \left( \bigcap_{\ell-2 \leq i \leq \ell} H_i \right),$$

which is of rank 3 in  $L(\mathcal{A}'')$ . According to (9), the multiplicity  $\kappa_X$  of the localization  $((\mathcal{A}'')_X, \kappa_X)$  satisfies  $\kappa_X \equiv 1$ . Thus, it follows from the construction and the hypotheses  $\ell \geq 5$  and  $0 \leq k \leq \ell - 3$  that the localization  $((\mathcal{A}'')_X, \kappa_X)$  is isomorphic to the simple reflection arrangement  $\mathcal{A}(G(r, r, 3))$ . Owing to [7, Prop. 3.2], the latter is not inductively free. Therefore,  $(\mathcal{A}'', \kappa)$  is not inductively free, by Theorem 1.3.

By the definition of  $(\mathcal{A}, \delta)$  and the fact that  $\ell \geq 5$  and  $0 \leq k \leq \ell - 3$ , we have  $\delta_Z \equiv 1$ , and so  $(\mathcal{A}_Z, \delta_Z)$  is isomorphic to the simple reflection arrangement  $\mathcal{A}(G(r, r, 3))$ . So again by [7, Prop. 3.2], the latter is not inductively free. Therefore,  $(\mathcal{A}, \delta)$  is not inductively free either, thanks to Theorem 1.3.

**Acknowledgements** We acknowledge support from the DFG-priority program SPP1489 ‘‘Algorithmic and Experimental Methods in Algebra, Geometry, and Number Theory’’.

## References

1. T. Abe, K. Nuida, Y. Numata, Signed-eliminable graphs and free multiplicities on the braid arrangement. *J. Lond. Math. Soc. (2)* **80**(1), 121–134 (2009)
2. T. Abe, H. Terao, M. Wakefield, The Euler multiplicity and addition-deletion theorems for multiarrangements. *J. Lond. Math. Soc. (2)* **77**(2), 335–348 (2008)
3. T. Abe, M. Yoshinaga, Free arrangements and coefficients of characteristic polynomials. *Math. Z.* **275**(3–4), 911–919 (2013)
4. N. Amend, T. Hoge, G. Röhrle, On inductively free restrictions of reflection arrangements. *J. Algebra* **418**, 197–212 (2014)
5. M. Barakat, M. Cuntz, Coxeter and crystallographic arrangements are inductively free. *Adv. Math.* **229**, 691–709 (2012)
6. M. Cuntz, T. Hoge, Free but not recursively free arrangements. *Proc. Am. Math. Soc.* **143**, 35–40 (2015)
7. T. Hoge, G. Röhrle, On inductively free reflection arrangements. *J. Reine Angew. Math.* **701**, 205–220 (2015)
8. P. Mücksch, On recursively free reflection arrangements. *J. Algebra* **474**, 24–48 (2017)
9. P. Orlik, L. Solomon, Arrangements defined by unitary reflection groups. *Math. Ann.* **261**, 339–357 (1982)
10. P. Orlik, H. Terao, *Arrangements of Hyperplanes* (Springer, Berlin, 1992)
11. M. Schulze, Freeness and multirestriction of hyperplane arrangements. *Compos. Math.* **148**(3), 799–806 (2012)
12. H. Terao, Arrangements of hyperplanes and their freeness I, II. *J. Fac. Sci. Univ. Tokyo* **27**, 293–320 (1980)
13. H. Terao, Generalized exponents of a free arrangement of hyperplanes and Shepherd-Todd-Brieskorn formula. *Invent. Math.* **63**(1), 159–179 (1981)
14. H. Terao, Free arrangements of hyperplanes over an arbitrary field. *Proc. Jpn. Acad. Ser. A Math. Sci.* **59**(7), 301–303 (1983)

15. M. Yoshinaga, Characterization of a free arrangement and conjecture of Edelman and Reiner. *Invent. Math.* **157**(2), 449–454 (2004)
16. M. Yoshinaga, On the freeness of 3-arrangements. *Bull. Lond. Math. Soc.* **37**(1), 126–134 (2005)
17. M. Yoshinaga, Freeness of hyperplane arrangements and related topics. *Ann. Fac. Sci. Toulouse Math. (6)* **23**(2), 483–512 (2014)
18. G. Ziegler, Multiarrangements of hyperplanes and their freeness, in *Singularities (Iowa City, IA, 1986)*. *Contemporary Mathematics*, vol. 90 (American Mathematical Society, Providence, RI, 1989), pp. 345–359
19. G. Ziegler, Matroid representations and free arrangements. *Trans. Am. Math. Soc.* **320**, 525–541 (1990)

# Toric Ext and Tor in `polymake` and `Singular`: The Two-Dimensional Case and Beyond



Lars Kastner

**Abstract** It is an open problem to describe the Ext and Tor groups for two torus-invariant Weil divisors on a toric variety, using only the combinatorial data of the underlying objects from toric geometry. We will give a survey on this description for the case of two-dimensional cyclic quotient singularities, in particular how this description is related with the continued fraction associated to a cyclic quotient singularity. Furthermore, we will elaborate on the applications of these modules and expectations of how to generalize the results to higher dimensions, highlighted by examples.

Studying the above problems sparked software development in `polymake` and `Singular`, heavily using the interface between the systems. Examples are accompanied by code snippets to demonstrate the functionality added and to illustrate how one may approach similar problems in toric geometry using these software packages.

**Keywords** Toric geometry • Cyclic quotient singularities • Continued fractions • Polymake • Singular

**Subject Classification** 14M25

## 1 Introduction

Given an affine toric variety  $X$ , it is an open question to determine all maximal Cohen-Macaulay divisor classes in the class group  $\text{Cl } X$  from the combinatorics of  $X$ . Choosing a torus-invariant representative  $D$  for a divisor class, we can consider its polyhedron of global sections  $P_D$  (see [9, (4.3.2)]). The torus invariant global

---

L. Kastner (✉)

Institut für Mathematik, Technische Universität Berlin, Str. des 17. Juni 136, 10623 Berlin, Germany

e-mail: [kastner@math.tu-berlin.de](mailto:kastner@math.tu-berlin.de)

sections of  $\mathcal{O}(D)$  correspond to the lattice points of  $P_D$ . The tail cone of  $P_D$  is the weight cone of  $X$ . Thus, it gives rise to a fractional ideal  $H_D$  over the coordinate ring  $R$  of  $X$ , which is isomorphic to the global sections of the coherent sheaf  $\mathcal{O}(D)$ . The divisor  $D$  being maximal Cohen-Macaulay is the same as  $H_D$  being maximal Cohen-Macaulay over  $R$ , which is a problem of multigraded or combinatorial commutative algebra.

Instead of using the criterion for maximal Cohen-Macaulayness of a module in terms of depth and dimension, we will use the following homological criterion of Yoshino [27, Cor 1.13] as a definition:

**Definition 1.1** A finitely generated  $R$ -module  $A$ , where  $R$  is a Cohen-Macaulay ring with canonical module  $\omega_R$ , is Maximal Cohen-Macaulay (MCM), if and only if

$$\mathrm{Ext}^i(A, \omega_R) = 0 \quad \forall i > 0.$$

In our setting, for  $R$  being the coordinate ring of a toric variety  $X$ , the module  $\omega_R$  arises from a certain torus invariant divisor  $K = K_X$ , i.e. we have  $\omega_R = H_K$ . The coordinate ring  $R$  itself is Cohen-Macaulay due to Hochster's theorem.

The coordinate ring  $R$  of  $X$  has a grading by  $M$ , the character lattice of the torus acting on  $X$ , corresponding to the torus action on  $X$ . The divisors  $K$  and  $D$  being torus-invariant means that the modules  $H_K$  and  $H_D$  are  $M$ -graded as well. At this point it is important to note that for any Weil divisor  $D$ , the module  $H_D$  is finitely generated. Hence, the functor  $\mathrm{Hom}_R(H_D, \bullet)$  maps  $M$ -graded modules to  $M$ -graded modules, for torus-invariant  $D$ . Thus, the modules  $\mathrm{Ext}_R^i(H_D, H_K)$  are  $M$ -graded as well.

There are two approaches for computing  $\mathrm{Ext}(A, B)$ : We may resolve  $A$  projectively, apply  $\mathrm{Hom}(\bullet, B)$  to the free resolution and then take cohomology. Or, we resolve  $B$  injectively, apply  $\mathrm{Hom}(A, \bullet)$  and take homology of the resulting complex. Here, we will discuss the first variant, for the following reasons: First, for the  $M$ -graded module, we can assume its free resolution to be  $M$ -graded as well. In particular, the kernel of every differential is itself isomorphic to a direct sum of other divisorial  $M$ -graded ideals. For  $X$  being Gorenstein, Eisenbud [11] demonstrated that the minimal free resolution of MCM modules over  $R$  becomes periodic of length 2. With the previous statement this implies a certain recursion of the Ext-modules. Applying this to our situation it seems possible to check [Theorem 1.1](#) in finite time, at least in the Gorenstein setting.

In this article we will mainly deal with the first class of affine toric singularities, quotients of  $\mathbb{C}^2$  by a finite group action, so-called cyclic quotient singularities (CQS). Combinatorially these arise from two-dimensional cones in two-dimensional space. Previous work by Wunram [25, 26] states a property for a Weil divisor to be special MCM and gives a theorem for a divisor to satisfy this condition. Recent work by Wemyss and Iyama gives a condition for a Weil divisor (or module) to be special MCM in terms of Ext. The idea behind the special MCM property in the setting of CQS is to extend the McKay correspondence to the non-Gorenstein case [24].



We will survey the results of [18, 19], to give a combinatorial understanding of the results of Wunram. On a CQS, every divisor is MCM, which reflects the fact that every Weil divisor is  $\mathbb{Q}$ -Cartier. The combinatorial construction for Ext gives a new interpretation for this. Furthermore one can generalize the previously mentioned periodicity observation of Eisenbud to the non-Gorenstein case. Using this generalization one can extend the description of Buchweitz [7] of the Yoneda product in the Ext-algebra of the MCM-modules to non-Gorenstein CQS. Many of the results presented for CQS can be applied to affine toric varieties that are products of  $\mathbb{A}^n$  with a CQS as well.

Cyclic quotient singularities have several properties that general affine toric varieties do not share. For example, their defining cone is simplicial and the singularity of the variety is isolated. Gorenstein CQS form a special subclass of all CQS and we will elaborate on the implications of this property. In the end, we will give two examples that elaborate on the pitfalls when trying to extend the results to more general affine toric varieties.

Special emphasis will be put on the accompanying software development in `polymake` [14] and **Singular** [10]. One main goal of the priority program SPP1489 was to build interfaces between these software packages. It is this interface that made the above research and discoveries possible. The general purpose portions of the code have been migrated to the `polymake` core, the other methods can be found in the appendix of [18].

The structure of the paper is as follows: We start by giving the general definitions from the toric world needed to understand the computational problem. After introducing cyclic quotient singularities, we give the full combinatorial description of  $\text{Ext}^i$  and  $\text{Tor}_i$  of two torus-invariant Weil divisors on a CQS. The connection of CQS with continued fractions gives a recursive algorithm for computing  $\text{Ext}^1$ . This is followed by a description of the involved software packages and an overview of the implementation. We conclude by discussing two examples in dimension three, in order to provide intuition for which results might generalize to higher dimensions.

## 2 Preliminaries

In this section we will revise the ingredients needed from toric geometry. We will follow the notation of Cox et al. [9] closely.

Let  $N$  be a lattice and define  $M := \text{Hom}_{\mathbb{Z}}(N, \mathbb{Z})$  to be the dual lattice. Then we define the associated  $\mathbb{Q}$ -vector space  $N_{\mathbb{Q}}$  to be  $N \otimes_{\mathbb{Z}} \mathbb{Q}$ . Analogously we define  $M_{\mathbb{Q}}$ . We will denote the pairing between  $N$  and  $M$  by  $\langle \bullet, \bullet \rangle$ . The extension of the pairing to  $N_{\mathbb{Q}}$  and  $M_{\mathbb{Q}}$  will be denoted by the same symbols. We will denote elements of  $N$  and  $N_{\mathbb{Q}}$  in round brackets  $(\dots)$  and elements of  $M$  and  $M_{\mathbb{Q}}$  in square brackets  $[\dots]$ .

Let  $\sigma \subseteq N_{\mathbb{Q}}$  be a convex polyhedral cone, then the associated semigroup ring and toric variety are

$$R := \mathbb{C}[\sigma^{\vee} \cap M] \text{ and } \text{TV}(\sigma) := \text{Spec } R,$$

where  $\sigma^\vee$  denotes the dual cone of  $\sigma$ , i.e.

$$\sigma^\vee := \{u \in M_{\mathbb{Q}} \mid \langle u, v \rangle \geq 0 \ \forall v \in \sigma\}.$$

This means that the rays of  $\sigma^\vee$  are the inner facet normals of  $\sigma$ . Note that throughout this paper, we assume  $\sigma$  to be pointed and full-dimensional, implying the same for  $\sigma^\vee$ .

*Example 2.1* Take for example the cone  $\sigma$  generated by the four rays

$$(0, 0, 1), (0, 1, 0), (1, 0, -1) \text{ and } (1, -1, 0).$$

Then the dual cone  $\sigma^\vee$  is generated by the following four rays

$$[1, 0, 0], [1, 1, 0], [1, 0, 1] \text{ and } [1, 1, 1].$$

These four primitive ray generators also form the Hilbert basis of the cone  $\sigma^\vee$ . If we label them with the variable names  $w, x, y$  and  $z$ , then  $\text{TV}(\sigma)$  is the hypersurface in  $\mathbb{A}^4$  given by the equation  $wz - xy$ . It is the cone over  $\mathbb{P}^1 \times \mathbb{P}^1$ .

Due to the orbit-cone correspondence [9, 3.2], codimension  $c$  faces of  $\sigma$  correspond to dimension  $c$  orbits of  $\text{TV}(\sigma)$ . Hence the rays of  $\sigma$  give us exactly the orbits of codimension 1. If  $\sigma$  is given as a cone over the rays  $\rho^0, \dots, \rho^n$ , then a torus-invariant Weil divisor  $D$  is a formal linear combination of the closures of the orbits corresponding to these rays:

$$D = \sum_{i=0}^n a_i \cdot \text{orb } \rho^i, \ a_i \in \mathbb{Z}.$$

To such a divisor we can associate its polyhedron of global sections:

$$P_D := \{u \in M_{\mathbb{Q}} \mid \langle u, \rho^i \rangle \geq -a_i \ \forall i = 1, \dots, n\}.$$

This is a polyhedron with tail cone  $\sigma^\vee$  and it gives rise to the  $R$ -module

$$H_D := \bigoplus_{u \in P_D \cap M} \mathbb{C} \cdot \chi^u.$$

The  $R$ -module  $H_D$  is  $M$ -graded and isomorphic to the global sections of  $\mathcal{O}(D)$ . Since we are working on an affine variety and the sheaf associated to a Weil divisor is coherent, computing  $\text{Ext}_{\text{TV}(\sigma)}^i(\mathcal{O}(D), \mathcal{O}(D'))$  is the same as computing  $\text{Ext}_R^i(H_D, H_{D'})$ . Thus, from now on we will write  $\text{Ext}^i(D, D')$ . Note that sometimes 0 will appear as a divisor, meaning that  $H_0 = R$ .

Most of the modules throughout this paper are completely determined through their support, i.e. they are  $M$ -graded and determined by their non-zero degrees in  $M$ .

Hence we will use the following short-hand notation from [21]. Let  $P \subseteq M_{\mathbb{Q}}$  be a subset, then we define

$$\mathbb{C}\{P\} := \bigoplus_{u \in P \cap M} \mathbb{C} \cdot \chi^u,$$

with the  $R$ -multiplication

$$x^w \cdot \chi^u := \begin{cases} \chi^{u+w} & u + w \in P \\ 0 & \text{else} \end{cases},$$

where  $w \in \sigma^\vee \cap M$  and  $x^w$  denotes the corresponding monomial in  $R$ .

Note that additional conditions on  $P$  are necessary to make this a module, but we will not go further into detail. Furthermore, note that  $P' \subsetneq P \subseteq M_{\mathbb{Q}}$  does not necessarily imply that  $\mathbb{C}\{P'\}$  is a submodule of  $\mathbb{C}\{P\}$ , due to  $\mathbb{C}\{P'\}$  having torsion elements that  $\mathbb{C}\{P\}$  does not have.

### 2.1 Two-Dimensional Cyclic Quotient Singularities

Two-dimensional cyclic quotient singularities arise as quotients of  $\mathbb{C}^2$  by a finite cyclic group  $\mathbb{Z}/n\mathbb{Z}$  acting via

$$\begin{pmatrix} \xi & 0 \\ 0 & \xi^q \end{pmatrix},$$

where  $\xi$  is a primitive  $n$ -th root of unity and  $q$  is a positive integer which is coprime to  $n$ . We observe that we may assume  $q < n$ .

**Definition 2.2** As toric varieties, CQS arise from  $N = \mathbb{Z}^2$ , with the cones  $\sigma$  and  $\sigma^\vee$  given as

$$\sigma = \text{cone}\{(1, 0), (-q, n)\} \text{ and } \sigma^\vee = \text{cone}\{[0, 1], [n, q]\}.$$

Here we identify  $M = \mathbb{Z}^2$  and pick the standard scalar product as the pairing of  $N$  and  $M$ . The CQS  $X$  is then the toric variety  $\text{TV}(\sigma)$ .

CQS are closely related to continued fraction expansions.

**Definition 2.3** Let  $c_1, \dots, c_n \in \mathbb{Z}_{>0}$ , then the continued fraction expansion  $[c_1, \dots, c_n]$  is defined as

$$[c_1, \dots, c_n] := c_1 - 1/[c_2, \dots, c_n] \text{ and } [c_n] := c_n.$$

For a CQS, there are two relevant continued fractions, namely

$$\underline{a} := [a_1, \dots, a_s] = \frac{n}{n-q} \text{ and } \underline{c} := [c_1, \dots, c_r] = \frac{n}{q}.$$

Continued fractions as a tool for studying cyclic quotient singularities were introduced by Riemenschneider [22], in order to describe their equations. Work of Christophersen [8] and Stevens [23] combinatorially describes the versal deformation of a CQS, based on continued fractions. This has in turn been used by Altmann [2] to describe the so-called p-resolutions introduced by Kollár and Shepherd-Barron [20] on a CQS and by Ilten [16] to compute Milnor number of a CQS. Recent work by Altmann and Kollár [3] studies certain infinitesimal deformations of CQS, called qG-deformations, in terms of continued fractions, among other things, with the aim to understand the infinitesimal structures of different versions of their moduli space.

We will mainly work with the cone  $\sigma^\vee$ , which is connected to the continued fraction  $\underline{a}$  in the following way: Let  $b^0, \dots, b^{r+1} \subseteq \sigma^\vee \cap M$  be the Hilbert basis of the cone  $\sigma^\vee$ , sorted with respect to the first coordinate, then we have the following equations:

$$b^{i-1} + b^{i+1} = a_i \cdot b^i \quad \forall i = 1, \dots, r. \quad (1)$$

As observed before, we have  $b^0 = [0, 1]$  and  $b^{r+1} = [n, q]$ . Together with the above equations this already completely determines the Hilbert basis. Furthermore, since we assume  $q < n$ , we have  $b^1 = [1, 1]$ .

*Example 2.4* Pick  $n = 8$  and  $q = 5$ , then we can access the continued fraction  $\underline{a}$  and the Hilbert basis of  $\sigma^\vee$  in `polymake` in the following way:

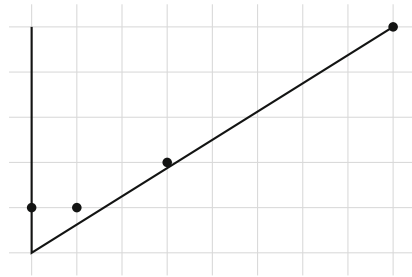
```
fulton > $cqs = new CyclicQuotient (N=>8, Q=>5);

fulton > print $cqs->DUAL_CONTINUED_FRACTION;
3 3
fulton > print $cqs->WEIGHT_CONE->HILBERT_BASIS;

8 5
0 1
3 2
1 1
```

One notes that the Hilbert basis is not yet sorted. After sorting it, we see the equations coming from the continued fraction  $\underline{a}$ .

The cone  $\sigma^\vee$  looks as in the following picture, where the dots indicate its Hilbert basis:



The class group of a CQS  $X$  can be computed via the following exact sequence:

$$0 \longrightarrow M \xrightarrow{A := \begin{pmatrix} 1 & 0 \\ -q & n \end{pmatrix}} \mathbb{Z}^2 \longrightarrow \text{Cl} X \longrightarrow 0 \quad (2)$$

In particular, the class group is finite and hence, there are only finitely many divisorial ideals in  $R$  up to isomorphism. In conjunction with the following theorem this shows that we only need to describe  $\text{Ext}^1$ .

**Theorem 2.5 ([18, Thm 5.16])** *Given a CQS  $X$  with class group  $\text{Cl} X \cong \mathbb{Z}/n\mathbb{Z}$  and two torus-invariant Weil divisors  $D$  and  $D'$  on  $X$ , there is a quiver  $\mathcal{Q}$  with vertices in  $\text{Cl} X$  such that*

$$\text{Ext}_R^{i+1}(D, D') \cong \bigoplus_{[G \rightarrow D] \in \mathcal{Q}} \text{Ext}_R^i(G, D') \text{ for all } i \geq 1.$$

In words,  $\text{Ext}^{i+1}$  can be computed from  $\text{Ext}^i$  by taking the sources of the incoming arrows to  $D$  in the quiver  $\mathcal{Q}$ . One needs to be a little careful, since this quiver actually at first has vertices all of the torus-invariant Weil divisors. If one identifies divisors via linear equivalence and instead adds labels on the arrows for the grading, one gets the precise  $M$ -graded version of the above theorem.

If  $X$  is Gorenstein, i.e. a hypersurface, then the quiver consists of disjoint cycles of length at most 2, resembling the previously mentioned 2-periodicity discovered by Eisenbud. Thus the generalization of this discovery for CQS is the quiver  $\mathcal{Q}$ . It is not clear how to proceed in higher dimensions. First of all the class group may be infinite. Still one will obtain a quiver as the right object in some cases, for example for the cone over  $\mathbb{P}^1 \times \mathbb{P}^1$ , see Sect. 5. This is the hypersurface case again, but the quiver we describe in this case can deal with all torus-invariant divisors, instead of just treating the MCM classes. The drawback is that it has infinitely many vertices. Considering examples like Sect. 6, a quiver does not suffice.

### 3 Ext<sup>1</sup>

As previously discussed, it is enough to combinatorially describe Ext<sup>1</sup> on a CQS. This section will give the combinatorial description together with its connection to continued fractions. This connection combinatorially describes how Ext<sup>1</sup> changes along the flat and proper maps of certain toric blow-ups. It has yet to be understood in terms of algebraic geometry.

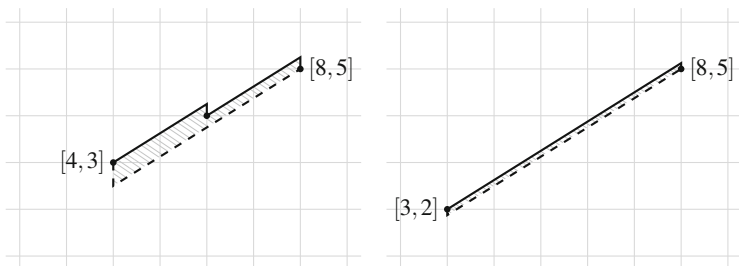
Let us start by giving the combinatorial description of Ext<sup>1</sup>. First we define an invariant of a torus invariant divisor  $D$ , derived from its polyhedron of global sections:

**Definition 3.1** For a torus-invariant divisor  $D$ , denote by  $G(D)$  the set of lattice points associated to the minimal generators of  $H_D$  and define

$$E(D) := \text{int}(P_D) \setminus \bigcup_{u \in G(D)} u + \text{int}(\sigma^\vee).$$

The set  $E(D)$  consists of the interior of  $P_D$  where we cut off the interior of  $\sigma^\vee$  shifted to the generators of  $H_D$ . By construction it contains all generators except the right- and leftmost ones.

*Example 3.2* We will illustrate what the sets  $E(\bullet)$  look like for two divisors from our running example:



The first divisor is  $-4 \cdot \text{orb}(\rho^0)$  and the second is  $-3 \cdot \text{orb}(\rho^0)$  where  $\rho^0 = (1, 0)$  denotes the zeroth ray of  $\sigma$ . The dashed line on the lower and leftmost boundary of the sets indicate that these do not belong to the sets  $E(\bullet)$ .

At this point it is necessary to note that every Weil divisor is  $\mathbb{Q}$ -Cartier on a CQS. For the torus invariant ones this means that their polyhedron of global sections is a shift of  $\sigma^\vee$  by a rational vector which we will denote by the symbol  $(\bullet)$ , i.e. we have the equation

$$P_D = (D) + \sigma^\vee \subseteq M_{\mathbb{Q}},$$

with the summation being the Minkowski sum of a rational vector and a cone. A divisor is trivial if and only if its vertex ( $\bullet$ ) is in the lattice  $M$ .

This can then be used directly to compute  $\text{Ext}^1(D, D')$  of two torus-invariant Weil divisors  $D$  and  $D'$  on  $X$ :

**Theorem 3.3 ([18, Prop 6.8])** *For two Weil divisors  $D$  and  $D'$  define*

$$\text{ext}(D, D') := -(E(D) - (D')).$$

*Then*

$$\text{Ext}^1(D, D') = \mathbb{C}\{\text{ext}(D, D')\}.$$

In general this can be used to compute  $\text{Ext}^1$  for non-torus-invariant Weil divisors as well. In this case the resulting module will not necessarily be graded, so the result is only unique up to non-canonical isomorphism.

Since the CQS  $X$  has an isolated singularity at the origin, the  $M$ -graded  $R$ -modules  $\text{Ext}_R^1(D, D')$  will actually be finite dimensional as  $\mathbb{C}$ -vector spaces. Choose the divisors

$$E^i := -i \cdot \text{orb}(\rho^0), \quad i = 1, \dots, n,$$

where  $\rho^0 = (1, 0)$  denotes the zeroth ray of  $\sigma$ , as a system of representatives for the class group  $\text{Cl} X$  of  $X$ . We define the following matrix  $\mathcal{E}$  containing all possible  $\mathbb{C}$ -dimensions of  $\text{Ext}^1$ :

$$\mathcal{E} := (e_{ij} = \dim_{\mathbb{C}} \text{Ext}_R^1(E^i, K_X - E^j))_{i,j=1,\dots,n}.$$

*Example 3.4* We can compute the matrix  $\mathcal{E}$  from  $n$  and  $q$  in the following way with `polymake`:

```
fulton > $cqs = new CyclicQuotient(N=>8, Q=>5);

fulton > print $cqs->EXT1_MATRIX;
2 2 1 2 2 1 1 0
2 2 1 2 2 1 1 0
1 1 1 1 1 0 0 0
2 2 1 3 2 1 1 0
2 2 1 2 2 1 1 0
1 1 0 1 1 0 0 0
1 1 0 1 1 0 1 0
0 0 0 0 0 0 0 0
```

One can already see that this matrix is very structured: It is symmetric and seems to repeat itself.

Before we discuss the recursive way to compute this matrix, let us have a look at some other properties. First of all, it is symmetric and one can ask, whether this is always the case. In particular, this may not only be a symmetry of  $\mathcal{E}$ , but even an isomorphism of  $\text{Ext}^1$ -modules. Using the incidence matrix  $\mathcal{I}$  of the quiver  $\mathcal{Q}$ , one can compute the dimensions of  $\text{Ext}^n$  taking  $\mathcal{I}^{n-1} \cdot \mathcal{E}$ , using the recursive description of  $\text{Ext}^n$  from [Theorem 2.5](#). If we define the matrix  $\mathcal{T}$  for  $\text{Tor}_1$  in almost the same way

$$\mathcal{T} := (t_{ij} = \dim_{\mathbb{C}} \text{Tor}_1^R(E^i, E^j))_{i,j=1,\dots,n},$$

we observe almost the same structure and arrive at the equation  $\mathcal{I}^2 \cdot \mathcal{E} = \mathcal{T}$ . Using the combinatorial description of  $\text{Ext}^1$  and `polymake` one then checks these conjectures for  $0 < q < n \leq 100$  and arrives at the following statements whose proofs can be found in [\[18\]](#):

**Theorem 3.5** ([\[18, 6.2, 7.2\]](#)) *For any two torus-invariant Weil-divisors  $D$  and  $D'$  on a cyclic quotient singularity  $X$  one has*

$$\begin{aligned} \dim_{\mathbb{C}} \text{Ext}^1(D, 0) &= \#\{\text{minimal generators of } H_D\} - 2 \text{ for } D \neq 0 \\ \text{Ext}^1(D, K - D') &= \text{Ext}^1(D', K - D) \\ \text{Ext}^3(D, K - D') &= (\text{Tor}_1(D, D'))^\vee \\ \text{Ext}^i(D, K) &= 0 \quad \forall i > 0, \end{aligned}$$

where we assume  $D$  to be non-trivial for the first equation,  $K$  denotes the canonical divisor on  $X$  and  $(\bullet)^\vee$  denotes the Matlis dual, i.e.  $\text{Hom}(\bullet, E(\mathbb{C}))$  the homomorphisms into the injective hull of  $\mathbb{C}$  over  $R$ .

The first equation is exactly the desired reformulation of Wunram’s theorem. His theorem classified the special MCM divisors as those with exactly two generators. Using the criterion of Iyama and Wemyss that  $D$  is special MCM if and only if

$$\text{Ext}^1(D, 0) = \text{Ext}^1(H_D, R) = 0,$$

we see that the first equation connects these two statements. Furthermore, we can observe, what happens in the Gorenstein case: Here, every non-trivial  $H_D$  is generated by exactly two elements, i.e. every divisor is special MCM. The special MCM divisors correspond 1-1 to the exceptional curves in the minimal resolution of the singularity [\[24–26\]](#).

The last equation states that all Weil divisors are MCM and combinatorially reformulates a theorem of Bruns and Gubeladze that all  $\mathbb{Q}$ -Cartier divisors are MCM [\[5, Cor. 6.68\]](#).

Assume now that the cone  $\sigma^\vee$  has the Hilbert basis  $\text{HB}(\sigma^\vee)$

$$b^0 = [0, 1], \quad b^1 = [1, 1], \dots, \quad b^s = [\tilde{n}, \tilde{q}], \quad b^{s+1} = [n, q].$$



From this we build two other cones  $\tilde{\sigma}^\vee$  and  $\sigma^{\vee'}$  determined by their Hilbert bases

$$\begin{aligned}
 HB(\tilde{\sigma}^\vee) &= \{b^0, b^1, \dots, b^s\}, \\
 HB((\sigma')^\vee) &= \{b^0, b^1, \dots, b^{s+1} - b^s\}.
 \end{aligned}$$

That means we have the following chains of inclusions:

$$\tilde{\sigma}^\vee \subseteq \sigma^\vee \subseteq (\sigma')^\vee \quad \text{and} \quad \tilde{\sigma} \supseteq \sigma \supseteq \sigma'.$$

Looking at the Hilbert bases, we obtain the following two continued fractions associated to the new CQS:

$$\tilde{a} := [a_1, \dots, a_{s-1}] \quad \text{and} \quad \underline{a}' := [a_1, \dots, a_s - 1].$$

Here we make the convention that

$$[a_1, \dots, a_s, 1] = [a_1, \dots, a_s - 1],$$

i.e. if the last entry of a continued fraction is 1, we can compress this continued fraction. Furthermore we say that

$$[] = [1] = 0$$

and we say that this is the continued fraction associated to  $X = \mathbb{A}_{\mathbb{C}}^2$ .

If we consider  $\mathcal{E}$  as a function taking a continued fraction to a matrix, we can state the following algorithmic theorem:

**Theorem 3.6 ([18, Thm 6.26])** *The matrix  $\mathcal{E}(\underline{a})$  is completely determined by the matrices  $\mathcal{E}(\tilde{a})$  and  $\mathcal{E}(\underline{a}')$ . This yields a recursive algorithm for computing  $\mathcal{E}$ .*

We omit the details and illustrate this theorem for an example instead.

*Example 3.7*

$$\begin{pmatrix} 2 & 2 & 1 & 2 & 2 & 1 & 1 & 0 \\ 2 & 2 & 1 & 2 & 2 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 2 & 2 & 1 & 3 & 2 & 1 & 1 & 0 \\ 2 & 2 & 1 & 2 & 2 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 0 & | & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & | & 1 & 1 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & | & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & | & 2 & 2 & 1 & 1 & 0 \\ 1 & 1 & 0 & | & 2 & 2 & 1 & 1 & 0 \\ \hline 0 & 0 & 0 & | & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & | & 1 & 1 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & | & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The matrix on the left is  $\mathcal{E}([3, 3])$ . The upper left block in the right-most matrix is  $\mathcal{E}([3])$  and the lower right block is  $\mathcal{E}([3, 2])$ . One can already guess how the

non quadratic blocks arise as submatrices of the bigger quadratic block. In general the non-quadratic blocks are always determined by the bigger of the two quadratic blocks. The smaller quadratic block determines a smaller part of the non-quadratic blocks and on this part the two quadratic blocks have to agree.

This theorem is proven by choosing certain systems of representatives for the class groups of all three involved CQS and then connecting the sets  $\text{ext}(\bullet, \bullet) \cap M$  for all three CQS. Thus there is actually a stronger version of this theorem that recursively constructs the support  $\text{ext}(\bullet, \bullet)$  of the module  $\text{Ext}^1(\bullet, \bullet)$ . The algebro-geometric interpretation would be that these CQS are connected via blow-ups and we have a combinatorial description on how  $\text{Ext}^1$  should behave on these blow-ups.

If we think of the divisors  $E^j$  previously introduced as a system of representatives for the class group  $X$ , then one notes that the lattice points

$$(\text{ext}(E^j, E^1) \cap M) \cup \{[n, q]\}, j = 1, \dots, n,$$

corresponds exactly to the minimal homogeneous generators of  $H_{E^j}$ . This is a generalization of the equations of Eq. (1) giving us similar relations for the generators of a divisorial ideal.

## 4 polymake and Singular

For computations in toric geometry one needs software for both combinatorics and algebraic geometry. Of course every problem can be formulated purely algebraic, but then we would lose the advantage that the combinatorial nature of toric varieties gives us. Not only does it vastly increase performance of algorithms, it also provides us with a new understanding of our problem.

For example the algorithm for finding a Hilbert basis of a cone is the same as computing the integral closure of a semigroup ring spanned by monomials corresponding to the primitive ray generators of the cone in the polynomial ring. Using **Normaliz** or **4ti2** we can find Hilbert bases with several hundred elements. On the algebraic side this corresponds to a ring that is the quotient of a polynomial ring with the same amount of variables, i.e. several hundred, a setting which makes computer algebra systems break instantly.

For the combinatorial side we chose the software framework **polymake** [14]. It already comes with a huge bulk of code for toric geometry in the application **fulton**. Furthermore it has an interface to **Singular**, the computer algebra system of our choice. In particular, it allows executing any **Singular** command via the `singular_eval` method. Furthermore the method `singular_get_var` allows us recovering certain kinds of variables, such as integers, vectors of integers and matrices of integers.

This section will elaborate on how one can use these methods in a proof of concept style for our particular problem of computing Ext and Tor of two Weil divisors on a toric variety.

## 4.1 `polymake`

The software framework `polymake` focuses on combinatorial problems. The objects of interest are matroids, graphs, cones and fans, polytopes and polyhedral complexes. It provides a large pool of instruments for the combinatorial parts of algebraic geometry, such as tropical geometry and toric geometry. The interpreter language of `polymake` is `perl`. Optionally can write and attach C++ code, in order to increase performance amongst other things.

Many computations are outsourced via interfaces to other software. For example dualising cones is done via `lrs` or `cdd` [4, 12]. Computing the Hilbert basis is done by `4ti2` or `Normaliz` [6, 28].

The main application for toric geometry is the application `fulton` named after the famous book by Fulton on toric varieties [13]. Another application of interest for us is the application `ideal`. This application comes to life when `polymake` is build with the **Singular** library version. Then one can use the object `ideal` from this application and all computations with ideals will be done in the background by **Singular**.

## 4.2 *Singular*

**Singular** is one of the leading computer algebra systems. It allows custom monomial orders when defining a ring and even computes with local rings. This makes it interesting for tropical geometry. Recently new interfaces to **GFan** [17] for tropical geometry and to `polymake` have been developed by Yue Ren from Kaiserslautern.

In our setting we will use **Singular** for the commutative algebra side of our problem. Let us give an example code snippet:

```
LIB "homolog.lib";
ring r = 0, (w,x,y,z), dp;
ideal toric = wy-xz;
qring q = std(toric);
ideal d1 = w3, w2x, wx2, x3;
ideal d2 = x2, xy, y2;
module H = Ext(5, syz(d1), syz(d2));
H = std(H);
dim(H);
vdim(H);
```

Here, we first load the **Singular** library `homolog.lib` for homological algebra [15]. Then we create the toric ring of the hypersurface singularity of the cone over  $\mathbb{P}^1 \times \mathbb{P}^1$ , define to divisorial ideals in it and finally compute  $\text{Ext}^5$  of these ideals.

This is the kind of code snippets that `polymake` will automatically create and run in the background. One advantage is that **Singular** will continue running in the background, so the objects stay alive and e.g. the standard basis of the toric ideal does not have to be recomputed every time we want to determine a dimension of an Ext-module.

### 4.3 *Interfacing `polymake` and `Singular`*

At this point most objects from commutative algebra only make sense for **Singular**. There is an “ideal” object in `polymake`, but quotient rings and modules only exist on the **Singular** side. Since **Singular** will do the calculations, all `polymake` needs to know are the names of the objects, so it can autogenerate the **Singular** code which is then executed by the `singular_eval` command. Let us illustrate this in one example:

```
object_specialization
  NormalToricVariety::AffineNormalToricVariety {

  property SINGULAR_TORIC_RING : String;

  property DIVISOR {

    property SINGULAR_IDEAL : String;

    property SINGULAR_SZYZYGIES : String;

  }

}
```

Here we add properties to affine normal toric varieties and divisors on these in `polymake`. All of the properties are Strings representing the names of these properties in **Singular**. The affine toric variety now has the name of its quotient ring in **Singular** and a divisor knows the name of its ideal and the syzygies thereof. These properties are computed via relatively straight forward rules, for example **4ti2** can compute the binomial exponents of the toric ideal of a cone and turning these into a string which is executed in **Singular**, we get `SINGULAR_TORIC_RING`.

Let us focus on the problem at hand and write a method computing the dimension and vector space dimension of an arbitrary  $\text{Ext}^i$ -module as follows:

```
user_method
  singular_exti_dimension( $ , TDivisor, TDivisor){
  my $toric_variety = $_[0];
  my $i = $_[1];
  my $divisor1 = $_[2];
  my $divisor2 = $_[3];
  my $ringname = $toric_variety->SINGULAR_TORIC_RING;
  my $syzygies1 = $divisor1->SINGULAR_SZYZYGIES;
```

```

my $syzygies2 = $divisor2->SINGULAR_SYZYGIES;
singular_eval("setring r_.$ringname.");
load_singular_library("homolog.lib");
singular_eval("module M =
  Ext(".$i.", syz_.$syzygies1.",
      syz_.$syzygies2.");");
singular_eval("M = std(M);");
singular_eval("int d = dim(M);");
singular_eval("int vd = vdim(M);");
return new Vector(singular_get_var("d"),
                  singular_get_var("vd"));
}

```

This method gets the integer  $i$  and two toric divisors as input. The first seven lines extract the necessary variables. Then we switch to **Singular**, make sure the right ring is set and compute  $\text{Ext}^i$ . Afterwards we extract the dimension and vector space dimension via `singular_get_var`.

## 5 Cone over $\mathbb{P}^1 \times \mathbb{P}^1$

This section continues the example introduced in [Theorem 2.1](#). The cone  $\sigma$  is obtained as the cone over a square at height one (with a non-standard height function). The resulting singularity  $X$  is Gorenstein and isolated. The class group of this singularity is  $\mathbb{Z}$ . We discuss this example to demonstrate why we hope for similar structures in higher dimensions as in the CQS case.

The following code snippet sets up this example in `polymake` and returns some Ext and Tor dimensions:

```

$c = new Cone(INPUT_RAYS=>[[1,0,0],[1,0,1],
                           [1,1,0],[1,1,1]]);
$dc = new Cone(INPUT_RAYS=>$c->FACETS);
$tv = new NormalToricVariety($dc);
print $tv->GORENSTEIN; # Is Gorenstein
$ccoeff = new Vector(-1,-1,-1,-1);
$div1Coeff = new Vector(7,-1,-1,-1);
$div1 = $tv->add("DIVISOR", COEFFICIENTS=>$div1Coeff);
$canonical = $tv->add("DIVISOR",
                     COEFFICIENTS=>$ccoeff);
$div2Coeff = new Vector(-5,0,0,0);
$div2 = $tv->add("DIVISOR", COEFFICIENTS=>$div2Coeff);
$kmd1 = $tv->add("DIVISOR",
               COEFFICIENTS=>$ccoeff - $div1Coeff);
$kmd2 = $tv->add("DIVISOR",
               COEFFICIENTS=>$ccoeff - $div2Coeff);

print $tv->singular_exti_dimension(2, $div1, $kmd2);
print $tv->singular_exti_dimension(2, $div2, $kmd1);

print $tv->singular_exti_dimension(3, $div2, $kmd1);

```

```
print $tv->singular_exti_dimension(3, $div1, $kmd2);

print $tv->singular_exti_dimension(4, $div1, $kmd2);
print $tv->singular_exti_dimension(4, $div2, $kmd1);
print $tv->singular_tori_dimension(1, $div2, $div1);
```

One observes that the last three lines yield the same output. Just as the two pairs of lines before these. This indicates a 2-periodicity for Ext, as well as the following relation between Ext and Tor

$$\dim_{\mathbb{C}} \text{Ext}^{i+3}(D, K - D') = \dim_{\mathbb{C}} \text{Tor}_i(D, D') \text{ for } i > 0$$

in this case. Apparently one can connect  $\text{Ext}^{i+d}$  and  $\text{Tor}_i$  in certain settings, where  $d$  denotes the dimension of the toric variety. Using **Singular**, we realize that the only two non-trivial MCM divisor classes are represented by

$$D_1 := [-1, 0, 0, 0] \text{ and } D_2 := [0, -1, 0, 0],$$

In the class group these correspond to  $\pm 1$ . Using **Singular** to take a closer look at the free resolutions of the divisorial ideals, we realize that for any  $D \in \text{Cl } X$  there is a short exact sequence

$$0 \rightarrow \bigoplus_{n < \infty} D_i \rightarrow R^n \rightarrow H_D \rightarrow 0,$$

where  $i$  is either 1 or 2 and  $R$  denotes the coordinate ring of  $X$ . In particular, the syzygies of  $H_{D_1}$  are  $H_{D_2}$  and vice versa, once more demonstrating the 2-periodicity observed by Eisenbud. In terms of [Theorem 2.5](#), we see that all free resolutions are encoded in an infinite quiver with maximal cycle length 2.

## 6 Cone over a Trapezoid

In this example the cone  $\sigma^\vee$  is the cone over a trapezoid height one, yielding a Hilbert basis of  $\sigma^\vee$  with five elements. The resulting singularity  $X$  is isolated, but not Gorenstein. We will give a code snippet and discuss the output.

```
$c = new Cone(INPUT_RAYS=>[[1,0,1], [1,1,0], [1,2,0],
                           [1,1,1], [1,0,2]]);
$dc = new Cone(INPUT_RAYS=>$c->FACETS);
$tv = new NormalToricVariety($dc);
$ccoeff = new Vector(-1,-1,-1,-1);
$div1Coeff = new Vector(7,-1,-1,-1);
$div1 = $tv->add("DIVISOR", COEFFICIENTS=>$div1Coeff);
$canonical = $tv->add("DIVISOR", COEFFICIENTS=>$ccoeff);
$div2Coeff = new Vector(-5,0,0,0);
$div2 = $tv->add("DIVISOR", COEFFICIENTS=>$div2Coeff);
```

```

$km d1 = $tv->add("DIVISOR",
                COEFFICIENTS=>$ccoeff - $div1Coeff);
$km d2 = $tv->add("DIVISOR",
                COEFFICIENTS=>$ccoeff - $div2Coeff);
print $tv->singular_exti_dimension(2, $div1, $km d2);
print $tv->singular_exti_dimension(2, $div2, $km d1);

print $tv->singular_exti_dimension(3, $div2, $km d1);
print $tv->singular_exti_dimension(3, $div1, $km d2);

print $tv->singular_exti_dimension(4, $div1, $km d2);
print $tv->singular_exti_dimension(4, $div2, $km d1);
print $tv->singular_tori_dimension(1, $div2, $div1);

```

The output one gets, shows that the two  $\text{Ext}^2$  have different dimensions, a hint that higher Ext might not be symmetric as well, thereby obliterating the hope for a connection of Ext and Tor. This is made even clearer by the last three lines. Even though one of the two Ext-modules has the same dimension as the Tor-module, one cannot really hope for an isomorphism of modules, since Tor is symmetric and would immediately imply a symmetry of Ext as well. This is interesting, since being Gorenstein did not play a role for the connection of Ext and Tor in the CQS case.

## 7 Conclusion

We will conclude this survey with a code sample: Take the hexagon singularity, which has been introduced and discussed exhaustively in [1]. The cone  $\sigma$  is given as the cone over a hexagon at height one. The Hilbert basis of the dual cone  $\sigma^\vee$  has seven elements and the associated toric ideal has eleven generators.

We will take two arbitrary divisors and compare their pairwise  $\text{Ext}^i$  and  $\text{Tor}_i$ . Since this singularity is isolated, both spaces will be finite dimensional  $\mathbb{C}$ -vector spaces, so it makes sense to start computing their dimensions. As discussed previously, we expect  $\text{Tor}_1$  to be connected to  $\text{Ext}^4$ . First we construct the toric variety and the divisors  $D_1$ ,  $D_2$ ,  $K_X$  and  $K_X - D_i$  in **polymake**. Then we use **Singular** to compute  $\text{Ext}^4$  and  $\text{Tor}_1$ .

```

$C = new Cone(INPUT_RAYS=>[[1,0,0],[1,1,0],[1,2,1],
                          [1,2,2],[1,1,2],[1,0,1]]);
$tv = new NormalToricVariety($C);
$ccoeff = new Vector<Integer>(-1,-1,-1,-1,-1,-1);
$canonical = $tv->add("DIVISOR",
                    COEFFICIENTS=>$ccoeff);
print $canonical->MODULE_GENERATORS;
$d1coeff = new Vector<Integer>(-3,0,0,0,0,0);
$d1 = $tv->add("DIVISOR", COEFFICIENTS=>$d1coeff);
$d2coeff = new Vector<Integer>(0,-4,0,0,0,0);
$d2 = $tv->add("DIVISOR", COEFFICIENTS=>$d2coeff);
$cmind2 = $tv->add("DIVISOR",

```

```

      COEFFICIENTS=>$ccoeff - $d2coeff);
$cmind1 = $tv->add("DIVISOR",
      COEFFICIENTS=>$ccoeff - $d1coeff);
print $tv->singular_exti_dimension(4, $d1, $cmind2);
print $tv->singular_exti_dimension(4, $d2, $cmind1);
print $tv->singular_tori_dimension(1, $d1, $d2);

```

This code will fail with an error indicating that **Singular** cannot create a matrix of the desired size. With help from the **Singular** team one can disable this error—only to see all memory being consumed by the calculation. One can even observe this phenomenon in the CQS case for large  $n$ . This demonstrates exactly why toric methods are important: The combinatorial algorithms make solving such problems feasible, thereby providing deeper understanding. Algebraic geometry in general is connected with toric geometry in many ways, think of deformation theory, toric degenerations and many more. Understanding this problem in the toric case provides understanding for many general cases as well.

**Acknowledgements** The author is supported by the DFG (German research foundation) priority program SPP 1489 ‘Computeralgebra’ and the thematic program ‘Combinatorial Algebraic Geometry’ of The Fields Institute in Toronto.

## References

1. K. Altmann, The versal deformation of an isolated toric Gorenstein singularity. *Invent. Math.* **128**(3), 443–479 (1997) (English)
2. K. Altmann, P-resolutions of cyclic quotients from the toric viewpoint, in *Singularities. The Brieskorn Anniversary Volume. Proceedings of the Conference Dedicated to Egbert Brieskorn on his 60th Birthday, Oberwolfach, July 1996* (Birkhäuser, Basel, 1998), pp. 241–250 (English)
3. K. Altmann, J. Kollár, The dualizing sheaf on first-order deformations of toric surface singularities (2016, to appear in *Crelle*). arXiv:1601.07805v2. <https://doi.org/10.1515/crelle-2016-0063>
4. D. Avis, G. Roumanis, A portable parallel implementation of the lrs vertex enumeration code, in *Combinatorial Optimization and Applications* (Springer, Heidelberg, 2013), pp. 414–429
5. W. Bruns, J. Gubeladze, *Polytopes, Rings, and K-Theory* (Springer, New York, NY, 2009) (English)
6. W. Bruns, B. Ichim, T. Römer, R. Sieg, C. Söger, *Normaliz. algorithms for rational cones and affine monoids*. Available at <https://www.normaliz.uni-osnabrueck.de>
7. R.-O. Buchweitz, *Maximal Cohen-Macaulay modules and Tate-cohomology over Gorenstein rings*.
8. J.A. Christophersen, *On the components and discriminant of the versal base space of cyclic quotient singularities.*, Symmetric Lagrangian singularities and Gauss maps of theta divisors (1991), pp. 81–92 (English)
9. D.A. Cox, J.B. Little, H.K. Schenck, *Toric Varieties* (American Mathematical Society (AMS), Providence, RI, 2011) (English)
10. W. Decker, G.-M. Greuel, G. Pfister, H. Schönemann, SINGULAR 4-0-2 – A computer algebra system for polynomial computations (2015). <http://www.singular.uni-kl.de>
11. D. Eisenbud, Homological algebra on a complete intersection, with an application to group representations. *Trans. Am. Math. Soc.* **260**(1), 35–64 (1980)



12. K. Fukuda, *The CDD Program*. Available from World Wide Web ([http://www.ifor.math.ethz.ch/~fukuda/cdd\\_home/cdd.html](http://www.ifor.math.ethz.ch/~fukuda/cdd_home/cdd.html)) (2003)
13. W. Fulton, *Introduction to Toric Varieties. The 1989 William H. Roever lectures in Geometry* (Princeton University Press, Princeton, NJ, 1993) (English)
14. E. Gawrilow, M. Joswig, polymake: a framework for analyzing convex polytopes, in *Polytopes – Combinatorics and Computation*, ed. by G. Kalai, G.M. Ziegler (Birkhäuser, Basel, 2000), pp. 43–74
15. G.-M. Greuel, B. Martin, C. Lossen, *homolog.lib. a SINGULAR 4.0.3 library for procedures for homological algebra*
16. N.O. Itten, Calculating Milnor numbers and versal component dimensions from p-resolution fans (2008). arxiv preprint arXiv:0801.2900
17. A.N. Jensen, **GFan**, a software system for Gröbner fans and tropical varieties, version 0.5 (2011). Available at <http://home.imf.au.dk/jensen/software/gfan/gfan.html>
18. L. Kastner, *Ext on affine toric varieties*, Ph.D. thesis. Freie Universität Berlin (2015). Available at [http://www.diss.fu-berlin.de/diss/receive/FUDISS\\_thesis\\_000000101520](http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000101520)
19. L. Kastner, Ext and Tor on two-dimensional cyclic quotient singularities. ArXiv e-prints (2016)
20. J. Kollár, N.I. Shepherd-Barron, Threefolds and deformations of surface singularities. *Invent. Math.* **91**(2), 299–338 (1988) (English)
21. E. Miller, B. Sturmfels, *Combinatorial Commutative Algebra* (Springer, New York, NY, 2005) (English)
22. O. Riemenschneider, Zweidimensionale Quotientensingularitäten: Gleichungen und Syzygien. *Arch. Math.* **37**, 406–417 (1981) (German)
23. J. Stevens, On the versal deformation of cyclic quotient singularities. *Symmetric Lagrangian singularities and Gauss maps of theta divisors* (1991), pp. 302–319 (English)
24. M. Wemyss, The  $GL(2, \mathbb{C})$  McKay correspondence. *Math. Ann.* **350**(3), 631–659 (2011) (English)
25. J. Wunram, Reflexive modules on cyclic quotient surface singularities, in *Singularities, Representation of Algebras, and Vector Bundles* (Springer, Berlin, 1987), pp. 221–231
26. J. Wunram, Reflexive modules on quotient surface singularities. *Math. Ann.* **279**(4), 583–598 (1988) (English)
27. Y. Yoshino, *Maximal Cohen-Macaulay Modules Over Cohen-Macaulay Rings*, vol. 146 (Cambridge University Press, Cambridge, 1990)
28. 4ti2 team, *4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces*. Available at [www.4ti2.de](http://www.4ti2.de)

# The Differential Dimension Polynomial for Characterizable Differential Ideals



Markus Lange-Hegermann

**Abstract** We generalize the notion of a differential dimension polynomial of a prime differential ideal to that of a characterizable differential ideal. Its computation is algorithmic, its degree and leading coefficient remain differential birational invariants, and it decides equality of characterizable differential ideals contained in each other.

**Keywords** Dimension polynomial • Characterizable ideal

**Subject Classifications** 12H05, 35A01, 35A10, 34G20

## 1 Introduction

Many systems of differential equations do not admit closed form solutions or any other finite representation of all solutions. Hence, such systems cannot be solved symbolically. Despite this, increasingly good and efficient heuristics to find solutions symbolically have been developed and are implemented in computer algebra systems [4, 5]. Of course, such algorithms can at best produce the subset that admits a closed form of the full set of solutions. Given such a set of closed form solutions returned by a computer algebra system, the natural question remains whether this set is a complete solution set (cf. Example 5.4).

Classical measures, e.g. the Cartan characters [3] and Einstein's strength [7], describe the size of such solution sets. However, they have a drawback: one can easily find two systems  $S_1$  and  $S_2$  of differential equations such that the solution set of  $S_1$  is a proper subset of the solution set of  $S_2$ , but these two solution sets have identical measures (cf. Example 5.3). In particular, if  $S_1$  is given by a solver of

---

M. Lange-Hegermann (✉)

Lehrstuhl B für Mathematik, RWTH Aachen University, 52062 Aachen, Germany  
e-mail: [markus.lange.hegermann@rwth-aachen.de](mailto:markus.lange.hegermann@rwth-aachen.de)

differential equations, these measures cannot detect whether this is the full set  $S_2$  of solutions.

Kolchin introduced the differential dimension polynomial to solve this problem for solution sets of systems of differential equations corresponding to prime differential ideals [12–15]. This polynomial generalizes the Cartan characters and strength by counting the number of freely choosable power series coefficients of an analytical solution. Recently, Levin generalized the differential dimension polynomial to describe certain subsets of the full solution set of a prime differential ideal [18].

Even though decomposing the radical differential ideal generated by a set of differential equations into prime differential ideals is theoretically possible, it is expensive in practice (cf. [2, §6.2]). Thus, there is a lack of practical methods which decide whether a subset of the solution set of a system of differential equations is a proper subset. This paper solves this problem for greater generality than solution sets of prime differential ideals. It generalizes the differential dimension polynomial to characterizable differential ideals and thereby gives a necessary condition for completeness of solution sets. Such ideals can be described by differential regular chains, and there exist reasonably fast algorithms that decompose a differential ideal into such ideals [1, 2].

To formulate the main theorem, we give some preliminary definitions; the missing definitions are given in Sect. 2. Denote by  $F\{U\}$  a differential polynomial ring in  $m$  differential indeterminates for  $n$  commuting derivations over a differential field  $F$  of characteristic zero. For a differential ideal  $I$  in  $F\{U\}$  let  $I_{\leq \ell} := I \cap F\{U\}_{\leq \ell}$ , where  $F\{U\}_{\leq \ell}$  is the subring of  $F\{U\}$  of elements of order at most  $\ell$ . We define the differential dimension function using the Krull dimension as

$$\Omega_I : \mathbb{Z}_{\geq 0} \mapsto \mathbb{Z}_{\geq 0} : \ell \mapsto \dim(F\{U\}_{\leq \ell} / I_{\leq \ell}) .$$

By the following theorem, this function is eventually polynomial for large  $\ell$  if  $I$  is characterizable. Such polynomials mapping  $\mathbb{Z}$  to  $\mathbb{Z}$  are called numerical polynomials, and there exists a natural total order  $\leq$  on them.

**Theorem 1.1** *Let  $I \subset F\{U\}$  be a characterizable differential ideal.*

1. *There exists a unique numerical polynomial  $\omega_I(\ell) \in \mathbb{Q}[\ell]$ , called differential dimension polynomial, with  $\omega_I(\ell) = \Omega_I(\ell)$  for sufficiently big  $\ell \in \mathbb{Z}_{\geq 0}$ .*
2.  *$0 \leq \omega_I(\ell) \leq m \binom{\ell+n}{n}$  for all  $\ell \in \mathbb{Z}_{\geq 0}$ . In particular,  $d_I := \deg_\ell(\omega_I) \leq n$ .*
3. *When writing  $\omega_I(\ell) = \sum_{i=0}^n a_i \binom{\ell+i}{i}$  with  $a_i \in \mathbb{Z}$  for all  $i \in \{0, \dots, n\}$ , the degree  $d_I$  and the coefficients  $a_i$  for  $i \geq d_I$  are differential birational invariants, i.e., they are well-defined on the isomorphism class of total quotient ring of  $F\{U\}/I$ .*
4. *The coefficient  $a_n$  is the differential dimension of  $F\{U\}/I$ , as defined below.*

*Let  $I \subseteq J \subset F\{U\}$  be another characterizable differential ideal.*

5. *Then  $\omega_J \leq \omega_I$ .*

Assume  $\omega_I = \omega_J$ , and let  $S$  respectively  $S'$  be differential regular chains with respect to an orderly differential ranking  $<$  that describe  $I$  respectively  $J$ .

- 6. The sets of leaders of  $S$  and  $S'$  coincide, and
- 7.  $I = J$  if and only if  $\deg_x(S_x) = \deg_x(S'_x)$  for all leaders  $x$  of  $S$ , where  $S_x$  is the unique element in  $S$  of leader  $x$ .

This theorem can be slightly strengthened, as  $I \subseteq J$  and  $\omega_I = \omega_J$  already imply  $\deg_x(S_x) \leq \deg_x(S'_x)$  for all leader  $x$  of  $S$  (cf. Lemma 3.5). Thus  $I = J$  if and only if  $\prod_x \deg_x(S_x) = \prod_x \deg_x(S'_x)$ . It would be interesting to have a version of Theorem 1.1, where this product is an intrinsic value, similar to the leading differential degree [9].

The importance of characterisable differential ideals and their connection to differential dimension polynomials appear in [6, §3.2], building on Lazard’s lemma [2]. In particular, the invariance conditions were implicitly observed. To the best of the author’s knowledge, testing equality by means of invariants does not appear in the literature. Testing equality of differential ideals is connected to Ritt’s problem of finding a minimal prime decomposition of differential ideals.<sup>1</sup>

Recently, the author introduced the differential counting polynomial [16, 17]. It gives a more detailed description of the set of solutions than the differential dimension polynomial, in fact so detailed that it seems not to be computable algorithmically. In particular, it provides a necessary criterion of completeness of solution sets, whereas the differential counting polynomial only provides a sufficient criterion. The intention of this paper is a compromise of giving a description of the size of the set of solutions that is detailed enough to be applicable to many problems, but that is still algorithmically computable.

A more detailed description of the content of this paper in the language of simple systems is a part of the author’s thesis [17].

Section 3 proves Theorem 1.1, Sect. 4 discusses the computation of the differential dimension polynomial, and Sect. 5 gives examples.

---

<sup>1</sup>It is easy to test equality of two prime differential ideals given by a characteristic set (cf. exercise 1 in [14, §IV.10]). However, the unsolved Ritt problem states that there is no algorithm known to find a *minimal* decomposition of a differential ideal given by a set of generators into prime ideals given by characteristic sets [20], [14, §IV.9]. Under mild conditions, Ritt’s problem is equivalent to several other problems, among them (1) deciding whether a differential ideal given by a set of generators is prime, (2) finding a set of generators of a prime differential ideal given by a characteristic set, and (3) given the characteristic sets of two prime differential ideals  $I_1$  and  $I_2$  determine whether  $I_1 \subseteq I_2$  [10].

## 2 Preliminaries

### 2.1 Squarefree Regular Chains

Let  $F$  be a field of characteristic zero,  $\overline{F}$  its algebraic closure, and  $R := F[y_1, \dots, y_n]$  a polynomial ring. We fix the total order, called ranking,  $y_1 < y_2 < \dots < y_n$  on  $\{y_1, \dots, y_n\}$ . The  $<$ -greatest variable  $\text{ld}(p)$  occurring in  $p \in R \setminus F$  is called the leader of  $p$ . The coefficient  $\text{ini}(p)$  of the highest power of  $\text{ld}(p)$  in  $p$  is called the initial of  $p$ . We denote the separant  $\frac{\partial p}{\partial \text{ld}(p)}$  of  $p$  by  $\text{sep}(p)$ .

Let  $S \subset R \setminus F$  be finite. Define  $\text{ld}(S) := \{\text{ld}(p) \mid p \in S\}$  and similarly  $\text{ini}(S)$  and  $\text{sep}(S)$ . The set  $S$  is called triangular if  $|\text{ld}(S)| = |S|$ ; in this case denote by  $S_x \in S$  the unique polynomial with  $\text{ld}(S_x) = x$  for  $x \in \text{ld}(S)$ . We call the ideal  $\mathcal{I}(S) := \langle S \rangle : \text{ini}(S)^\infty \subseteq R$  the ideal associated to  $S$ . Let  $S_{<x} := \{p \in S \mid \text{ld}(p) < x\}$  for each  $x \in \{y_1, \dots, y_n\}$ . The set  $S$  is called a squarefree regular chain if it is triangular and neither  $\text{ini}(S_x)$  is a zero divisor modulo  $\mathcal{I}(S_{<x})$  nor  $\text{sep}(S_x)$  is a zero divisor modulo  $\mathcal{I}(S)$  for each  $x \in \text{ld}(S)$ .

**Proposition 2.1 ([11, Prop. 5.8])** *Let  $S$  be a squarefree regular chain in  $R$  and  $1 \leq i \leq n$ . Then  $\mathcal{I}(S_{<y_i}) \cap F[y_1, \dots, y_{i-1}] = \mathcal{I}(S) \cap F[y_1, \dots, y_{i-1}]$ . Furthermore, if  $p \in F[y_1, \dots, y_{i-1}]$  is not a zero-divisor modulo  $\mathcal{I}(S_{<y_i})$ , then  $p$  is not a zero-divisor modulo  $\mathcal{I}(S)$ .*

Note that the last sentence follows easily using that the zero divisors (and zero) are the union of the associated primed, cf. [8, Thm. 3.1].

**Theorem 2.2 (Lazard’s lemma, [11, Thm. 4.4, Coro. 7.3, Thm. 7.5], [2, Thm. 1])** *Let  $S$  be a squarefree regular chain in  $R$ . Then  $\mathcal{I}(S)$  is a radical ideal in  $R$ , and the set  $\{y_1, \dots, y_n\} \setminus \text{ld}(S)$  forms a transcendence basis for every associated prime of  $\mathcal{I}(S)$ . Let such an associated prime  $\mathcal{I}(S')$  be given by a squarefree regular chain  $S'$ . Then  $\text{ld}(S) = \text{ld}(S')$  and, in particular,  $R/\mathcal{I}(S)$  is equidimensional of dimension  $n - |\text{ld}(S)|$ .*

### 2.2 Differential Algebra

Let  $F$  be a differential field of characteristic zero with pairwise commuting derivations  $\Delta = \{\partial_1, \dots, \partial_n\}$ . Let  $U := \{u^{(1)}, \dots, u^{(m)}\}$  be a set of differential indeterminates and define  $u_\mu^{(j)} := \partial^\mu u^{(j)}$  for  $\partial^\mu := \partial_1^{\mu_1} \dots \partial_n^{\mu_n}$ ,  $\mu = (\mu_1, \dots, \mu_n) \in (\mathbb{Z}_{\geq 0})^n$ . For any set  $S$  let  $\{S\}_\Delta := \{\partial^\mu s \mid s \in S, \mu \in (\mathbb{Z}_{\geq 0})^n\}$ . The differential polynomial ring  $F\{U\}$  is the infinitely generated polynomial ring in the indeterminates  $\{U\}_\Delta$ . The derivations  $\partial_i : F \rightarrow F$  extend to  $\partial_i : F\{U\} \rightarrow F\{U\}$  by setting  $\partial_i \partial_1^{\mu_1} \dots \partial_n^{\mu_n} u^{(j)} = \partial_1^{\mu_1} \dots \partial_i^{\mu_i+1} \dots \partial_n^{\mu_n} u^{(j)}$  ( $1 \leq i \leq n$ ,  $1 \leq j \leq m$ ) via additivity and Leibniz rule. We denote the differential ideal generated by  $p_1, \dots, p_t \in F\{U\}$  by  $\langle p_1, \dots, p_t \rangle_\Delta$ .

A ranking of the differential polynomial ring  $F\{U\}$  is a total ordering  $<$  on the set  $\{U\}_\Delta$  satisfying additional properties (cf. e.g. [14, p. 75]). A ranking  $<$  is called orderly if  $|\mu| < |\mu'|$  implies  $u_\mu^{(i)} < u_{\mu'}^{(j)}$ , where  $|\mu| := \mu_1 + \dots + \mu_n$ . In what follows, we fix an orderly ranking  $<$  on  $F\{U\}$ . The concepts of leader, initial and separant carry over to elements in the polynomial ring  $F\{U\}$ .

Let  $R$  be a residue class ring of a differential polynomial ring by a differential ideal. A differential transcendence basis  $\{p_1, \dots, p_d\} \subset R$  is a maximal set such that  $\bigcup_{i=1}^d \{p_i\}_\Delta$  is algebraically independent over  $F$ . The differential dimension of  $R$  is the corresponding cardinality  $d$ .

A finite set  $S \subset F\{U\} \setminus F$  is called (weakly) triangular if  $\text{ld}(p)$  is not a derivative of  $\text{ld}(q)$  for all  $p, q \in S, p \neq q$ . Define  $S_{<x}$  and  $S_x$  as in the algebraic case. We call  $\mathcal{S}(S) := \langle S \rangle_\Delta : (\text{ini}(S) \cup \text{sep}(S))^\infty \subseteq F\{U\}$  the differential ideal associated to  $S$ . The set  $S$  is called coherent if the  $\Delta$ -polynomials of  $S$  are reduced to zero with respect to  $S$  [21], and it is called a differential regular chain if it is triangular, coherent, and if neither  $\text{ini}(S_x)$  is a zero divisor modulo  $\mathcal{S}(S_{<x})$  nor  $\text{sep}(S_x)$  is a zero-divisor module  $\mathcal{S}(S)$  for each  $x \in \text{ld}(S)$ . An ideal  $\mathcal{S}(S)$  is called characterizable if  $S$  is a differential regular chain.

Let  $S$  be a differential regular chain in  $F\{U\}$ ,  $\ell \in \mathbb{Z}_{\geq 0}$ , and  $L := \{\partial^\mu y \mid y \in \text{ld}(S)\} \cap F\{U\}_{\leq \ell}$  be the set of derivatives of leaders of elements in  $S$  of order at most  $\ell$ . For each  $x \in L$  there exists a  $\mu_{[x]} \in \mathbb{Z}_{\geq 0}^n$  and a  $p_{[x]} \in S$  such that  $\text{ld}(\partial^{\mu_{[x]}} p_{[x]}) = x$ . Define an algebraic triangular set associated to  $S$  as  $S_{\leq \ell} := \{\partial^{\mu_{[x]}} p_{[x]} \mid x \in L\}$ . Although  $S_{\leq \ell}$  depends on the choice of  $\mu_{[x]}$  and  $p_{[x]}$ , it has properties independent of the choice.

**Lemma 2.3 (Rosenfeld’s Lemma)** *Let  $S$  be a differential regular chain in  $F\{U\}$ ,  $\ell \in \mathbb{Z}_{\geq 0}$ , and  $<$  orderly. Then  $S_{\leq \ell}$  is a squarefree regular chain and  $\mathcal{S}_{F\{U\}_{\leq \ell}}(S_{\leq \ell}) = \mathcal{S}(S)_{\leq \ell}$ .*

The idea is due to [21]. For a detailed proof cf. [17, Lemma 1.93].

### 2.3 Numerical Polynomials

Numerical polynomials are elements in the free  $\mathbb{Z}$ -module  $\left\{ \binom{\ell+k}{k} \in \mathbb{Q}[\ell] \mid 0 \leq k \leq n \right\}$ , i.e., rational polynomials that map an integer to an integer. They are totally ordered by  $p \leq q$  if  $p(\ell) \leq q(\ell)$  for all  $\ell$  sufficiently large. Then  $p \leq q$  if and only if either  $p = q$  or there is a  $j \in \{0, \dots, d\}$  such that  $a_k = b_k$  for all  $k > j$  and  $a_j < b_j$ , where  $p = \sum_{k=0}^d a_k \binom{\ell+k}{k}$  and  $q = \sum_{k=0}^d b_k \binom{\ell+k}{k}$ .

### 3 Proofs

#### 3.1 Proof of Existence and Elementary Properties

We prove Theorem 1.1.(1), (2), (4), and (5). Therefore, let  $I \subseteq J \subset F\{U\}$  be characterizable differential ideals,  $S$  be a differential regular chain with respect to an orderly differential ranking  $<$  with  $\mathcal{S}(S) = I$ , and  $\ell \in \mathbb{Z}_{\geq 0}$  be sufficiently big.

Lemma 2.3 implies  $I_{\leq \ell} = \mathcal{S}(S_{\leq \ell})$  and Theorem 2.2 states that the dimension  $\dim(F\{U\}_{\leq \ell}/I_{\leq \ell})$  can be read off from the number of polynomials in  $S_{\leq \ell}$ , which only depends on  $\text{ld}(S)$ . Thus, to prove Theorem 1.1.(1) and 1.1.(2) we may assume  $S = \text{ld}(S)$ . In this case  $\mathcal{S}(S)$  is a prime differential ideal, and hence the statements follow from Kolchin’s original theorem [14, §II.12].

For the proof of Theorem 1.1.(4) note that the transcendence bases of all associated primes of  $\mathcal{S}(S)$  are equal by Theorem 2.2, and for each of these associated prime the claim follows from Kolchin’s original theorem.

To prove Theorem 1.1.(5) note that  $I \subseteq J$  implies  $I_{\leq \ell} \subseteq J_{\leq \ell}$  for all  $\ell \geq 0$ . In particular, the map from  $F\{U\}_{\leq \ell}/I_{\leq \ell}$  to  $F\{U\}_{\leq \ell}/J_{\leq \ell}$  is surjective and, thus,  $\dim(F\{U\}_{\leq \ell}/I_{\leq \ell}) \geq \dim(F\{U\}_{\leq \ell}/J_{\leq \ell})$ . ■

#### 3.2 Invariance Proof

The differential polynomial ring  $F\{U\}$  is filtered by the finitely generated  $F$ -algebras  $F\{U\}_{\leq \ell}$ . This filtration induces a filtration on  $F\{U\}/I$  for a differential ideal  $I$ . To prove the invariance statement in Theorem 1.1.(3) we show that this filtration extends to  $\text{K}(F\{U\}/I)$  if  $I$  is characterizable, where  $\text{K}$  denotes the total quotient ring. Thereby, standard techniques of filtrations can be adapted from Kolchin’s proof.

*Example 3.1* Consider  $\Delta = \{\partial_t\}$ ,  $U = \{u, v\}$ , and  $I := \langle u_0 \cdot v_1 \rangle_{\Delta}$ . Then  $u_0$  is not a zero-divisor in  $F\{U\}_{\leq 0}/I_{\leq 0} \cong F[u_0, v_0]$ , but  $u_0 \cdot v_1 = 0$  in  $F\{U\}/I$ . So, even though the inclusion  $\alpha : F\{U\}_{\leq 0}/I_{\leq 0} \hookrightarrow F\{U\}_{\leq 1}/I_{\leq 1}$  is injective, the image of this map under the total quotient ring functor  $\text{K}$  is no longer injective, as  $\text{K}(\alpha) : \text{K}(F\{U\}_{\leq 0}/I_{\leq 0}) \rightarrow \text{K}(F\{U\}_{\leq 1}/I_{\leq 1}) = \text{K}(F[u_0, v_0, u_1, v_1]/\langle u_0 \cdot v_1 \rangle)$  maps  $u_0$  to zero, as zero divisors become zero in the total quotient ring, cf. e.g. [8, Prop. 2.1].

**Lemma 3.2** *Let  $I \subseteq F\{U\}$  be a characterizable differential ideal and  $\ell \in \mathbb{Z}_{\geq 0}$ . Then,  $F\{U\}_{\leq \ell}/I_{\leq \ell} \hookrightarrow F\{U\}_{\leq \ell+1}/I_{\leq \ell+1}$  induces an inclusion*

$$\text{K}(F\{U\}_{\leq \ell}/I_{\leq \ell}) \hookrightarrow \text{K}(F\{U\}_{\leq \ell+1}/I_{\leq \ell+1}) .$$

*Proof* Any non-zero-divisor in  $F\{U\}_{\leq \ell}/I_{\leq \ell}$  is a non-zero-divisor when considered in  $F\{U\}_{\leq \ell+1}/I_{\leq \ell+1}$  (cf. Proposition 2.1), and thus a unit in  $\text{K}(F\{U\}_{\leq \ell+1}/I_{\leq \ell+1})$ . Hence,  $F\{U\}_{\leq \ell}/I_{\leq \ell} \rightarrow \text{K}(F\{U\}_{\leq \ell+1}/I_{\leq \ell+1})$  factors over  $\text{K}(F\{U\}_{\leq \ell}/I_{\leq \ell})$  by the universal property of localizations. This induces a map  $\iota : \text{K}(F\{U\}_{\leq \ell}/I_{\leq \ell}) \rightarrow$

$\mathbb{K}(F\{U\}_{\leq \ell+1}/I_{\leq \ell+1})$ . Now,  $\ker \iota \cap (F\{U\}_{\leq \ell}/I_{\leq \ell})$  is zero, since it is the kernel of the composition  $F\{U\}_{\leq \ell}/I_{\leq \ell} \hookrightarrow F\{U\}_{\leq \ell+1}/I_{\leq \ell+1} \hookrightarrow \mathbb{K}(F\{U\}_{\leq \ell+1}/I_{\leq \ell+1})$  of monomorphisms. By [8, Prop. 2.2] there is an injection<sup>2</sup> from the set of ideals in  $\mathbb{K}(F\{U\}_{\leq \ell}/I_{\leq \ell})$  into the set of ideals in  $F\{U\}_{\leq \ell}/I_{\leq \ell}$ . This implies  $\ker \iota = 0$ . ■

This filtration is well-behaved under differential isomorphisms.

**Lemma 3.3** *Let  $I \subseteq F\{U\}$  and  $J \subseteq F\{V\}$  be characterizable differential ideals. Let  $\varphi : \mathbb{K}(F\{U\}/I) \rightarrow \mathbb{K}(F\{V\}/J)$  be a differential isomorphism. Then there exists an  $\ell_0 \in \mathbb{Z}_{\geq 0}$  such that*

$$\varphi(\mathbb{K}(F\{U\}_{\leq \ell}/I_{\leq \ell})) \subseteq \mathbb{K}(F\{V\}_{\leq \ell+\ell_0}/J_{\leq \ell+\ell_0}).$$

*Proof*  $F\{U\}/I$  is a (left)  $F[\Delta]$ -module for every differential ideal  $I \subset F\{U\}$ , where  $F[\Delta]$  is the ring of linear differential operators with coefficients in  $F$ . The filtration of  $F[\Delta]$  by the linear differential operators  $F[\Delta]_{\leq k}$  of order  $\leq k$  is compatible with the filtration of  $F\{U\}$  in the sense that  $F[\Delta]_{\leq k}(F\{U\}_{\leq \ell}/I_{\leq \ell}) \subseteq F\{U\}_{\leq \ell+k}/I_{\leq \ell+k}$ . Note that the canonical image of  $F[\Delta]_{\leq \ell}(F\{U\}_{\leq 0}/I_{\leq 0})$  in  $F\{U\}_{\leq \ell}/I_{\leq \ell}$  generates the latter as an  $F$ -algebra. Abusing notation, given any  $F$ -module  $M$  of an  $F$ -algebra, denote by  $\mathbb{K}(M)$  the total quotient ring of the  $F$ -algebra generated by  $M$ . In particular,  $\mathbb{K}(F[\Delta]_{\leq \ell}(F\{U\}_{\leq 0}/I_{\leq 0})) = \mathbb{K}(F\{U\}_{\leq \ell}/I_{\leq \ell})$ .

There exists an  $\ell_0 \in \mathbb{Z}_{\geq 0}$  with  $\varphi(F\{U\}_{\leq 0}/I_{\leq 0}) \subseteq \mathbb{K}(F\{V\}_{\leq \ell_0}/J_{\leq \ell_0})$ , as  $F\{V\}/J = \bigcup_{\ell \in \mathbb{Z}_{\geq 0}} F\{V\}_{\leq \ell}/J_{\leq \ell}$ . Now

$$\begin{aligned} \varphi(\mathbb{K}(F\{U\}_{\leq \ell}/I_{\leq \ell})) &= \varphi(\mathbb{K}(F[\Delta]_{\leq \ell}(F\{U\}_{\leq 0}/I_{\leq 0}))) \\ &= \mathbb{K}(F[\Delta]_{\leq \ell} \varphi(F\{U\}_{\leq 0}/I_{\leq 0})) \\ &\subseteq \mathbb{K}(F[\Delta]_{\leq \ell} \mathbb{K}(F\{V\}_{\leq \ell_0}/J_{\leq \ell_0})) \\ &\subseteq \mathbb{K}(F\{V\}_{\leq \ell+\ell_0}/J_{\leq \ell+\ell_0}) \end{aligned}$$

■

The Krull-dimension changes when passing to total quotient rings. Instead, we use  $\dim_F(R) := \max_{P \in \text{Ass}(R)} \text{trdeg}_F(\mathbb{K}(R/P))$  as notion of dimension for  $F$ -algebras  $R$ . Then,  $\dim(R) = \dim_F(R) = \dim_F(\mathbb{K}(R))$  allows to prove the invariance condition.

*Proof of Theorem 1.1.(3)* Let  $\varphi$  be as in Lemma 3.3. Then,

$$\mathbb{K}(F\{U\}_{\leq \ell}/I_{\leq \ell}) \cong \varphi(\mathbb{K}(F\{U\}_{\leq \ell}/I_{\leq \ell})) \subseteq \mathbb{K}(F\{V\}_{\leq \ell+\ell_0}/J_{\leq \ell+\ell_0})$$

with the  $\ell_0 \in \mathbb{Z}_{\geq 0}$  from Lemma 3.3, and thus

$$\begin{aligned} \dim(F\{U\}_{\leq \ell}/I_{\leq \ell}) &= \dim_F(\mathbb{K}(F\{U\}_{\leq \ell}/I_{\leq \ell})) \\ &\leq \dim_F(\mathbb{K}(F\{V\}_{\leq \ell+\ell_0}/J_{\leq \ell+\ell_0})) \\ &= \dim(F\{V\}_{\leq \ell+\ell_0}/J_{\leq \ell+\ell_0}). \end{aligned}$$

<sup>2</sup>The image consists of those ideals which do not contain any zero divisors.



Thus  $\omega_I(\ell) \leq \omega_I(\ell + \ell_0)$  and by symmetry  $\omega_I(\ell) \leq \omega_I(\ell + \ell_0)$ . Now, an elementary argument implies that the degrees and leading coefficients of  $\omega_I$  and  $\omega_J$  are the same. ■

### 3.3 Comparison Proof

The proof of Theorem 1.1.(6) and (7) uses two propositions, which relate ideals and squarefree regular chains. The first proposition is a direct corollary to Lazard’s Lemma (Theorem 2.2).

**Proposition 3.4** *Let  $S, S'$  be squarefree regular chains in  $F[y_1, \dots, y_n]$  with  $\mathcal{I}(S) \subseteq \mathcal{I}(S')$  and  $|S| = |S'|$ . Then  $\text{ld}(S) = \text{ld}(S')$ .*

The following lemma is used to prove the second proposition. It captures an obvious property of the pseudo reduction with respect to a squarefree regular chain  $S$ : if a polynomial  $p$  can be reduced to zero by  $S$ , but  $\text{ini}(p)$  cannot be reduced to zero, then there must be a suitable element in  $S$  to reduce the highest power of  $\text{ld}(p)$ .

**Lemma 3.5** *Let  $S$  be a squarefree regular chain and  $p \in F[y_1, \dots, y_n]$  with  $\text{ld}(p) = x$ ,  $p \in \mathcal{I}(S)$ , and  $\text{ini}(p) \notin \mathcal{I}(S)$ . Then  $S$  has an element of leader  $x$  and  $\text{deg}_x(S_x) \leq \text{deg}_x(p)$ .*

**Proposition 3.6** *Let  $S$  and  $S'$  be squarefree regular chains in  $R = F[y_1, \dots, y_n]$  with  $\mathcal{I}(S) \subseteq \mathcal{I}(S')$  and  $|S| = |S'|$ . Then,  $\mathcal{I}(S) = \mathcal{I}(S')$  if and only if  $\text{deg}_x(S_x) = \text{deg}_x(S'_x)$  for all  $x \in \text{ld}(S) = \text{ld}(S')$ .*

*Proof* Let  $\text{deg}_x(S_x) = \text{deg}_x(S'_x)$  for all  $x \in \text{ld}(S)$ . We show  $\mathcal{I}(S) \supseteq \mathcal{I}(S')$  by a Noetherian induction. The statement is clear for the principle ideals  $\mathcal{I}(S_{<y_2})$  and  $\mathcal{I}(S'_{<y_2})$ . Let  $p \in \mathcal{I}(S')$  with  $\text{ld}(p) = y_i$  and  $\text{deg}_{y_i}(p) = j$ . Assume by induction that  $q \in \mathcal{I}(S')$  implies  $q \in \mathcal{I}(S)$  for all  $q$  with  $\text{ld}(q) < y_i$  or  $\text{ld}(q) = y_i$  and  $\text{deg}_{y_i}(q) < j$ . Without loss of generality  $\text{ini}(p) \notin \mathcal{I}(S'_{<y_i}) = \mathcal{I}(S)_{<y_i}$ , as otherwise  $p$  has a lower degree in  $y_i$  or a lower ranking leader when substituting  $\text{ini}(p)$  by zero. Now, Lemma 3.5 implies  $y_i \in \text{ld}(S)$  and  $\text{deg}_{y_i}(p) \geq \text{deg}_{y_i}(S_{y_i})$ . Then,

$$r := \text{ini}(S_{y_i}) \cdot p - \text{ini}(p) \cdot y_i^{\text{deg}_{y_i}(p) - \text{deg}_{y_i}(S_{y_i})} \cdot S_{y_i}$$

is in  $\mathcal{I}(S)$  if and only if  $p \in \mathcal{I}(S)$  is, but  $r$  is of lower degree or of lower ranking leader than  $p$ . The claim follows by induction.

Let  $\mathcal{I}(S) = \mathcal{I}(S')$  and  $x \in \text{ld}(S)$ . This implies  $\text{ini}(S_x) \notin \mathcal{I}(S')$ , and thus  $\text{deg}_x(S'_x) \leq \text{deg}_x(S_x)$  by Lemma 3.5. By symmetry  $\text{deg}_x(S'_x) \geq \text{deg}_x(S_x)$ , and thus  $\text{deg}_x(S_x) = \text{deg}_x(S'_x)$ . ■

*Proof of Theorem 1.1.(6) and (7)* Lemma 2.3 reduces the statements to the algebraic case. In this case, Proposition 3.4 implies Theorems 1.1, and 1.1 follows from Proposition 3.6, because all polynomials in  $S_{\leq \ell}$  ( $\ell \in \mathbb{Z}_{\geq 0}$ ) of degree greater than one in their respective leader already lie in  $S$ . ■

## 4 Computation of the Differential Dimension Polynomial

To compute the differential dimension polynomial  $\omega_{\mathcal{S}(S)}$  of a characterizable differential ideal  $\mathcal{S}(S) \subseteq F\{U\}$  for a differential regular chain  $S$  we may assume  $S = \text{ld}(S)$  (cf. Sect. 3.1). This assumption implies that  $\mathcal{S}(S)$  is a prime differential ideal, and for this case there exist well-known combinatorial algorithms for  $\omega_{\mathcal{S}(S)}$  [15].

Alternatively, the differential dimension polynomial  $\omega_{\mathcal{S}(S)}$  can be read off the set of equations  $S$  of a simple differential system [1]. Such a set  $S$  is almost a differential regular chain, except that weak triangularity is replaced by the Janet decomposition, which associates a subset of  $\Delta$  of cardinality  $\zeta_p$  to each  $p \in S$ . Then, the differential dimension polynomial is given by the closed formula

$$\omega_{\mathcal{S}(S)}(l) = m \binom{n + l}{n} - \sum_{p \in S} \binom{\zeta_p + l - \text{ord}(\text{ld}(p))}{\zeta_p},$$

involving only the cardinalities  $\zeta_p$  and the orders  $\text{ord}(\text{ld}(p))$ .

## 5 Examples

For each prime differential ideal  $I$  there exists a differential regular chain  $S$  with  $I = \mathcal{S}(S)$ . Thus, the differential dimension polynomial defined in Theorem 1.1 includes the version of Kolchin. However, the following example shows that Theorem 1.1 is more general.

*Example 5.1* Consider  $U = \{u, v\}$ ,  $\Delta = \{\partial_t\}$ ,  $p = u_1^2 - v$ , and  $q = v_1^2 - v$ . The characterizable differential ideal  $I := \mathcal{S}(\{p, q\})$  is not prime, as  $p - q = (u_1 - v_1)(u_1 + v_1)$ .

Prime differential ideals  $I \subseteq J$  are equal if and only if  $\omega_I = \omega_J$  by Kolchin's theorem. By the following example, this is wrong for characterizable ideals and any generalization to such ideals needs to consider the degrees of polynomials in a differential regular chain.

*Example 5.2* Consider  $\langle u_0^2 - u_0 \rangle_{\Delta} = \mathcal{S}(\{u_0^2 - u_0\}) \subsetneq \langle u_0 \rangle_{\Delta} = \mathcal{S}(\{u_0\})$  in  $F\{u\}$  for  $|\Delta| = 1$ . Both differential ideals are characterizable and have the differential dimension polynomial 0. However, they are not equal.

The next example shows that the Cartan characters and other invariants do not suffice to prove that two solution sets are unequal.

*Example 5.3* For  $\Delta = \{\partial_x, \partial_y\}$  consider the regular chains  $S_1 = \{u_{1,0}\}$  and  $S_2 = \{u_{2,0}, u_{1,1}\}$  in  $\mathbb{C}\{u\}$ . Then  $\mathcal{S}(S_2) \subseteq \mathcal{S}(S_1)$ . The strength and first Cartan character are one and the second Cartan character and differential dimension are zero for both

ideals (in any order high enough), i.e., these values are the same for both ideals. However,  $\mathcal{J}(S_2) \neq \mathcal{J}(S_1)$ , as  $\omega_{\mathcal{J}(S_1)}(\ell) = \ell + 1 \neq \ell + 2 = \omega_{\mathcal{J}(S_2)}(\ell)$ .

In the last example, the differential dimension polynomial proves that a symbolic differential equation solver does not find all solutions.

*Example 5.4* Let  $U = \{u\}$  and  $\Delta = \{\frac{\partial}{\partial t}, \frac{\partial}{\partial x}\}$ . The viscous BURGERS' equation  $b = u_{0,2} - u_{1,0} - 2u_{0,1} \cdot u_{0,0}$  has the differential dimension polynomial  $2\ell + 1$ . MAPLE's `pdsolve` [19] finds the set

$$T := \left\{ c_1 \tanh(c_1 x + c_2 t + c_3) - \frac{c_2}{2c_1} \mid c_1, c_2, c_3 \in \mathbb{C}, c_1 \neq 0 \right\}$$

of solutions, which only depends on three parameters. The differential dimension polynomial shows that the set of solutions is infinite dimensional, and hence  $T$  is only a small subset of all solutions.

**Acknowledgements** The author was partly supported by Schwerpunkt SPP 1489 and Graduiertenkolleg Experimentelle und konstruktive Algebra of the DFG.

## References

1. T. Bächler, V.P. Gerdt, M. Lange-Hegermann, D. Robertz, Algorithmic Thomas decomposition of algebraic and differential systems. *J. Symb. Comput.* **47**(10), 1233–1266 (2012). [arXiv:1108.0817](https://arxiv.org/abs/1108.0817)
2. F. Boulier, D. Lazard, F. Ollivier, M. Petitot, Computing representations for radicals of finitely generated differential ideals. *Appl. Algebra Eng. Commun. Comput.* **20**(1), 73–121 (2009)
3. É. Cartan, Sur la théorie des systèmes en involution et ses applications à la relativité. *Bull. Soc. Math. Fr.* **59**, 88–118 (1931)
4. E.S. Cheb-Terrab, A.D. Roche, Hypergeometric solutions for third order linear odes (2008). [arXiv:0803.3474](https://arxiv.org/abs/0803.3474)
5. E.S. Cheb-Terrab, K. von Bülow, A computational approach for the analytical solving of partial differential equations. *Comput. Phys. Commun.* **90**(1), 102–116 (1995)
6. T. Cluzeau, E. Hubert, Resolvent representation for regular differential ideals. *Appl. Algebra Eng. Commun. Comput.* **13**(5), 395–425 (2003)
7. A. Einstein, *Supplement to Appendix II of "The Meaning of Relativity, 4th ed."* (Princeton University Press, Princeton, NJ, 1953)
8. D. Eisenbud, *Commutative Algebra with a View Toward Algebraic Geometry*. Graduate Texts in Mathematics, vol. 150 (Springer, New York, 1995)
9. X.-S. Gao, W. Li, C.-M. Yuan, Intersection theory in differential algebraic geometry: generic intersections and the differential Chow form. *Trans. Am. Math. Soc.* **365**(9), 4575–4632 (2013)
10. O.D. Golubitsky, M.V. Kondratieva, A.I. Ovchinnikov, On the generalized Ritt problem as a computational problem. *Fundam. Prikl. Mat.* **14**(4), 109–120 (2008)
11. E. Hubert, Notes on triangular sets and triangulation-decomposition algorithms. I. Polynomial systems. In: *Symbolic and Numerical Scientific Computation (Hagenberg, 2001)*. Lecture Notes in Computer Science, vol. 2630 (Springer, Berlin, 2003), pp. 1–39
12. J. Johnson, Differential dimension polynomials and a fundamental theorem on differential modules. *Am. J. Math.* **91**, 239–248 (1969)

13. E.R. Kolchin, The notion of dimension in the theory of algebraic differential equations. *Bull. Am. Math. Soc.* **70**, 570–573 (1964)
14. E.R. Kolchin, *Differential Algebra and Algebraic Groups*. Pure and Applied Mathematics, vol. 54 (Academic, New York, 1973).
15. M.V. Kondratieva, A.B. Levin, A.V. Mikhalev, E.V. Pankratiev, *Differential and Difference Dimension Polynomials*. Mathematics and Its Applications, vol. 461 (Kluwer Academic Publishers, Dordrecht, 1999)
16. M. Lange-Hegermann, The differential counting polynomial. *Found. Comput. Math.* (accepted). [arXiv:1407.5838](https://arxiv.org/abs/1407.5838)
17. M. Lange-Hegermann, Counting solutions of differential equations. PhD thesis, RWTH Aachen (2014). Available at [http://darwin.bth.rwth-aachen.de/opus3/frontdoor.php?source\\_opus=4993](http://darwin.bth.rwth-aachen.de/opus3/frontdoor.php?source_opus=4993)
18. A. Levin, Dimension polynomials of intermediate fields and Krull-type dimension of finitely generated differential field extensions. *Math. Comput. Sci.* **4**(2–3), 143–150 (2010)
19. Maple 17.00, Maplesoft, a division of Waterloo Maple Inc., Waterloo, Ontario
20. J.F. Ritt, *Differential Equations from the Algebraic Standpoint*. (American Mathematical Society, Providence, 1932)
21. A. Rosenfeld, Specializations in differential algebra. *Trans. Am. Math. Soc.* **90**, 394–407 (1959)

# Factorization of $\mathbb{Z}$ -Homogeneous Polynomials in the First $q$ -Weyl Algebra



Albert Heinle and Viktor Levandovskyy

**Abstract** Factorization of elements of noncommutative rings is an important problem both in theory and applications. For the class of domains admitting nontrivial grading, we have recently proposed an approach, which utilizes the grading in order to factor general elements. This is heavily based on the factorization of graded elements. In this paper, we present algorithms to factorize weighted homogeneous (graded) elements in the polynomial first  $q$ -Weyl and Weyl algebras, which are both viewed as  $\mathbb{Z}$ -graded rings. We show that graded polynomials have finite number of factorizations. Moreover, the factorization of such can be almost completely reduced to commutative univariate factorization over the same base field with some additional uncomplicated combinatorial steps. This allows to deduce the complexity of our algorithms in detail, which we prove to be polynomial-time. Furthermore, we show, that for a graded polynomial  $p$ , irreducibility of  $p$  in the polynomial first Weyl algebra implies its irreducibility in the localized (rational) Weyl algebra, which is not true for general polynomials. We report on our implementation in the computer algebra system SINGULAR. For graded polynomials, it outperforms currently available implementations for factoring in the first Weyl algebra—in speed as well as in elegance of the results.

**Keywords** Factorization • Noncommutative factorization • Weyl algebra •  $q$ -Weyl algebra

**Subject Classifications** 68W30, 16Z05, 68-04, 68W40

---

A. Heinle  
David R. Cheriton School of Computer Science, University of Waterloo, Waterloo,  
ON N2L 3G1, Canada  
e-mail: [aheinle@uwaterloo.ca](mailto:aheinle@uwaterloo.ca)

V. Levandovskyy (✉)  
Lehrstuhl B für Mathematik, RWTH Aachen University, 52064 Aachen, Germany  
e-mail: [viktor.levandovskyy@math.rwth-aachen.de](mailto:viktor.levandovskyy@math.rwth-aachen.de)

## 1 Introduction

Factorization of polynomials in commutative rings is a classical and well-studied topic, central to modern computer algebra and widely used in advanced applications.

### 1.1 Problems and Questions

When talking on the factorization of elements in noncommutative rings, a couple of ambiguities are immediately present. Even by restricting our attention to linear partial differential operators, by addressing the term “factorization”, one has to specify at least two more pieces of information:

1. Since linear partial differential operators can be defined over a general differential ring, one needs to say what is the ground ring. A very typical examples are  $\mathbb{K}[x]$  and  $\mathbb{K}(x)$ , where  $\mathbb{K}$  is a fixed computable field. Later on we will see, how different are the properties of the corresponding algebras of operators.
2. Indeed, there are several notions of elements being associated and therefore several notions of a factorization itself. It has an immediate impact on the number of possible factorizations, which can be even infinite.

Let us consider the latter question in detail. In classical ring theory (following, for instance N. Jacobson and P. M. Cohn), two elements  $a, b$  from a ring  $R$  are called *left similar* [4] and denoted by  $a \approx b$ , if  $R/Ra \cong R/Rb$  as left  $R$ -modules. Then it can be proved that some noncommutative domains are unique factorization domains with respect to  $\approx$ , i.e. if  $a_1 \cdot \dots \cdot a_s = b_1 \cdot \dots \cdot b_t$  and  $a_i, b_j$  are non-units from a domain  $R$ , which cannot be written as a product of two units, then  $s = t$  and there exists a permutation  $\sigma$  such that  $\forall i a_i \approx b_{\sigma(i)}$  holds. In particular, both the rational first Weyl algebra  $B_1$  (see Sect. 2.3) and the free associative algebra  $\mathbb{K}\langle x_1, \dots, x_n \rangle$  enjoy this property, which we shortly denote as  $\text{UFD}_{\approx}$ .

On the contrary, more natural generalization of a commutative factorization, which is mostly needed in practice, is given in Definition 1.10. Shortly, the association relation there is stronger:  $a \sim b$  if and only if  $a$  differs from  $b$  by a multiplication with a nonzero central invertible element. With this definition, a single element might have several and even infinitely many different factorizations.

Example 2.12 shows, that the rational first Weyl algebra  $B_1$ , which is an  $\text{UFD}_{\approx}$ , admits *infinitely many* different factorizations with respect to  $\sim$ . At the same time, the free associative algebra  $\mathbb{K}\langle x_1, \dots, x_n \rangle$  is another  $\text{UFD}_{\approx}$ ; from [3] it follows, that any element of it admits only finitely many factorizations with respect to  $\sim$ . However, it is not an  $\text{UFD}_{\sim}$ , as  $xyx - x = x \cdot (yx - 1) = (xy - 1) \cdot x$  demonstrates.

So, a natural question arises in connection with it: given a  $\mathbb{K}$ -algebra, is it a finite factorization domain (FFD) with respect to  $\sim$ ?

## 1.2 Our Methodology and Results

Recently, the two authors together with Bell [3] have answered previous question for a vast family of noncommutative algebras, including polynomial ( $q$ -)Weyl and ( $q$ -)shift algebras. Notably, in the current paper we prove the finiteness of the number  $N(p)$  of factorizations of a graded polynomial  $p$  by elementary methods and give a better upper bound for  $N(p)$ , compared to [3].

Moreover, in [10] the two authors together with Mark Giesbrecht have proposed a novel method, having no analogon in the commutative case, to factorize a general polynomial in a graded algebra by utilizing the factorization of graded polynomials. The algorithms have been implemented for  $n$ th ( $q$ -)Weyl algebras [10] and, recently [16], for a big family of  $G$ -algebras and demonstrated very good performance.

In this work we deal with the class of graded polynomials by describing our methods in detail (for these methods are fundamental for a number of further algorithms). Among other, this is needed for deriving a complexity estimate (which was not investigated in the mentioned papers [10] and [16]) for the factorization by using our method in the case, where the underlying field is computable.

In the next section we report on existing algorithms and implementations for factorizations, which are able to factorize a large number of polynomials. However, as we will see in this paper, there exists a large class of polynomials that seem to form the worst case for the mentioned algorithms. By using our approach one can obtain a factorization of such polynomials very quickly.

We state another main result in Theorem 2.14. There, we prove that irreducible graded polynomials in the polynomial first Weyl algebra stay irreducible when considering them as elements in the rational first Weyl algebra. This is rather unexpected, as this statement is not true for general inhomogeneous polynomials. Moreover, the same statement holds true for the  $q$ -Weyl algebra.

Our algorithms are implemented in the computer algebra system

SINGULAR:PLURAL [6, 11, 12], and since version 3-1-3 they belong to the distribution of SINGULAR as the library `ncfactor.lib`. Since then, a number of additions and enhancements has been done in this library.

## 1.3 Historical Developments

Algebras of linear partial functional operators, such as ( $q$ -)Weyl and ( $q$ -)shift algebras are important objects in mathematics. In particular, they allow algebraization of properties of the solution spaces of systems of equations in the mentioned operators. Especially concerning the problem of finding the solutions of a linear ordinary ( $q$ -)differential equation, the preconditioning step of factorizing this operator may come in helpful.

Often such algebras of operators are noncommutative polynomial rings, and a factorization of an element in those algebras is neither unique in the classical sense (i.e. unique up to multiplication by a unit), nor easy to compute at all in general.

Nevertheless, a lot has been done in this field in the past. Tsarev has studied the form, number and the properties of the factors of a linear differential operator in [32] and [33], where he uses and extends the work presented in [21] and [22].

A very general approach to noncommutative algebras and their properties, including factorization, is also done in [4]. The authors provide several algorithms and introduce various points of views when dealing with noncommutative polynomial algebras.

In his dissertation van Hoeij developed an algorithm to factorize a linear differential operator [34]. There were several papers following that dissertation using and extending those techniques (e.g. [35, 36] and [37]), and nowadays this algorithm is implemented in the `DETOOLS` package of `MAPLE` [25] as the standard algorithm for factorization of those operators.

For the finite field case, Giesbrecht and Zhang have developed a polynomial time algorithm to factor polynomials in  $\mathbb{F}_q(t)[\mathcal{D}; \sigma, \delta]$  [9]. This includes the Weyl algebras with rational function coefficients over a finite field. The applied methodology extends the results in [8].

From the more algebraic point of view and dealing only with strictly polynomial noncommutative algebras, i.e. all units are in the center of the algebra, Melenk and Apel developed a package for the computer algebra system `REDUCE` [24]. This package provides tools to deal with noncommutative polynomial algebras and also contains a factorization algorithm for the supported algebras.

In the computer algebra system `ALLTYPES` [28], which is based on `REDUCE` and solely accessible as a web-service, Schwarz and Grigoriev have implemented the algorithm for factoring differential operators they introduced in [14].

Beals and Kartashova [2] consider the problem of finding a first-order left hand factor of an element from the second Weyl algebra over a computable differential field, where they are able to deduce parametric factors. Similarly, Shemyakova studied factorization properties of linear partial differential operators in [29, 30] and [31]. Concerning special classes of polynomials in algebras of operators, the paper [7] deals with factorization of fourth-order differential equations satisfied by certain Laguerre-Hahn orthogonal polynomials [26].

## 1.4 Preliminaries

We will start by introducing the first  $q$ -Weyl algebra and the first Weyl algebra. By  $\mathbb{K}$ , we always denote an arbitrary field. All algebras are unital associative  $\mathbb{K}$ -algebras. For the complexity discussions, we assume that

1.  $\mathbb{K}$  is computable and its arithmetics have polynomial costs with respect to the bit-size of the elements in  $\mathbb{K}$ .



2. There exists a norm  $|\cdot| : \mathbb{K} \rightarrow \mathbb{R}$ . The representation size in bits for an element  $k \in \mathbb{K}$  is bounded by  $\lceil \log |k| \rceil$ .

The role of the invertible parameter  $q$  can be different: from  $q \in \mathbb{K}$  to  $q$  being transcendental over  $\mathbb{K}$ . We use the unified notation  $\mathbb{K}(q)$  for all these cases. Moreover, for  $m \in \mathbb{N}$  we denote by  $\underline{m}$  the set  $\{1, \dots, m\}$ .

**Definition 1.1** The **polynomial first  $q$ -Weyl algebra**  $Q_1$  is defined as

$$Q_1 := \mathbb{K}(q)\langle x, \partial \mid \partial x = qx\partial + 1 \rangle.$$

For the special case where  $q = 1$  we have the **polynomial first Weyl algebra**  $A_1$ .

*Remark 1.2* The first  $q$ -Weyl algebra can be viewed as an algebra associated to the operator

$$\partial_q : f(x) \mapsto \frac{f(qx) - f(x)}{(q - 1)x},$$

also known as the  $q$ -derivative, where  $f$  is a univariate function in  $x$  (cf. [17]).

For  $q = 1$ , the operator is still well defined. This can be seen in the following way. Let  $f = \sum_{i=0}^n a_i x^i$ , where  $n \in \mathbb{N}_0$  and  $a_i \in \mathbb{K}$ . Then

$$f(qx) - f(x) = \sum_{i=0}^n a_i (qx)^i - \sum_{i=0}^n a_i x^i = \sum_{i=0}^n a_i x^i (q^i - 1).$$

The expression  $q - 1$  is clearly a divisor of  $q^i - 1$  for all  $i \geq 1$ , and we obtain

$$\frac{f(qx) - f(x)}{(q - 1)x} = \sum_{i=1}^n a_i x^{i-1} \left( \sum_{j=0}^{i-1} q^j \right).$$

The first ( $q$ -)Weyl algebra possesses a nontrivial  $\mathbb{Z}$ -grading—utilized, in particular, by Kashiwara and Malgrange in a broader context of the so-called  $V$ -filtration [19, 23]—using the weight vector  $[-v, v]$  for non-zero  $v \in \mathbb{Z}$  on the tuple  $[x, \partial]$ . For simplicity, we will choose  $v := 1$ . In what follows,  $\text{deg}$  denotes the total degree induced by this weight vector. We will write  $\text{deg}_x$  and  $\text{deg}_\partial$  for the degree of a polynomial in  $Q_1$  resp.  $A_1$  with respect to  $x$  and  $\partial$ . From now on, we mean by *homogeneous* or *graded* a polynomial, which is homogeneous with respect to the weight vector  $[-1, 1]$ .

*Example 1.3* We have  $\text{deg}(\partial x) = \text{deg}(x\partial + 1) = \text{deg}(x\partial) = 0$ . Consider

$$p = x\partial^2 + x^4\partial^5 + \partial = (x\partial + x^4\partial^4 + 1)\partial.$$

Then  $\text{deg}_x(p) = 4$ ,  $\text{deg}_\partial(p) = 5$  and  $p$  is  $[-1, 1]$ -homogeneous of degree one.

For  $n \in \mathbb{Z}$ , the  $n$ th graded part (cf. 2.2 for more detailed description) of  $Q_1$  and analogously the  $n$ th graded part of  $A_1$  is given by

$$Q_1^{(n)} := \left\{ \sum_{j-i=n} r_{i,j} x^i \partial^j \mid i, j \in \mathbb{N}_0, r_{i,j} \in \mathbb{K} \right\}.$$

Concerning this choice of degree, the so called **Euler operator**  $\theta := x\partial$ , which is homogeneous of degree 0, will play an important role as we will see soon.

First of all, let us investigate some commutation rules the Euler operator has with  $x$  and  $\partial$ . For  $Q_1$ , in order to abbreviate the size of our formulas, we introduce the so called  $q$ -bracket.

**Definition 1.4** For  $n \in \mathbb{N}$ , we define the  $q$ -bracket  $[n]_q$  by

$$[n]_q := \frac{1 - q^n}{1 - q} = \sum_{i=0}^{n-1} q^i.$$

**Lemma 1.5 (Compare with [27])** In  $A_1$ , for  $n \in \mathbb{N}$  there are commutation rules

$$\begin{aligned} \theta x^n &= x^n (\theta + n), \\ \theta \partial^n &= \partial^n (\theta - n). \end{aligned}$$

More generally, in  $Q_1$  the following commutation rules do hold for  $n \in \mathbb{N}$ :

$$\begin{aligned} \theta x^n &= x^n (q^n \theta + [n]_q), \\ \theta \partial^n &= \frac{\partial^n}{q} \left( \frac{\theta - 1}{q^{n-1}} - \frac{q^{-n+2} - q}{1 - q} \right). \end{aligned}$$

**Remark 1.6** If the characteristic of  $\mathbb{K}$  is  $p > 0$ , the elements  $x^{ap}$  (resp.  $\partial^{ap}$ ) for all  $a \in \mathbb{N}_0$  commute with  $\theta$  in  $A_1$ .

For any  $\mathbb{K}$ , suppose that  $q$  is an  $m$ th root of unity, then  $x^{am}$  (resp.  $\partial^{am}$ ) for all  $a \in \mathbb{N}_0$  commute with  $\theta$  in  $Q_1$ .

**Remark 1.7** With the help of the Lemma above one can also easily see that the polynomial first **shift algebra**

$$\mathbb{K}\langle n, s \mid sn = (n + 1)s \rangle$$

is a subalgebra of the first Weyl algebra  $A_1$ . One of possible embeddings can be realized via the homomorphism of  $\mathbb{K}$ -algebras,  $n \mapsto \theta, s \mapsto \partial$ .

Analogously, though less known, the polynomial first  **$q$ -shift algebra**

$$\mathbb{K}\langle a, b \mid ba = q \cdot ab \rangle$$

can be realized as a subalgebra of the first  $q$ -Weyl algebra  $Q_1$  via the homomorphism  $a \mapsto x, b \mapsto (q - 1)x\partial + 1$ .

Therefore, the factorization techniques developed here for the first Weyl algebra can also be applied to the first shift und  $q$ -shift algebras.

The commutation rules in Lemma 1.5 extend to arbitrary polynomials in  $\theta$ .

**Corollary 1.8** Consider  $f(\theta) := f \in \mathbb{K}[\theta], \theta := x\partial$ . Then, in  $Q_1$ , for all  $n \in \mathbb{N}$

$$f(\theta)x^n = x^n f(q^n \theta + [n]_q),$$

$$f(\theta)\partial^n = \partial^n f\left(\frac{1}{q}\left(\frac{\theta - 1}{q^{n-1}} - \frac{q^{-n+2} - q}{1 - q}\right)\right),$$

whereas in  $A_1$  we have

$$f(\theta)x^n = x^n f(\theta + n),$$

$$f(\theta)\partial^n = \partial^n f(\theta - n).$$

Those are the basic tools we need to explain our approach for factoring homogeneous polynomials in the first Weyl and the first  $q$ -Weyl algebra.

For the complexity discussion, let us define some constants we will utilize in order to estimate the operations needed to perform our methods.

**Definition 1.9** Let us denote by  $\omega_q(n, c)$ , for  $n, c \in \mathbb{N}_0$ , the number of bit operations that an algorithm for factoring a polynomial of degree  $n$  in a univariate polynomial ring over  $\mathbb{K}(q)$ , where each coefficient has at most bit-size  $c$ , needs to perform.

We denote for  $n, c \in \mathbb{N}_0$  by  $\rho_q(n, c)$  the number of bit operations needed to multiply two polynomials in a univariate polynomial ring over  $\mathbb{K}(q)$ , where each polynomial has degree at most  $n$  and where  $c$  is the maximal bit size of each coefficient in the two polynomials.

We will write  $\mathcal{S}_q(n, k, c, \sigma)$ ,  $n, c \in \mathbb{N}_0, k \in \mathbb{Z}, \sigma \in \text{Aut}(\mathbb{K}[x])$ , for the number of bit operations needed for computing  $f(\sigma^k(x))$  for a polynomial  $f$  in  $\mathbb{K}(q)[x]$  of degree  $n$ , where  $x$  is an indeterminate and transcendental over  $\mathbb{K}(q)$  and each coefficient of  $f$  has at most bit-size  $c$ .

If we deal with the case  $q = 1$ , we will omit writing the subscript.

For a detailed complexity discussion, we need to specify the expected output of our factorization algorithms.

**Definition 1.10** Let  $A$  be a polynomial algebra over a field  $\mathbb{K}$  and  $f \in A \setminus \mathbb{K}$  be a polynomial. For a fixed totally ordered monomial  $\mathbb{K}$ -basis of  $A$ , the leading coefficient  $\text{lc}(f)$  of  $f$  is uniquely defined. A **nontrivial factorization** of  $f$  is a tuple  $(c, f_1, \dots, f_m)$ , where  $c \in \mathbb{K} \setminus \{0\}, f_1, \dots, f_m \in A \setminus \{1\}$  are **monic** (i.e. they satisfy  $\text{lc}(f_i) = 1$ ) and  $f = c \cdot f_1 \cdot \dots \cdot f_m$ .

By a slight abuse of notation, we may omit the first element in the tuple if  $c = 1$ .

The following lemma provides a complexity estimate of the cost of testing whether a polynomial in  $Q_1$  resp.  $A_1$  is homogeneous.

**Lemma 1.11** *In order to determine whether a polynomial  $p \in Q_1$  resp.  $p \in A_1$  is homogeneous, it requires  $\#\{\text{Terms in } p\}$  integer additions and comparisons.*

*Proof* A polynomial  $p$  is homogeneous with respect to our definition if and only if in every term the difference between the degree in  $x$  and the degree in  $\partial$  is the same.

Graded elements enjoy nice properties, in particular regarding factorizations.

**Lemma 1.12** *Let  $(\Gamma, +)$  be a monoid, totally ordered by  $<$ , such that  $a < b \Rightarrow a + c < b + c$  for all  $a, b, c \in \Gamma$ . Moreover, let  $D$  be a domain over a field  $\mathbb{K}$ , nontrivially graded by  $\Gamma$ , that is  $D = \bigoplus_{\gamma \in \Gamma} D_\gamma$  for  $\mathbb{K}$ -vector spaces  $D_\gamma$  and  $\forall \alpha, \beta \in \Gamma$  one has  $D_\alpha \cdot D_\beta \subseteq D_{\alpha+\beta}$ .*

*Consider  $d \in D \setminus \{0\}$ . If there is  $m \geq 1$  and  $d_i \in D$ , such that  $d = d_1 \cdot \dots \cdot d_m$ , then  $d$  is  $\Gamma$ -graded if and only if  $d_1, \dots, d_m$  are  $\Gamma$ -graded.*

*Proof* The  $\Leftarrow$  direction follows by the definition of grading, so it remains to prove the  $\Rightarrow$  direction. For an element  $f \in D \setminus \{0\}$ , let us denote by  $\alpha(f) \in \Gamma$  resp. by  $\omega(f) \in \Gamma$  the degree of the highest resp. the lowest nonzero graded part of  $f$ . Note, that  $\omega(f) \leq \alpha(f)$ . Thus  $f = f_{\alpha(f)} + \dots + f_{\omega(f)}$  and, moreover,  $f$  is graded if and only if  $f = f_{\alpha(f)} = f_{\omega(f)}$ .

Suppose  $d = bc$ , where  $b = b_{\alpha(b)} + \dots + b_{\omega(b)}$  and  $c = c_{\alpha(c)} + \dots + c_{\omega(c)}$ . Then  $bc = b_{\alpha(b)}c_{\alpha(c)} + \dots + b_{\omega(b)}c_{\omega(c)}$  is the graded decomposition of  $d = bc$ , and  $(bc)_{\alpha(bc)} = (bc)_{\alpha(b)+\alpha(c)} = b_{\alpha(b)}c_{\alpha(c)}$  since  $D$  is a domain. Analogously  $(bc)_{\omega(bc)} = b_{\omega(b)}c_{\omega(c)}$ . Since  $d = bc$  is graded one has thus  $\alpha(bc) = \omega(bc)$ , that is  $\alpha(b) + \alpha(c) = \omega(b) + \omega(c)$ . Together with  $\alpha(b) \geq \omega(b), \alpha(c) \geq \omega(c)$  this delivers  $\alpha(b) = \omega(b)$  and  $\alpha(c) = \omega(c)$ , proving the claim.

## 2 A New Approach for Factoring Homogeneous Polynomials in the First ( $q$ -)Weyl Algebra

The main idea of our factorization technique lies in the reduction to a natural commutative univariate polynomial subring of  $A_1$  resp.  $Q_1$ , namely  $\mathbb{K}[\theta]$ . We will show that there are only two monic irreducible elements in  $\mathbb{K}[\theta]$ , that are reducible in  $A_1$  resp.  $Q_1$ . Hence, factoring graded elements in  $A_1$  (which belong to fin. gen.  $\mathbb{K}[\theta]$ -modules) can be reduced to factoring in  $\mathbb{K}[\theta]$ , identifying these two elements in a given list of factors, and interchanging using commutation rules.

We will start with discussing how to find one factorization of a given homogeneous polynomial, which, in the process, also leads us to the answer of the question how to find all possible factorizations.

### 2.1 Factoring Homogeneous Polynomials of Degree Zero

**Lemma 2.1** (Compare with [27], Lemma 1.3.1 for  $A_1$ ) *In  $A_1$ , we have the following identity for  $n \in \mathbb{N}$ :*

$$x^n \partial^n = \prod_{i=0}^{n-1} (\theta - i).$$

*In  $Q_1$ , one can rewrite  $x^n \partial^n$  as element in  $\mathbb{K}[\theta]$  and it is equal to*

$$\frac{1}{q^{T_{n-1}}} \prod_{i=0}^{n-1} \left( \theta - \sum_{j=0}^{i-1} q^j \right) = \frac{1}{q^{T_{n-1}}} \prod_{i=0}^{n-1} (\theta - [i]_q),$$

where  $T_i$  denotes the  $i$ th triangular number  $\sum_{j=0}^i j = \frac{i(i+1)}{2}$  for all  $i \in \mathbb{N}_0$ .

Therefore the factorization of a homogeneous polynomial  $p$  of degree zero can be done by rewriting  $p$  as element in  $\mathbb{K}[\theta]$  and by factoring it in  $\mathbb{K}[\theta]$ , which is implemented in every computer algebra system for practical choices of  $\mathbb{K}$ .

Of course, this does not already yield a complete factorization, since there are still elements irreducible in  $\mathbb{K}[\theta]$ , but reducible in  $Q_1$  resp.  $A_1$ . An obvious example is  $\theta = x\partial$  itself. Fortunately, there are only two monic polynomials irreducible in  $\mathbb{K}[\theta]$ , but reducible in  $A_1$  resp.  $Q_1$ . This is shown by Lemma 2.3, which requires the following proposition for its proof.

**Proposition 2.2**  $Q_1^{(0)}$  is a  $\mathbb{K}$ -algebra, generated by the element  $\theta := x\partial$ . The graded direct summands  $Q_1^{(k)}$  are cyclic  $Q_1^{(0)}$ -bimodules, generated by the element  $x^{-k}$ , if  $k < 0$ , or by  $\partial^k$ , if  $k > 0$ . Literally the same holds, when  $Q_1$  is replaced by  $A_1$ .

*Proof* The first statement can be seen using Lemma 2.1, as we can identify  $Q_1^{(0)}$  resp.  $A_1^{(0)}$  with  $\mathbb{K}[\theta]$ . For the second statement recall that being homogeneous of degree  $k \in \mathbb{Z}$  for a polynomial  $p \in Q_1^{(k)}$  resp.  $p \in A_1^{(k)}$  means, that every monomial is – for a certain  $n \in \mathbb{N}_0$  – of the form  $x^n \partial^{n+k}$ , if  $k \geq 0$ , or of the form  $x^{n-k} \partial^n$ , if  $k < 0$ . Since we can transform  $x^n \partial^n$  into an expression in  $\mathbb{K}[\theta]$  via Lemma 2.1 and use the commutation rules in Lemma 1.5, we can move  $x^{-k}$  resp.  $\partial^k$  to the right and the left and hence obtain the desired bimodule structure.

**Lemma 2.3** *The polynomials  $\theta$  and  $\theta + \frac{1}{q}$  are the only irreducible monic elements in  $\mathbb{K}[\theta]$  that are reducible in  $Q_1$ . For  $A_1$ , the polynomials  $\theta$  and  $\theta + 1$  are the only irreducible monic elements in  $\mathbb{K}[\theta]$  that are reducible in  $A_1$ .*

*Proof* We will only consider the proof for  $Q_1$ , as the proof for  $A_1$  is done in an analogue way. Let  $f \in \mathbb{K}[\theta]$  be a monic polynomial. Assume that it is irreducible in  $\mathbb{K}[\theta]$ , but reducible in  $Q_1$ . Let  $\varphi, \psi$  be elements in  $Q_1$  with  $\varphi\psi = f$ . Then  $\varphi$  and  $\psi$  are homogeneous and  $\varphi \in Q_1^{(-k)}$ ,  $\psi \in Q_1^{(k)}$  for a  $k \in \mathbb{Z} \setminus \{0\}$ . As for the case where

$k$  is negative a similar argument is applicable, we assume without loss of generality that  $k$  is positive.

Due to Proposition 2.2, we have for some  $\tilde{\varphi}, \tilde{\psi} \in \mathbb{K}[\theta]$

$$\varphi = \tilde{\varphi}(\theta)x^k, \quad \psi = \tilde{\psi}(\theta)\partial^k.$$

Using Corollary 1.8, we obtain

$$f = \tilde{\varphi}(\theta)x^k\tilde{\psi}(\theta)\partial^k = \tilde{\varphi}(\theta)x^k\partial^k\tilde{\psi} \left( \frac{1}{q} \left( \frac{\theta - 1}{q^{n-1}} - \frac{q^{-n+2} - q}{1 - q} \right) \right).$$

As we know from Lemma 1.5 the equation

$$x^k\partial^k = \frac{1}{q^{\overline{k-1}}} \prod_{i=0}^{k-1} \left( \theta - \sum_{j=0}^{i-1} q^j \right)$$

holds. Thus, because we assumed  $f$  to be irreducible in  $\mathbb{K}[\theta]$ , we must have  $\tilde{\varphi}, \tilde{\psi} \in \mathbb{K}$  and  $k = 1$  due to Lemma 1.5. Because  $f$  is monic, we must also have  $\tilde{\varphi} = \tilde{\psi}^{-1}$ .

As a result, the only possible  $f$  is  $f = \theta$ . If we originally had chosen  $k$  to be negative, the only possibility for  $f$  would be  $f = \theta + \frac{1}{q}$ . This completes the proof.

Therefore, we have a procedure for factoring a homogeneous polynomial  $p \in A_1$  (resp.  $p \in Q_1$ ) of degree zero in  $\mathbb{K}[\theta]$ . It is done using the following steps.

1. Rewrite  $p$  as an element in  $\mathbb{K}[\theta]$ ;
2. Factorize  $p$  in  $\mathbb{K}[\theta]$  using commutative methods, i.e. obtain a list  $[c, p_1, \dots, p_\ell] \in \mathbb{K} \times \mathbb{K}[\theta]^\ell$ ,  $\ell \in \mathbb{N}$ , where  $c \cdot p_1 \cdot \dots \cdot p_\ell = p$ .
3. For every  $p_j, j \in \underline{\ell}$  which is equal to  $\theta$  or  $\theta + 1$  (resp.  $\theta + \frac{1}{q}$ ), remove  $p_j$  from the list and insert into position  $j$  and  $j + 1$  the elements  $x_i, \partial_i$  resp.  $\partial_i, x_i$ .
4. Replace for every element in the list from the previous step  $\theta$  by  $x \cdot \partial$ . Return the resulting list.

Let us consider the complexity of the above steps to factor a homogeneous element of degree zero in  $A_1$ .

**Ad step 1:** The polynomial  $p$  has, due to the assumption of being homogeneous of degree zero, the form

$$p = \sum_{i=0}^n p_i x^i \partial^i, \quad n \in \mathbb{N}, p_i \in \mathbb{K} \text{ (resp. } \mathbb{K}(q)). \tag{1}$$

In order to transform it into an element in  $\mathbb{K}[\theta]$ , we have to apply the rewriting rule stated in Lemma 2.1 for every term  $x^i \partial^i$  in  $p$ . For that, one makes use of the identity

$$x^{n+1} \partial^{n+1} = x^n \partial^n \cdot (\theta - n).$$

Thus, in order to perform step 1, we need to perform for every  $i \in \underline{n}$  a multiplication of a polynomial in  $\mathbb{K}[\theta]$  of degree  $i$  with a polynomial of degree 1.

**Ad step 2:** Unfortunately, the factorization problem even in the univariate case does not have polynomial complexity in general. One might face exponential complexity with respect to the bit-length of the coefficients in  $\mathbb{K}$  or it might even be undecidable, depending on the choice of  $\mathbb{K}$ .

An example for a polynomial-time complexity with respect to the bit-length of the coefficients would be  $\mathbb{K} = \mathbb{Q}$ , due to the famous LLL algorithm by Lenstra, Lenstra and Lovász developed in 1982 [20]. For certain classes of fields, including algebraic ones, polynomial time algorithms have been discovered in [5] and [13]. For further reading on the complexity of the factorization problem we also recommend [18] and [38]. As in Definition 1.9, we simply write  $\omega(n)$  resp.  $\omega_q(n)$  for the amount of bit operations needed for factoring a univariate polynomial of degree  $n$ .

**Ad step 3:** In order to find and identify the polynomials, it does not require any operations on the polynomials other than comparisons.

**Ad step 4:** For each monomial in each factor that has degree zero, we need to replace  $\theta$  by  $x \cdot \partial$  and bring it into the normal form, i.e. each monomial in the end must have the form  $x^i \partial^i$  for  $i \leq n$ . This can be calculated, up to a constant factor, with the same number of operations as performed for step 1, since we only need to reverse the mapping outlined there.

Thus, we can formulate the following corollary.

**Corollary 2.4** *Given  $p$  as in (1), let  $b$  be the maximal coefficient in  $p$  with respect to its bit-size. In order to obtain one factorization of  $p$  over  $Q_1$ , it requires*

$$O(n \cdot \rho_q(n, \lceil \log |n!| \rceil) + \omega_q(n, \lceil \log |b \cdot n!| \rceil)) \tag{2}$$

*bit operations.*

*Example 2.5* Let  $\mathbb{K} := \mathbb{Q}$  and  $p := x^3 \partial^3 + 4x^2 \partial^2 + 3x \partial \in A_1$ . Clearly  $p$  is homogeneous of degree zero; rewritten in  $\mathbb{K}[\theta]$ , one obtains  $p = \theta^3 + \theta^2 + \theta$ . This polynomial factorizes in  $\mathbb{K}[\theta]$  to  $\theta \cdot (\theta^2 + \theta + 1)$ , which further factorizes as  $\theta$  is reducible to  $x \cdot \partial \cdot (\theta^2 + \theta + 1) \in A_1$ . To get more (in fact, as we will see in the next subsection, all) possible factorizations of  $p$ , we apply the commutation rules with  $x$  resp.  $\partial$  and obtain the following factorizations:

$$x \cdot \partial \cdot (\theta^2 + \theta + 1) = x \cdot (\theta^2 + 3\theta + 3) \cdot \partial = (\theta^2 + \theta + 1) \cdot x \cdot \partial.$$

## 2.2 Factoring Homogeneous Polynomials of Arbitrary Degree

Fortunately, the hard work is already done and factoring of homogeneous polynomials of arbitrary degree is just a small further step.

The reason is Proposition 2.2, which leads to the following steps to obtain one factorization of a homogeneous polynomial  $p \in Q_1^{(k)}$  resp.  $p \in A_1^{(k)}$  of degree  $k \in \mathbb{Z}$ .

1. Represent  $p$  as  $\tilde{p}x^{-k}$  resp.  $\tilde{p}\partial^k$ , where  $\tilde{p}$  in  $A_1^{(0)}$ , written as polynomial in  $\mathbb{K}[\theta]$ . We need  $O(d^2 \cdot \rho_q(d, \lceil \log |b \cdot d!| \rceil))$ , where  $d := \min\{\deg_x(p), \deg_\partial(p)\}$  and  $b \in \mathbb{K}$  denotes the maximal coefficient in  $p$  with respect to the bit-size, operations to obtain this  $\tilde{p}$ . Afterwards, if  $k < 0$ , one additional application of a  $k$ -shift to  $\tilde{p}$  is required.
2. Factorize  $\tilde{p}$  (which is homogeneous of degree zero) by using the steps shown in the previous subsection.

Now we have everything we need to formulate an algorithm to find one factorization of a homogeneous element in  $A_1$  resp.  $Q_1$ , namely Algorithm 1 which can be found below. The next corollary states a complexity estimate of the Algorithm 1.

**Corollary 2.6** *Let  $p \in Q_1$  be homogeneous of degree  $k \in \mathbb{Z}$ , and let all the coefficients in  $p$  have bit size at most  $b \in \mathbb{N}_0$ . Then, due to Proposition 2.2,  $p$  can be written in the form  $p = p_0\varphi^{|k|}$ , where  $p_0$  is a polynomial of degree  $n \in \mathbb{N}_0$  in  $\mathbb{K}(q)[\theta]$  and  $\varphi \in \{x, \partial\}$ . Obtaining one factorization in  $Q_1$  of  $p$  requires*

$$O(n \cdot \rho_q(n, \lceil \log |n!| \rceil) + \omega_q(n, \lceil \log |b \cdot n!| \rceil) + \mathcal{S}_q(n, k, \lceil \log |b \cdot n!| \rceil, \sigma))$$

*bit operations, where  $\sigma(x) = x + 1$  if  $q = 1$ , and  $\sigma(x) = q \cdot x + 1$  otherwise.*

We also would like to address the topic how to obtain all possible factorizations of a homogeneous polynomial. As mentioned before, the factorization of a polynomial in a noncommutative ring is generally not unique in the classical sense, i.e. up to multiplication by units or up to interchanging factors. Thus several different factorizations can occur. For the homogeneous case, they can fortunately be easily characterized by the commutation rules from Lemma 1.5 and the identities from Lemma 2.3. This is proven by the following Lemma.

**Lemma 2.7** *Let  $z \in \mathbb{Z}$  and  $p \in A_1^{(z)}$ , resp.  $p \in Q_1^{(z)}$ , is monic. Suppose, that one factorization of  $p$  has been constructed following Proposition 2.2 and has the form  $Q(\theta) \cdot T(\theta) \cdot \psi^{|z|}$ , where*

- $T(\theta) = (x\partial)^t(\partial x)^s$ ,  $t, s \in \mathbb{N}_0$ , is a product of irreducible factors in  $\mathbb{K}[\theta]$ , which are reducible in  $A_1$ , resp.  $Q_1$ ,
- $Q(\theta)$  is the product of irreducible factors in both  $\mathbb{K}[\theta]$  and  $A_1$  (resp.  $Q_1$ ), and
- $\psi = x$ , if  $z < 0$ , and  $\psi = \partial$  otherwise.

*Let  $p_1 \cdot \dots \cdot p_m$  for  $m \in \mathbb{N}$  be another nontrivial factorization of  $p$ . Then this factorization can be derived from  $Q(\theta) \cdot T(\theta) \cdot \psi^{|z|}$  by using two operations, namely (i) “swapping”, that is interchanging two adjacent factors according to the commutation rules and (ii) “rewriting” of occurring  $\theta$  resp.  $\theta + 1$  ( $\theta + \frac{1}{q}$  in the  $q$ -Weyl case) by  $x \cdot \partial$  resp.  $\partial \cdot x$ .*



---

**Algorithm 1** HomogFac: factorization of a homogeneous polynomial in the first  $(q)$ -Weyl algebra
 

---

*Input:*  $h \in A_1^{(m)}$  (resp.  $h \in Q_1^{(m)}$ ), where  $m \in \mathbb{Z}$

*Output:*  $(f_1, \dots, f_n) \in A_1^n$  resp.  $(f_1, \dots, f_n) \in Q_1^n$ , such that  $f_1 \cdot \dots \cdot f_n = h$ ,  $n \in \mathbb{N}$

*Assumption:*  $h$  is normalized, i.e. the leading coefficient is 1.

```

1: if  $m \neq 0$  then
2:   if  $m < 0$  then
3:     Determine  $\hat{h} \in A_1^{(0)}$  such that  $h = \hat{h}x^{-m}$ 
4:      $factor := \underbrace{(x, \dots, x)}_{-m \text{ times}}$ 
5:   else
6:     Determine  $\hat{h}$  such that  $h = \hat{h}\partial^m$ 
7:      $factor := \underbrace{(\partial, \dots, \partial)}_{m \text{ times}}$ 
8:   end if
9: else
10:   $\hat{h} := h$ 
11:   $factor := 1$ 
12: end if
13:  $(\hat{f}_1, \dots, \hat{f}_l) :=$  Factorization of  $\hat{h}$  as element in  $\mathbb{K}[\theta]$  ( $l \in \mathbb{N}$ )
14:  $(\hat{f}_1, \dots, \hat{f}_l) :=$  Substitute  $\theta$  by  $x \cdot \partial$  in  $(\hat{f}_1, \dots, \hat{f}_l)$ 
15:  $result := ()$ 
16: for  $i$  from 1 to  $l$  do
17:   if  $\hat{f}_i = x \cdot \partial$  then
18:     Append  $x$  and  $\partial$  to  $result$ 
19:   else
20:     if  $\hat{f}_i = \partial \cdot x$  then
21:       Append  $\partial$  and  $x$  to  $result$ 
22:     else
23:       Append  $\hat{f}_i$  to  $result$ 
24:     end if
25:   end if
26: end for
27: Append each element in  $factor$  to  $result$ 
28: return  $result$ 

```

---

*Proof* Since  $p$  is homogeneous, all  $p_i$  for  $i \in \underline{m}$  are homogeneous. Thus each of them can be written in the form  $p_i = \tilde{p}_i(\theta) \cdot \psi_{e_i}$ , where  $e_i \in \mathbb{Z}$ , and  $\psi_{e_i} = x^{-e_i}$ , if  $e_i < 0$  and  $\psi_{e_i} = \partial^{e_i}$  otherwise. With respect to the commutation rules as stated in Corollary 1.8, we can swap the  $\tilde{p}_i(\theta)$  to the left for any  $2 \leq i \leq m$ . Note that it is possible for them to be transformed to the form  $\theta$  resp.  $\theta + 1$  ( $\theta + \frac{1}{q}$  in the  $q$ -Weyl case), after performing these swapping steps. I.e., we have commuting factors, both belonging to  $Q(\theta)$ , as well as to  $T(\theta)$  at the left. Our resulting product is thus  $\tilde{Q}(\theta)\tilde{T}(\theta) \prod_{j=1}^m \psi_{e_j}$ , where the factors in  $\tilde{Q}(\theta)$ , resp.  $\tilde{T}(\theta)$ , contain a subset of the factors of  $Q(\theta)$  resp.  $T(\theta)$ . By our assumption of  $p$  having degree  $z$ , we are able to swap  $\psi_z$  to the right in  $F := \prod_{j=1}^m \psi_{e_j}$ , i.e.,  $F = \tilde{F}\psi_z$  for  $\tilde{F} \in A_1^{(0)}$ . This step may involve combining  $x$  and  $\partial$  to  $\theta$  resp.  $\theta + 1$  ( $\theta + \frac{1}{q}$  in the  $q$ -Weyl case). Afterwards,

this is also done to the remaining factors in  $\tilde{F}$  that are not yet polynomials in  $\mathbb{K}[\theta]$  using the swapping operation. These polynomials are the remaining factors that belong to  $Q(\theta)$ , resp.  $T(\theta)$ , and can be swapped commutatively to their respective positions. Since reverse engineering of those steps is possible, we can derive the factorization  $p_1 \cdot \dots \cdot p_m$  from  $Q(\theta) \cdot T(\theta) \cdot \psi_z$  as claimed.

With the help of the above lemma, we are also able to formulate an algorithm to find all factorizations of a given homogeneous polynomial in  $A_1$ , namely Algorithm 2 as stated below.

In order to discuss the complexity of finding all factorizations of a homogeneous element in  $A_1$  resp.  $Q_1$ , we need an upper bound on the number of possible factorizations. With the lemma it becomes clear, that unlike the rational ( $q$ -)Weyl algebras (see Example 2.12), homogeneous elements in  $A_1$  resp.  $Q_1$  always have a finite number of factorizations.

---

**Algorithm 2** HomogFacAll: all factorizations of a homogeneous polynomial in the first ( $q$ -)Weyl algebra

---

*Input:*  $h \in A_1^{(m)}$  (resp.  $h \in Q_1^{(m)}$ ), where  $m \in \mathbb{Z}$

*Output:*  $\{(f_1, \dots, f_n) \in A_1^n \mid f_1 \cdot \dots \cdot f_n = h, n \in \mathbb{N}\}$

*Assumption:*  $h$  is normalized, i.e. the leading coefficient is 1.

```

1:  $(f_1, \dots, f_v, g, \dots, g) := \text{HomogFac}(h)$  without lines 16 – 26
    $\{v \in \mathbb{N}_0, g \in \{x, \partial\}, f_i \in A_1^{(0)}\}$ 
2: Rewrite each  $f_i$  as element in  $\mathbb{K}[\theta]$ 
3:  $result := \{\text{Permutations of } (f_1, \dots, f_v, g, \dots, g) \text{ with respect to the commutation rules}\}$ 
4: for  $(g_1, \dots, g_n) \in result$  do
5:   for  $i$  from 1 to  $n$  do
6:     if  $g_i = \theta$  then
7:        $g_i := x, \partial$ 
8:        $leftpart := \{(g_1, \dots, g_k, x, g_{k+1}(\theta + 1), \dots, g_{i-1}(\theta + 1)) \mid k \leq i - 1, g_j \in A_1^{(0)} \text{ for all } k < j \leq i - 1\}$ 
9:        $rightpart := \{(g_{i+1}(\theta + 1), \dots, g_{k-1}(\theta + 1), \partial, g_k, \dots, g_n) \mid k \geq i + 1, g_j \in A_1^{(0)} \text{ for all } i + 1 \leq j < k\}$ 
10:      Append each element in  $\{(l_1, \dots, l_j, r_1 \dots r_k) \mid (l_1, \dots, l_j) \in leftpart, (r_1, \dots, r_k) \in rightpart \text{ for } j, k \in \mathbb{N}\}$  to  $result$ .
11:     end if
12:     if  $g_i = \theta + 1$  (resp.  $g_i = \theta + \frac{1}{q}$ ) then
13:        $g_i := \partial, x$ 
14:        $leftpart := \{(g_1, \dots, g_k, \partial, g_{k+1}(\theta - 1), \dots, g_{i-1}(\theta - 1)) \mid k \leq i - 1, g_j \in A_1^{(0)} \text{ for all } k < j \leq i - 1\}$ 
15:        $rightpart := \{(g_{i+1}(\theta - 1), \dots, g_{k-1}(\theta - 1), x, g_k, \dots, g_n) \mid k \geq i + 1, g_j \in A_1^{(0)} \text{ for all } i + 1 \leq j < k\}$ 
16:      Append each element in  $\{(l_1, \dots, l_j, r_1 \dots r_k) \mid (l_1, \dots, l_j) \in leftpart, (r_1, \dots, r_k) \in rightpart \text{ for } j, k \in \mathbb{N}\}$  to  $result$ .
17:     end if
18:   end for
19: end for
20: return  $result$ 

```

---

**Lemma 2.8** *Let  $p = p_0 \cdot \varphi^k$  be a homogeneous polynomial in  $A_1$  resp.  $Q_1$ , where  $k \in \mathbb{N}$ ,  $p_0 \in \mathbb{K}[\theta]$  and  $\varphi \in \{x, \partial\}$ . Furthermore let  $n := \deg_\theta(p_0)$ . Then the number of different factorizations of  $p$  is at most*

$$\left(\left\lfloor \frac{n}{2} \right\rfloor + 1\right)^2 \cdot n! \cdot \binom{n+k}{k}.$$

*Proof* Let us assume that  $p_0$  decomposes in  $\mathbb{K}[\theta]$  into  $\tilde{n} \in \mathbb{N}$  factors, where  $\tilde{n} \leq n$ . As all of these factors commute, there are up to  $\tilde{n}!$  different possibilities to rearrange them. For every such arrangement of the factors of  $p_0$ , we can place the  $k$  available  $\varphi$  at any position (with applied shift to the respective factors of  $p_0$ ), which leads to  $\binom{\tilde{n}+k}{k}$  possibilities each time. Finally, due to Lemma 2.3, the linear factors of  $p_0$  might split into  $f_1 \cdot f_2$ , where  $(f_1, f_2) \in \{(x, \partial), (\partial, x)\}$ . The element  $f_1$  can be swapped into up to  $j \leq \tilde{n} + 1$  positions to its left, and the element  $f_2$  can be swapped into  $\tilde{n} - j + 1$  positions to its right. The possibilities maximize if  $j = \left\lfloor \frac{\tilde{n}}{2} \right\rfloor + 1$ , which we can consider as upper bound. Hence, this would add for each instance at most  $\left(\left\lfloor \frac{\tilde{n}}{2} \right\rfloor + 1\right)^2$  new distinct factorizations. As  $p_0$  factors at most into linear factors, we can assume  $\tilde{n} = n$  and obtain the stated upper bound.

*Remark 2.9* In [3] we prove that in the case of the polynomial  $n$ th ( $q$ -)Weyl algebra, a nonzero polynomial has only finitely many different factorizations. In yet another recent paper [10] we have developed an algorithm for computing all factorizations of a given polynomial in the  $n$ th ( $q$ -)Weyl algebra.

The termination of Algorithms 1 and 2 is clear, as we only iterate over finite sets. The correctness follows by our preliminary work.

**Corollary 2.10** *With the notations as in Corollary 2.6, the number of different factorizations of  $p$  by Lemma 2.8 is bounded by*

$$\left(\left\lfloor \frac{n}{2} \right\rfloor + 1\right)^2 \cdot n! \cdot \binom{n+|k|}{|k|}.$$

*In order to obtain all these different factorizations, it would require*

$$\begin{aligned} &O\left(n \cdot \rho_q(n, \lceil \log |n!| \rceil) + \omega_q(n, b + \lceil \log |n!| \rceil) \right. \\ &\quad \left. + \left(n^2 + \left(\left\lfloor \frac{n}{2} \right\rfloor + 1\right)^2 \cdot n! \cdot \binom{n+|k|}{|k|}\right) \mathcal{L}_q(n, 1, \lceil \log |b \cdot n!| \rceil, \sigma) \right) \end{aligned}$$

*bit operations, where  $\sigma(x) = x + 1$  if  $q = 1$ , and  $\sigma(x) = q \cdot x + 1$  otherwise.*

### 2.3 Application to the Rational First Weyl Algebra

In practice, one is often interested in ordinary differential equations over the field of rational functions in the indeterminate  $x$ . We refer to the corresponding algebra of operators as the **first rational Weyl algebra** and denote it as  $B_1$ . The commutation rules over  $B_1$  are extended from those in  $A_1$ , that is  $\partial g(x) = g(x)\partial + \frac{\partial g(x)}{\partial x}$  for  $g(x) \in \mathbb{K}(x)$ . Unlike in the polynomial Weyl algebra, an infinite number of nontrivial factorizations of an element is possible. The easiest example is the polynomial  $\partial^2 \in A_1$ , having except  $\partial \cdot \partial$  a family of nontrivial factorizations  $(\partial + \frac{1}{x+c})(\partial - \frac{1}{x+c})$  for all  $c \in \mathbb{K}$  over  $B_1$ ; the only factorization in  $A_1$  is  $\partial \cdot \partial$ . Thus, at first glance, the factorization problem in both the rational and the polynomial Weyl algebras seems to be distinct in general. But there are still many things in common.

The formalism of the **Ore localization** of a ring (cf. e. g. [4]) can be briefly recalled as follows. Let  $R$  be a domain and  $\{0\} \subsetneq S \subset R$  be a multiplicatively closed **Ore set** in  $R$ , i. e. the **Ore condition** holds for  $S$  and  $R$  (the condition will appear below). Then there exists a localized ring, denoted by  $S^{-1}R$  together with the classical embedding  $\iota : R \rightarrow S^{-1}R, r \mapsto 1^{-1}r$ , such that  $\iota(S) \subset S^{-1}R$  becomes invertible. Note, that the presentation of a left fraction  $s^{-1}r \in S^{-1}R$  via the tuple  $(s, r) \in S \times R$  defines an equivalence class and is by no means unique.

Rational ( $q$ -)Weyl algebras can be recognized as Ore localizations of polynomial ( $q$ -)Weyl algebras with respect to the multiplicatively closed set  $S := \mathbb{K}[x] \setminus \{0\}$ , which can be proven to be an Ore set both in  $A_1$  and in  $Q_1$ . Let us clarify the connection between factorizations in an algebra and in its Ore localization.

**Lemma 2.11** *Let  $R$  be a domain and  $S \subset R$  be an Ore set in  $R$ . Then for any  $m \in \mathbb{N}$  and for any  $h_1, \dots, h_m \in S^{-1}R \setminus \{0\}$  there exist  $q \in S$  and  $\tilde{h}_1, \dots, \tilde{h}_m \in R \setminus \{0\}$  such that  $q \cdot h_1 \cdot \dots \cdot h_m = \tilde{h}_1 \cdot \dots \cdot \tilde{h}_m$ .*

*Proof* Suppose that  $h = h_1 h_2 = (s_1^{-1} r_1) \cdot (s_2^{-1} r_2)$  for  $r_i \in R, s_i \in S$ . Then by the Ore condition  $\exists \hat{r}_1 \in R, \hat{s}_2 \in S$  such that  $r_1 s_2^{-1} = \hat{s}_2^{-1} \hat{r}_1$ . Thus  $h = s_1^{-1} \hat{s}_2^{-1} \hat{r}_1 r_2$  and for  $q = \hat{s}_2 s_1 \in S$  and  $\tilde{h}_1 = \hat{r}_1, \tilde{h}_2 = r_2 \in R$  one has  $qh = \tilde{h}_1 \tilde{h}_2 \in R$ . The rest follows by induction.

Thus we can lift any factorization from the ring  $S^{-1}R$  to a factorization in  $R$  by a left multiplication with an element of  $S$ .

*Example 2.12* As it was mentioned before, in the first rational Weyl algebra one has  $\partial^2 = (\partial + \frac{1}{x+c})(\partial - \frac{1}{x+c})$  for all  $c \in \mathbb{K}$ . Let us fix  $c$  and analyze the lifting.

$$(\partial + (x+c)^{-1})(\partial - (x+c)^{-1}) = (x+c)^{-1} \cdot ((x+c)\partial + 1) \cdot (x+c)^{-1} \cdot ((x+c)\partial - 1)$$

Since  $\partial \cdot (x+c) = (x+c)\partial + 1$ , one has  $((x+c)\partial + 1) \cdot (x+c)^{-1} = \partial$  and thus

$$\partial^2 = (x+c)^{-1} \cdot \partial \cdot ((x+c)\partial - 1),$$

from which we read off the corresponding factorization

$$(x + c) \cdot \partial^2 = \partial \cdot ((x + c)\partial - 1)$$

in the polynomial first Weyl algebra. In the notation of the preceding Lemma  $q = x + c, \tilde{h}_1 = \partial, \tilde{h}_2 = (x + c)\partial - 1$ . In particular, the infinite family of factorizations we started with does not propagate to the polynomial case: as we see, the parameter  $c$  is present in the lifted polynomial  $(x + c)\partial^2$ . By our approach we can prove, that for any  $c \in \mathbb{K}$  the only factorizations of  $x\partial^2 + c\partial^2$  in  $A_1$  are

$$(x + c) \cdot \partial^2 = \partial \cdot ((x + c)\partial - 1).$$

**Proposition 2.13** *Let  $U := \{r \in R \mid 1^{-1}r \in S^{-1}R \text{ is invertible}\} \subset R$ . Then*

1.  $r \in U \subseteq R \Leftrightarrow \exists w \in R : wr \in S$ .
2. Let  $R \in \{A_1, Q_1\}$  and  $S = \mathbb{K}[x] \setminus \{0\}$ . Let  $h = h_1 \cdot \dots \cdot h_m$  be a factorization of a fraction  $h \in S^{-1}R \setminus \{0\}$ . Then there exist  $q \in S$  and an associated factorization  $R \ni qh = \tilde{h}_1 \cdot \dots \cdot \tilde{h}_m$ , where  $\tilde{h}_i \in R \setminus \{0\}$  and  $\deg_{\partial} h_i = \deg_{\partial} \tilde{h}_i$  holds.
3. Let  $r \in R$  and  $1^{-1}r$  be an irreducible element in  $S^{-1}R$ . Then in any factorization  $r = pq$ , where  $p, q \in R \setminus U(R)$  one has  $p \in U$  or  $q \in U$ , i. e.  $r$  is not necessarily irreducible in  $R$ .
4. If  $r \in R$  is irreducible in  $R$ , then  $1^{-1}r$  is not necessarily irreducible in  $S^{-1}R$ .

Surprisingly, irreducible  $[-1, 1]$ -homogeneous polynomials remain irreducible in the rational ( $q$ -)Weyl algebra, as the following Theorem shows.

**Theorem 2.14** *Let  $p$  be an irreducible  $[-1, 1]$ -homogeneous polynomial in  $A_1$ . Then, in the first rational Weyl algebra  $B_1$ ,  $1^{-1}p$  is irreducible up to an invertible multiple.*

*Proof* The following monic homogeneous polynomials are irreducible in  $A_1$ :

1.  $\partial$ , which is also irreducible over  $B_1$ ,
2.  $x$ , which is a unit in  $B_1$ ,
3. a monic irreducible  $p$  over  $\mathbb{K}[\theta], p \notin \{\theta, \theta + 1\}$ .

Therefore, the only interesting case is the third one. Now let  $p$  be a monic irreducible element in  $A_1^{(0)} \setminus \{\theta, \theta + 1\}$ . From now on we identify  $p$  with  $1^{-1}p \in B_1$ . Suppose, that  $p$  is nontrivially reducible over  $B_1$ , say  $p = p_1 \cdot p_2$  for  $p_1, p_2 \in B_1 \setminus A_1$ , both non-invertible, thus  $\deg_{\partial}(p_1), \deg_{\partial}(p_2) \geq 1$  and therefore  $\deg_{\partial}(p) \geq 2$ . By Lemma 2.11, there exist  $q \in \mathbb{K}[x], \tilde{p}_1, \tilde{p}_2 \in A_1 \setminus \mathbb{K}[x]$ , such that  $qp = \tilde{p}_1\tilde{p}_2$ .

**Case 1**  $q = x^k, k \in \mathbb{N}; q$  is homogeneous.

Then all possible factorizations of  $x^k \cdot p$  in  $A_1$  are due to Lemma 1.5 of the form

$$x^{k-\ell}p(\theta - \ell)x^{\ell}, \ell \in \mathbb{N}_0, \ell \leq k.$$

As shifts of irreducible elements in a univariate commutative polynomial ring  $\mathbb{K}[\theta]$  are irreducible (see e.g. [1], Section 4.2) and  $\deg_\partial(p) \geq 2$ , we see that  $\tilde{p}_1$  and  $\tilde{p}_2$  as supposed above do not exist.

**Case 2**  $q = \sum_{i=0}^n q_i x^i$ ,  $n \geq 1$ ,  $q_i \in \mathbb{K}$ ,  $q_n \neq 0$ ;  **$q$  is not homogeneous.**

Note, that the product  $qp$  in this case is not homogeneous with respect to the  $[-1, 1]$ -grading. Let  $m \in \mathbb{N}$ ,  $m < n$  be minimal, satisfying  $q_m \neq 0$ , then the sum in  $qp = \sum_{i=m}^n q_i x^i p$  coincides with the graded decomposition of  $qp$ .

With notations from the proof of Lemma 1.12, suppose that  $\alpha(\tilde{p}_1) = \eta \in \mathbb{Z}$  and  $\alpha(\tilde{p}_2) = \mu \in \mathbb{Z}$ . Then

$$q_m x^m p = (qp)_{\alpha(qp)} = (\tilde{p}_1 \tilde{p}_2)_{\alpha(\tilde{p}_1 \tilde{p}_2)} = (\tilde{p}_1)_\eta (\tilde{p}_2)_\mu.$$

Since  $q_m \neq 0$ , we can proceed like in Case 1, where two kinds of factorization are possible. Let us first write  $(\tilde{p}_1)_\eta = x^{m-\ell} p(\theta - \ell)$  for some  $0 \leq \ell \leq m$  and  $(\tilde{p}_2)_\mu = q_m x^\ell$ , then  $\deg_\partial(\tilde{p}_1) \geq \deg_\partial(\tilde{p}_1)_\eta = \deg_\partial(p) = \deg_\partial(qp) = \deg_\partial(\tilde{p}_1 \tilde{p}_2) = \deg_\partial(\tilde{p}_1) + \deg_\partial(\tilde{p}_2)$ , indicating that  $\deg_\partial(\tilde{p}_2) = 0$  and  $\deg_\partial(\tilde{p}_1) = \deg_\partial(p)$ . That is,  $\tilde{p}_2$  must be in  $\mathbb{K}[x]$  and therefore cannot be as supposed above. The second case, where  $\deg_\partial(\tilde{p}_2)_\mu = \deg_\partial(p)$  is analogous and thus the proof is completed.

Analogously to the case of Wel algebra, it turns out that the set  $S = \mathbb{K}[x] \setminus \{0\}$  is an Ore set in the first polynomial  $q$ -Weyl algebra  $Q_1$ . Therefore  $W_1 := S^{-1}Q_1$  is called the **first rational  $q$ -Weyl algebra**.

**Corollary 2.15** *Let  $p$  be an irreducible  $[-1, 1]$ -homogeneous polynomial in  $Q_1$ . Then, in the first rational  $q$ -Weyl algebra  $W_1$ ,  $1^{-1}p$  is irreducible up to an invertible multiple.*

*Proof* By inspecting the proof of Theorem 2.14 above, we see that we only have to replace  $\theta + 1$  with  $\theta + q^{-1}$  in the list of monic irreducible elements; the rest of arguments hold verbatim.

### 3 Implementation and Benchmarking

We implemented the presented algorithms in SINGULAR:PLURAL, and since version 3-1-3 they are part of the distribution of SINGULAR. The following example shows how to use the library containing them.

*Example 3.1* Let  $\mathbb{K} = \mathbb{Q}$  and  $h \in Q_1$  be the polynomial

$$\begin{aligned} h := & q^{25} x^{10} \partial^{10} + q^{16} (q^4 + q^3 + q^2 + q + 1)^2 x^9 \partial^9 \\ & + q^9 (q^{13} + 3q^{12} + 7q^{11} + 13q^{10} + 20q^9 + 26q^8 \\ & + 30q^7 + 31q^6 + 26q^5 + 20q^4 + 13q^3 + 7q^2 + 3q + 1) x^8 \partial^8 \end{aligned}$$

$$\begin{aligned}
& +q^4(q^9 + 2q^8 + 4q^7 + 6q^6 + 7q^5 + 8q^4 + 6q^3 + 4q^2 + 2q + 1) \\
& (q^4 + q^3 + q^2 + q + 1)(q^2 + q + 1)x^7\partial^7 \\
& +q(q^2 + q + 1)(q^5 + 2q^4 + 2q^3 + 3q^2 + 2q + 1) \\
& (q^4 + q^3 + q^2 + q + 1)(q^2 + 1)(q + 1)x^6\partial^6 \\
& +(q^{10} + 5q^9 + 12q^8 + 21q^7 + 29q^6 + 33q^5 \\
& +31q^4 + 24q^3 + 15q^2 + 7q + 12)x^5\partial^5 + 6x^3\partial^3 + 24.
\end{aligned}$$

We can use SINGULAR to obtain all of its factorizations in the following way.

```

LIB "ncfactor.lib";
ring R = (0,q), (x,d), dp;
def r = nc_algebra (q,1);
setring(r);
poly h = ... //See the polynomial defined above.
homogfacFirstQWeyl_all(h);
[1]:
  [1]:
    1
  [2]:
    x5d5+x3d3+4
  [3]:
    x5d5+6

[2]:
  [1]:
    1
  [2]:
    x5d5+6
  [3]:
    x5d5+x3d3+4

```

As one can see here, the output is a list containing lists containing elements in  $Q_1$ . Those elements in  $Q_1$  are factors of  $h$ , and each list represents one possible factorization of  $h$ .

The command `homogfacFirstQWeyl` can be used if the user is interested in just one factorization. The output is just one list containing elements in  $Q_1$ .

The calculation was run on a computer with a 4-core Intel CPU (Intel® Core™i7-3520M CPU with 2.90 GHz, two physical cores, two hardware threads, 32 K L1[i,d], 256 K L2, 4 MB L3 cache), 16 GB RAM and Ubuntu 12.04LTS as operating system. The computation time was 0.62 s.

*Remark 3.2* The factorization of products of homogeneous elements in  $A_1$  can be observed to be faster than the factorization of the same products in  $Q_1$ . The product of factors in the example above, i.e.  $(x^5\partial^5 + 6)(x^5\partial^5 + x^3\partial^3 + 4)$ , viewed as an element in  $A_1$ , takes 0.08 s to factorize compared to 0.62 s in the  $q$ -Weyl case. This

seems to be way slower considering that both algorithms have the same complexity. But this slowdown is not due to more steps that need to be done in the algorithm for the  $q$ -Weyl algebra, but due to the parameter  $q$  and the speed of calculating in  $\mathbb{Q}(q)$  as the basefield instead of just in  $\mathbb{Q}$ .

In fact, there is no computer algebra system known to the authors that can factor polynomials in the first  $q$ -Weyl algebra  $Q_1$ . Therefore, we cannot compare our algorithms in this case to other implementations.

For the first Weyl algebra  $A_1$ , there exist other implementations. We can draw a comparison to the `DFactor` method in the `DETools` package of `MAPLE` and the `nc_factorize_all` method in the `NCPoly` library of `REDUCE`. Furthermore, we were provided with a wrapper for the algorithm “Coprime Index 1 Factorizations” (`CP1F`) mentioned in [36] dealing with polynomials of the form  $\mathbb{K}[x][\theta]$  in order to be able to compare it to the algorithm for this special case explicitly. This guarantees a fair evaluation on a core level for an intersection with homogeneous polynomials that does not invoke the complete factorization machinery implemented in `DFactor`.

In the next subsection, we will only compare `DFactor` and `nc_factorize_all` to our implementation. Later on, we will compare the wrapper of `CP1F` implemented in `MAPLE` to our implementation, as we have to choose for the comparison a special set of polynomials, namely the homogeneous ones supported by `CP1F`.

### 3.1 Comparison to `DFactor` and `nc_factorize_all`

We used version 17 of `MAPLE` and version 3.8 of `REDUCE`. In order to make our benchmarks reproducible, we utilized the `SDEVAL` framework presented in [15]. The sources and the results of the computations can be downloaded from [https://cs.uwaterloo.ca/~aheinle/software\\_projects.html](https://cs.uwaterloo.ca/~aheinle/software_projects.html).

*Remark 3.3* As mentioned before, the algorithm `DFactor` implemented in `MAPLE` factorizes over the rational Weyl algebra, i.e. the variable  $x$  is a rational argument having adjusted commutation rules with  $\partial$ . This is a weaker assumption on the input since the ring that is dealt with there is larger. The comparison is still valid, since we have shown in Theorem 2.14 that a factorization of a homogeneous polynomial into irreducible elements over  $A_1$  cannot be further refined in the first Weyl algebra with rational coefficients.

We will not go into detail about how the algorithm in `MAPLE` works. The interested reader can find details in [35]. It works with collections of exponential parts and their multiplicities at all singularities of a given differential operator  $f$  and subsequent calculation of left and right hand factors.

The algorithm implemented in `REDUCE` is also working with the polynomial Weyl algebra. In fact, the algorithm written there can be applied to a broad class of polynomial noncommutative rings.



Details about the functionality of the algorithm in REDUCE are unfortunately not available. In order to understand these, we have analyzed the code that is given open source. It makes an Ansatz for coefficients of elements of smaller total degree, obtains a system of commutative polynomial equations and, at the end, uses several Gröbner basis computations in order to find its solutions.

*Example 3.4* Consider again (cf. Remark 3.2) the element

$$h := (x^5\partial^5 + 6) \cdot (x^5\partial^5 + x^3\partial^3 + 4) \in A_1$$

in the expanded form.

- SINGULAR: Found two factorizations in less than a second.
- MAPLE: Found one factorization after 29 s; The factors are huge (size of the output file is around 100 KB).
- REDUCE: Did not terminate after 9 h of calculation.

*Example 3.5* We experimented with other randomly generated products of two homogeneous polynomials in the first Weyl algebra. The following collection is representative.

$$\begin{aligned} f_0 &= (x^{10}\partial^{10} + 5x\partial + 7) \cdot x^2 \cdot (x^{11}\partial^{11} + 3x^7\partial^7 + x\partial + 4), \\ f_1 &= (x^5\partial^5 + 6) \cdot (x^5\partial^5 + x^3\partial^3 + 4) \cdot \partial^{10}, \\ f_2 &= (5x^{10}\partial^{10} + 7x^9\partial^9 + 8x^8\partial^8 + 9x^7\partial^7 + 6x^6\partial^6 + 5x^5\partial^5 + 8x^4\partial^4 + 5x^3\partial^3 + 9x^2\partial^2 + 9x\partial + 6) \cdot \partial^{20}, \\ f_3 &= (7x^{15}\partial^{15} + x^{13}\partial^{13} - x^{12}\partial^{12} - 3x^{10}\partial^{10} + 2x^9\partial^9 + x^8\partial^8 + x^7\partial^7 - x^5\partial^5 - 9x^4\partial^4 + x\partial - 1) \cdot (8x^{13}\partial^{13} + 3x^{12}\partial^{12} + x^{11}\partial^{11} - 2x^{10}\partial^{10} + 10x^8\partial^8 - 3x^7\partial^7 + 2x^5\partial^5 + x^4\partial^4 + 38x\partial + 1) \cdot \partial^6, \\ f_4 &= (x^{10}\partial^{10} + 23x^9\partial^9 + 3x^8\partial^8 - 9x^7\partial^7 - x^5\partial^5 + 3x^4\partial^4 + 6x^3\partial^3 + 4x\partial + 1) \cdot (-x^8\partial^8 + 4x^7\partial^7 - x^6\partial^6 + 4x^5\partial^5 - 5x^4\partial^4 + x^2\partial^2 - 7x\partial - 10) \cdot x^{10}, \\ f_5 &= (-2x^{24}\partial^{24} + x^{23}\partial^{23} + 4x^{22}\partial^{22} - 110x^{21}\partial^{21} + x^{20}\partial^{20} + x^{19}\partial^{19} + x^{18}\partial^{18} + x^{17}\partial^{17} + 5x^{16}\partial^{16} - 7x^{15}\partial^{15} + 4x^{14}\partial^{14} - x^{13}\partial^{13} + x^{12}\partial^{12} - 2x^{11}\partial^{11} + x^9\partial^9 + 5x^8\partial^8 + x^7\partial^7 + 6x^5\partial^5 + x^4\partial^4 + 2x^3\partial^3 + 219x^2\partial^2 + x\partial - 1) \cdot (-x^{25}\partial^{25} + x^{24}\partial^{24} - 32x^{23}\partial^{23} + x^{22}\partial^{22} + 7x^{21}\partial^{21} + 61x^{20}\partial^{20} - 2x^{18}\partial^{18} + x^{16}\partial^{16} + 2x^{15}\partial^{15} - 2x^{14}\partial^{14} - x^{12}\partial^{12} - 3x^{11}\partial^{11} + 2x^{10}\partial^{10} + 2x^8\partial^8 - 9x^7\partial^7 - x^6\partial^6 + x^5\partial^5 + 4x^3\partial^3 + x^2\partial^2), \\ f_6 &= (x^{10}\partial^{10} + 13x^9\partial^9 - x^8\partial^8 + 4x^7\partial^7 + 13x^6\partial^6 - 3x^5\partial^5 - 37x^4\partial^4 - x^3\partial^3 + x^2\partial^2 + x\partial - 1) \cdot (-x^{10}\partial^{10} - 23x^9\partial^9 + 3x^8\partial^8 + x^7\partial^7 - x^6\partial^6 - 2x^5\partial^5 - 2x^4\partial^4 + 2x^3\partial^3 - x^2\partial^2 - 2x\partial - 2), \\ f_7 &= (98x^{15}\partial^{15} + 40x^{14}\partial^{14} + 98x^{13}\partial^{13} + 44x^{12}\partial^{12} + 55x^{11}\partial^{11} + 96x^{10}\partial^{10} + 95x^9\partial^9 + 7x^8\partial^8 + 56x^7\partial^7 + 56x^6\partial^6 + 40x^5\partial^5 + 11x^4\partial^4 + 40x^3\partial^3 + 78x^2\partial^2 + 13x\partial + 19) \cdot (61x^{15}\partial^{15} + 50x^{14}\partial^{14} + 83x^{13}\partial^{13} + 11x^{12}\partial^{12} + 89x^{11}\partial^{11} + 55x^{10}\partial^{10} + 81x^9\partial^9 + 63x^8\partial^8 + 22x^7\partial^7 + 10x^6\partial^6 + 35x^5\partial^5 + 90x^4\partial^4 + 60x^3\partial^3 + 20x^2\partial^2 + 30x\partial + 43), \\ f_8 &= (85x^{20}\partial^{20} + 80x^{19}\partial^{19} + 27x^{18}\partial^{18} + 74x^{17}\partial^{17} + 49x^{16}\partial^{16} + 95x^{15}\partial^{15} + 96x^{14}\partial^{14} + 37x^{13}\partial^{13} + 26x^{12}\partial^{12} + 93x^{11}\partial^{11} + 39x^{10}\partial^{10} + 19x^9\partial^9 + 48x^8\partial^8 + 82x^7\partial^7 + 26x^6\partial^6 + 26x^5\partial^5 + 7x^4\partial^4 + 61x^3\partial^3 + 8x^2\partial^2 + 81x\partial + 88)^2. \end{aligned}$$

The results concerning the factorization of  $f_i$  are listed in the next table. An entry labeled with “– NT –” stands for “no termination after 2 h”.

Polynomial	SINGULAR	MAPLE	REDUCE
$f_0$	0.08 s; 12 factorizations	– NT –	SEGFAULT
$f_1$	0.77 s; 132 factorizations	11.18 s; 1 factorization	– NT –
$f_2$	0.18 s; 21 factorizations	– NT –	– NT –
$f_3$	5.88 s; 504 factorizations	– NT –	– NT –
$f_4$	0.76 s; 132 factorizations	– NT –	– NT –
$f_5$	28.23 s; 230 factorizations	– NT –	– NT –
$f_6$	0.06 s; 6 factorizations	– NT –	– NT –
$f_7$	0.08 s; 2 factorizations	– NT –	– NT –
$f_8$	0.08 s; 1 factorization	– NT –	– NT –

The conclusion we can draw at this point is: Even if homogeneous polynomials seem to be easy objects to factorize according to the algorithm we propose, they seem to form a worst case class for the implementations in REDUCE and MAPLE.

Therefore, with our algorithm we are now able to factorize more polynomials using computer algebra systems: homogeneous polynomials in  $Q_1$  in general, and for  $A_1$  we have broadened the range of polynomials that can be factorized in a feasible amount of time or even sometimes at all.

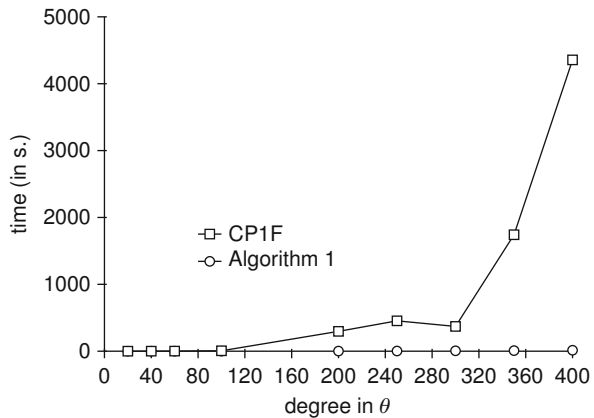
Moreover, our approach can be used to enhance existing algorithms and their implementations as follows. Namely, the check of a given polynomial for the homogeneity is a very cheap procedure as we have seen in Lemma 1.11. Moreover, for the case of a homogeneous polynomial our algorithm can be applied, hence factoring homogeneous polynomials—appearing, for instance, as factors of a bigger polynomial—can be eliminated from further computations.

### 3.2 Comparison to CP1F

As indicated before, we were provided a wrapper to the function implemented in MAPLE that represents CP1F, whose supported input polynomials are of the form  $\mathbb{K}[x][\theta]$ . Hence, there is a nontrivial intersection with homogeneous polynomials in  $A_1$ . Comparing it to the implementation of our Algorithm 1 on homogeneous polynomials of  $\theta$ -degree between 20 and 400, we obtain the following timings.

Example	Algorithm 1	CP1F
Degree 20	0.04 s	0.17 s
Degree 40	0.07 s	0.61 s
Degree 60	0.11 s	1.66 s
Degree 100	0.26 s	6.39 s
Degree 200	2.03 s	296.78 s
Degree 250	2.86 s	454.17 s
Degree 300	5.9 s	370.49 s
Degree 350	8.78 s	1741.53 s
Degree 400	14.62 s	4355.32 s

**Fig. 1** Visualization of asymptotic behaviour of CP1F and Algorithm 1



We can derive from this table that for small degrees, the timings are close to each other. With increasing degree though, the difference in performance becomes more visible, and one observes also different asymptotic behaviours, as Fig. 1 visualizes.

## 4 Conclusion

With this paper, we contributed an algorithm for the factorization problem considering  $[-1, 1]$ -homogeneous polynomials in the first  $q$ -Weyl algebra over an arbitrary field  $\mathbb{K}$ . For computable fields, we discussed a complexity estimate for our approach. Our approach is implemented as the library `ncfactor.lib`, which is distributed with the computer algebra system SINGULAR:PLURAL since version 3-1-3, and several improved and extended versions have been released since then.

Furthermore, we also considered the special case of the first Weyl algebra and showed that our algorithm beats for the large class of  $[-1, 1]$ -homogeneous polynomials current implementations in terms of speed and elegance of the solutions. Due to Theorem 2.14, we can even state that the factorizations that our algorithm finds cannot be further refined when factoring over the rational Weyl algebra. This result

is interesting by itself and could play a role for future research on the question how to characterize arbitrary irreducible elements in the polynomial first Weyl algebra, that become reducible after localization.

We can construct a family of polynomials where the implementation in SINGULAR:PLURAL is the only one that is able to factorize those elements in a feasible amount of time and memory consumption. Since our techniques are easy to implement, they can be used to extend existing implementations in order to broaden the range of polynomials in the first Weyl algebra that we are nowadays able to factorize using a computer algebra system.

**Acknowledgements** We would like to thank to Dima Grigoriev for discussions on the subject, and Mark van Hoeij for his expert opinion. Many thanks to Mark Giesbrecht for his helpful suggestions and comments. We are grateful to Wolfram Koepf and Martin Lee for providing us with interesting examples, and to Daniel Rettstadt and Johannes Hoffmann for stimulating exchange of opinions.

We acknowledge the helpful suggestions and comments of the anonymous referees.

## References

1. J.A. Beachy, W.D. Blair, *Abstract Algebra* (Waveland Press, Long Grove, IL, 2006)
2. R. Beals, E. Kartashova, Constructively factoring linear partial differential operators in two variables. *Theor. Math. Phys.* **145**(2), 1511–1524 (2005). <http://link.springer.com/article/10.1007/s11232-005-0178-7>
3. J.P. Bell, A. Heinle, V. Levandovskyy, On noncommutative finite factorization domains. *Trans. Am. Math. Soc.* **369**, 2675–2695 (2017). <http://doi.org/10.1090/tran/6727>
4. J. Bueso, J. Gómez-Torrecillas, A. Verschoren, *Algorithmic Methods in Non-Commutative Algebra. Applications to Quantum Groups* (Kluwer Academic Publishers, Dordrecht, 2003)
5. A.L. Chistov, Algorithm of polynomial complexity for factoring polynomials and finding the components of varieties in subexponential time. *J. Sov. Math.* **34**(4), 1838–1882 (1986)
6. W. Decker, G.M. Greuel, G. Pfister, H. Schönemann, SINGULAR 4-1-0 – a computer algebra system for polynomial computations (2016). <http://www.singular.uni-kl.de>
7. M. Foupouagnigni, W. Koepf, A. Ronveaux, Factorization of fourth-order differential equations for perturbed classical orthogonal polynomials. *J. Comput. Appl. Math.* **162**(2), 299–326 (2004)
8. M. Giesbrecht, Factoring in skew-polynomial rings over finite fields. *J. Symb. Comput.* **26**(4), 463–486 (1998)
9. M. Giesbrecht, Y. Zhang, Factoring and decomposing ore polynomials over  $\text{fq}(t)$ , in *Proceedings of the 2003 International Symposium on Symbolic and Algebraic Computation, ISSAC '03* (ACM, New York, NY, 2003), pp. 127–134. <http://doi.acm.org/10.1145/860854.860888>
10. M. Giesbrecht, A. Heinle, V. Levandovskyy, Factoring linear differential operators in  $n$  variables. *J. Symb. Comput.* **75**, 127–148 (2016). <http://dx.doi.org/10.1016/j.jsc.2015.11.011>
11. G.M. Greuel, G. Pfister, *A Singular Introduction to Commutative Algebra*. With contributions by O. Bachmann, C. Lossen, H. Schönemann, 2nd extended ed. (Springer, Berlin, 2007)
12. G.M. Greuel, V. Levandovskyy, A. Motsak, H. Schönemann, PLURAL. A SINGULAR 4-1-0 Subsystem for Computations with Non-commutative Polynomial Algebras (Centre for Computer Algebra, TU Kaiserslautern, 2016). <http://www.singular.uni-kl.de>
13. D. Grigoriev, Factoring polynomials over a finite field and solving systems of algebraic equations. *Zapiski Nauchnykh Seminarov POMI* **137**, 20–79 (1984)

14. D. Grigoriev, F. Schwarz, Factoring and solving linear partial differential equations. *Computing* **73**(2), 179–197 (2004). <https://doi.org/10.1007/s00607-004-0073-3>
15. A. Heinle, V. Levandovskyy, The SDEval benchmarking toolkit. *ACM Commun. Comput. Algebra* **49**(1), 1–9 (2015). <http://doi.org/10.1145/2768577.2768578>
16. A. Heinle, V. Levandovskyy, A factorization algorithm for  $G$ -algebras and applications, in *Proceedings of the International Symposium on Symbolic and Algebraic Computation (ISSAC'16)* (ACM, New York, NY, 2016). <http://doi.org/10.1145/2930889.2930906>
17. V. Kac, P. Cheung, *Quantum Calculus* (Springer, New York, NY, 2002)
18. E. Kaltofen, On the complexity of factoring polynomials with integer coefficients. Ph.D. thesis, Rensselaer Polytechnic Institute (1982)
19. M. Kashiwara, Vanishing cycle sheaves and holonomic systems of differential equations, in *Algebraic Geometry*. Springer Lecture Notes in Mathematics, vol. 1016 (Springer, Berlin, 1983), pp. 134–142
20. A.K. Lenstra, H.W. Lenstra, L. Lovász, Factoring polynomials with rational coefficients. *Math. Ann.* **261**(4), 515–534 (1982)
21. A. Loewy, Über reduzible lineare homogene Differentialgleichungen. *Math. Ann.* **56**, 549–584 (1903). <http://dx.doi.org/10.1007/BF01444307>
22. A. Loewy, Über vollständig reduzible lineare homogene Differentialgleichungen. *Math. Ann.* **62**, 89–117 (1906). <http://dx.doi.org/10.1007/BF01448417>
23. B. Malgrange, Polynômes de Bernstein-Sato et cohomologie evanescente. *Astérisque* **101–102**, 243–267 (1983)
24. H. Melenk, J. Apel, *REDUCE Package NCPOLY: Computation in Non-Commutative Polynomial Ideals* (Konrad-Zuse-Zentrum (ZIB), Berlin, 1994)
25. M.B. Monagan, K.O. Geddes, K.M. Heal, G. Labahn, S.M. Vorkoetter, J. McCarron, P. DeMarco, *Maple Introductory Programming Guide* (Maplesoft, Waterloo, ON, 2008)
26. A.F. Nikiforov, V.B. Uvarov, *Special Functions of Mathematical Physics: A Unified Introduction with Applications* (Birkhäuser, Boston, MA, 1988)
27. M. Saito, B. Sturmfels, N. Takayama, *Gröbner Deformations of Hypergeometric Differential Equations* (Springer, Berlin, 2000)
28. F. Schwarz, Alltypes in the web. *ACM Commun. Comput. Algebra* **42**(3), 185–187 (2009). <http://doi.acm.org/10.1145/1504347.1504379>
29. E. Shemyakova, Parametric factorizations of second-, third- and fourth-order linear partial differential operators with a completely factorable symbol on the plane. *Math. Comput. Sci.* **1**(2), 225–237 (2007). <http://dx.doi.org/10.1007/s11786-007-0019-1>
30. E. Shemyakova, Multiple factorizations of bivariate linear partial differential operators, in *Proceedings of the CASC 2009* (Springer, Berlin, 2009), pp. 299–309
31. E. Shemyakova, Refinement of two-factor factorizations of a linear partial differential operator of arbitrary order and dimension. *Math. Comput. Sci.* **4**, 223–230 (2010). <http://dx.doi.org/10.1007/s11786-010-0052-3>
32. S.P. Tsarev, Problems that appear during factorization of ordinary linear differential operators. *Program. Comput. Softw.* **20**(1), 27–29 (1994)
33. S.P. Tsarev, An algorithm for complete enumeration of all factorizations of a linear ordinary differential operator, in *Proceedings of the 1996 International Symposium on Symbolic and Algebraic Computation* (ACM, New York, 1996), pp. 226–231
34. M. van Hoeij, Factorization of linear differential operators. Nijmegen (1996). <http://books.google.de/books?id=rEmjPgAACAAJ>
35. M. van Hoeij, Factorization of differential operators with rational functions coefficients. *J. Symb. Comput.* **24**(5), 537–561 (1997). <http://dx.doi.org/10.1006/jsc.1997.0151>
36. M. van Hoeij, Formal solutions and factorization of differential operators with power series coefficients. *J. Symb. Comput.* **24**(1), 1–30 (1997). <http://dx.doi.org/10.1006/jsc.1997.0110>

37. M. van Hoeij, Q. Yuan, Finding all Bessel type solutions for linear differential equations with rational function coefficients, in *Proceedings of the 2010 International Symposium on Symbolic and Algebraic Computation, ISSAC '10* (ACM, New York, NY, 2010), pp. 37–44. <http://doi.acm.org/10.1145/1837934.1837948>
38. J. von zur Gathen, J. Gerhard, *Modern Computer Algebra* (Cambridge University Press, Cambridge, 2013)

# Complexity of Membership Problems of Different Types of Polynomial Ideals



Ernst W. Mayr and Stefan Toman

**Abstract** We survey degree bounds and complexity classes of the word problem for polynomial ideals and related problems. The word problem for general polynomial ideals is known to be exponential space-complete, but there are several interesting subclasses of polynomial ideals that allow for better bounds. We review complexity results for polynomial ideals with low degree, toric ideals, binomial ideals, and radical ideals. Previously known results as well as recent findings in our project “Degree Bounds for Gröbner Bases of Important Classes of Polynomial Ideals and Efficient Algorithms” are presented.

**Keywords** Polynomial ideal • Binomial ideal • Gröbner basis • Degree bound • Radical • Thue system • Cellular decomposition • Computational complexity

**Subject Classifications** 13P10, 14Q20, 03D40, 08A50, 03D03, 06B10

## 1 Introduction

Solving systems of polynomial equations is one of the most common problems in mathematics. Objects are modelled as polynomial equations and the solutions of these equations themselves or properties of them need to be found. These problems turn out to be inherently hard and it is a common technique to first deduce bases with certain properties that make solving the problems easier. The goal of our project “Degree Bounds for Gröbner Bases of Important Classes of Polynomial Ideals and Efficient Algorithms” in the priority program SPP 1489 “Algorithmic

---

E.W. Mayr • S. Toman (✉)  
Institut für Informatik, TU München, München, Germany  
e-mail: [mayr@in.tum.de](mailto:mayr@in.tum.de); [toman@tum.de](mailto:toman@tum.de)

© Springer International Publishing AG, part of Springer Nature 2017  
G. Böckle et al. (eds.), *Algorithmic and Experimental Methods  
in Algebra, Geometry, and Number Theory*,  
[https://doi.org/10.1007/978-3-319-70566-8\\_20](https://doi.org/10.1007/978-3-319-70566-8_20)

481

and Experimental Methods in Algebra, Geometry, and Number Theory” of Deutsche Forschungsgemeinschaft (DFG) was to find upper and lower complexity bounds for solving different subclasses of polynomial equations. In this paper we report on results that were previously known and the ones found during this project.

The special case of linear equations is well-understood. These systems can be transformed to row-echelon form in polynomial time using the Gaussian algorithm. Several problems like finding solutions of the system or the word problem can be solved easily once the system is in row-echelon form. The corresponding problems for non-linear systems of equations are inherently much harder. According to the Abel-Ruffini theorem solutions of these systems cannot even be expressed in general using simple formulas involving roots only [1]. Nevertheless, there is also a normal form of these systems called Gröbner bases that allows for easier computations of many problems.

## 2 General Polynomial Ideals

Gröbner bases were introduced in 1965 by Buchberger in his PhD thesis [4]. They can be employed to solve many problems in computer algebra, the most immediate being the word problem for polynomial ideals. This problem is given a list of multivariate polynomial equations to decide whether another polynomial equation is contained in the ideal they span, i.e. whether the latter is already implied by these equations. The word problem can be solved using Gröbner bases and Buchberger also presented an algorithm for this problem using the so-called Buchberger criterion.

At first, it was only known that Buchberger’s algorithm runs in finite time without having better space or runtime bounds. In 1982 Mayr and Meyer proved a lower bound on the worst-case space usage of each algorithm solving the word problem for polynomial ideals that is exponential in the number of variables appearing in the equations [18].

**Theorem 2.1 ([18])** *There is a constant  $\epsilon \in \mathbb{Q}$  with  $\epsilon > 0$  such that any algorithm which is able to decide the word problem for polynomial ideals contained in  $\mathbb{Q}[x_1, \dots, x_n]$  for some  $n \in \mathbb{N}_{>0}$  requires space exceeding  $2^{\epsilon n}$  on infinitely many instances of this problem with different sizes.*

Their result was slightly improved in 1991 by Yap who changed the constant in the exponent [27].

The time and space requirements of algorithms computing Gröbner bases heavily depend on the number of indeterminates of the polynomial ring. For this reason it is important to have degree bounds, for instance the ones by Hermann [10] and Dubé [6]. They found degree bounds double-exponential in the number of indeterminates.



**Theorem 2.2 ([6])** *Let  $f_1, \dots, f_s \in R[x_1, \dots, x_n]$  be polynomials with  $\deg(f_i) \leq d$  for all  $i \in \{1, \dots, s\}$  over a ring  $R$  for some  $d, s \in \mathbb{N}_{>0}$ . Every reduced Gröbner basis of  $\langle f_1, \dots, f_s \rangle$  consists of polynomials  $g_1, \dots, g_r \in R[x_1, \dots, x_n]$  for some  $r \in \mathbb{N}_{>0}$  with*

$$\deg(g_i) \leq 2 \left( \frac{d^2}{2} + d \right)^{2^{n-1}}$$

for all  $i \in \{1, \dots, r\}$ .

Using those bounds, Kühnle and Mayr showed in 1996 that Gröbner bases can indeed be computed using exponential space in the number of indeterminates [15]. Thus, the lower and upper bounds for the word problem for polynomial ideals coincided and the problem was proven to be **EXPSpace**-complete in the number of indeterminates. There are several surveys on further complexity results for the computation of Gröbner bases, for instance the one presented by Mayr [17].

Since the computation of Gröbner bases is that important and hard it is a natural question to ask whether there are special subclasses of polynomial ideals that allow for faster computations of them.

### 3 Polynomial Ideals with Low Dimension

One class of polynomial ideals that allows easier computations of their Gröbner bases is the set of zero-dimensional polynomial ideals. The dimension of a polynomial ideal is the maximum size of a set of indeterminates such that no leading monomial of a polynomial contained in this ideal consists of these indeterminates only. Equivalently, the dimension of a polynomial ideal is the size of the biggest set of indeterminates that is unrelated modulo the ideal. This means that zero-dimensional ideals have many relations between their indeterminates and there is even a relation for each indeterminate alone. This additional structure may be used to find improved degree bounds.

In 1983 Faugère et al. presented an algorithm that is much faster in practice for zero-dimensional polynomial ideals than Buchberger's algorithm [8]. It was also proven that there is a single-exponential bound on the degree of Gröbner basis elements for zero-dimensional polynomial ideals by Dickenstein et al. which enables better algorithms [5].

**Theorem 3.1 ([5])** *Let  $f_1, \dots, f_s \in k[x_1, \dots, x_n]$  be polynomials with  $\deg(f_i) \leq d$  for all  $i \in \{1, \dots, s\}$  over a field  $k$  for  $d, n, s \in \mathbb{N}_{>0}$  such that  $\langle f_1, \dots, f_s \rangle$  has dimension 0 and let  $g \in \langle f_1, \dots, f_s \rangle$  be a polynomial. There are polynomials*

$$g_1, \dots, g_s \in k[x_1, \dots, x_n]$$

such that  $g = \sum_{i=1}^s f_i g_i$  and

$$\deg(f_i g_i) \leq nd^{2n} + d^n + d + \deg(f)$$

for all  $i \in \{1, \dots, s\}$ .

Since the special case of zero-dimensional polynomial ideals is much easier than the general problem one could expect polynomial ideals with low dimension to have better algorithms, too.

All degree bounds given above are dependent on the number of indeterminates as this turned out to be a very significant parameter of polynomial ideals to describe their inherent complexity. The following bound does not use the number of indeterminates as a parameter but the degree of the polynomial ideal which may result in better bounds for special subsets of polynomial ideals.

Using a new degree bound by Kratzer [14], Mayr and Ritscher were able to find an algorithm to compute Gröbner bases whose space is bounded exponentially in the dimension of the polynomial ideal [19].

**Theorem 3.2 ([19])** *Let  $f_1, \dots, f_s \in k[x_1, \dots, x_n]$  be polynomials with  $\deg(f_i) \leq d$  for all  $i \in \{1, \dots, s\}$  over an infinite field  $k$  for  $d, n, s \in \mathbb{N}_{>0}$ . Let  $m \in \mathbb{N}_0$  be the dimension of  $\langle f_1, \dots, f_s \rangle$ . Every reduced Gröbner basis of  $\langle f_1, \dots, f_s \rangle$  with respect to an admissible monomial ordering consists of polynomials  $g_1, \dots, g_r \in R[x_1, \dots, x_n]$  for some  $r \in \mathbb{N}_{>0}$  with*

$$\deg(g_i) \leq 2 \left( \frac{1}{2} \left( d^{2(n-m)^2} + d \right) \right)^{2^m}$$

for all  $i \in \{1, \dots, r\}$ .

This theorem is proven using an algorithm based on a cone decomposition of the space of polynomials. The construction of this decomposition is based on a similar decomposition presented by Dubé [6].

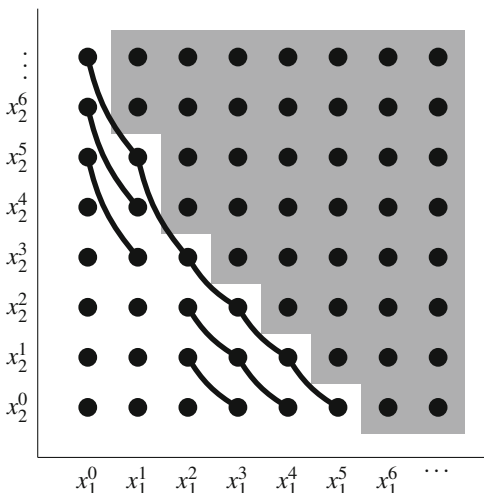
They also presented an incremental version of their algorithm that does not use degree bounds. The space bound of this algorithm uses the degree of the actual problem instance instead of a worst-case instance. Therefore, the algorithm does not require the knowledge of any a-priori degree bounds.

Both findings improved the known space bounds for polynomial ideals with low degree in comparison to the general bounds. Later, they also proved a matching lower bound [20] which finished the complexity analysis of computing Gröbner bases depending on the degree of the polynomial ideal.

## 4 Binomial Ideals

Another interesting subclass are binomial ideals and pure binomial ideals which are polynomial ideals that can be generated by binomials respectively pure binomials only. Binomials are polynomials with at most two terms while pure binomials

**Fig. 1** The equivalence classes of monomials modulo  $I = \langle x_1^3 - x_1^2x_2, x_1x_2^3 - x_2^5 \rangle$ . Monomials are represented by points. Two of them are in the same equivalence class if they are connected by a line or both are in an area with gray background color



are polynomials with exactly two terms and coefficients 1 and  $-1$ , respectively. Pure binomial ideals can be thought of a partition of the set of monomials into equivalence classes. Two monomials are in the same equivalence class if and only if a pure binomial in the ideal uses those monomials as terms. This means that each of the monomials can be replaced by the other one modulo the ideal. A visualization of these equivalence classes for an example is shown in Fig. 1.

The example Mayr and Meyer presented to prove the exponential space lower bound on the word problem for general polynomial ideals consists of pure binomial ideals only, which implies that the word problem for pure binomial ideals is as hard as the word problem for general polynomial ideals.

**Theorem 4.1 ([18])** *The word problem for pure binomial ideals is EXPSPACE-complete.*

It is an interesting finding that binomial ideals do already contain the full complexity of the general case whereas the word problem for monomial ideals is solvable in polynomial time. An overview of the complexity of the word problem for different classes of polynomial ideals is listed in Table 1.

An application from theoretical computer science where pure binomial ideals occur are commutative Thue systems. These systems are term replacement systems with the additional properties that the order of the characters may be changed at every time and all replacements can also be executed backwards [23, 24].

**Definition 4.2 ([23, 24])** A commutative Thue system consists of a finite set of congruences  $\mathcal{P} = \{\alpha_i \equiv \beta_i \mid i \in \{1, \dots, s\}\}$  for some  $s \in \mathbb{N}_{>0}$  between words over a finite alphabet  $\Sigma$ . Two words  $\gamma, \delta \in \Sigma^*$  are equivalent modulo the commutative Thue system  $\mathcal{P}$  if and only if there are  $\lambda_1, \dots, \lambda_r \in \Sigma^*$  for some  $r \in \mathbb{N}_{>0}$  with  $\gamma = \lambda_1, \delta = \lambda_r$  and  $\lambda_j$  can be transformed to  $\lambda_{j+1}$  for all  $j \in \{1, \dots, r - 1\}$  by

**Table 1** Complexity classes known for the (radical) word problem of several classes of polynomial ideals

Type of ideals	Word problem	Radical word problem
Polynomial ideals	<b>EXPSpace</b> -complete [18]	<b>PSPACE</b> [3, 11]
Binomial ideals	<b>EXPSpace</b> -complete [18]	<b>coNP</b> -complete [25]
Pure binomial ideals	<b>EXPSpace</b> -complete [18]	<b>coNP</b> -complete [25]
Toric ideals	<b>L</b> [22]	<b>L</b> [22]
Monomial ideals	<b>P</b>	<b>P</b>

reordering letters or applying a congruence  $\alpha_i \equiv \beta_i$  for some  $i \in \{1, \dots, s\}$ , i.e. replacing the subword  $\alpha_i$  by the subword  $\beta_i$  or vice versa.

It turns out that there is a bijection from the set of pure binomial ideals to the set of commutative Thue systems preserving the structure of these objects.

**Definition 4.3** Let  $\Sigma = \{\sigma_1, \dots, \sigma_n\}$  be a finite set with  $n \in \mathbb{N}_0$  and let

$$\mathcal{P} = \left\{ \alpha_i \equiv \beta_i \mid i \in \{1, \dots, s\} \right\}$$

be a commutative Thue system for  $s \in \mathbb{N}_{>0}$  and  $\alpha_i, \beta_i \in \Sigma^*$  for  $i \in \{1, \dots, s\}$ . Let  $\Phi : \Sigma^* \rightarrow \mathbb{N}_0^s$  be the Parikh mapping, i.e. the  $i$ -th entry of  $\Phi(\gamma)$  is the number of occurrences of  $\sigma_i$  in  $\gamma$  for all  $\gamma \in \Sigma^*$  and  $i \in \{1, \dots, n\}$ . For each ring  $R$  we define the polynomial ideal

$$\mathcal{I}_R(\mathcal{P}) := \left\langle \underline{x}^{\Phi(\alpha_i)} - \underline{x}^{\Phi(\beta_i)} \mid i \in \{1, \dots, s\} \right\rangle \trianglelefteq R[x_1, \dots, x_n]$$

Each instance of the equivalence problem for commutative Thue systems can be mapped to an instance of the word problem for pure binomial ideals and vice versa.

**Theorem 4.4 ([18])** For all commutative Thue Systems  $\mathcal{P}$  we have

$$\underline{x}^{\Phi(\gamma)} - \underline{x}^{\Phi(\delta)} \in \mathcal{I}_{\mathbb{Z}}(\mathcal{P}) \Leftrightarrow \underline{x}^{\Phi(\gamma)} - \underline{x}^{\Phi(\delta)} \in \mathcal{I}_{\mathbb{Q}}(\mathcal{P}) \Leftrightarrow \gamma \equiv \delta(\mathcal{P})$$

for all  $\gamma, \delta \in \Sigma^*$ .

The lower bound by Mayr and Meyer [18] was proven by giving a reduction of the **EXPSpace**-complete problem to decide whether three-counter machines terminate with a computation bounded double-exponentially by the input size to the equivalence problem of commutative Thue systems.

## 5 Toric Ideals

Toric ideals are another important special case of polynomial ideals that often occur in practice. They appear in particular as the kernel of maps from polynomial rings to rings of Laurent polynomials. Toric ideals are the same as saturated pure binomial ideals or the extension of pure binomial ideals into the ring of Laurent polynomials [25]. Over algebraically closed coefficient fields toric ideals are the same as binomial prime ideals [7].

There is a polynomial time algorithm to solve the word problem for toric ideals and binomials by using Gaussian elimination on the Macaulay matrix of the toric ideal. This means that the inherent complexity of toric ideals is much lower than the one of binomial ideals.

**Theorem 5.1** *Let  $f_1, \dots, f_s \in R[x_1, \dots, x_n]$  be pure binomials over a ring  $R$  with  $n, s \in \mathbb{N}_{>0}$  such that  $\langle f_1, \dots, f_s \rangle$  is a toric ideal. Let  $\phi$  be a map from the set of pure binomials contained in  $R[x_1, \dots, x_n]$  to the  $\mathbb{Z}$ -vector space  $\mathbb{Z}^n$  defined by  $\phi(\underline{x}^\alpha - \underline{x}^\beta) := \alpha - \beta$  for all  $\alpha, \beta \in \mathbb{N}_0^n$ . Then we have*

$$g \in \langle f_1, \dots, f_s \rangle \trianglelefteq R[x_1, \dots, x_n] \Leftrightarrow \phi(g) \in \text{span} \{ \phi(f_1), \dots, \phi(f_s) \} \subseteq \mathbb{Z}^n$$

for all pure binomials  $g \in R[x_1, \dots, x_n]$ .

Using this approach and a memory-efficient algorithm for solving linear systems of equations Ritscher proved an upper space bound for the membership problem for toric ideals [22]. He was also able to extend the algorithm to test for the membership of general polynomials instead of pure binomials in toric ideals.

**Theorem 5.2 ([22])** *Let  $f_1, \dots, f_s \in k[x_1, \dots, x_n]$  be pure binomials over a well-endowed field  $k$  with  $n, s \in \mathbb{N}_{>0}$  such that  $\langle f_1, \dots, f_s \rangle$  is a toric ideal and let  $g \in k[x_1, \dots, x_n]$  be a polynomial with  $t \in \mathbb{N}_{>0}$  terms. Let  $q \in \mathbb{N}_{>0}$  be an upper bound on the bitsize of all coefficients and exponents of  $g$  and  $f_1, \dots, f_s$ . The word problem to check whether  $g \in \langle f_1, \dots, f_s \rangle$  can be decided in space  $\mathcal{O}(\log^2((n + s + t)q))$ .*

The word problem for toric ideals is therefore known to be contained in  $\mathbf{L}$ , the complexity class of all problems solvable in logarithmic space.  $\mathbf{L}$  is known to be contained in  $\mathbf{P}$ , the class of all problems solvable in polynomial time, but it is unknown whether both classes are actually the same.

## 6 Radical Ideals

The radical of a polynomial ideal is constructed by adding all polynomials to the ideal such that a power of them is included in the original polynomial ideal. Thus, the radical is a superset of the original polynomial ideal. Growing the polynomial ideal in the described way does not change its variety, i.e. the set of

common solutions of all polynomials contained in the polynomial ideal, but just the multiplicities of the roots. The radical therefore contains all geometric information about a polynomial ideal. In contrast to pure binomial ideals, the degree bounds for radical ideals are better than the ones for the general case, although there are much less results for radical ideals. Containing the full geometric information but having better degree bounds makes radical ideals interesting objects to study.

Brownawell proved a single-exponential bound for the degrees of the coefficients of a polynomial's representation contained in radical ideals [3]. Kollár improved this bound 1 year later [11].

**Theorem 6.1 ([11])** *Let  $f_1, \dots, f_s \in k[x_1, \dots, x_n]$  be polynomials with  $\deg(f_i) \leq d$ ,  $\deg(f_i) \neq 2$  for all  $i \in \{1, \dots, s\}$  over a field  $k$  for some  $d, n, s \in \mathbb{N}_{>0}$  and let  $g \in \sqrt{\langle f_1, \dots, f_s \rangle}$ . There are  $r \in \mathbb{N}_{>0}$  and  $g_1, \dots, g_s \in k[x_1, \dots, x_n]$  such that*

$$g^r = \sum_{i=1}^s f_i g_i$$

with  $s \leq d^n$  and  $\deg(f_i g_i) \leq (1 + \deg(g))d^n$  for all  $i \in \{1, \dots, s\}$ .

Using those degree bounds the word problem for radical ideals can be solved in polynomial space and exponential time by enumerating all possible  $g_1, \dots, g_s$ . There are also several algorithms that compute the actual radical of a polynomial ideal, for instance the one presented by Laplagne in 2006 [16], but all known algorithms that compute radicals of all polynomial ideals need at least exponential space and double-exponential time which are the same bounds as for Gröbner basis computations and the word problem for general polynomial ideals.

## 7 Radical Binomial Ideals

The radical word problem is given a polynomial ideal and a polynomial to solve the word problem for the radical of this ideal and the polynomial. This problem is a generalization of the word problem for radical ideals. Algorithms solving the radical word problem might be more efficient than **EXSPACE** which is needed to compute the radical since they do not need to actually compute a basis of the radical ideal. This problem is interesting because the result of the radical word problem is true if and only if the given polynomial holds for all solutions of the polynomial ideal, which means that using this problem one can deduce information about the solutions of a system of equations without actually computing the full solution.

Radicals of (pure) binomial ideals and toric ideals have even more structure as proven by Gilmer in 1984 [9] and Eisenbud and Sturmfels in 1996 [7].

**Theorem 7.1 ([7, 9])** *Let  $k$  be a field and  $n \in \mathbb{N}_{>0}$ . The radical of each binomial ideal contained in  $k[x_1, \dots, x_n]$  is a binomial ideal again. Similarly, the radical of each pure binomial ideal contained in  $k[x_1, \dots, x_n]$  is a pure binomial ideal again.*

**Theorem 7.2 ([7])** *Let  $k$  be an algebraically closed field with  $\text{char}(k) = 0$  and  $n \in \mathbb{N}_{>0}$  and let  $I \leq k[x_1, \dots, x_n]$  be a toric ideal.  $I$  is a radical ideal.*

This means the radical operation is closed under binomial ideals and pure binomial ideals. For toric ideals computing the radical does not change the ideal at all. It is therefore a common technique to reduce computations of radicals to toric ideals which are already radical.

In the case of binomial ideals over fields with characteristic 0 there is a special tool available to compute radicals doing this. In 1996 Eisenbud and Sturmfels introduced the cellular decomposition of binomial ideals [7]. They suggested to partition the variety of the binomial ideal into cells where points are in the same cell if they have the same components being non-zero. The intersection of the ideals corresponding to each cell is the radical of the original ideal.

**Theorem 7.3 ([7])** *Let  $k$  be a field with  $\text{char}(k) = 0$ ,  $n \in \mathbb{N}_0$  and  $I \leq k[x_1, \dots, x_n]$  be a binomial ideal. Then*

$$\sqrt{I} = \bigcap_{\Delta \subseteq \{x_1, \dots, x_n\}} I_\Delta : \left( \prod_{x_i \in \Delta} x_i \right)^\infty + \langle \{x_i \mid x_i \notin \Delta\} \rangle$$

where  $I_\Delta$  is the image of  $I$  under the ring endomorphism on  $k[x_1, \dots, x_n]$  defined by

$$1 \mapsto 1, x_i \mapsto \begin{cases} x_i & \text{if } x_i \in \Delta \\ 0 & \text{else} \end{cases}$$

for all  $i \in \{1, \dots, n\}$  and  $\Delta \subseteq \{x_1, \dots, x_n\}$ .

Even though the radical of a binomial ideal over a field with characteristic 0 is binomial again, the intermediate results do not have to be binomial since the intersection of two binomial ideals is not binomial in general. Nevertheless, in 1997 Becker, Grobe, and Niermann proved that the intersections of the cellular decomposition can be executed in an order such that all intermediate results are binomial [2]. This result implies that all intermediate results of the cellular decomposition of pure binomial ideals can be interpreted as commutative Thue systems, too.

Mayr and Toman presented an algorithm to solve the radical word problem for pure binomial ideals in **coNP** [21]. They used the cellular decomposition of binomial ideals and the polynomial time algorithm to solve the word problem for toric ideals. They also showed how to encode the coefficients of binomials to solve the radical word problem for non-pure binomial ideals in the same complexity class. Additionally, they proved a matching lower bound for the radical word problem of pure binomial ideals by giving a reduction from the **TAUTOLOGY** problem. This showed that the radical word problem for binomial ideals is **coNP**-complete.

It is interesting to note that this complexity class is characterized by the time needed for running the machine instead of its space consumption. All other complexity classes listed in Table 1 for general polynomial ideals or subclasses

mentioned in this report use space bounds for the machines. It is known that **coNP** is contained in **PSPACE**, which is the complexity of the radical word problem for general polynomial ideals, but it is still unknown whether there are problems contained in **PSPACE** but not in **coNP**.

We have seen that the bijection between pure binomial ideals and commutative Thue systems provides versatile tools for systems of pure binomial ideals. Operations on pure binomial ideals like the sum, product, intersection, quotient, and saturation each can be equivalently defined in terms of commutative Thue systems. As opposed to this, it is not possible to directly define radicals of commutative Thue systems since this definition involves powers of pure binomials which are no pure binomials anymore and therefore have no corresponding objects in terms of commutative Thue systems. In his PhD thesis Toman suggests a definition of radicals of commutative Thue systems not involving polynomial ideals [26].

To do this one needs a way to represent powers of binomials as binomials again. Squares of binomials for instance can be split up to two different binomials.

**Theorem 7.4 ([26])** *Let  $I \trianglelefteq k[x_1, \dots, x_n]$  be a pure binomial ideal over a field  $k$  with  $\text{char}(k) = 0$  for some  $n \in \mathbb{N}_{>0}$ . Let  $u, v \in \mathbb{N}_{>0}^n$ . We have*

$$(\underline{x}^u - \underline{x}^v)^2 \in I \Leftrightarrow \underline{x}^{2u} - \underline{x}^{u+v} \in I, \underline{x}^{u+v} - \underline{x}^{2v} \in I$$

For a binomial  $g$  this theorem implies that  $g^2$  is contained in the pure binomial ideal  $I$  if and only if all monomials of  $g^2$  are equivalent modulo  $I$ . The latter property can be easily expressed using commutative Thue systems whereas the former involves polynomials that are no pure binomials and can therefore not be expressed in terms of commutative Thue systems. A similar theorem is true for higher powers of the binomial.

**Theorem 7.5 ([26])** *Let  $f_1, \dots, f_s \in R[x_1, \dots, x_n]$  be pure binomials over a ring  $R$  for some  $d, n, s \in \mathbb{N}_{>0}$ . Let*

$$g \in \sqrt{\langle f_1, \dots, f_s \rangle} \trianglelefteq R[x_1, \dots, x_n]$$

*be a pure binomial. There is an  $r \in \mathbb{N}_{>0}$  such that all terms of  $g^r$  are equivalent modulo  $\langle f_1, \dots, f_s \rangle$ .*

We were also able to find a degree bound on the exponent  $r$  that makes all terms of  $g^r$  equivalent that is only slightly bigger than all known degree bounds on the exponent  $t$  such that  $g^t \in \langle f_1, \dots, f_s \rangle$ .

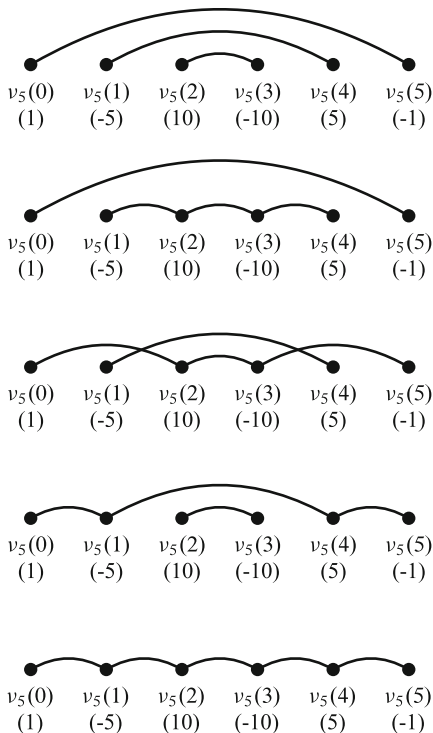
**Theorem 7.6 ([26])** *In the setting of Theorem 7.5 with the additional property that  $2 < \text{deg}(f_i) \leq d$  for all  $i \in \{1, \dots, s\}$  we can choose*

$$r := 2d^n \left( \log 2d^n + \log \log 2d^n \right) + 1 \in \mathbb{N}_{>0}$$

All terms of a polynomial contained in a pure binomial ideal can be partitioned into equivalence classes modulo the ideal where the sum of the coefficients in each



**Fig. 2** All possible equivalence classes of terms of a fifth power of a pure binomial  $(x^u - x^v)^5 \in I$  modulo a pure binomial ideal  $I$ . We use the notation  $v_i(j) = x^{ju+(i-j)v}$  and the coefficients of the terms are given in brackets



equivalence class is to 0. All possible configurations of those equivalence classes of the fifth power of a pure binomial are visualized in Fig. 2.

The theorems above imply that for powers of pure binomials we additionally only get one equivalence class for exponents of the given size. Those theorems allow the following definition of a radical of commutative Thue systems by translating them form pure binomial ideals to commutative Thue systems since they only contain statements on pure binomials only.

**Theorem 7.7 ([26])** *Let  $\Sigma$  be a finite alphabet and let  $\mathcal{P}$  be a commutative Thue system over  $\Sigma^*$ . We iteratively define  $\mathcal{P}_0 := \mathcal{P}$  and  $\mathcal{P}_i$  to be the commutative Thue system over  $\Sigma^*$  generated by all equivalences  $\alpha \equiv_{\mathcal{P}_i} \beta$  with  $\alpha, \beta \in \Sigma^*$  and*

$$\alpha\alpha \equiv_{\mathcal{P}_{i-1}} \alpha\beta \text{ as well as } \alpha\beta \equiv_{\mathcal{P}_{i-1}} \beta\beta$$

for  $i \in \mathbb{N}_{>0}$ . There is an  $s \in \mathbb{N}_{>0}$  with  $\mathcal{P}_s = \mathcal{P}_i$  for all  $i \in \mathbb{N}_{>0}, i \geq s$  and

$$\mathcal{I}_{\mathbb{Q}}(\mathcal{P}_s) = \sqrt{\mathcal{I}_{\mathbb{Q}}(\mathcal{P})}$$

Using this approach one can compute the radical of a commutative Thue system in **EXPSpace**. This is similar to numerous other problems on commutative Thue

systems which have a complexity of **EXPSpace** like the coverability, the subword, the containment, and the equivalence problems [12, 13].

$\mathcal{P}_s$  is a radical of  $\mathcal{P}$  defined purely in terms of commutative Thue systems. This construction provides another tool for finding better degree bounds for radical ideals.

## 8 Future Research

There are numerous open questions related to degree bounds of polynomial ideals that can be answered in future research. The known upper and lower bounds for the degree of the generators of a Gröbner bases of the radical of a given polynomial ideal do not match.

Likewise, the algorithms presented above for the radical word problem of binomial ideals do not compute an actual basis of the radical ideal. It is an open question whether the complexity for computing the basis of the radical of a binomial ideal is different to the general case.

Some of the results presented above only work for the rationals as base field of the polynomial ring. Similar bounds for positive characteristics of the base field are often unknown.

**Acknowledgements** This work was partially supported by Deutsche Forschungsgemeinschaft (DFG) through priority program SPP 1489 “Algorithmic and Experimental Methods in Algebra, Geometry, and Number Theory” in the project “Degree Bounds for Gröbner Bases of Important Classes of Polynomial Ideals and Efficient Algorithms”, TUM Graduate School, and TopMath, a graduate program of the Elite Network of Bavaria. We are grateful for their support.

## References

1. N.H. Abel, Démonstration de l'impossibilité de la résolution algébrique des équations générales qui passent le quatrième degré. *J. Reine Angew. Math.* **1**, 65–96 (1826)
2. E. Becker, R. Grobe, M. Niermann, Radicals of binomial ideals. *J. Pure Appl. Algebra* **117–118**, 41–79 (1997)
3. W.D. Brownawell, Bounds for the degrees in the nullstellensatz. *Ann. Math.* **126**(3), 577 (1987)
4. B. Buchberger, Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal (An algorithm for finding the basis elements in the residue class ring modulo a zero dimensional polynomial ideal). Ph.D. Thesis, Mathematical Institute, University of Innsbruck (1965)
5. A. Dickstein, N. Fitchas, M. Giusti, C. Sessa, The membership problem for unmixed polynomial ideals is solvable in single exponential time. *Discret. Appl. Math.* **33**(1), 73–94 (1991)
6. T.W. Dubé, The structure of polynomial ideals and Gröbner bases. *SIAM J. Comput.* **19**(4), 750–773 (1990)
7. D. Eisenbud, B. Sturmfels, Binomial ideals. *Duke Math. J.* **84**(1), 1–45 (1996)

8. J.C. Faugère, P. Gianni, D. Lazard, T. Mora, Efficient computation of zero-dimensional gröbner bases by change of ordering. *J. Symb. Comput.* **16**(4), 329–344 (1993)
9. R. Gilmer, *Commutative Semigroup Rings*. Chicago Lectures in Mathematics (University of Chicago Press, Chicago, 1984)
10. G. Hermann, Die Frage der endlich vielen Schritte in der Theorie der Polynomideale. *Math. Ann.* **95**(1), 736–788 (1926)
11. J. Kollár, Sharp effective Nullstellensatz. *J. Am. Math. Soc.* **1**(4), 963 (1988)
12. U. Koppenhagen, E.W. Mayr, Optimal algorithms for the coverability, the subword, the containment, and the equivalence problems for commutative semigroups. *Inf. Comput.* **158**(2), 98–124 (2000)
13. U. Koppenhagen, E.W. Mayr, An optimal algorithm for constructing the reduced Gröbner basis of binomial ideals, and applications to commutative semigroups. *J. Symb. Comput.* **31**(1–2), 259–276 (2001)
14. M. Kratzer, Computing the dimension of a polynomial ideal and membership in low-dimensional ideals. Master's Thesis, TU München (2008)
15. K. Kühnle, E.W. Mayr, Exponential space computation of Gröbner bases. In: *Proceedings of the 1996 International Symposium on Symbolic and Algebraic Computation - ISSAC '96* (ACM, New York, 1996), pp. 63–71
16. S. Laplagne, An algorithm for the computation of the radical of an ideal. In: *Proceedings of the 2006 International Symposium on Symbolic and Algebraic Computation - ISSAC '06* (ACM, New York, 2006), p. 191
17. E.W. Mayr, Some complexity results for polynomial ideals. *J. Complexity* **13**(3), 303–325 (1997)
18. E.W. Mayr, A.R. Meyer, The complexity of the word problems for commutative semigroups and polynomial ideals. *Adv. Math.* **46**(3), 305–329 (1982)
19. E.W. Mayr, S. Ritscher, Space-efficient Gröbner basis computation without degree bounds. In: *Proceedings of the 36th International Symposium on Symbolic and Algebraic Computation (ISSAC '11)* (ACM, New York, 2011), pp. 257–264
20. E.W. Mayr, S. Ritscher, Dimension-dependent bounds for Gröbner bases of polynomial ideals. *J. Symb. Comput.* **49**, 78–94 (2013). The International Symposium on Symbolic and Algebraic Computation
21. E.W. Mayr, S. Toman, The complexity of the membership problem for radical binomial ideals, in *International Conference Polynomial Computer Algebra '2015*, ed. by N.N. Vassiliev (Euler International Mathematical Institute/VVM Publishing, Saint Petersburg, 2015), pp. 61–64
22. S. Ritscher, Degree bounds and complexity of Gröbner bases of important classes of polynomial ideals. Ph.D. Thesis, TU München (2012)
23. A. Thue, Die Lösung eines Spezialfalles eines generellen logischen Problems. *Skrifter udg. af Videnskabs-Selskabet i Christiania. I, Math. Naturv. Klasse* **8** (1910)
24. A. Thue, Problem über Veränderungen von Zeichenreihen nach gegebenen Regeln. *Skrifter udg. af Videnskabs-Selskabet i Christiania. I, Math. Naturv. Klasse* **10** (1914)
25. S. Toman, The radical word problem for binomial ideals. Master's Thesis, TU München (2015)
26. S. Toman, Radicals of binomial ideals and commutative Thue systems. Ph.D. Thesis, TU München (2017)
27. C.K. Yap, A new lower bound construction for commutative Thue systems with applications. *J. Symb. Comput.* **12**(1), 1–27 (1991)

# Localizations of Inductively Factored Arrangements



Tilman Möller and Gerhard Röhrle

**Abstract** We show that the class of inductively factored arrangements is closed under taking localizations. We illustrate the usefulness of this with an application.

**Keywords** Nice arrangement • Inductively factored arrangement • Localization of arrangement

**Subject Classifications** Primary 52C35, 14N20; Secondary 51D20

## 1 Introduction

The notion of a nice arrangement is due to Terao [12]. This class generalizes the class of supersolvable arrangements, [9] (cf. [10, Thm. 3.81]). There is an inductive version of this class, so called inductively factored arrangements, due to Jambu and Paris [6], see Definition 2.7. This inductive class (properly) contains the class of supersolvable arrangements and is (properly) contained in the class of inductively free arrangements, see [3, Rem. 3.33].

For an overview on properties of nice and inductively factored arrangements, and for their connection with the underlying Orlik-Solomon algebra, see [10, §3], [6], and [3]. In [3], Hoge and the second author proved an addition-deletion theorem for nice arrangements, see Theorem 2.6 below. This is an analogue of Terao's celebrated Addition-Deletion Theorem 2.1 for free arrangements for the class of nice arrangements.

The class of free arrangements is known to be closed under taking localizations, [10, Thm. 4.37]. It is also known that this property restricts to various stronger notions of freeness, see [5, Thm. 1]. It is clear that the class of nice arrangements also satisfies this property, see Remark 2.5 below. Inductive arguments by means of

---

T. Möller • G. Röhrle (✉)

Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany  
e-mail: [tilman.moeller@rub.de](mailto:tilman.moeller@rub.de); [gerhard.roehrle@rub.de](mailto:gerhard.roehrle@rub.de)

localizations are a pivotal technique in the theory of arrangements. Therefore, it is natural to investigate this question for the stronger property of inductively factored arrangements as well. Here is the main result of our note.

**Theorem 1.1** *The class of inductively factored arrangements is closed under taking localizations.*

Theorem 1.1 readily extends to the class of hereditarily inductively factored arrangements, see Remark 3.3.

We show the utility of Theorem 1.1 in Example 3.4 which in turn is used in the classification of all inductively factored restrictions of reflection arrangements in [7].

## 2 Recollections and Preliminaries

### 2.1 Hyperplane Arrangements

Let  $\mathbb{K}$  be a field and let  $V = \mathbb{K}^\ell$  be an  $\ell$ -dimensional  $\mathbb{K}$ -vector space. A *hyperplane arrangement*  $\mathcal{A}$  in  $V$  is a finite collection of hyperplanes in  $V$ . We also use the term  $\ell$ -arrangement for  $\mathcal{A}$ . The empty  $\ell$ -arrangement is denoted by  $\Phi_\ell$ .

The *lattice*  $L(\mathcal{A})$  of  $\mathcal{A}$  is the set of subspaces of  $V$  of the form  $H_1 \cap \dots \cap H_i$  where  $\{H_1, \dots, H_i\}$  is a subset of  $\mathcal{A}$ . For  $X \in L(\mathcal{A})$ , we have two associated arrangements, firstly  $\mathcal{A}_X := \{H \in \mathcal{A} \mid X \subseteq H\} \subseteq \mathcal{A}$ , the *localization of  $\mathcal{A}$  at  $X$* , and secondly, the *restriction of  $\mathcal{A}$  to  $X$* ,  $(\mathcal{A}^X, X)$ , where  $\mathcal{A}^X := \{X \cap H \mid H \in \mathcal{A} \setminus \mathcal{A}_X\}$ . Note that  $V$  belongs to  $L(\mathcal{A})$  as the intersection of the empty collection of hyperplanes and  $\mathcal{A}^V = \mathcal{A}$ . The lattice  $L(\mathcal{A})$  is a partially ordered set by reverse inclusion:  $X \leq Y$  provided  $Y \subseteq X$  for  $X, Y \in L(\mathcal{A})$ .

If  $0 \in H$  for each  $H$  in  $\mathcal{A}$ , then  $\mathcal{A}$  is called *central*. If  $\mathcal{A}$  is central, then the *center*  $T_{\mathcal{A}} := \bigcap_{H \in \mathcal{A}} H$  of  $\mathcal{A}$  is the unique maximal element in  $L(\mathcal{A})$  with respect to the partial order. We have a *rank function* on  $L(\mathcal{A})$ :  $r(X) := \text{codim}_V(X)$ . The *rank*  $r := r(\mathcal{A})$  of  $\mathcal{A}$  is the rank of a maximal element in  $L(\mathcal{A})$ . Throughout, we only consider central arrangements.

More generally, for  $U$  an arbitrary subspace of  $V$ , we can define  $\mathcal{A}_U := \{H \in \mathcal{A} \mid U \subseteq H\} \subseteq \mathcal{A}$ , the *localization of  $\mathcal{A}$  at  $U$* , and  $\mathcal{A}^U := \{U \cap H \mid H \in \mathcal{A} \setminus \mathcal{A}_U\}$ , a subarrangement in  $U$ .

### 2.2 Free Hyperplane Arrangements

Let  $S = S(V^*)$  be the symmetric algebra of the dual space  $V^*$  of  $V$ . Let  $\text{Der}(S)$  be the  $S$ -module of  $\mathbb{K}$ -derivations of  $S$ . Since  $S$  is graded,  $\text{Der}(S)$  is a graded  $S$ -module.

Let  $\mathcal{A}$  be a central arrangement in  $V$ . Then for  $H \in \mathcal{A}$  we fix  $\alpha_H \in V^*$  with  $H = \ker \alpha_H$ . The *defining polynomial*  $Q(\mathcal{A})$  of  $\mathcal{A}$  is given by  $Q(\mathcal{A}) := \prod_{H \in \mathcal{A}} \alpha_H \in S$ .

The *module of  $\mathcal{A}$ -derivations* of  $\mathcal{A}$  is defined by

$$D(\mathcal{A}) := \{\theta \in \text{Der}(S) \mid \theta(Q(\mathcal{A})) \in Q(\mathcal{A})S\}.$$

We say that  $\mathcal{A}$  is *free* if  $D(\mathcal{A})$  is a free  $S$ -module, cf. [10, §4].

If  $\mathcal{A}$  is a free arrangement, then the  $S$ -module  $D(\mathcal{A})$  admits a basis of  $n$  homogeneous derivations, say  $\theta_1, \dots, \theta_n$ , [10, Prop. 4.18]. While the  $\theta_i$ 's are not unique, their polynomial degrees  $\text{pdeg } \theta_i$  are unique (up to ordering). This multiset is the set of *exponents* of the free arrangement  $\mathcal{A}$  and is denoted by  $\text{exp } \mathcal{A}$ .

Terao's celebrated *Addition-Deletion Theorem* which we recall next plays a pivotal role in the study of free arrangements, [10, §4]. For  $\mathcal{A}$  non-empty, let  $H_0 \in \mathcal{A}$ . Define  $\mathcal{A}' := \mathcal{A} \setminus \{H_0\}$ , and  $\mathcal{A}'' := \mathcal{A}^{H_0} = \{H_0 \cap H \mid H \in \mathcal{A}'\}$ , the restriction of  $\mathcal{A}$  to  $H_0$ . Then  $(\mathcal{A}, \mathcal{A}', \mathcal{A}'')$  is a *triple* of arrangements, [10, Def. 1.14].

**Theorem 2.1** ([11]) *Suppose that  $\mathcal{A} \neq \Phi_\ell$ . Let  $(\mathcal{A}, \mathcal{A}', \mathcal{A}'')$  be a triple of arrangements. Then any two of the following statements imply the third:*

- (i)  $\mathcal{A}$  is free with  $\text{exp } \mathcal{A} = \{b_1, \dots, b_{\ell-1}, b_\ell\}$ ;
- (ii)  $\mathcal{A}'$  is free with  $\text{exp } \mathcal{A}' = \{b_1, \dots, b_{\ell-1}, b_\ell - 1\}$ ;
- (iii)  $\mathcal{A}''$  is free with  $\text{exp } \mathcal{A}'' = \{b_1, \dots, b_{\ell-1}\}$ .

There are various stronger notions of freeness which we discuss in the following sections.

### 2.3 Inductively Free Arrangements

Theorem 2.1 motivates the notion of *inductively free* arrangements, see [11] or [10, Def. 4.53].

**Definition 2.2** The class  $\mathcal{IF}$  of *inductively free* arrangements is the smallest class of arrangements subject to

- (i)  $\Phi_\ell \in \mathcal{IF}$  for each  $\ell \geq 0$ ;
- (ii) if there exists a hyperplane  $H_0 \in \mathcal{A}$  such that both  $\mathcal{A}'$  and  $\mathcal{A}''$  belong to  $\mathcal{IF}$ , and  $\text{exp } \mathcal{A}'' \subseteq \text{exp } \mathcal{A}'$ , then  $\mathcal{A}$  also belongs to  $\mathcal{IF}$ .

Free arrangements are closed with respect to taking localizations, e.g. see [10, Thm. 4.37]. This also holds for the class  $\mathcal{IF}$ .

**Theorem 2.3** ([5, Thm. 1]) *If  $\mathcal{A}$  is inductively free, then so is  $\mathcal{A}_U$  for every subspace  $U$  in  $V$ .*

### 2.4 Nice and Inductively Factored Arrangements

The notion of a *nice* or *factored* arrangement is due to Terao [12]. It generalizes the concept of a supersolvable arrangement, see [9, Thm. 5.3] and [10, Prop. 2.67, Thm. 3.81]. Terao’s main motivation was to give a general combinatorial framework to deduce factorizations of the underlying Orlik-Solomon algebra, see also [10, §3.3]. We recall the relevant notions from [12] (cf. [10, §2.3]):

**Definition 2.4** Let  $\pi = (\pi_1, \dots, \pi_s)$  be a partition of  $\mathcal{A}$ .

- (a)  $\pi$  is called *independent*, provided for any choice  $H_i \in \pi_i$  for  $1 \leq i \leq s$ , the resulting  $s$  hyperplanes are linearly independent, i.e.  $r(H_1 \cap \dots \cap H_s) = s$ .
- (b) Let  $X \in L(\mathcal{A})$ . The *induced partition*  $\pi_X$  of  $\mathcal{A}_X$  is given by the non-empty blocks of the form  $\pi_i \cap \mathcal{A}_X$ .
- (c)  $\pi$  is *nice* for  $\mathcal{A}$  or a *factorization* of  $\mathcal{A}$  provided
  - (i)  $\pi$  is independent, and
  - (ii) for each  $X \in L(\mathcal{A}) \setminus \{V\}$ , the induced partition  $\pi_X$  admits a block which is a singleton.

If  $\mathcal{A}$  admits a factorization, then we also say that  $\mathcal{A}$  is *factored* or *nice*.

*Remark 2.5* The class of nice arrangements is closed under taking localizations. For, if  $\mathcal{A}$  is non-empty and  $\pi$  is a nice partition of  $\mathcal{A}$ , then the non-empty parts of the induced partition  $\pi_X$  form a nice partition of  $\mathcal{A}_X$  for each  $X \in L(\mathcal{A}) \setminus \{V\}$ ; cf. the proof of [12, Cor. 2.11].

Following Jambu and Paris [6], we introduce further notation. Suppose  $\mathcal{A}$  is not empty. Let  $\pi = (\pi_1, \dots, \pi_s)$  be a partition of  $\mathcal{A}$ . Let  $H_0 \in \pi_1$  and let  $(\mathcal{A}, \mathcal{A}', \mathcal{A}'')$  be the triple associated with  $H_0$ . Then  $\pi$  induces a partition  $\pi'$  of  $\mathcal{A}'$ , i.e. the non-empty subsets  $\pi_i \cap \mathcal{A}'$ . Note that since  $H_0 \in \pi_1$ , we have  $\pi_i \cap \mathcal{A}' = \pi_i$  for  $i = 2, \dots, s$ . Also, associated with  $\pi$  and  $H_0$ , we define the *restriction map*

$$\varrho := \varrho_{\pi, H_0} : \mathcal{A} \setminus \pi_1 \rightarrow \mathcal{A}'' \text{ given by } H \mapsto H \cap H_0$$

and set

$$\pi''_i := \varrho(\pi_i) = \{H \cap H_0 \mid H \in \pi_i\} \text{ for } 2 \leq i \leq s.$$

In general,  $\varrho$  need not be surjective nor injective. However, since we are only concerned with cases when  $\pi'' = (\pi''_2, \dots, \pi''_s)$  is a partition of  $\mathcal{A}''$ ,  $\varrho$  has to be onto and  $\varrho(\pi_i) \cap \varrho(\pi_j) = \emptyset$  for  $i \neq j$ .

The following analogue of Terao’s Addition-Deletion Theorem 2.1 for free arrangements for the class of nice arrangements is proved in [3, Thm. 3.5].

**Theorem 2.6** *Suppose that  $\mathcal{A} \neq \Phi_\ell$ . Let  $\pi = (\pi_1, \dots, \pi_s)$  be a partition of  $\mathcal{A}$ . Let  $H_0 \in \pi_1$  and let  $(\mathcal{A}, \mathcal{A}', \mathcal{A}'')$  be the triple associated with  $H_0$ . Then any two of the following statements imply the third:*

- (i)  $\pi$  is nice for  $\mathcal{A}$ ;
- (ii)  $\pi'$  is nice for  $\mathcal{A}'$ ;
- (iii)  $\varrho : \mathcal{A} \setminus \pi_1 \rightarrow \mathcal{A}''$  is bijective and  $\pi''$  is nice for  $\mathcal{A}''$ .

Note the bijectivity condition on  $\varrho$  in Theorem 2.6 is necessary, cf. [3, Ex. 3.3]. Theorem 2.6 motivates the following stronger notion of factorization, cf. [6], [3, Def. 3.8].

**Definition 2.7** The class  $\mathcal{IFAC}$  of *inductively factored* arrangements is the smallest class of pairs  $(\mathcal{A}, \pi)$  of arrangements  $\mathcal{A}$  together with a partition  $\pi$  subject to

- (i)  $(\Phi_\ell, ()) \in \mathcal{IFAC}$  for each  $\ell \geq 0$ ;
- (ii) if there exists a partition  $\pi$  of  $\mathcal{A}$  and a hyperplane  $H_0 \in \pi_1$  such that for the triple  $(\mathcal{A}, \mathcal{A}', \mathcal{A}'')$  associated with  $H_0$  the restriction map  $\varrho = \varrho_{\pi, H_0} : \mathcal{A} \setminus \pi_1 \rightarrow \mathcal{A}''$  is bijective and for the induced partitions  $\pi'$  of  $\mathcal{A}'$  and  $\pi''$  of  $\mathcal{A}''$  both  $(\mathcal{A}', \pi')$  and  $(\mathcal{A}'', \pi'')$  belong to  $\mathcal{IFAC}$ , then  $(\mathcal{A}, \pi)$  also belongs to  $\mathcal{IFAC}$ .

If  $(\mathcal{A}, \pi)$  is in  $\mathcal{IFAC}$ , then we say that  $\mathcal{A}$  is *inductively factored with respect to*  $\pi$ , or else that  $\pi$  is an *inductive factorization* of  $\mathcal{A}$ . Sometimes we simply say  $\mathcal{A}$  is *inductively factored* without reference to a specific inductive factorization of  $\mathcal{A}$ .

*Remark 2.8* If  $\pi$  is an inductive factorization of  $\mathcal{A}$ , then there exists an *induction of factorizations* by means of Theorem 2.6 as follows. This procedure amounts to choosing a total order on  $\mathcal{A}$ , say  $\mathcal{A} = \{H_1, \dots, H_n\}$ , so that each of the pairs  $(\mathcal{A}_0 = \Phi_\ell, ())$ ,  $(\mathcal{A}_i := \{H_1, \dots, H_i\}, \pi_i := \pi|_{\mathcal{A}_i})$ , and  $(\mathcal{A}_i'' := \mathcal{A}_i^{H_i}, \pi_i'')$  for each  $1 \leq i \leq n$ , belongs to  $\mathcal{IFAC}$  see [3, Rem. 3.16].

The connection with the previous notions is as follows.

**Proposition 2.9** ([3, Prop. 3.11]) *If  $\mathcal{A}$  is supersolvable, then  $\mathcal{A}$  is inductively factored.*

**Proposition 2.10** ([6, Prop. 2.2], [3, Prop. 3.14]) *Let  $\pi = (\pi_1, \dots, \pi_r)$  be an inductive factorization of  $\mathcal{A}$ . Then  $\mathcal{A}$  is inductively free with  $\exp \mathcal{A} = \{0^{\ell-r}, |\pi_1|, \dots, |\pi_r|\}$ .*

**Definition 2.11** We say that  $\mathcal{A}$  is *hereditarily inductively factored* provided  $\mathcal{A}^Y$  is inductively factored for every  $Y \in L(\mathcal{A})$ .

### 3 Proof of Theorem 1.1

Theorem 1.1 follows from our next theorem which asserts that an inductive factorization of an arrangement affords one for any localization.

**Theorem 3.1** *For  $U$  a proper, non-trivial subspace of  $V$ , if  $(\mathcal{A}, \pi)$  belongs to  $\mathcal{IFAC}$ , then so does  $(\mathcal{A}_U, \pi_U)$ .*



*Proof* We argue by induction on  $|\mathcal{A}|$ . If  $\mathcal{A} = \Phi_\ell$  there is nothing to show. So assume  $|\mathcal{A}| > 0$  and that the result holds for arrangements with fewer than  $|\mathcal{A}|$  hyperplanes. Since  $\mathcal{A}$  is non-empty and inductively factored, there is a partition  $\pi$  of  $\mathcal{A}$  and  $H_0 \in \mathcal{A}$  so that  $(\mathcal{A}, \pi)$ ,  $(\mathcal{A}', \pi')$  and  $(\mathcal{A}'', \pi'')$  belong to  $\mathcal{I}\mathcal{F}\mathcal{A}\mathcal{C}$ .

Suppose that  $U \not\subseteq H_0$ . Then  $(\mathcal{A}_U, \pi_U) = ((\mathcal{A}')_U, \pi'_U)$ . As  $|\mathcal{A}'| < |\mathcal{A}|$  and  $(\mathcal{A}', \pi')$  belongs to  $\mathcal{I}\mathcal{F}\mathcal{A}\mathcal{C}$ , so does  $((\mathcal{A}')_U, \pi'_U)$ , by our induction hypothesis. So  $(\mathcal{A}_U, \pi_U) \in \mathcal{I}\mathcal{F}\mathcal{A}\mathcal{C}$ .

Now suppose that  $U \subseteq H_0$ . Then we have that  $((\mathcal{A}_U)', (\pi_U)') = ((\mathcal{A}')_U, (\pi')_U)$  and also  $((\mathcal{A}_U)'', (\pi_U)'') = ((\mathcal{A}'')_U, (\pi'')_U)$ . Therefore, since  $|\mathcal{A}'|, |\mathcal{A}''| < |\mathcal{A}|$ , both  $((\mathcal{A}_U)', (\pi_U)')$  and  $((\mathcal{A}_U)'', (\pi_U)'')$  belong to  $\mathcal{I}\mathcal{F}\mathcal{A}\mathcal{C}$ , as both  $(\mathcal{A}', \pi')$  and  $(\mathcal{A}'', \pi'')$  do.

Moreover, as  $\pi_U$  is nice for  $\mathcal{A}_U$  and  $\pi'_U$  is nice for  $\mathcal{A}'_U$ , by Remark 2.5, it follows from Theorem 2.6 that the corresponding restriction map is bijective. Therefore, by Definition 2.7,  $(\mathcal{A}_U, \pi_U)$  belongs to  $\mathcal{I}\mathcal{F}\mathcal{A}\mathcal{C}$ , as claimed.

*Remark 3.2* An alternative proof of Theorem 3.1 consists in choosing an inductive chain of  $(\mathcal{A}, \pi)$  and intersecting it with  $\mathcal{A}_U$ . One then shows that (after removing redundant terms) that this then affords an inductive chain of  $(\mathcal{A}_U, \pi_U)$ .

*Remark 3.3* Theorem 1.1 readily extends to hereditarily inductively factored arrangements. For, let  $\mathcal{A}$  be hereditarily inductively factored and let  $Y \leq X$  in  $L(\mathcal{A})$ . Then, since  $\mathcal{A}^Y$  is inductively factored, so is  $(\mathcal{A}^Y)_X$ , by Theorem 1.1. Finally, since  $(\mathcal{A}_X)^Y = (\mathcal{A}^Y)_X$ , it follows that  $(\mathcal{A}_X)^Y$  is inductively factored.

The following example shows the utility of the results above. In particular, this example is used in the classifications of the nice and inductively factored restrictions of reflection arrangements in [7].

*Example 3.4* Let  $V = \mathbb{C}^\ell$  be an  $\ell$ -dimensional  $\mathbb{C}$ -vector space. Orlik and Solomon defined intermediate arrangements  $\mathcal{A}_\ell^k(r)$  in [8, §2] (cf. [10, §6.4]) which interpolate between the reflection arrangements  $\mathcal{A}(G(r, 1, \ell))$  and  $\mathcal{A}(G(r, r, \ell))$  of the complex reflection groups  $G(r, 1, \ell)$  and  $G(r, r, \ell)$ . For  $\ell, r \geq 2$  and  $0 \leq k \leq \ell$ , the defining polynomial of  $\mathcal{A}_\ell^k(r)$  is

$$Q(\mathcal{A}_\ell^k(r)) = x_1 \cdots x_k \prod_{\substack{1 \leq i < j \leq \ell \\ 0 \leq n < r}} (x_i - \zeta^n x_j),$$

where  $\zeta$  is a primitive  $r$ th root of unity, so that  $\mathcal{A}_\ell^\ell(r) = \mathcal{A}(G(r, 1, \ell))$  and  $\mathcal{A}_\ell^0(r) = \mathcal{A}(G(r, r, \ell))$ . Note that for  $1 < k < \ell$ ,  $\mathcal{A}_\ell^k(r)$  is not a reflection arrangement.

Each of these arrangements is known to be free, cf. [10, Prop. 6.85]. The supersolvable and inductively free cases among them were classified in [2], and [1], respectively.

If  $k \in \{\ell - 1, \ell\}$ , then  $\mathcal{A}_\ell^k(r)$  is supersolvable, by [2, Thm. 1.3], and so  $\mathcal{A}_\ell^k(r)$  is inductively factored, by Proposition 2.9. Let  $\ell \geq 4$ . We claim that  $\mathcal{A}_\ell^k(r)$  is not nice for  $0 \leq k \leq \ell - 4$  and moreover  $\mathcal{A}_\ell^k(r)$  is not inductively factored for  $0 \leq k \leq \ell - 3$ .

For  $k = 0$ , this follows from [4, Thm. 1.3]. So let  $1 \leq k \leq \ell - 3$  and set  $\mathcal{A} = \mathcal{A}_\ell^k(r)$ . Define

$$X := \bigcap_{\substack{k+1 \leq i < j \leq \ell \\ 0 \leq n < r}} \ker(x_i - \zeta^n x_j).$$

Then one checks that

$$\mathcal{A}_X \cong \mathcal{A}_{\ell-k}^0(r) = \mathcal{A}(G(r, r, \ell - k)).$$

For  $1 \leq k \leq \ell - 4$ , it follows from [4, Thm. 1.3] that  $\mathcal{A}(G(r, r, \ell - k))$  is not nice. Consequently, neither is  $\mathcal{A}_\ell^k(r)$ , by Remark 2.5. For  $k = \ell - 3$ , we have  $\mathcal{A}_X \cong \mathcal{A}(G(r, r, 3))$ . By Hoge and Röhrle [4, Cor. 1.4], the latter is not inductively factored, thus neither is  $\mathcal{A}_\ell^{\ell-3}(r)$ , thanks to Theorem 1.1.

**Acknowledgements** We acknowledge support from the DFG-priority program SPP1489 “Algorithmic and Experimental Methods in Algebra, Geometry, and Number Theory”.

We are grateful to the referee for some helpful comments which lead to a simpler argument for the proof of Theorem 1.1.

## References

1. N. Amend, T. Hoge, G. Röhrle, On inductively free restrictions of reflection arrangements. *J. Algebra* **418**, 197–212 (2014)
2. N. Amend, T. Hoge, G. Röhrle, Supersolvable restrictions of reflection arrangements. *J. Comb. Theory Ser. A* **127**, 336–352 (2014)
3. T. Hoge, G. Röhrle, Addition-deletion theorems for factorizations of Orlik-Solomon algebras and nice arrangements. *Eur. J. Comb.* **55**, 20–40 (2016)
4. T. Hoge, G. Röhrle, Nice reflection arrangements. *Electron. J. Comb.* **23**(2), Paper 2.9, 24 pp. (2016)
5. T. Hoge, G. Röhrle, A. Schauenburg, Inductive and recursive freeness of localizations of multiarrangements, in *Algorithmic and Experimental Methods in Algebra, Geometry, and Number Theory* (Springer, Berlin, 2017)
6. M. Jambu, L. Paris, Combinatorics of inductively factored arrangements. *Eur. J. Comb.* **16**, 267–292 (1995)
7. T. Möller, G. Röhrle, Nice restrictions of reflection arrangements. *Electron. J. Comb.* **24**(3), Paper 3.44 (2017)
8. P. Orlik, L. Solomon, Arrangements defined by unitary reflection groups. *Math. Ann.* **261**, 339–357 (1982)
9. P. Orlik, L. Solomon, H. Terao, Arrangements of hyperplanes and differential forms, in *Combinatorics and Algebra (Boulder, CO, 1983)*. Contemporary Mathematics, vol. 34 (American Mathematical Society, Providence, 1984), pp. 29–65
10. P. Orlik, H. Terao, *Arrangements of Hyperplanes* (Springer, Berlin, 1992)
11. H. Terao, Arrangements of hyperplanes and their freeness I, II. *J. Fac. Sci. Univ. Tokyo* **27**, 293–320 (1980)
12. H. Terao, Factorizations of the Orlik-Solomon algebras. *Adv. Math.* **92**, 45–53 (1992)

# One Class Genera of Lattice Chains Over Number Fields



Markus Kirschmer and Gabriele Nebe

**Abstract** We classify all one-class genera of admissible lattice chains of length at least 2 in hermitian spaces over number fields. If  $L$  is a lattice in the chain and  $\mathfrak{p}$  the prime ideal dividing the index of the lattices in the chain, then the  $\{\mathfrak{p}\}$ -arithmetic group  $\text{Aut}(L_{\{\mathfrak{p}\}})$  acts chamber transitively on the corresponding Bruhat-Tits building. So our classification provides a step forward to a complete classification of these chamber transitive groups which has been announced 1987 (without a detailed proof) by Kantor, Liebler and Tits. In fact we find all their groups over number fields and one additional building with a discrete chamber transitive group.

**Keywords** Genus of lattice • Class number • Affine buildings • Lattice chains

**Subject Classifications** 11E41, 20G30, 20G25

## 1 Introduction

Kantor et al. [11] classified discrete groups  $\Gamma$  with a type preserving chamber transitive action on the affine building  $\mathcal{B}^+$  of a simple adjoint algebraic group of relative rank  $r \geq 2$ . Such groups are very rare and hence this situation is an interesting phenomenon. Except for two cases in characteristic 2 [11, case (v)] and the exceptional group  $G_2(\mathbb{Q}_2)$  ([11, case (iii)], Sect. 5.4) the groups arise from classical groups  $U_p$  over  $\mathbb{Q}_p$  for  $p = 2, 3$ . Moreover  $\Gamma$  is a subgroup of the  $S$ -arithmetic group  $\Gamma_{\max} := \text{Aut}(L \otimes_{\mathbb{Z}} \mathbb{Z}[\frac{1}{p}])$  (so  $S = \{p\}$ ) for a suitable lattice  $L$  in some hermitian space  $(V, \Phi)$  and  $U_p = \text{U}(V_p, \Phi)$  is the completion of the unitary

---

M. Kirschmer • G. Nebe (✉)

Lehrstuhl D für Mathematik, RWTH Aachen University, 52056 Aachen, Germany  
e-mail: [markus.kirschmer@math.rwth-aachen.de](mailto:markus.kirschmer@math.rwth-aachen.de); [nebe@math.rwth-aachen.de](mailto:nebe@math.rwth-aachen.de)

group  $U(V, \Phi)$  (see Remark 2.3). This paper uses the classification of one- and two-class genera of hermitian lattices in [16] to obtain these  $S$ -arithmetic groups  $\Gamma_{max}$ .

Instead of the thick building  $\mathcal{B}^+$  we start with the affine building  $\mathcal{B}$  of admissible lattice chains as defined in [1]. The points in the building  $\mathcal{B}$  correspond to homothety classes of certain  $\mathbb{Z}_p$ -lattices in  $V_p$ . The lattices form a simplex in  $\mathcal{B}$ , if and only if representatives in these classes can be chosen to form an admissible chain of lattices in  $V_p$ . In particular the maximal simplices of  $\mathcal{B}$  (the so called chambers) correspond to the fine admissible lattice chains in  $V_p$  (for the thick building  $\mathcal{B}^+$  one might have to apply the oriflamme construction as explained in Remark 4.6).

Any fine admissible lattice chain  $\mathcal{L}_p$  in  $V_p$  arises as the completion of a lattice chain  $\mathcal{L}'$  in  $(V, \Phi)$ . After rescaling and applying the reduction operators from Sect. 2.3 we obtain a fine  $p$ -admissible lattice chain  $\mathcal{L} = (L_0, \dots, L_r)$  in  $(V, \Phi)$  (see Definition 3.4) such that  $\text{Aut}(\mathcal{L}) \supseteq \text{Aut}(\mathcal{L}')$  and such that the completion of  $\mathcal{L}$  at  $p$  is  $\mathcal{L}_p$ . The  $S$ -arithmetic group

$$\text{Aut}(L_0 \otimes \mathbb{Z}[\frac{1}{p}]) = \text{Aut}(L_i \otimes \mathbb{Z}[\frac{1}{p}]) =: \text{Aut}(\mathcal{L} \otimes \mathbb{Z}[\frac{1}{p}])$$

contains  $\text{Aut}(\mathcal{L}' \otimes \mathbb{Z}[\frac{1}{p}])$ . Therefore we call this group closed.

The closed  $\{p\}$ -arithmetic group  $\text{Aut}(L_0 \otimes \mathbb{Z}[\frac{1}{p}])$  acts chamber transitively on  $\mathcal{B}$ , if the lattice  $L_0$  represents a genus of class number one and  $\text{Aut}(L_0)$  acts transitively on the fine flags of (isotropic) subspaces in the hermitian space  $\overline{L_0}$  (see Theorem 4.4). If we only impose chamber transitivity on the thick building  $\mathcal{B}^+$ , then we also have to take two-class genera of lattices  $L_0$  into account. To obtain a complete classification of all chamber transitive actions of closed  $S$ -arithmetic groups on the thick building  $\mathcal{B}^+$  using this strategy there are two ingredients missing:

- (a) By Theorem 4.9 we need the still unknown classification of proper special genera of lattices  $L_0$  with class number one (see also Proposition 4.5 and [27, Proposition 1]).
- (b) We should also include the skew hermitian forms over quaternion algebras for which a classification of one-class genera is still unknown.

Already taking only the one-class genera of lattices  $L_0$  we find all the groups from [11] and one additional case (described in Proposition 5.3 (1)). Hence our computations correct an omission in the classification of [11]. A list of the corresponding buildings and groups  $U_p$  is given in Sect. 6.

## 2 Lattices in Hermitian Spaces

Let  $K$  be a number field. Further, let  $E/K$  be a field extension of degree at most 2 or let  $E$  be a quaternion skewfield over  $K$ . The canonical involution of  $E/K$  will be denoted by  $\sigma: E \rightarrow E$ . In particular,  $K$  is the fixed field of  $\sigma$  and hence the involution  $\sigma$  is the identity if and only if  $K = E$ . A hermitian space over  $E$  is

a finitely generated (left) vector space  $V$  over  $E$  equipped with a non-degenerate sesquilinear form  $\Phi: V \times V \rightarrow E$  such that

- $\Phi(x + x', y) = \Phi(x, y) + \Phi(x', y)$  for all  $x, x', y \in V$ .
- $\Phi(\alpha x, \beta y) = \alpha \Phi(x, y) \sigma(\beta)$  for all  $x, y \in V$  and  $\alpha, \beta \in E$ .
- $\Phi(y, x) = \sigma(\Phi(x, y))$  for all  $x, y \in V$ .

The unitary group  $U(V, \Phi)$  of  $\Phi$  is the group of all  $E$ -linear endomorphisms of  $V$  that preserve the hermitian form  $\Phi$ . The *special unitary group* is defined as

$$SU(V, \Phi) := \{g \in U(V, \Phi) \mid \det(g) = 1\}$$

if  $E$  is commutative and  $SU(V, \Phi) := U(V, \Phi)$  if  $E$  is a quaternion algebra.

We denote by  $\mathbb{Z}_K$  the ring of integers of the field  $K$  and we fix some maximal order  $\mathcal{M}$  in  $E$ . Further, let  $d$  be the dimension of  $V$  over  $E$ .

**Definition 2.1** An  $\mathcal{M}$ -lattice in  $V$  is a finitely generated  $\mathcal{M}$ -submodule of  $V$  that contains an  $E$ -basis of  $V$ . If  $L$  is an  $\mathcal{M}$ -lattice in  $V$  then its *automorphism group* is

$$\text{Aut}(L) := \{g \in U(V, \Phi) \mid Lg = L\}.$$

### 2.1 Completion of Lattices and Groups

Let  $\mathfrak{P}$  be a maximal two sided ideal of  $\mathcal{M}$  and let  $\mathfrak{p} = \mathfrak{P} \cap K$ . The completion  $U_{\mathfrak{p}} := U(V \otimes_K K_{\mathfrak{p}}, \Phi)$  is an algebraic group over the  $\mathfrak{p}$ -adic completion  $K_{\mathfrak{p}}$  of  $K$ .

Let  $L \leq V$  be some  $\mathcal{M}$ -lattice in  $V$ . We define the  $\mathfrak{p}$ -adic completion of  $L$  as  $L_{\mathfrak{p}} := L \otimes_{\mathbb{Z}_K} \mathbb{Z}_{K_{\mathfrak{p}}}$  and we let

$$L(\mathfrak{p}) := \{X \leq V \mid X_{\mathfrak{q}} = L_{\mathfrak{q}} \text{ for all prime ideals } \mathfrak{q} \neq \mathfrak{p}\},$$

be the set of all  $\mathcal{M}$ -lattices in  $V$  whose  $\mathfrak{q}$ -adic completion coincides with the one of  $L$  for all prime ideals  $\mathfrak{q} \neq \mathfrak{p}$ .

*Remark 2.2* By the local global principle, given a lattice  $X$  in  $V_{\mathfrak{p}}$ , there is a unique lattice  $M \in L(\mathfrak{p})$  with  $M_{\mathfrak{p}} = X$ .

To describe the groups  $U_{\mathfrak{p}}$  in the respective cases, we need some notation: Let  $R$  be one of  $E, K, \mathbb{Z}_K, \mathcal{M}$  or a suitable completion. A hermitian module  $\mathbb{H}(R)$  with  $R$ -basis  $(e, f)$  satisfying  $\Phi(e, f) = 1, \Phi(e, e) = \Phi(f, f) = 0$  is called a *hyperbolic plane*. By Kneser [18, Theorem (2.22)] any hermitian space over  $E$  is either anisotropic (i.e.  $\Phi(x, x) \neq 0$  for all  $x \neq 0$ ) or it has a hyperbolic plane as an orthogonal direct summand.

*Remark 2.3* In our situation the following cases are possible:

- $E = K$ : Then  $(V \otimes_K K_{\mathfrak{p}}, \Phi)$  is a quadratic space and hence isometric to  $\mathbb{H}(K_{\mathfrak{p}})^r \perp (V_0, \Phi_0)$  with  $(V_0, \Phi_0)$  anisotropic. The rank of  $U_{\mathfrak{p}}$  is  $r$ . The group that acts type

preservingly on the thick Bruhat-Tits building  $\mathcal{B}^+$  defined in Sect. 4.3 is

$$U_p^+ := \{g \in U_p \mid \det(g) = 1, \theta(g) \in K^2\}$$

the subgroup of the special orthogonal group with trivial spinor norm  $\theta$ .

- $\mathfrak{A} \neq \sigma(\mathfrak{A})$ . Then  $E \otimes_K K_p \cong K_p \oplus K_p$  where the involution interchanges the two components and  $U_p \cong GL_d(K_p)$  has rank  $r = d - 1$ . As  $\mathfrak{A}$  is assumed to be a maximal 2-sided ideal of  $\mathcal{M}$ , the case that  $E$  is a quaternion algebra is not possible here. Here we let

$$U_p^+ = \{g \in U_p \mid \det(g) = 1\} = SL_d(K_p).$$

- $[E : K] = 4$  and  $\mathfrak{A} = \mathfrak{p}\mathcal{M}$ . Then  $E_p \cong K_p^{2 \times 2}$  and for  $x \in E_p$ ,  $\sigma(x)$  is simply the adjugate of  $x$  as  $\sigma(x)x \in K$ . Let  $e^2 = e \in E_p$  such that  $\sigma(e) = 1 - e$ . Then  $V_p = eV_p \oplus (1 - e)V_p$ . The hermitian form  $\Phi$  gives rise to a skew-symmetric form

$$\begin{aligned} \Psi: eV_p \times eV_p &\rightarrow eE_p(1 - e) \cong K_p, \\ (ex, ey) &\mapsto \Phi(ex, ey) = e\Phi(x, y)(1 - e). \end{aligned}$$

From  $E_p = E_p e E_p$  we conclude that  $V_p = E_p e E_p V$ . Hence we can recover the form  $\Phi$  from  $\Psi$  and thus  $U_p \cong U(eV, \Psi) \cong Sp_{2d}(K_p)$  has rank  $r = d$ . Here the full group  $U_p$  acts type preservingly on  $\mathcal{B}^+$  and we put  $U_p^+ := U_p$ .

- In the remaining cases  $E \otimes K_p = E_{\mathfrak{A}_p}$  is a skewfield, which is ramified over  $K_p$  if and only if  $\mathfrak{A}^2 = \mathfrak{p}\mathcal{M}$ . In all cases  $U_p$  is isomorphic to a unitary group over  $E_{\mathfrak{A}_p}$ . Hence it admits a decomposition  $\mathbb{H}(E_{\mathfrak{A}_p})^r \perp (V_0, \Phi_0)$  with  $(V_0, \Phi_0)$  anisotropic where  $r$  is the rank of  $U_p$ . If  $E_p$  is commutative, we define

$$U_p^+ := \{g \in U_p \mid \det(g) = 1\} = SU_p$$

and put  $U_p^+ = SU_p := U_p$  in the non-commutative case.

## 2.2 The Genus of a Lattice

To shorten notation, we introduce the adelic ring  $A = A(K) = \prod_v K_v$  where  $v$  runs over the set of all places of  $K$ . We denote the adelic unitary group of the  $A \otimes_K E$ -module  $V_A = A \otimes_K V$  by  $U(V_A, \Phi)$ . The normal subgroup

$$U^+(V_A, \Phi) := \{(g_p)_p \in U(V_A, \Phi) \mid g_p \in U_p^+\} \leq U(V_A, \Phi)$$

is called the special adelic unitary group.

The adelic unitary group acts on the set of all  $\mathcal{M}$ -lattices in  $V$  by letting  $Lg = L'$  where  $L'$  is the unique lattice in  $V$  such that its  $\mathfrak{p}$ -adic completion  $(L')_{\mathfrak{p}} = L_{\mathfrak{p}}g_{\mathfrak{p}}$  for all maximal ideals  $\mathfrak{p}$  of  $\mathbb{Z}_K$ .

**Definition 2.4** Let  $L$  be an  $\mathcal{M}$ -lattice in  $V$ . Then

$$\text{genus}(L) := \{Lg \mid g \in U(V_A, \Phi)\}$$

is called the *genus* of  $L$ .

Two lattices  $L$  and  $M$  are said to be *isometric* (respectively *properly isometric*), if  $L = Mg$  for some  $g \in U(V, \Phi)$  (resp.  $g \in \text{SU}(V, \Phi)$ ).

Two lattices  $L$  and  $M$  are said to be in the same *proper special genus*, if there exist  $g \in \text{SU}(V, \Phi)$  and  $h \in U^+(V_A, \Phi)$  such that  $Lgh = M$ . The proper special genus of  $L$  will be denoted by  $\text{genus}^+(L)$ .

Let  $L$  be an  $\mathcal{M}$ -lattice in  $V$ . It is well known that  $\text{genus}(L)$  is a finite union of isometry classes, c.f. [4, Theorem 5.1]. The number of isometry classes in  $\text{genus}(L)$  is called the class number  $h(L)$  of (the genus of)  $L$ . Similarly the proper special genus is a finite union of proper isometry classes, the proper class number will be denoted by  $h^+(L)$ .

### 2.3 Normalised Genera

**Definition 2.5** Let  $L$  be an  $\mathcal{M}$ -lattice in  $V$ . Then

$$L^{\#} = \{x \in V \mid \Phi(x, L) \subseteq \mathcal{M}\}$$

is called the *dual* lattice of  $L$ . If  $\mathfrak{p}$  is a maximal ideal of  $\mathbb{Z}_K$ , then the unique  $\mathcal{M}$ -lattice  $X \in L(\mathfrak{p})$  such that  $X_{\mathfrak{p}} = L^{\#}_{\mathfrak{p}}$  is called the *partial dual* of  $L$  at  $\mathfrak{p}$ . It will be denoted by  $L^{\#, \mathfrak{p}}$ .

**Definition 2.6** Let  $L$  be an  $\mathcal{M}$ -lattice in  $V$ . Further, let  $\mathfrak{A}$  be a maximal two sided ideal of  $\mathcal{M}$  and set  $\mathfrak{p} = \mathfrak{A} \cap K$ . If  $E_{\mathfrak{p}} \cong K_{\mathfrak{p}} \oplus K_{\mathfrak{p}}$  then  $L_{\mathfrak{p}}$  is called *square-free* if  $L_{\mathfrak{p}} = L^{\#}_{\mathfrak{p}}$ . In all other cases,  $L_{\mathfrak{p}}$  is called square-free if  $\mathfrak{A}L^{\#}_{\mathfrak{p}} \subseteq L_{\mathfrak{p}} \subseteq L^{\#}_{\mathfrak{p}}$ . The lattice  $L$  is called square-free if  $L_{\mathfrak{p}}$  is square-free for all maximal ideals  $\mathfrak{p}$  of  $\mathbb{Z}_K$ .

Given a maximal two sided ideal  $\mathfrak{A}$  of  $\mathcal{M}$ , we define an operator  $\rho_{\mathfrak{A}}$  on the set of all  $\mathcal{M}$ -lattices as follows:

$$\rho_{\mathfrak{A}}(L) = \begin{cases} L + (\mathfrak{A}^{-1}L \cap L^{\#}) & \text{if } \mathfrak{A} \neq \sigma(\mathfrak{A}), \\ L + (\mathfrak{A}^{-1}L \cap \mathfrak{A}L^{\#}) & \text{otherwise.} \end{cases}$$

The operators generalise the maps defined by Gerstein in [7] for quadratic spaces. They are similar in nature to the *p-mappings* introduced by Watson in [32]. The maps satisfy the following properties:

*Remark 2.7* Let  $L$  be an  $\mathcal{M}$ -lattice in  $V$ . Let  $\mathfrak{P}$  be a maximal two sided ideal of  $\mathcal{M}$  and set  $\mathfrak{p} = \mathfrak{P} \cap \mathbb{Z}_K$ .

1.  $\rho_{\mathfrak{P}}(L) \in L(\mathfrak{p})$ .
2. If  $L_{\mathfrak{p}}$  is integral, then  $(\rho_{\mathfrak{P}}(L))_{\mathfrak{p}} = L_{\mathfrak{p}} \iff L_{\mathfrak{p}}$  is square-free.
3. If  $\Omega$  is a maximal two sided ideal of  $\mathcal{M}$ , then  $\rho_{\mathfrak{P}} \circ \rho_{\Omega} = \rho_{\Omega} \circ \rho_{\mathfrak{P}}$ .
4. If  $L$  is integral, there exist a sequence of not necessarily distinct maximal two sided ideals  $\mathfrak{P}_1, \dots, \mathfrak{P}_s$  of  $\mathcal{M}$  such that

$$L' := (\rho_{\mathfrak{P}_1} \circ \dots \circ \rho_{\mathfrak{P}_s})(L)$$

is square-free. Moreover, the genus of  $L'$  is uniquely determined by the genus of  $L$ .

**Proposition 2.8** *Let  $L$  be an  $\mathcal{M}$ -lattice in  $V$  and let  $\mathfrak{P}$  be a maximal two sided ideal of  $\mathcal{M}$ . Then the class number of  $\rho_{\mathfrak{P}}(L)$  is at most the class number of  $L$ .*

*Proof* The definition of  $\rho_{\mathfrak{P}}(L)$  only involves taking sums and intersections of multiples of  $L$  and its dual. Hence  $\rho_{\mathfrak{P}}(L)g = \rho_{\mathfrak{P}}(Lg)$  for all  $g \in U(V, \Phi)$  and similar for  $g \in U(V_A, \Phi)$ . In particular,  $\rho_{\mathfrak{P}}$  maps lattices in the same genus (isometry class) to ones in the same genus (isometry class). The result follows.  $\square$

**Definition 2.9** Let  $\mathfrak{A}$  be a two sided  $\mathcal{M}$ -ideal. An  $\mathcal{M}$ -lattice  $L$  is called  $\mathfrak{A}$ -maximal, if  $\Phi(x, x) \in \mathfrak{A}$  for all  $x \in L$  and no proper overlattice of  $L$  has that property. Similarly, one defines maximal lattices in  $V_{\mathfrak{p}}$  for a maximal ideal  $\mathfrak{p}$  of  $\mathbb{Z}_K$ .

**Definition 2.10** Let  $\mathfrak{P}$  be a maximal two sided ideal of  $\mathcal{M}$  and set  $\mathfrak{p} = \mathfrak{P} \cap K$ . We say that an  $\mathcal{M}$ -lattice  $L$  is  $\mathfrak{p}$ -normalised if  $L$  satisfies the following conditions:

- $L$  is square-free.
- If  $E = K$  then  $L_{\mathfrak{p}} \cong \mathbb{H}(\mathbb{Z}_{K_{\mathfrak{p}}})^r \perp M_0$  where  $M_0 = \rho_{\mathfrak{P}}^{\infty}(M)$  and  $M$  denotes a  $2\mathbb{Z}_{K_{\mathfrak{p}}}$ -maximal lattice in an anisotropic quadratic space over  $K_{\mathfrak{p}}$ .
- If  $E_{\mathfrak{p}}/K_{\mathfrak{p}}$  is a quadratic field extension with different  $\mathcal{D}(E_{\mathfrak{p}}/K_{\mathfrak{p}})$ , then  $L_{\mathfrak{p}} \cong \mathbb{H}(\mathcal{M}_{\mathfrak{p}})^r \perp M_0$  where  $M_0 = \rho_{\mathfrak{P}}^{\infty}(M)$  for some  $\mathcal{D}(E_{\mathfrak{p}}/K_{\mathfrak{p}})$ -maximal lattice  $M$  in an anisotropic hermitian space over  $E_{\mathfrak{p}}$ .
- If  $[E : K] = 4$ , then  $L_{\mathfrak{p}} = L_{\mathfrak{p}}^{\#}$ .

Here  $\rho_{\mathfrak{P}}^{\infty}(M)$  denotes the image of  $M$  under repeated application of  $\rho_{\mathfrak{P}}$  until this process becomes stable.

*Remark 2.11* Let  $\mathfrak{P}, \mathfrak{p}$  and  $L$  be as in Definition 2.10. Then the isometry class of  $L_{\mathfrak{p}}$  is uniquely determined by  $(V_{\mathfrak{p}}, \Phi)$ .

*Proof* There is nothing to show if  $[E : K] = 4$ . Suppose now  $E = K$ . The space  $KM_0$  is a maximal anisotropic subspace of  $(V_{\mathfrak{p}}, \Phi)$ . By Witt's theorem [25, Theorem 42:17] its isometry type is uniquely determined by  $(V_{\mathfrak{p}}, \Phi)$ . Further,  $M_0$  is the unique  $2\mathbb{Z}_{K_{\mathfrak{p}}}$ -maximal  $\mathbb{Z}_{K_{\mathfrak{p}}}$ -lattice in  $KM_0$ , see [25, Theorem 91:1]. Hence the isometry type of  $\rho_{\mathfrak{P}}^{\infty}(M_0)$  depends only on  $(V_{\mathfrak{p}}, \Phi)$ . The case  $[E : K] = 2$  is proved similarly.  $\square$



### 3 Genera of Lattice Chains

**Definition 3.1** Let  $\mathcal{L} := (L_1, \dots, L_m)$  and  $\mathcal{L}' := (L'_1, \dots, L'_m)$  be two  $m$ -tuples of  $\mathcal{M}$ -lattices in  $V$ . Then  $\mathcal{L}$  and  $\mathcal{L}'$  are *isometric*, if there is some  $g \in U(V, \Phi)$  such that  $L_i g = L'_i$  for all  $i = 1, \dots, m$ . They are in the same *genus* if there is such an element  $g \in U(V_A, \Phi)$ . Let

$$[\mathcal{L}] := \{\mathcal{L}' \mid \mathcal{L}' \text{ is isometric to } \mathcal{L}\}$$

and

$$\text{genus}(\mathcal{L}) := \{\mathcal{L}' \mid \mathcal{L}' \text{ and } \mathcal{L} \text{ are in the same genus}\}$$

denote the isometry class and the genus of  $\mathcal{L}$ , respectively. The *automorphism group* of  $\mathcal{L}$  is the stabiliser of  $\mathcal{L}$  in  $U(V, \Phi)$ , i.e.

$$\text{Aut}(\mathcal{L}) = \bigcap_{i=1}^m \text{Aut}(L_i).$$

It is well known [4, Theorem 5.1] that any genus of a single lattice contains only finitely many isometry classes. This is also true for finite tuples of lattices in  $V$ :

**Lemma 3.2** *Let  $\mathcal{L} = (L_1, \dots, L_m)$  be an  $m$ -tuple of  $\mathcal{M}$ -lattices in  $V$ . Then  $\text{genus}(\mathcal{L})$  is the disjoint union of finitely many isometry classes. The number of isometry classes in  $\text{genus}(\mathcal{L})$  is called the class number of  $\mathcal{L}$ .*

*Proof* The case  $m = 1$  is the classical case. So assume that  $m \geq 2$  and let  $\text{genus}(L_1) := [M_1] \uplus \dots \uplus [M_h]$ , with  $M_i = L_1 g_i$  for suitable  $g_i \in U(V_A, \Phi)$ . We decompose  $\text{genus}(\mathcal{L}) = \mathcal{G}_1 \uplus \dots \uplus \mathcal{G}_h$  where

$$\mathcal{G}_i := \{(L'_1, \dots, L'_m) \in \text{genus}(\mathcal{L}) \mid L'_1 \cong M_i\}.$$

It is clearly enough to show that each  $\mathcal{G}_i$  is the union of finitely many isometry classes. By construction, any isometry class in  $\mathcal{G}_i$  contains a representative of the form  $(M_i, L'_2, \dots, L'_m)$  for some lattices  $L'_j$  in the genus of  $L_j$ . As all the  $L_j$  are lattices in the same vector space  $V$ , there are  $a, b \in \mathbb{Z}_K$  such that

$$bL_1 \subseteq L_j \subseteq \frac{1}{a}L_1 \text{ for all } 1 \leq j \leq m.$$

As  $(M_i, L'_2, \dots, L'_m) = \mathcal{L}g$  for some  $g \in U(V_A, \Phi)$  we also have

$$bM_i \subseteq L'_j \subseteq \frac{1}{a}M_i \text{ for all } 2 \leq j \leq m.$$

So there are only finitely many possibilities for such lattices  $L'_j$ . Hence the set of all  $m$ -tuples  $(M_i, L'_2, \dots, L'_m) \in \text{genus}(\mathcal{L})$  is finite and so is the class number.  $\square$

*Remark 3.3* If  $\mathcal{L}' \subseteq \mathcal{L}$  then the class number of  $\mathcal{L}'$  is at most the class number of  $\mathcal{L}$ .

### 3.1 Admissible Lattice Chains

**Definition 3.4** Let  $\mathfrak{A}$  be a maximal 2-sided ideal of  $\mathcal{M}$  and  $\mathfrak{p} := K \cap \mathfrak{A}$ . A lattice chain

$$\mathcal{L} := \{L_0 \supset L_1 \supset \dots \supset L_{m-1} \supset L_m\}$$

is called *admissible* for  $\mathfrak{A}$ , if

1.  $L_0 \subseteq L_0^{\#, \mathfrak{p}}$ ,
2.  $\mathfrak{A}L_0 \subset L_m$ ,
3.  $\mathfrak{A}L_m^{\#, \mathfrak{p}} \subseteq L_m$  if  $\mathfrak{A} = \sigma(\mathfrak{A})$ .

We call a  $\mathfrak{A}$ -admissible chain *fine*, if  $L_0$  is normalised for  $\mathfrak{p}$  in the sense of Definition 2.10,  $L_i$  is a maximal sublattice of  $L_{i-1}$  for all  $i = 1, \dots, m$  and either

- (a)  $\mathfrak{A} = \sigma(\mathfrak{A})$  and  $L_m/\mathfrak{A}L_m^{\#, \mathfrak{p}}$  is an anisotropic space over  $\mathcal{M}/\mathfrak{A}$
- (b)  $\mathfrak{A} \neq \sigma(\mathfrak{A})$  and  $\mathfrak{A}L_0$  is a maximal sublattice of  $L_m$ .

*Remark 3.5* In the case that  $\mathfrak{A} \neq \sigma(\mathfrak{A})$  the *length*  $m$  of a fine admissible lattice chain is just  $m = r = \dim_E(V) - 1$ . Also if  $\mathfrak{A} = \sigma(\mathfrak{A})$ , then  $m = r$ , where  $r$  is the rank of the  $\mathfrak{p}$ -adic group defined in Remark 2.3.

Note that any admissible chain  $\mathcal{L}$  contains a unique maximal integral lattice which we will always denote by  $L_0$ .

*Remark 3.6* Let  $\mathcal{L} = (L_0, \dots, L_r)$  be a fine admissible lattice chain for  $\mathfrak{A}$ .

- (a) If  $\mathfrak{A} = \sigma(\mathfrak{A})$  then  $\overline{L_0} := L_0/\mathfrak{A}L_0^{\#, \mathfrak{p}}$  is a hermitian space over  $\mathcal{M}/\mathfrak{A}$  and the spaces  $V_j := \mathfrak{A}L_j^{\#, \mathfrak{p}}/\mathfrak{A}L_0^{\#, \mathfrak{p}}$  ( $j = 1, \dots, r$ ) define a maximal chain of isotropic subspaces of this hermitian space. We call the chain  $(V_1, \dots, V_{r-1})$  *truncated*.
- (b) If  $\mathfrak{A} \neq \sigma(\mathfrak{A})$  then  $\overline{L_0} := L_0/\mathfrak{A}L_0$  is a vector space over  $\mathcal{M}/\mathfrak{A}$  and the spaces  $V_j := L_j/\mathfrak{A}L_0$  ( $j = r, \dots, 1$ ) form a maximal chain of subspaces. Here we call the chain  $(V_{r-1}, \dots, V_1)$  *truncated*.

For the different hermitian spaces  $\overline{L_0}$ , the number of such chains of isotropic subspaces can be found by recursively applying the formulas in [29, Exercises 8.1, 10.4, 11.3].

**Lemma 3.7** *The fine admissible lattice chain  $\mathcal{L}$  represents a one-class genus of lattice chains if and only if  $L_0$  represents a one-class genus of lattices and  $\text{Aut}(L_0)$  is transitive on the maximal chains of (isotropic) subspaces of  $\overline{L_0}$ .*

*Proof* If  $\mathcal{L}$  has class number one, so has any lattice in the chain  $\mathcal{L}$ . Suppose now  $L_0$  has class number one. Let  $\mathcal{L}'$  be any other lattice chain in the genus of  $\mathcal{L}$ . We have to show that  $\mathcal{L}$  and  $\mathcal{L}'$  are isometric. To that end, let  $L'_0$  be the unique maximal integral lattice in  $\mathcal{L}'$ . Then  $L_0$  and  $L'_0$  are isometric, as they are in the same genus. So without loss of generality,  $L_0 = L'_0$ . Then  $\mathcal{L}$  and  $\mathcal{L}'$  correspond to unique maximal chains of (isotropic) subspaces of  $\overline{L_0}$ . Since  $\text{Aut}(L_0)$  acts transitively on these chains of subspaces, it yields an isometry from  $\mathcal{L}$  to  $\mathcal{L}'$ .  $\square$

## 4 Chamber Transitive Actions on Affine Buildings

Kantor et al. [11] classified discrete groups acting chamber transitively and type preservingly on the affine building of a simple adjoint algebraic group of relative rank  $\geq 2$  over a locally compact local field. Such groups are very rare and hence this situation is an interesting phenomenon, further studied in [9, 10, 12, 19, 22], and [21] (and many more papers by these authors) where explicit constructions of the groups are given. One major disadvantage of the existing literature is that the proof in [11] is very sketchy, essentially the authors limit the possibilities that need to be checked to a finite number.

From the classification of the one-class genera of admissible fine lattice chains in Sect. 5, we obtain a number theoretic construction of the groups in [11] over fields of characteristic 0. It turns out that we find essentially all these groups and that our construction allows to find one more case: The building of  $U_5(\mathbb{Q}_3(\sqrt{-3}))$  of type  $C - BC_2$ , see Proposition 5.3 (1), which, to our best knowledge, has not appeared in the literature before.

### 4.1 *S*-Arithmetic Groups

We assume that  $(V, \Phi)$  is a totally positive definite hermitian space, i.e.  $K$  is totally real and  $\Phi(x, x) \in K$  is totally positive for all non-zero  $x \in V$ .

Let  $S = \{\mathfrak{p}_1, \dots, \mathfrak{p}_m\}$  be a finite set of prime ideals of  $\mathbb{Z}_K$ . For a prime ideal  $\mathfrak{p}$  we denote by  $\nu_{\mathfrak{p}}$  the  $\mathfrak{p}$ -adic valuation of  $K$ . Then the ring of  $S$ -integers in  $K$  is

$$\mathbb{Z}_S := \{a \in K \mid \nu_{\mathfrak{q}}(a) \geq 0 \text{ for all prime ideals } \mathfrak{q} \notin S\}.$$

Let  $L$  be some  $\mathcal{M}$ -lattice in  $(V, \Phi)$  and put  $L_S := L \otimes_{\mathbb{Z}_K} \mathbb{Z}_S$ . Then the group

$$\text{Aut}(L_S) := \{g \in U(V, \Phi) \mid L_S g = L_S\}$$

is an  $S$ -arithmetic subgroup of  $U(V, \Phi)$ .

*Remark 4.1* For any prime ideal  $\mathfrak{p}$ , the group  $U(V, \Phi)$  (being a subgroup of  $U_{\mathfrak{p}}$ ) acts on the Bruhat-Tits building  $\mathcal{B}$  of the group  $U_{\mathfrak{p}}$  defined in Remark 2.3. Assume that the rank of  $U_{\mathfrak{p}}$  is at least 1. The action of the subgroup  $\text{Aut}(L_S)$  is discrete and cocompact on  $\mathcal{B}$ , if and only if  $\mathfrak{p} \in S$  and  $(V_{\mathfrak{q}}, \Phi)$  is anisotropic for all  $\mathfrak{p} \neq \mathfrak{q} \in S$ .

### 4.2 The Action on the Building of $U_{\mathfrak{p}}$

In the following we fix a prime ideal  $\mathfrak{p}$  and assume that  $S = \{\mathfrak{p}\}$ .

A lattice class model for the affine building  $\mathcal{B}$  has been described in [1]. Note that [1] imposes the assumption that the residue characteristic of  $K_{\mathfrak{p}}$  is  $p \neq 2$ . This is only necessary to obtain a proof of the building axioms that is independent from Bruhat-Tits theory. For  $p = 2$ , the dissertation [6] contains the analogous description of the Bruhat-Tits building for orthogonal groups. For all residue characteristics, the chambers in  $\mathcal{B}$  correspond to certain fine lattice chains in the natural  $U_{\mathfrak{p}}$ -module  $W_{\mathfrak{p}}$ .

Let  $L$  be a fixed  $\mathfrak{p}$ -normalised lattice in  $V$  and put  $V_{\mathfrak{p}} := V \otimes_K K_{\mathfrak{p}}$ .

In the case that  $E \otimes_K K_{\mathfrak{p}}$  is a skewfield, we decompose the completion

$$L_{\mathfrak{p}} = \mathbb{H}(\mathcal{M}_{\mathfrak{p}})^r \perp M_0 = \bigoplus_{i=1}^r \langle e_i, f_i \rangle_{\mathcal{M}_{\mathfrak{p}}} \perp M_0$$

as in Definition 2.10. Then  $V_{\mathfrak{p}} = V_0 \perp \langle e_1, \dots, e_r, f_1, \dots, f_r \rangle_{K_{\mathfrak{p}}}$  where  $V_0 = K_{\mathfrak{p}}M_0$  is anisotropic. Then the standard chamber corresponding to  $L$  and the choice of this hyperbolic basis is represented by the admissible fine lattice chain

$$\mathcal{L} = (L = L_0, L_1, \dots, L_r)$$

where  $L_j \in L(\mathfrak{p})$  is the unique lattice in  $V$  such that

$$(L_j)_{\mathfrak{p}} = \bigoplus_{i=1}^j \langle \pi e_i, f_i \rangle_{\mathcal{M}_{\mathfrak{p}}} \perp \bigoplus_{i=j+1}^r \langle e_i, f_i \rangle_{\mathcal{M}_{\mathfrak{p}}} \perp M_0.$$

Now assume that  $E \otimes_K K_{\mathfrak{p}} \cong K_{\mathfrak{p}}^{2 \times 2}$  and  $W_{\mathfrak{p}} = eV_{\mathfrak{p}}$  for some primitive idempotent  $e$  such that  $\sigma(e) = 1 - e$  as in Remark 2.3. Then  $W_{\mathfrak{p}}$  carries a symplectic form  $\Psi$  and the lattice  $L_{\mathfrak{p}}e$  has a symplectic basis  $(e_1, f_1, \dots, e_r, f_r)$ , i.e.

$$L_{\mathfrak{p}}e = \bigoplus_{i=1}^r \langle e_i, f_i \rangle_{\mathbb{Z}_{K_{\mathfrak{p}}}}$$

with  $\Psi(e_i, f_i) = 1$ . The standard chamber corresponding to  $L$  and the choice of this symplectic basis is represented by the admissible fine lattice chain

$$\mathcal{L} = (L = L_0, L_1, \dots, L_r)$$

where  $L_j \in L(\mathfrak{p})$  is the unique lattice in  $V$  such that

$$(L_j)_{\mathfrak{p}} = \bigoplus_{i=1}^j \langle \pi e_i, f_i \rangle_{\mathcal{M}_{\mathfrak{p}}} \perp \bigoplus_{i=j+1}^r \langle e_i, f_i \rangle_{\mathcal{M}_{\mathfrak{p}}}.$$

In the last and most tricky case  $E \otimes_K K_{\mathfrak{p}} \cong K_{\mathfrak{p}} \oplus K_{\mathfrak{p}}$ . Then  $W_{\mathfrak{p}} = V_{\mathfrak{p}} e_{\mathfrak{P}}$  for any of the two maximal ideals  $\mathfrak{P}$  of  $\mathcal{M}$  that contain  $\mathfrak{p}$ ,  $U_{\mathfrak{p}} \supseteq \text{SL}(W_{\mathfrak{p}})$  and  $M_{\mathfrak{p}} := L_{\mathfrak{p}} e_{\mathfrak{P}}$  is a lattice in  $W_{\mathfrak{p}}$ . To define the standard chamber fix some  $\mathbb{Z}_{K_{\mathfrak{p}}}$ -basis  $(e_1, \dots, e_r)$  of  $M_{\mathfrak{p}}$ . Then the fine admissible lattice chain

$$\mathcal{L} = (L = L_0, L_1, \dots, L_r)$$

where  $L_j$  is the unique lattice in  $V$  such that

- $(L_j)_{\Omega} = L_{\Omega}$  for all prime ideals  $\Omega \neq \mathfrak{P}$  of  $\mathcal{M}$
- $(L_j)_{\mathfrak{P}} = \bigoplus_{i=1}^j \langle \pi e_i \rangle_{\mathcal{M}_{\mathfrak{P}}} \oplus \bigoplus_{i=j+1}^r \langle e_i \rangle_{\mathcal{M}_{\mathfrak{P}}}.$

**Lemma 4.2** *Assume that  $\mathfrak{P} \neq \sigma(\mathfrak{P})$ , so  $E \otimes_K K_{\mathfrak{p}} \cong K_{\mathfrak{p}} \oplus K_{\mathfrak{p}}$  and keep the notation from above. Let  $M$  be some  $\mathcal{M}$ -lattice in  $V$ . Then*

$$\{X \in M(\mathfrak{p}) \mid e_{\mathfrak{P}} X_{\mathfrak{p}} = e_{\mathfrak{P}} M_{\mathfrak{p}}\}$$

*contains a unique lattice  $Y$  with  $Y = Y^{\# \cdot \mathfrak{p}}$ .*

*Proof* As  $Y \in M(\mathfrak{p})$  it is enough to define  $Y_{\mathfrak{p}} = e_{\mathfrak{P}} M_{\mathfrak{p}} \oplus (1 - e_{\mathfrak{P}}) X_{\mathfrak{p}}$ . This  $\mathcal{M}_{\mathfrak{p}}$ -lattice is unimodular if and only if

$$(1 - e_{\mathfrak{P}}) X_{\mathfrak{p}} = \{x \in (1 - e_{\mathfrak{P}}) V \mid \Phi(e_{\mathfrak{P}} M_{\mathfrak{p}}, x) \subseteq \mathcal{M}_{\mathfrak{p}}\}.$$

The result follows. □

Thus for  $\mathfrak{P} \neq \sigma(\mathfrak{P})$  the stabiliser in the  $S$ -arithmetic group  $\text{Aut}(L_S)$  of a vertex in the building  $\mathcal{B}$  is the automorphism group of a  $\mathfrak{p}$ -unimodular lattice. Also if  $\mathfrak{P} = \sigma(\mathfrak{P})$ , any vertex in the building  $\mathcal{B}$  corresponds to a unique homothety class of lattices  $[M_{\mathfrak{p}}] = \{a M_{\mathfrak{p}} \mid a \in K_{\mathfrak{p}}^*\}$ . So by Remark 2.2 there is a unique lattice  $X \in L(\mathfrak{p})$  with  $X_{\mathfrak{p}} = M_{\mathfrak{p}}$ . Hence the stabilisers of the vertices in  $\mathcal{B}$  are exactly the automorphism groups of the respective lattices in  $V$ . In particular these are finite groups.

*Remark 4.3* As  $U_{\mathfrak{p}}$  acts transitively on the chambers of  $\mathcal{B}$ , any other chamber (i.e.  $r$ -dimensional simplex) in  $\mathcal{B}$  corresponds to some lattice chain in the genus of  $\mathcal{L} =$

$(L_0, \dots, L_r)$ . The  $(r - 1)$ -dimensional simplices are the  $U_{\mathfrak{p}}$ -orbits of the subchains  $\mathcal{L}_j := (L_i \mid i \neq j)$  of  $\mathcal{L}$  for  $j = 0, \dots, r$ . We call these simplices *panels* and  $j$  the cotype of the panel  $\mathcal{L}_j$ .

**Theorem 4.4** *Let  $\mathcal{L} = (L_0, \dots, L_r)$  be a fine admissible lattice chain for  $\mathfrak{P}$  of class number one. Put  $L := L_0$  and  $S := \{\mathfrak{p}\}$ . Then  $\text{Aut}(L_S)$  acts chamber transitively on the (weak) Bruhat-Tits building  $\mathcal{B}$  of the completion  $U_{\mathfrak{p}}$ .*

*Proof* We use the characterisation of Lemma 3.7. Let  $\mathcal{C}$  be the chamber of  $\mathcal{B}$  that corresponds to  $\mathcal{L}$  by the construction above and let  $\mathcal{D}$  be some other chamber in  $\mathcal{B}$ . Then there is some element  $g \in U_{\mathfrak{p}}$  with  $\mathcal{C}g = \mathcal{D}$ . As the genus of  $L$  consists only of one class, there is some  $h \in \text{Aut}(L_S)$  such that  $gh \in U_{\mathfrak{p}}$  stabilises the vertex  $v$  that corresponds to  $L$ . So  $gh \in \text{Stab}_{U_{\mathfrak{p}}}(L_{\mathfrak{p}})$  and  $\mathcal{D}h$  is some chamber in  $\mathcal{B}$  containing the vertex  $v$ . Now  $\text{Aut}(L)$  acts transitively on the set of all fine admissible lattice chains for  $\mathfrak{P}$  starting in  $L$ , so there is some  $h' \in \text{Aut}(L)$  such that  $\mathcal{D}hh' = \mathcal{C}$ . Thus the element  $hh' \in \text{Aut}(L_S)$  maps  $\mathcal{D}$  to  $\mathcal{C}$ . □

As in [27, Proposition 1] we obtain the following if and only if statement:

**Proposition 4.5** *The group  $\text{Aut}^+(L_S)$  acts chamber transitively on the (weak) Bruhat-Tits building  $\mathcal{B}$  if and only if the special class number  $h^+(\mathcal{L}) = 1$  or equivalently if  $h^+(L_0) = 1$  and  $\text{Aut}^+(L_0)$  is transitive on the maximal chains of (isotropic) subspaces of  $\overline{L_0}$ .*

For the maximal  $S$ -arithmetic group  $\text{Aut}(L_S)$  an if and only if statement is technically more involved due to the fact that  $U(V, \Phi)$  is not necessarily connected and so we do not have strong approximation for this group. Here we obtain that  $\text{Aut}(L_S)$  acts chamber transitively on  $\mathcal{B}$  if and only if  $\text{Aut}(L_0)$  acts transitively on the maximal chains of (isotropic) subspaces of  $\overline{L_0}$  (see Lemma 3.7) and all  $\mathfrak{p}$ -neighbours of  $L_0$  (i.e. all lattices  $L$  in the genus of  $L_0$  with  $L/(L \cap L_0) \cong \mathcal{M}/\mathfrak{P}$  for some maximal two-sided ideal  $\mathfrak{P}$  of  $\mathcal{M}$  over  $\mathfrak{p}$ ) are isometric to  $L_0$ .

For the orthogonal groups we can further characterise the transitivity of  $\text{Aut}(L_S)$  on  $\mathcal{B}$ : Let  $g \in U(V, \Phi)$  be some isometry of determinant  $-1$ . Then the union of the proper special genera of  $L$  and  $g(L)$  consists of exactly  $h^+(L)$  isometry classes. Let

$$\mathcal{N}_{\mathfrak{p}}^+(L) := \{Mh \mid M \text{ is an iterated } \mathfrak{p}\text{-neighbour, } h \in U^+(V, \Phi)\}.$$

Then by Benham and Hsia [2] the set  $\mathcal{N}_{\mathfrak{p}}^+(L)$  consists of  $a \leq 2$  proper special genera. The exact value of  $a$  is given by some local condition, see [2, Equation (1.1)]. In particular, the union of all isometry classes of iterated  $\mathfrak{p}$ -neighbours is the following union of proper special genera

$$\text{genus}^+(L) \cup \text{genus}^+(Lg) \cup \text{genus}^+(L') \cup \text{genus}^+(L'g).$$

where  $L'$  denotes any  $\mathfrak{p}$ -neighbour of  $L$ . The above union consists of a single isometry class, if and only if  $h^+(L) = 1$  and  $a = 1$ .

### 4.3 The Oriflamme Construction

The buildings  $\mathcal{B}$  described above are in general not thick buildings, i.e. there are panels that are only contained in exactly two chambers. Such panels are called thin. To obtain a thick building  $\mathcal{B}^+$  (with a type preserving action by the group  $U_{\mathfrak{p}}^+$  defined in Remark 2.3) we need to apply a generalisation of the oriflamme construction as described in [1, Section 8]. In particular [1, Section 8.1] gives the precise situations which panels are thin for the case that  $p \neq 2$ . Also for  $p = 2$  only the panels of cotype 0 and  $r$  can be thin. We refrain from describing the situations for  $p = 2$  in general, but refer to the individual examples below.

*Remark 4.6* Assume that  $\mathfrak{A} = \sigma(\mathfrak{A})$ .

- (a) Assume that there are only two lattices  $L_0$  and  $L'_0$  in the genus of  $L_0$  such that

$$L_1 \subseteq L_0, L'_0 \subseteq L_1^{\#,\mathfrak{p}}.$$

Then  $\mathcal{L}$  and  $\mathcal{L}' := (L'_0, L_1, \dots, L_r)$  are the only chambers in  $\mathcal{B}$  that contain the panel  $\mathcal{L}_0 = (L_1, \dots, L_r)$  and hence this panel is thin. Then we replace the vertex represented by  $L_1$  by the one represented by  $L'_0$ .

- (b) Assume that there are only two lattices  $L_r$  and  $L'_r$  in the genus of  $L_r$  such that

$$\mathfrak{A}L_{r-1}^{\#,\mathfrak{p}} \subseteq L_r, L'_r \subseteq L_{r-1}.$$

Then  $\mathcal{L}$  and  $\mathcal{L}' := (L_0, L_1, \dots, L'_r)$  are the only chambers in  $\mathcal{B}$  that contain the panel  $\mathcal{L}_r = (L_0, \dots, L_{r-1})$  and hence this panel is thin. Then we replace the vertex represented by  $L_{r-1}$  by the one represented by  $L'_r$ .

- (c) After this construction the standard chamber  $\mathcal{L}^+$  in the thick building  $\mathcal{B}^+$  is either represented by  $\mathcal{L}$ ,  $(L_0, L'_0, L_2, \dots, L_r)$ ,  $(L_0, L_1, \dots, L_{r-2}, L_r, L'_r)$ , or  $(L_0, L'_0, L_2, \dots, L_{r-2}, L_r, L'_r)$ . Note that by construction the chain  $\mathcal{L}$  can be recovered from  $\mathcal{L}^+$ , so the stabiliser of  $\mathcal{L}$  is equal to the stabiliser of all lattices in  $\mathcal{L}^+$ . Moreover every element in  $U_{\mathfrak{p}}$  mapping the chain  $\mathcal{L}$  to some other chain  $\mathcal{L}'$  maps the chamber  $\mathcal{L}^+$  to the chamber  $(\mathcal{L}')^+$ .

For more details we refer to [1, Section 8.3].

In particular by part (c) of the previous remark we find the important corollary.

**Corollary 4.7** *In the situation of Theorem 4.4 the group  $\text{Aut}(L_S)$  also acts chamber transitively (not necessarily type preservingly) on the thick building  $\mathcal{B}^+$ .*

*Remark 4.8* Also in the situation where  $\mathfrak{A} \neq \sigma(\mathfrak{A})$ , i.e.  $E_{\mathfrak{p}} = K_{\mathfrak{p}} \oplus K_{\mathfrak{p}}$ , the stabilisers of the points in the building are not the stabilisers of the lattices in the lattice chain. By Lemma 4.2 the lattices  $L_i$  ( $i = 1, \dots, r$ ) need to be replaced by the uniquely defined lattices  $Y_i \in L_i(\mathfrak{p})$ , such that  $(Y_i)_{\mathfrak{p}}$  is unimodular (as in Lemma 4.2) and  $Y_i \cap L_0 = L_i$ . We refer to this construction as a variant of the oriflamme construction in the examples below.

**Theorem 4.9** *Let  $\mathcal{L} = (L_0, \dots, L_r)$  be a fine  $\mathfrak{P}$ -admissible lattice chain for some maximal two sided ideal  $\mathfrak{P}$  of  $\mathcal{M}$  such that  $\sigma(\mathfrak{P}) = \mathfrak{P}$ . Suppose that the oriflammé construction replaces  $\mathcal{L}$  by some sequence of lattices  $\mathcal{L}^+$  which is one of*

$$\begin{aligned} \mathcal{L} &= (L_0, \dots, L_r), (L_0, L'_0, L_2, \dots, L_r), \\ &(L_0, L_1, \dots, L_{r-2}, L_r, L'_r) \text{ or } (L_0, L'_0, L_2, \dots, L_{r-2}, L_r, L'_r). \end{aligned}$$

*Then  $L_0$  and  $L'_0$  as well as  $L_r$  and  $L'_r$  are in the same genus but not in the same proper special genus. Put  $L := L_0$  and  $S := \{\mathfrak{p}\}$ . Then  $\text{Aut}^+(L_S) := \text{Aut}(L_S) \cap \text{SU}(V, \Phi)$  acts type preservingly on the thick building  $\mathcal{B}^+$ . This action is chamber transitive if and only if  $h^+(L) = 1$  and  $\text{Aut}^+(L)$  is transitive on the maximal chains (in the first two cases) respectively truncated maximal chains (in the last two cases) of isotropic subspaces of  $\overline{L}$  defined in Remark 3.6.*

*Proof* The proof that the action is chamber transitive in all cases is completely analogous to the proof of Theorem 4.4. We only need to show that  $h^+(L) = 1$ . So let  $M$  be some lattice in the same proper special genus as  $L$ . By strong approximation for  $U^+(V_A, \Phi)$  (see [17]), there is some element  $g \in U_{\mathfrak{p}}^+$  and  $h \in \text{SU}(V, \Phi)$  such that  $Mh = Lg$ . As  $\text{Aut}^+(L_S)$  is chamber transitive and type preserving, there is some  $f \in \text{Aut}^+(L_S)$  such that  $Lf = Mh$  so  $M = Lfh^{-1}$  is properly isometric to  $L$ .  $\square$

To obtain a classification of all chamber transitive discrete actions on  $\mathcal{B}^+$  we hence need a classification of all proper spinor genera with proper class number one. The thesis [16] only lists the genera of class number one and two. In some cases,  $h(L) = h^+(L)$  for every square-free lattice  $L$ , for example if:

- (a)  $E = K$ ,  $\dim(V) \geq 5$  and  $K$  has narrow class number one [25, Theorem 102.9],
- (b)  $[E : K] = 2$  and  $\dim_E(V)$  is odd [28],
- (c)  $[E : K] = 4$ .

It seems to be very unlikely that there are square-free lattices  $L$  with  $h^+(L) = 1$  and  $h(L) > 2$  that yield a chamber transitive action.

## 5 The One-Class Genera of Fine Admissible Lattice Chains

We split this section into three sections dealing with the different types of hermitian spaces ( $[E : K] = 1, 2, 4$ ). The fourth section comments on the exceptional groups.

Suppose  $\mathcal{L} = (L_0, \dots, L_r)$  is a fine  $\mathfrak{P}$ -admissible lattice chain of class number one, where  $\mathfrak{P}$  is a maximal two sided ideal of  $\mathcal{M}$ . Then  $\mathfrak{p} := \mathfrak{P} \cap \mathbb{Z}_K$  together with  $L_0$  determines the isometry class of  $\mathcal{L} := \mathcal{L}(L_0, \mathfrak{p})$ . Moreover  $L_0$  is a  $\mathfrak{p}$ -normalised lattice in  $(V, \Phi)$  of class number one and by Corollary 3.7 the finite group  $\text{Aut}(L_0)$  acts transitively on the fine chains of (isotropic) subspaces of  $\overline{L_0}$  as in Remark 3.6. The one- and two-class genera of lattices in hermitian spaces  $(V, \Phi)$  have been classified in [16]. For all such lattices  $L_0$  and all prime ideals  $\mathfrak{p}$ , for which  $L_0$  is



$\mathfrak{p}$ -normalised, we check by computer if  $\text{Aut}(L_0)$  acts transitively on the fine chains of (isotropic) subspaces of  $\overline{L}_0$ . Note that the number of such chains grows with the norm of  $\mathfrak{p}$ , so the order of  $\text{Aut}(L_0)$  gives us a bound on the possible prime ideals  $\mathfrak{p}$ . We also checked weaker conditions (similar to the ones in Theorem 4.9) that would imply a chamber transitive action on the thick building  $\mathcal{B}^+$ , i.e.  $h(L_0) \leq 2$  and transitivity only on the truncated maximal chains. The cases  $h(L_0) = 2$  never gave a transitive action on the chambers of  $\mathcal{B}^+$ .

For any non-empty subset  $T$  of  $\{1, 2, \dots, r\}$  we list the automorphism group  $G_T$  of the subchain  $(L_i)_{i \in T}$ . With our applications on the action on buildings in mind, we also give the order of

$$G_T^+ := G_T \cap U_{\mathfrak{p}}^+$$

where  $U_{\mathfrak{p}}^+$  is given in Remark 2.3. Note that we will always assume that the rank of the group  $U_{\mathfrak{p}}$  is  $r \geq 2$ .

### 5.1 Quadratic Forms

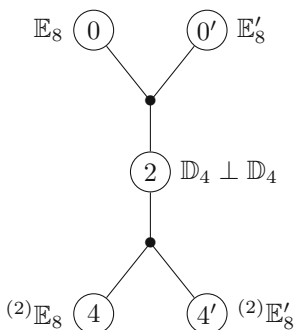
In this section suppose that  $E = K$ . We denote by  $\mathbb{A}_n, \mathbb{B}_n, \mathbb{D}_n, \mathbb{E}_n$  the root lattices of the same type over  $\mathbb{Z}_K$ . If  $L$  is a lattice and  $a \in K$  we denote by  ${}^{(a)}L$  the lattice  $L$  with form rescaled by  $a$ . Sometimes we identify lattices over number fields using the trace lattice. For instance  $(\mathbb{E}_8)_{\sqrt{-3}}$  denotes a hermitian lattice over  $\mathbb{Z}[\frac{1+\sqrt{-3}}{2}]$  of dimension 4 whose trace lattice over  $\mathbb{Z}$  is isometric to  $\mathbb{E}_8$ .

#### 5.1.1 Quadratic Forms in More than Four Variables

If  $E = K$ ,  $\dim_K(V) \geq 5$  and  $(V, \Phi)$  contains a one-class genus of lattices, then by Kirschmer [16, Section 7.4] either  $K = \mathbb{Q}$  or  $K = \mathbb{Q}[\sqrt{5}]$  where one has essentially one one-class genus of lattices of dimension 5 and 6 each. The rational lattices have been classified in [20] and are available electronically from [13].

**Proposition 5.1** *If  $E = K$ ,  $\dim_K(V) \geq 5$  and  $(V, \Phi)$  contains a fine  $\mathfrak{p}$ -admissible lattice chain  $\mathcal{L}(L_0, \mathfrak{p})$  of class number one for some prime ideal  $\mathfrak{p}$ , then  $K = \mathbb{Q}$  and  $\mathcal{L}(L_0, \mathfrak{p})$  is one of the following nine essentially different chains:*

1.  $\mathcal{L}(\mathbb{E}_8, 2) = (\mathbb{E}_8, \mathbb{D}_8, \mathbb{D}_4 \perp \mathbb{D}_4, {}^{(2)}\mathbb{D}_8^\#, {}^{(2)}\mathbb{E}_8)$ . After applying the oriflamme construction, the lattice chain becomes

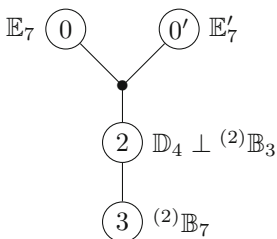


The automorphism groups are as follows

$T$	$G_T$	$\#G_T^+$
$\{i\}$	$2.O_8^+(2).2$	$2^{13} \cdot 3^5 \cdot 5^2 \cdot 7$
$\{2\}$	$\text{Aut}(\mathbb{D}_4) \wr C_2$	$2^{13} \cdot 3^4$
$\{i, j\}$	$2_+^{1+6}.S_8$	$2^{13} \cdot 3^2 \cdot 5 \cdot 7$
$\{2, i\}$	$N.(S_3 \times S_3 \wr C_2)$	$2^{13} \cdot 3^3$
$\{i, j, k\}$	$2_+^{1+6}.(C_3^2, \text{PSL}_2(7))$	$2^{13} \cdot 3 \cdot 7$
$\{2, i, j\}$	$N.(C_2 \times S_3 \wr C_2)$	$2^{13} \cdot 3^2$
$\{0, 0', 4, 4'\}$	$2_+^{1+6}.(C_2^3 : S_4)$	$2^{13} \cdot 3$
$\{2, i, j, k\}$	$N.(C_2^3 \times S_3)$	$2^{13} \cdot 3$
$\{0, 0', 2, 4, 4'\}$	$N.C_2^3$	$2^{13}$

where  $N = O_2(G_{\{2\}}) \cong 2_+^{1+4} \times 2_+^{1+4}$  and  $i, j, k \in \{0, 0', 2, 4, 4'\}$  with  $\#\{i, j, k\} = 3$ .

- $\mathcal{L}(\mathbb{E}_7, 2) = (\mathbb{E}_7, \mathbb{D}_6 \perp \mathbb{A}_1, \mathbb{D}_4 \perp ({}^2)\mathbb{B}_3, ({}^2)\mathbb{B}_7)$ . After applying the oriflamme construction, the lattice chain becomes



The automorphism groups are as follows

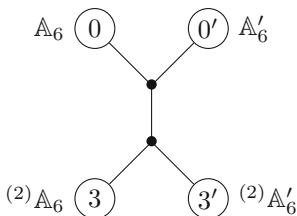
$T$	$G_T$	$\#G_T^+$
$\{i\}$	$C_2 \times \text{PSp}_6(2)$	$2^9 \cdot 3^4 \cdot 5 \cdot 7$
$\{2\}$	$\text{Aut}(\mathbb{D}_4) \times C_2 \wr S_3$	$2^9 \cdot 3^3$
$\{3\}$	$C_2 \wr S_7$	$2^9 \cdot 3^2 \cdot 5 \cdot 7$
$\{0, 0'\}$	$C_2^6 \cdot S_6$	$2^9 \cdot 3^2 \cdot 5$
$\{i, 2\}$	$N \cdot S_3^2$	$2^9 \cdot 3^2$
$\{i, 3\}$	$C_2^7 \cdot \text{PSL}_2(7)$	$2^9 \cdot 3 \cdot 7$
$\{2, 3\}$	$N \cdot (C_2 \times S_3^2)$	$2^9 \cdot 3^2$
$\{0, 0', 2\}, \{i, 2, 3\}$	$N \cdot D_{12}$	$2^9 \cdot 3$
$\{0, 0', 3\}$	$C_2^7 \cdot S_4$	$2^9 \cdot 3$
$\{0, 0', 2, 3\}$	$N \cdot C_2^2$	$2^9$

where  $N := O_2(G_{\{2\}}) \cong 2_+^{1+4} \times Q_8$  and  $i \in \{0, 0'\}$ . The one-class chain

$$\mathcal{L}(\mathbb{B}_7, 2) = \{\mathbb{B}_7, {}^{(2)}(\mathbb{D}_4^\# \perp B_3), {}^{(2)}\mathbb{D}_6^\# \perp \mathbb{B}_1, {}^{(2)}\mathbb{E}_7^\#\}$$

yields the same stabilisers.

- $\mathcal{L}(\mathbb{A}_6, 2) = \{\mathbb{A}_6, X, {}^{(2)}X^{\#2}, {}^{(2)}\mathbb{A}_6\}$ . Here  $X$  is an indecomposable lattice with  $\text{Aut}(X) = (C_2^4 \times C_3) \cdot D_{12}$ . After applying the oriflamme construction, the lattice chain becomes



The automorphism groups are as follows

$T$	$G_T$	$\#G_T^+$	sgdb
$\#T = 1$	$C_2 \times S_7$	$2^3 \cdot 3^2 \cdot 5 \cdot 7$	—
$\{0, 0'\}, \{3, 3'\}$	$C_2 \times S_3 \times S_4$	$2^3 \cdot 3^2$	43
$\{0, 3\}, \{0, 3'\}, \{0', 3\}, \{0', 3'\}$	$C_2 \times \text{PSL}_2(7)$	$2^3 \cdot 3 \cdot 7$	42
$\#T = 3$	$C_2 \times S_4$	$2^3 \cdot 3$	12
$\{0, 0', 3, 3'\}$	$C_2 \times D_8$	$2^3$	3

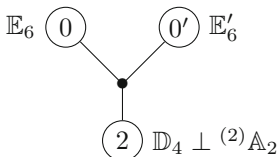
Here, and in the following tables, the column sgdb gives the label of  $G_T^+$  as defined by the small group database [3].

*The admissible one-class chain*

$$\mathcal{L}({}^{(7)}\mathbb{A}_6^\#, 2) = \{{}^{(7)}\mathbb{A}_6^\#, {}^{(7)}X^{\#,7}, {}^{(14)}X^\#, {}^{(14)}\mathbb{A}_6^\#\}$$

yields the same groups.

4.  $\mathcal{L}(\mathbb{E}_6, 2) = \{\mathbb{E}_6, Y_0, \mathbb{D}_4 \perp {}^{(2)}\mathbb{A}_2\}$ . Here  $Y_0$  is the even sublattice of  $\mathbb{B}_5 \perp {}^{(3)}\mathbb{B}_1$ . It is indecomposable and  $\text{Aut}(Y_0) = \text{Aut}(\mathbb{B}_5 \perp {}^{(3)}\mathbb{B}_1) \cong C_2 \times C_2 \wr S_5$ . After applying the oriflamme construction, the lattice chain becomes



*The automorphism groups are as follows*

$T$	$G_T$	$\#G_T^+$	$sgdb$
$\{i\}$	$C_2 \times U_4(2).2$	$2^6 \cdot 3^4 \cdot 5$	—
$\{2\}$	$\text{Aut}(\mathbb{D}_4) \times D_{12}$	$2^6 \cdot 3^3$	—
$\{0, 0'\}$	$C_2 \wr S_5$	$2^6 \cdot 3 \cdot 5$	11358
$\{i, 2\}$	$N.S_3^2$	$2^6 \cdot 3^2$	8277
$\{0, 0', 2\}$	$N.D_{12}$	$2^6 \cdot 3$	201

where  $N = O_2(G_3) \cong 2_+^{1+4} \times C_2$  and  $i \in \{0, 0'\}$ . The admissible one-class chains

$$\begin{aligned} \mathcal{L}(\mathbb{A}_2 \perp \mathbb{D}_4, 2) &= \{\mathbb{A}_2 \perp \mathbb{D}_4, {}^{(2)}Y^{\#,2}, {}^{(2)}\mathbb{E}_6\} \\ \mathcal{L}({}^{(3)}(\mathbb{A}_2^\# \perp \mathbb{D}_4), 2) &= \{{}^{(3)}(\mathbb{A}_2^\# \perp \mathbb{D}_4), {}^{(6)}Y^\#, {}^{(6)}\mathbb{E}_6\} \\ \mathcal{L}({}^{(3)}\mathbb{E}_6^\#, 2) &= \{{}^{(3)}\mathbb{E}_6^\#, {}^{(3)}Y^{\#,3}, {}^{(3)}(\mathbb{A}_2 \perp \mathbb{D}_4)^{\#,3}\} \end{aligned}$$

yield the same stabilisers.

5.  $\mathcal{L}(\mathbb{D}_6, 2) = \{\mathbb{D}_6, \mathbb{D}_4 \perp {}^{(2)}\mathbb{B}_2, {}^{(2)}\mathbb{B}_6\}$ . Here the application of the oriflamme construction is not necessary. The automorphism groups are as follows

$T$	$G_T$	$\#G_T^+$	$sgdb$
$\{0\}, \{2\}$	$C_2 \wr S_6$	$2^8 \cdot 3^2 \cdot 5$	—
$\{1\}$	$\text{Aut}(\mathbb{D}_4) \perp C_2 \wr S_2$	$2^8 \cdot 3^2$	—
$\{0, 1\}, \{1, 2\}$	$C_2^6.(C_2 \times S_4)$	$2^8 \cdot 3$	1086007
$\{0, 2\}$	$C_2^6.(C_2 \times S_4)$	$2^8 \cdot 3$	1088660
$\{0, 1, 2\}$	$C_2^6.(C_2 \times D_8)$	$2^8$	6331

6.  $\mathcal{L}(\mathbb{E}_6, 3) = \{\mathbb{E}_6, \mathbb{A}_2^3, {}^{(3)}\mathbb{E}_6\}$ . Here the application of the oriflamme construction is not necessary. The automorphism groups are as follows

$T$	$G_T$	$\#G_T^+$	sgdb
$\{0\}, \{2\}$	$C_2 \times U_4(2).2$	$2^6 \cdot 3^4 \cdot 5$	—
$\{1\}$	$D_{12} \wr S_3$	$2^5 \cdot 3^4$	—
$\{0, 2\}$	$3_+^{1+2} \cdot (C_2 \times GL_2(3))$	$2^3 \cdot 3^4$	533
$\{0, 1\}, \{1, 2\}$	$N \cdot (C_2^2 \times S_4)$	$2^3 \cdot 3^4$	704
$\{0, 1, 2\}$	$N \cdot (C_2^2 \times S_3)$	$2 \cdot 3^4$	10

where  $N = O_3(G_{\{1\}}) \cong C_3^3$ .

7.  $\mathcal{L}(\mathbb{B}_5 \perp {}^{(3)}\mathbb{B}_1, 3) = \{\mathbb{B}_5 \perp {}^{(3)}\mathbb{B}_1, \mathbb{B}_2 \perp \mathbb{A}_2 \perp {}^{(3)}\mathbb{B}_2; \mathbb{B}_1 \perp {}^{(3)}\mathbb{B}_5\}$ . Here the application of the oriflamme construction is not necessary. The automorphism groups are as follows

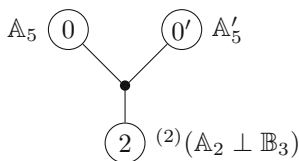
$T$	$G_T$	$\#G_T^+$	sgdb
$\{0\}, \{2\}$	$C_2 \times C_2 \wr S_5$	$2^6 \cdot 3 \cdot 5$	—
$\{1\}$	$C_2 \wr S_2 \times D_{12} \times C_2 \wr S_2$	$2^5 \cdot 3$	144
$\{0, 2\}$	$C_2^2 \times GL_2(3)$	$2^3 \cdot 3$	3
$\{0, 1\}, \{1, 2\}$	$C_2^2 \times D_8 \times S_3$	$2^3 \cdot 3$	8
$\{0, 1, 2\}$	$C_2^3 \times S_3$	$2 \cdot 3$	2

For  $0 \leq i \leq 2$  let  $Y_i$  be the even sublattice of  $\mathcal{L}(\mathbb{B}_5 \perp {}^{(3)}\mathbb{B}_1, 3)_i$ , see also part (4). Then the admissible one-class chains

$$\mathcal{L}(Y_0, 3) = \{Y_0, Y_1, Y_2\} \text{ and } \mathcal{L}({}^{(2)}Y_0^{\#2}, 3) = ({}^{(2)}Y_0^{\#2}, {}^{(2)}Y_1^{\#2}, {}^{(2)}Y_2^{\#2})$$

yield the same groups.

8.  $\mathcal{L}(\mathbb{A}_5, 2) = \{\mathbb{A}_5, {}^{(2)}\mathbb{B}_1 \perp Z, {}^{(2)}(\mathbb{A}_2 \perp \mathbb{B}_3)\}$ . Here  $Z$  is the even sublattice of  $\mathbb{B}_3 \perp {}^{(3)}\mathbb{B}_1$  and  $\text{Aut}(Z) = \text{Aut}(\mathbb{B}_3 \perp {}^{(3)}\mathbb{B}_1)$ . After applying the oriflamme construction, the lattice chain becomes



The automorphism groups are as follows

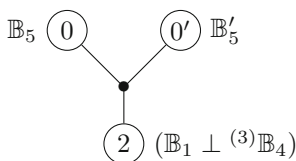
$T$	$G_T$	$\#G_T^+$	$sgdb$
$\{0\}, \{0'\}$	$C_2 \times S_6$	$2^3 \cdot 3^2 \cdot 5$	118
$\{2\}$	$D_{12} \times C_2 \wr S_3$	$2^3 \cdot 3^2$	43
$\#T = 2$	$C_2^2 \times S_4$	$2^3 \cdot 3$	12
$\{0, 0', 2\}$	$C_2^2 \times D_8$	$2^3$	3

The admissible one-class chains

$$\begin{aligned} \mathcal{L}({}^{(3)}\mathbb{A}_5^{\#,3}, 2) &= \{({}^{(3)}\mathbb{A}_5^{\#,3}, {}^{(6)}\mathbb{B}_1 \perp {}^{(3)}Z^{\#,3}, {}^{(6)}(\mathbb{A}_2^\# \perp \mathbb{B}_3)\} \\ \mathcal{L}(\mathbb{A}_2 \perp \mathbb{B}_3, 2) &= \{\mathbb{A}_2 \perp \mathbb{B}_3, \mathbb{B}_1 \perp {}^{(2)}Z^{\#,2}, {}^{(2)}\mathbb{A}_5^{\#,2}\} \\ \mathcal{L}({}^{(3)}(\mathbb{A}_2^\# \perp \mathbb{B}_3), 2) &= \{({}^{(3)}(\mathbb{A}_2^\# \perp \mathbb{B}_3), {}^{(3)}\mathbb{B}_1 \perp {}^{(6)}Z^\#, {}^{(6)}\mathbb{A}_5^\#\} \end{aligned}$$

yield the same stabilisers.

9.  $\mathcal{L}(\mathbb{B}_5, 3) = \{\mathbb{B}_5, \mathbb{B}_2 \perp \mathbb{A}_2 \perp {}^{(3)}\mathbb{B}_1, \mathbb{B}_1 \perp {}^{(3)}\mathbb{B}_4\}$ . After applying the oriflamme construction, the lattice chain becomes



$T$	$G_T$	$\#G_T^+$	$sgdb$
$\{i\}$	$C_2 \wr S_5$	$2^6 \cdot 3 \cdot 5$	11358
$\{2\}$	$C_2 \wr S_4 \times C_2$	$2^5 \cdot 3$	204
$\{0, 0'\}$	$(C_2 \times D_8) \times S_3$	$2^3 \cdot 3$	3
$\{i, 2\}$	$C_2 \times GL_2(3)$	$2^3 \cdot 3$	8
$\{0, 0', 2\}$	$C_2^2 \times S_3$	$2 \cdot 3$	2

where  $i \in \{0, 0'\}$ . The admissible one-class chain

$$\mathcal{L}(\mathbb{B}_4 \perp {}^{(3)}\mathbb{B}_1, 3) = \{\mathbb{B}_4 \perp {}^{(3)}\mathbb{B}_1, \mathbb{B}_1 \perp \mathbb{A}_2 \perp {}^{(3)}\mathbb{B}_2, {}^{(3)}\mathbb{B}_5\}$$

yields the same stabilisers.

### 5.1.2 Quadratic Forms in Four Variables

Now assume that  $K = E$  and  $\dim_K(V) = 4$ . By Kirschmer [16, Theorem 7.4.1] there are up to similarity exactly 481 one-class genera of lattices if  $K = \mathbb{Q}$  and additionally 607 such genera over 22 other base fields where the largest degree is

$[K : \mathbb{Q}] = 5$  [16, Theorem 7.4.2]. As we are only interested in the case where the rank of  $U_{\mathfrak{p}}$  is 2, we only need to consider pairs  $(L, \mathfrak{p})$  where  $L$  is one of these 1088 lattices and  $\mathfrak{p}$  a prime ideal such that  $V_{\mathfrak{p}} \cong \mathbb{H}(K_{\mathfrak{p}}) \perp \mathbb{H}(K_{\mathfrak{p}})$ . In this case the building  $\mathcal{B}$  of  $U_{\mathfrak{p}}$  is of type  $A_1 \oplus A_1$  and not connected even after oriflamme construction. We will not list the groups acting chamber transitively on  $\mathcal{B}^+$ , also because of the numerous cases of one-class lattice chains in this situation.

To list the lattices we need some more notation. We denote by  $\mathcal{Q} := \mathcal{O}_{\alpha, \infty, \mathfrak{p}_1, \dots, \mathfrak{p}_s}$  a definite quaternion algebra over  $K = \mathbb{Q}(\alpha)$  which ramifies exactly at the finite places  $\mathfrak{p}_1, \dots, \mathfrak{p}_s$  of  $K$ . Given an integral ideal  $\mathfrak{a}$  of  $\mathbb{Z}_K$  coprime to all  $\mathfrak{p}_i$ , then  $\mathcal{O}_{\alpha, \infty, \mathfrak{p}_1, \dots, \mathfrak{p}_s; \mathfrak{a}}$  denotes an Eichler order of level  $\mathfrak{a}$  in  $\mathcal{Q}$ .

We omit the subscript  $\alpha$  whenever  $K = \mathbb{Q}$ . Similarly, the subscript  $\mathfrak{a}$  is omitted, if  $\mathfrak{a} = \mathbb{Z}_K$ , i.e. the order is maximal.

Then  $\mathcal{O}_{\alpha, \infty, \mathfrak{p}_1, \dots, \mathfrak{p}_s; \mathfrak{a}}$  with the reduced norm form of  $\mathcal{Q}$  yields a quaternary lattice over  $\mathbb{Z}_K$ . By Nebe [24, Corollary 4.6] this lattice is unique in its genus, if and only if all Eichler orders of level  $\mathfrak{a}$  in  $\mathcal{Q}$  are conjugate.

Hence we identify such orders with their quaternary lattices.

**Proposition 5.2** *Let  $L$  be a  $\mathfrak{p}$ -normalised, quaternary lattice over  $\mathbb{Z}_K$  such that  $\mathcal{L}(L, \mathfrak{p})$  is a fine  $\mathfrak{p}$ -admissible lattice chain of length 2 and class number one. Then one of the following holds.*

1.  $K = \mathbb{Q}$  and either
  - $\mathfrak{p} = 2$  and  $L \cong \mathcal{O}_{\infty, 3} \cong \mathbb{A}_2 \perp \mathbb{A}_2$  or  $\mathcal{O}_{\infty, 5}$ .
  - $\mathfrak{p} \in \{3, 5, 11\}$  and  $L \cong \mathcal{O}_{\infty, 2} \cong \mathbb{D}_4$ .
  - $\mathfrak{p} = 3$  and  $L \cong \mathbb{B}_4$ .
2.  $K = \mathbb{Q}(\sqrt{5})$  and either
  - $\text{Nr}_{K/\mathbb{Q}}(\mathfrak{p}) \in \{4, 5, 9, 11, 19, 29, 59\}$  and  $L \cong \mathcal{O}_{\sqrt{5}, \infty}$ . This lattice is called  $H_4$  in [26].
  - $\text{Nr}_{K/\mathbb{Q}}(\mathfrak{p}) \in \{5, 11\}$  and  $L \cong \mathcal{O}_{\sqrt{5}, \infty; 2\mathbb{Z}_K} \cong \mathbb{D}_4$ .
  - $\mathfrak{p} = 2\mathbb{Z}_K$  and  $L \cong \mathcal{O}_{\sqrt{5}, \infty; \mathfrak{a}} \cong \mathcal{L}(\mathcal{O}_{\sqrt{5}, \infty}, \mathfrak{a})_2$  with  $\text{Nr}_{K/\mathbb{Q}}(\mathfrak{a}) \in \{5, 11\}$ .
3.  $K = \mathbb{Q}(\sqrt{2})$  and either
  - $\text{Nr}_{K/\mathbb{Q}}(\mathfrak{p}) \in \{2, 7, 23\}$  and  $L \cong \mathcal{O}_{\sqrt{2}, \infty}$ .
  - $\text{Nr}_{K/\mathbb{Q}}(\mathfrak{p}) = 7$  and  $L \cong \mathcal{O}_{\sqrt{2}, \infty; \sqrt{2}\mathbb{Z}_K} \cong \mathcal{L}(\mathcal{O}_{\sqrt{2}, \infty})_2$  or  $L$  is isometric to a unimodular lattice of norm  $\sqrt{2}\mathbb{Z}_K$  in  $(V, \Phi) \cong \langle 1, 1, 1, 1 \rangle$ . By O’Meara [25, IX:93], the genus of the latter lattice is uniquely determined and it has class number one by Kirschmer [16].
  - $\mathfrak{p} = \sqrt{2}\mathbb{Z}_K$  and  $L \cong \mathcal{O}_{\sqrt{2}, \infty; \mathfrak{a}} \cong \mathcal{L}(\mathcal{O}_{\sqrt{2}, \infty}, \mathfrak{a})_2$  with  $\text{Nr}_{K/\mathbb{Q}}(\mathfrak{a}) = 7$ .
4.  $K = \mathbb{Q}(\sqrt{3})$  and either
  - $\mathfrak{p} = \sqrt{3}\mathbb{Z}_K$  and  $L \cong \mathcal{O}_{\sqrt{3}, \infty; \mathfrak{p}_2}$  or  $L$  is isometric to a unimodular lattice of norm  $\mathfrak{p}_2$  in  $(V, \Phi) \cong \langle 1, 1, 1, 1 \rangle$ . Again, this lattice is unique up to isometry.
  - $\mathfrak{p} = \mathfrak{p}_2$  and  $L \cong \mathcal{O}_{\sqrt{3}, \infty; \sqrt{3}\mathbb{Z}_K}$ .

5.  $K = \mathbb{Q}(\sqrt{13})$  and  $\text{Nr}_{K/\mathbb{Q}}(\mathfrak{p}) = 3$  and  $L \cong \mathcal{O}_{\sqrt{13}, \infty}$ . This lattice is called  $D_4^\sim$  in [26].
6.  $K = \mathbb{Q}(\sqrt{17})$  and  $\text{Nr}_{K/\mathbb{Q}}(\mathfrak{p}) = 2$  and  $L \cong \mathcal{O}_{\sqrt{17}, \infty}$ . This lattice is called  $(2A_2)^\sim$  in [26].
7.  $K = \mathbb{Q}(\theta_9)$  is the maximal totally real subfield of the cyclotomic field  $\mathbb{Q}(\zeta_9)$  and  $\mathfrak{p} = 2\mathbb{Z}_K$  and  $L \cong \mathcal{O}_{\theta, \infty, \mathfrak{p}_3}$ .
8.  $K = \mathbb{Q}(\alpha) \cong \mathbb{Q}[X]/(X^3 - X^2 - 3X + 1)$  is the unique totally real number field of degree 3 and discriminant 148. Then either  $\mathfrak{p} = \mathfrak{p}_5$  and  $L \cong \mathcal{O}_{\alpha, \infty; \mathfrak{p}_2}$  or  $\mathfrak{p} = \mathfrak{p}_2$  and  $L \cong \mathcal{O}_{\alpha, \infty; \mathfrak{p}_5}$ .
9.  $K = \mathbb{Q}(\alpha) \cong \mathbb{Q}[X]/(X^3 - X^2 - 4X + 2)$  is the unique totally real number field of degree 3 and discriminant 316. Then  $\mathfrak{p} = \mathfrak{p}_2$  and  $L \cong \mathcal{O}_{\alpha, \infty; \mathfrak{p}_4}$ .
10.  $K = \mathbb{Q}(\alpha) \cong \mathbb{Q}[X]/(X^4 - X^3 - 3X^2 + X + 1)$  is the unique totally real number field of degree 4 and discriminant 725. Then  $L \cong \mathcal{O}_{\alpha, \infty}$  and  $\text{Nr}_{K/\mathbb{Q}}(\mathfrak{p}) \in \{11, 19\}$  or  $\mathfrak{p}$  is the ramified prime ideal of norm 29.
11.  $K = \mathbb{Q}(\alpha) \cong \mathbb{Q}[X]/(X^4 - 4X^2 - X + 1)$  is the unique totally real number field of degree 4 and discriminant 1957. Then  $\mathfrak{p} = \mathfrak{p}_3$  and  $L \cong \mathcal{O}_{\alpha, \infty}$ .
12.  $K = \mathbb{Q}(\alpha) \cong \mathbb{Q}[X]/(X^4 - X^3 - 4X^2 + X + 2)$  is the unique totally real number field of degree 4 and discriminant 2777. Then  $\mathfrak{p} = \mathfrak{p}_2$  and  $L \cong \mathcal{O}_{\alpha, \infty}$ .

Here  $\mathfrak{p}_q$  denotes a prime ideal of  $\mathbb{Z}_K$  of norm  $q$ . Conversely, in all these cases the chain  $\mathcal{L}(L, \mathfrak{p})$  is  $\mathfrak{p}$ -admissible and has class number one.

## 5.2 Hermitian Forms

In this section we treat the case that  $[E : K] = 2$ , so  $E$  is a totally complex extension of degree 2 of the totally real number field  $K$ . The automorphism groups of the hermitian lattices that occur in the tables below are strongly related to maximal finite symplectic matrix groups classified in [14]. We use the notation introduced in this thesis (see also [15]) to name the groups. All hermitian lattices with class number  $\leq 2$  are classified in [16, Section 8] and listed explicitly for  $n \geq 3$  in [16, pp. 129–140].

**Proposition 5.3** *Let  $\mathcal{L}(L_0, \mathfrak{p})$  be a fine  $\mathfrak{P}$ -admissible chain of class number one and of length at least 2. Then  $K = \mathbb{Q}$ ,  $d := \dim_E(V) \in \{3, 4, 5\}$  and one the following holds:*

1.  $E = \mathbb{Q}(\sqrt{-3})$ ,  $\mathfrak{p} = 3\mathbb{Z}$  and  $L_0 \cong \mathbb{B}_5 \otimes_{\mathbb{Z}} \mathbb{Z}[\frac{1+\sqrt{-3}}{2}] \cong (\mathbb{A}_2^5)_{\sqrt{-3}}$ :

$$\mathcal{L}(L_0, 3) = \{L_0, (\mathbb{A}_2^2 \perp {}^{(3)}\mathbb{B}_6^\#)_{\sqrt{-3}}, (\mathbb{A}_2 \perp \mathbb{E}_8)_{\sqrt{-3}}\}.$$

Here the application of the oriflamme construction is not necessary. The automorphism groups are as follows:

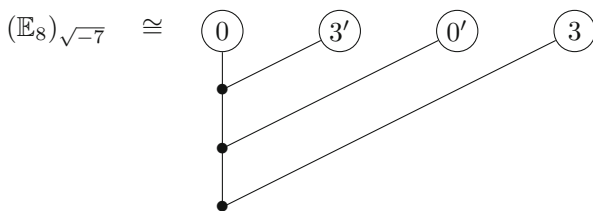


$T$	$G_T$	$\#G_T^+$
$\{0\}$	$C_6 \wr S_5$	$2^7 \cdot 3^5 \cdot 5$
$\{1\}$	$C_6 \wr S_2 \times \sqrt{-3}[\pm 3_+^{1+2} \cdot \text{SL}_2(3)]_3$	$2^6 \cdot 3^5$
$\{2\}$	$C_6 \times \sqrt{-3}[\text{Sp}_4(3) \times C_3]_4$	$2^7 \cdot 3^5 \cdot 5$
$\{0, 1\}$	$C_6 \wr S_2 \times \sqrt{-3}[\pm 3_+^{1+2} \cdot C_6]_3$	$2^4 \cdot 3^5$
$\{1, 2\}$	$C_6 \times \sqrt{-3}[\pm(3_+^{1+2} \cdot \text{SL}_2(3) \times C_3)]_4$	$2^4 \cdot 3^5$
$\{0, 1\}$	$C_6 \times \sqrt{-3}[\pm 3^3 : S_4 \times C_3]_4$	$2^4 \cdot 3^5$
$\{1, 2\}$	$C_6 \times \sqrt{-3}[\pm(3_+^{1+2} \cdot \text{SL}_2(3) \times C_3)]_4$	$2^4 \cdot 3^5$
$\{0, 1, 2\}$	$C_6 \times \sqrt{-3}[\pm 3_+^{1+2} \cdot C_6 \times C_3]_4$	$2^2 \cdot 3^5$

2.  $E = \mathbb{Q}(\sqrt{-7})$ ,  $\mathfrak{p} = 2\mathbb{Z}$  and  $L_0 \cong (\mathbb{E}_8)_{\sqrt{-7}}$ :

$$\mathcal{L}((\mathbb{E}_8)_{\sqrt{-7}}, 2) = \{(\mathbb{E}_8)_{\sqrt{-7}}, (\mathbb{D}_8)_{\sqrt{-7}}, (\mathbb{D}_4 \perp \mathbb{D}_4)_{\sqrt{-7}}, {}^{(2)}\mathbb{D}_8)_{\sqrt{-7}}\}.$$

After applying the variant of the oriflamme construction described in Remark 4.8, the lattice chain becomes



The automorphism groups are as follows:

$T$	$G_T$	$\#G_T^+$	$sgdb$
$\#T = 1$	$2.\text{Alt}_7$	$2^4 \cdot 3^2 \cdot 5 \cdot 7$	—
$\{0, 0'\}, \{3, 3'\}$	$\text{SL}_2(3) \times C_3 : 2$	$2^4 \cdot 3^2$	124
$\{0, 3\}, \{0, 3'\}, \{0', 3\}, \{0', 3'\}$	$\text{SL}_2(7)$	$2^4 \cdot 3 \cdot 7$	114
$\#T = 3$	$2.S_4$	$2^4 \cdot 3$	28
$\{0, 0', 3, 3'\}$	$Q_{16}$	$2^4$	9

3.  $E = \mathbb{Q}(\sqrt{-3})$ ,  $\mathfrak{p} = 2\mathbb{Z}$  and  $L_0 \cong (\mathbb{E}_8)_{\sqrt{-3}}$ :

$$\mathcal{L}((\mathbb{E}_8)_{\sqrt{-3}}, 2) = \{(\mathbb{E}_8)_{\sqrt{-3}}, (\mathbb{D}_4 \perp \mathbb{D}_4)_{\sqrt{-3}}, {}^{(2)}\mathbb{E}_8)_{\sqrt{-3}}\}.$$

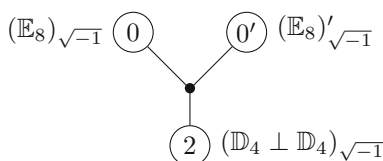
Here the application of the oriflamme construction is not necessary. The automorphism groups are as follows:

$T$	$G_T$	$\#G_T^+$
$\{0\}, \{2\}$	$\sqrt{-3}[\mathrm{Sp}_4(3) \times C_3]_4$	$2^7 \cdot 3^4 \cdot 5$
$\{1\}$	$\sqrt{-3}[\mathrm{SL}_2(3) \times C_3]_2^2$	$2^7 \cdot 3^3$
$\{0, 2\}$	$2_-^{1+4} \cdot \mathrm{Alt}_5 \times C_3$	$2^7 \cdot 3 \cdot 5$
$\{0, 1\}, \{1, 2\}$	$\mathrm{SL}_2(3) \wr C_2 \times C_3$	$2^7 \cdot 3^2$
$\{0, 1, 2\}$	$(Q_8 \wr S_2) : C_3 \times C_3$	$2^7 \cdot 3$

4.  $E = \mathbb{Q}(\sqrt{-1})$ ,  $\mathfrak{p} = 2\mathbb{Z}$  and  $L_0 \cong (\mathbb{E}_8)_{\sqrt{-1}}$ :

$$\mathcal{L}((\mathbb{E}_8)_{\sqrt{-1}}, 2) = \{(\mathbb{E}_8)_{\sqrt{-1}}, (\mathbb{D}_8)_{\sqrt{-1}}, (\mathbb{D}_4 \perp \mathbb{D}_4)_{\sqrt{-1}}\}.$$

Here the application of the oriflamme construction is not necessary. After applying the oriflamme construction, one obtains the following lattices

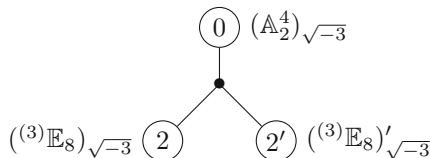


$T$	$G_T$	$\#G_T^+$
$\{0\}, \{0'\}$	$i[(2_+^{1+4}YC_4) \cdot S_6]_4$	$2^9 \cdot 3^2 \cdot 5$
$\{2\}$	$i[(D_8YC_4) \cdot S_3]_2^2$	$2^9 \cdot 3^2$
$\{0, 2\}, \{0', 2\}$		$2^9 \cdot 3$
$\{0, 0'\}$		$2^9 \cdot 3$
$\{0, 2, 0'\}$		$2^9$

5.  $E = \mathbb{Q}(\sqrt{-3})$ ,  $\mathfrak{p} = 3\mathbb{Z}$  and  $L_0 = \mathbb{B}_4 \otimes_{\mathbb{Z}} \mathbb{Z}[\frac{1+\sqrt{3}}{2}] \cong (\mathbb{E}_8)_{\sqrt{-3}}$ : Here the application of the oriflamme construction is not necessary.

$$\mathcal{L}(L_0, 3) = \{(\mathbb{A}_2^4)_{\sqrt{-3}}, (\mathbb{A}_2 \perp {}^{(3)}\mathbb{E}_6^{\#})_{\sqrt{-3}}, ({}^{(3)}\mathbb{E}_8)_{\sqrt{-3}}\}.$$

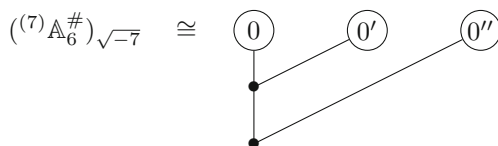
After applying the oriflamme construction, the chain becomes:



The automorphism groups are as follows:

$T$	$G_T$	$\#G_T^+$	$sgdb$
$\{0\}$	$C_6 \wr S_4$	$2^6 \cdot 3^4$	—
$\{2\}, \{2'\}$	$\sqrt{-3}[\mathrm{Sp}_4(3) \times C_3]_4$	$2^7 \cdot 3^4 \cdot 5$	—
$\{0, 2\}, \{0, 2'\}$	$(\pm C_3^4).S_4$	$2^4 \cdot 3^4$	3085
$\{2, 2'\}$	$(C_6 \times 3_+^{1+2}).S_3$	$2^4 \cdot 3^4$	2895
$\{0, 2, 2'\}$	$(C_6 \times C_3 \wr C_3).2$	$2^2 \cdot 3^4$	68

6.  $E = \mathbb{Q}(\sqrt{-7})$ ,  $\mathfrak{p} = 2\mathbb{Z}$  and  $L_0 = ({}^{(7)}\mathbb{A}_6^\#)_{\sqrt{-7}}$ . After applying the variant of the oriflamme construction described in Remark 4.8, the chain becomes:



The automorphism groups are as follows:

$T$	$G_T$	$\#G_T^+$
$\#T = 1$	$\pm C_7 : 3$	$3 \cdot 7$
$\#T = 2$	$C_6$	3
$\{0, 0', 0''\}$	$C_2$	1

### 5.3 Quaternionic Hermitian Forms

In this section we treat the case that  $[E : K] = 4$ , so  $E$  is a totally definite quaternion algebra over the totally real number field  $K$ . All quaternionic hermitian lattices with class number  $\leq 2$  are classified in [16, Section 9] and listed explicitly for  $n \geq 2$  in [16, pp. 147–150].

**Proposition 5.4** *Suppose  $E$  is a definite quaternion algebra and let  $\mathcal{L}(L_0, \mathfrak{p})$  be a fine  $\mathfrak{P}$ -admissible chain of length at least 2 and of class number one. Then  $K = \mathbb{Q}$ ,  $d := \dim_E(V) = 2$  and one of the following holds:*

1.  $E \cong \mathcal{Q}_{\infty, 2}$ , the rational quaternion algebra ramified at 2 and  $\infty$ ,  $\mathfrak{p} = 3\mathbb{Z}$  and  $L_0 \cong (\mathbb{E}_8)_{\infty, 2}$  is the unique  $\mathcal{M}$ -structure of the  $\mathbb{E}_8$ -lattice whose automorphism group is called  ${}_{\infty, 2}[2_-^{1+4}.\mathrm{Alt}_5]_2$  in [24]. The oriflamme construction is not necessary and the automorphism groups are

$T$	$G_T$	$\#G_T$	$sgdb$
$\{0\}, \{2\}$	$\infty_2[2_{-}^{1+4}.Alt_5]_2$	$2^7 \cdot 3 \cdot 5$	–
$\{1\}$	$Q_8 : SL_2(3)$	$2^6 \cdot 3$	1022
$\{0, 1\}, \{1, 2\}$	$C_2 \times SL_2(3)$	$2^4 \cdot 3$	32
$\{0, 2\}$	$C_3 : SD_{16}$	$2^4 \cdot 3$	16
$\#T = 3$	$C_2 \times C_6$	$2^2 \cdot 3$	9

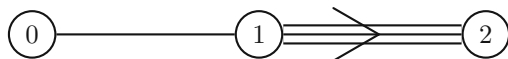
2.  $E \cong Q_{\infty,3}$  and  $\mathfrak{p} = 2\mathbb{Z}$  and  $L_0 \cong (\mathbb{E}_8)_{\infty,3}$  is the unique  $\mathcal{M}$ -structure of the  $\mathbb{E}_8$ -lattice whose automorphism group is called  $\infty_3[SL_2(9)]_2$  in [24]. The oriflamme construction is not necessary and the automorphism groups are

$T$	$G_T$	$\#G_T$	$sgdb$
$\{0\}, \{2\}$	$\infty_3[SL_2(9)]_2$	$2^4 \cdot 3^2 \cdot 5$	409
$\{1\}$	$SL_2(3).S_3$	$2^4 \cdot 3^2$	124
$\#T = 2$	$C_2.S_4$	$2^4 \cdot 3$	28
$\{0, 1, 2\}$	$Q_{16}$	$2^4$	9

Note that the above quaternion algebras only have one conjugacy class of maximal orders and for any such order  $\mathcal{M}$ , the above  $\mathcal{M}$ -lattice  $L_0$  is uniquely determined up to isometry.

### 5.4 The Exceptional Groups

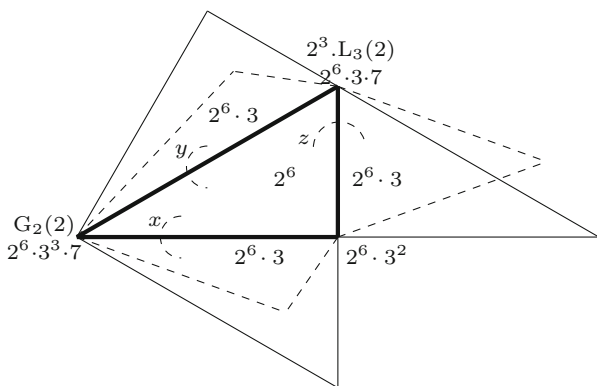
The exceptional groups have been dealt with in [16, Chapter 10], where it is shown that only the group  $G_2$  admits one-class genera defined by a coherent family of parahoric subgroups. In all cases the number field is the field of rational numbers. The one-class genera of lattice chains correspond to the coherent families of parahoric subgroups  $(P_q)_{q \text{ prime}}$  where for one prime  $p$  the parahoric subgroup  $P_p$  is the Iwahori subgroup, a stabiliser of a chamber in the corresponding  $p$ -adic building. Hence [16, Theorem 10.3.1] shows directly that there is a unique  $S$ -arithmetic group of type  $G_2$  with a discrete and chamber transitive action. It is given by the  $\mathbb{Z}$ -form  $\mathbf{G}_2$  where each parahoric subgroup  $P_q$  is hyperspecial. This integral model of  $G_2$  is described in [8] (see also [5] for more one-class genera of  $G_2$ ). Here  $\mathbf{G}_2(\mathbb{Z}) \cong G_2(2)$  and the  $S$ -arithmetic group is  $\mathbf{G}_2(\mathbb{Z}[\frac{1}{2}])$  (so  $S = \{(2)\}$ ). The extended Dynkin diagram of  $G_2$  is as follows.



The stabilisers  $G_T$  of the simplices  $T \subseteq \{0, 1, 2\}$  in the corresponding building of  $G_2(\mathbb{Q}_2)$  are given in [16, Section 10.3]:

T	$G_T$	$\#G_T$	sgdb
{0}	$G_2(2)$	$2^6 \cdot 3^3 \cdot 7$	—
{2}	$2^3 \cdot \text{GL}_3(2)$	$2^6 \cdot 3 \cdot 7$	814
{1}	$2^{1+4}_+ \cdot ((C_3 \times C_3) \cdot 2)$	$2^6 \cdot 3^2$	8282
{1,2}	$2^{1+4}_+ \cdot S_3$	$2^6 \cdot 3$	1494
{0,2}	$((C_4 \times C_4) \cdot 2) \cdot S_3$	$2^6 \cdot 3$	956
{0,1}	$2^{1+4}_+ \cdot S_3$	$2^6 \cdot 3$	988
{0,1,2}	$\text{Syl}_2(G_2(2))$	$2^6$	134

One may visualise the chamber transitive action of  $G_2(\mathbb{Z}[\frac{1}{2}])$  on the Bruhat-Tits building of  $G_2(\mathbb{Q}_2)$  by indicating the three generators  $x, y, z$  of  $G_2(\mathbb{Z}[\frac{1}{2}])$  of order 3 mapping the standard chamber to one of the (three times) two neighbours.



Using a suitable embedding  $G_2 \hookrightarrow O_7$  we find matrices for the three generators

$$x := \begin{pmatrix} 0 & 1 & 1 & -1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 0 & -1 & 1 \\ 1 & 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & -1 & 0 \end{pmatrix}, \quad y := \begin{pmatrix} 1 & 1 & 0 & -1 & -1 & -1 & 0 \\ 1 & 1 & -1 & -1 & 0 & -1 & 0 \\ 1 & 1 & -1 & 0 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 & 0 & -1 & 0 \\ 1 & 1 & 0 & -1 & 0 & -1 & -1 \\ 0 & 1 & 0 & -1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad z := \frac{1}{2} \begin{pmatrix} 2 & 2 & 0 & -1 & 0 & -2 & -1 \\ 1 & 0 & 2 & 0 & -1 & 0 & -1 \\ 2 & 2 & 2 & -3 & -2 & -2 & -1 \\ 2 & 2 & 0 & -1 & -2 & -2 & -1 \\ 0 & 0 & 4 & -2 & -2 & 0 & -2 \\ 2 & -2 & 0 & 1 & 2 & 0 & -1 \\ 0 & 2 & 0 & -1 & 0 & -2 & 1 \end{pmatrix}.$$

## 6 Chamber Transitive Actions on $p$ -Adic Buildings

In this section we tabulate the chamber transitive actions on the  $p$ -adic buildings obtained from the one-class genera of lattice chains given in the previous section.

We use the names and the local Dynkin diagrams as given in [31]. The name for  $U_p$  usually does not give the precise type of the  $p$ -adic group. For instance the

**Table 1** Buildings with chamber transitive discrete actions

$p$	$r$	$L_0$	ref	$U_p$	index	root system	local Dynkin diagram	Lit
2	4	$\mathbb{E}_8$	Prop. 3(1)	$O_8^+(\mathbb{Q}_2)$	${}^1D_{4,4}^{(1)}$	$\tilde{D}_4$	$  \begin{array}{c}  4 \quad \quad 0 \\  \diagdown \quad \diagup \\  2 \\  \diagup \quad \diagdown \\  4' \quad \quad 0'  \end{array}  $	[11] [9]
2	3	$\mathbb{E}_7$	Prop. 3(2)	$O_7(\mathbb{Q}_2)$	$B_{3,3}$	$\tilde{B}_3$	$  \begin{array}{c}  0 \\  \diagdown \quad \diagup \\  2 \\  \diagup \quad \diagdown \\  3 \quad \quad 0'  \end{array}  $	[11] [9]
2	3	$\mathbb{A}_6$ $(\mathbb{E}_8)_{\sqrt{-7}}$	Prop. 3(3) Prop. 5(2)	$O_6^+(\mathbb{Q}_2) \cong$ $SL_4(\mathbb{Q}_2)$	${}^1A_3$	$\tilde{A}_3$	$  \begin{array}{c}  0 \quad \quad 3 \\  \diagdown \quad \diagup \\  \quad \quad 0' \\  \diagup \quad \diagdown \\  3' \quad \quad 0'  \end{array}  $	[11] [10]
2	2	$\mathbb{E}_6$ $(\mathbb{E}_8)_{\sqrt{-3}}$	Prop. 3(4) Prop. 5(3)	$O_6^-(\mathbb{Q}_2) \cong$ $U_4(\mathbb{Q}_2(\sqrt{-3}))$	${}^2A_{3,2}^{(1)}$	$\tilde{C}_2$	$0 \rightleftharpoons 2 \rightleftharpoons 0'$	[11] [22]
2	2	$\mathbb{D}_6$ $(\mathbb{E}_8)_{\sqrt{-1}}$	Prop. 3(5) Prop. 5(4)	$O_6^-(\mathbb{Q}_2) \cong$ $U_4(\mathbb{Q}_2(\sqrt{-1}))$	${}^2D_{3,2}^{(1)}$	$C-B_2$	$0 \leftrightharpoons 2 \leftrightharpoons 0'$	[11] [9]
3	2	$\mathbb{E}_6$ $\mathbb{B}_5 \perp {}^{(3)}\mathbb{B}_1$ $(\mathbb{E}_8)_{\sqrt{-3}}$	Prop. 3(6) Prop. 3(7) Prop. 5(5)	$O_6^-(\mathbb{Q}_3) =$ $O_6^-(\mathbb{Q}_3) \cong$ $U_4(\mathbb{Q}_3(\sqrt{-3}))$	${}^2D_{3,2}^{(1)}$	$C-B_2$	$0 \leftrightharpoons 2 \leftrightharpoons 0'$	[11] [12]
2	2	$\mathbb{A}_5$ $(\mathbb{E}_8)_{\infty,3}$	Prop. 3(8) Prop. 6(1)	$O_5(\mathbb{Q}_2) \cong$ $Sp_4(\mathbb{Q}_2)$	$C_{2,2}^{(1)}$	$\tilde{C}_2$	$0 \rightleftharpoons 2 \rightleftharpoons 0'$	[11] [22]
3	2	$\mathbb{B}_5$ $(\mathbb{E}_8)_{\infty,2}$	Prop. 3(9) Prop. 6(2)	$O_5(\mathbb{Q}_3) \cong$ $Sp_4(\mathbb{Q}_3)$	$C_{2,2}^{(1)}$	$\tilde{C}_2$	$0 \rightleftharpoons 2 \rightleftharpoons 0'$	[11] [12]
2	2	${}^{(7)}\mathbb{A}_6^\#$ $(\mathbb{E}_8)_{\sqrt{-7}}$	Prop. 5(6)	$SL_3(\mathbb{Q}_2)$	${}^1A_2$	$\tilde{A}_2$	$  \begin{array}{c}  0 \\  \diagdown \quad \diagup \\  \quad \quad 0' \\  \diagup \quad \diagdown \\  0'' \quad \quad 0'  \end{array}  $	[11] [19] [23]
3	2	$(\mathbb{A}_2^5)_{\sqrt{-3}}$	Prop. 5(1)	$U_5(\mathbb{Q}_3(\sqrt{-3}))$	${}^2A_{4,2}^{(1)}$	$C-BC_2$	$0 \leftrightharpoons 2 \leftrightharpoons 0'$	new
2	2	$\mathbf{G}_2(\mathbb{Z}[\frac{1}{2}])$	Sec. 1.5.4	$G_2(\mathbb{Q}_2)$	$G_{2,2}$	$\tilde{G}_2$	$0 \text{ --- } 1 \rightleftharpoons 2$	[11] [9]

lattices  $\mathbb{E}_6$  and  $\mathbb{D}_6$  define two non isomorphic non-split forms of the algebraic group  $O_6$  over  $\mathbb{Q}_2$  which we both denote by  $O_6^-(\mathbb{Q}_2)$ . To distinguish these groups, we also give the Tits index as in [30] and [31, Section 4.4]. Note that the isomorphism  $O_6^- \cong U_4$  is given by the action of  $O_6$  on the even part of the Clifford algebra. So we find the one-class genera of lattice chains also in a hermitian geometry, for  $\mathcal{L}(\mathbb{E}_6, 2)$  (from 5.1 (3)) we get the same stabilisers as for  $\mathcal{L}((\mathbb{E}_8)_{\sqrt{-3}}, 2)$  (from 5.3 (2)) in the projective group. Such coincidences are indicated by listing the lattices  $L_0$  and the corresponding references (ref) in Table 1. The last column of Table 1 refers to a construction of the respective chamber transitive action in the literature. For a more detailed description of the different unitary groups  $U_p$  associated to the various types of local Dynkin diagrams we refer the reader to [31, Section 4.4].

**Acknowledgements** The research leading to this paper was supported by the DFG in the framework of the SPP1489. The authors thank Bill Kantor and Rudolf Scharlau for helpful discussions.

## References

1. P. Abramenko, G. Nebe, Lattice chain models for affine buildings of classical type. *Math. Ann.* **322**(3), 537–562 (2002)
2. J.W. Benham, J.S. Hsia, Spinor equivalence of quadratic forms. *J. Number Theory* **17**(3), 337–342 (1983)
3. H.U. Besche, B. Eick, E.A. O'Brien, The groups of order at most 2000. *Electron. Res. Announc. Am. Math. Soc.* **7**, 1–4 (2001)
4. A. Borel, Some finiteness properties of Adele groups over number fields. *Publ. Math. I.H.E.S.* **16**, 5–30 (1963)
5. A.M. Cohen, G. Nebe, W. Plesken, Maximal integral forms of the algebraic group  $G_2$  defined by finite subgroups. *J. Number Theory* **72**(2), 282–308 (1998)
6. W. Frisch, *The Cohomology of  $S$ -Arithmetic Spin Groups and Related Bruhat-Tits Buildings* (University of Göttingen, Mathematisch-Naturwissenschaftliche Fakultät, Göttingen, 2002)
7. L. Gerstein, The growth of class numbers of quadratic forms. *Am. J. Math.* **94**(1), 221–236 (1972)
8. B.H. Gross, Groups over  $\mathbb{Z}$ . *Invent. Math.* **124**(1–3), 263–279 (1996)
9. W.M. Kantor, Some exceptional 2-adic buildings. *J. Algebra* **92**(1), 208–223 (1985)
10. W.M. Kantor, Some locally finite flag-transitive buildings. *Eur. J. Comb.* **8**(4), 429–436 (1987)
11. W.M. Kantor, R.A. Liebler, J. Tits, On discrete chamber-transitive automorphism groups of affine buildings. *Bull. Am. Math. Soc. (N.S.)* **16**(1), 129–133 (1987)
12. W.M. Kantor, T. Meixner, M. Wester, Two exceptional 3-adic affine buildings. *Geom. Dedicata* **33**(1), 1–11 (1990)
13. M. Kirschmer, Genera of quadratic and hermitian lattices with small class number (2016). <http://www.math.rwth-aachen.de/homes/Markus.Kirschmer/forms/>
14. M. Kirschmer, Finite symplectic matrix groups. Ph.D. thesis, RWTH Aachen University, 2009
15. M. Kirschmer, Finite symplectic matrix groups. *Exp. Math.* **20**(2), 217–228 (2011)
16. M. Kirschmer, Definite quadratic and hermitian form with small class number. Habilitation, RWTH Aachen University, 2016
17. M. Kneser, Strong approximation, in *Algebraic Groups and Discontinuous Subgroups (Proceedings of Symposia in Pure Mathematics, Boulder, CO, 1965)* (American Mathematical Society, Providence, 1966), pp. 187–196
18. M. Kneser, *Quadratische Formen* (Springer, Berlin, 2002). Revised and edited in collaboration with Rudolf Scharlau
19. P. Köhler, T. Meixner, M. Wester, The 2-adic affine building of type  $\tilde{A}_2$  and its finite projections. *J. Comb. Theory Ser. A* **38**(2), 203–209 (1985)
20. D. Lorch, M. Kirschmer, Single-class genera of positive integral lattices. *LMS J. Comput. Math.* **16**, 172–186 (2013)
21. T. Meixner, Klassische Tits Kammersysteme mit einer transitiven Automorphismengruppe. *Mitt. Math. Sem. Giessen* **174**, x+115 (1986)
22. T. Meixner, M. Wester, Some locally finite buildings derived from Kantor's 2-adic groups. *Commun. Algebra* **14**(3), 389–410 (1986)
23. D. Mumford, An algebraic surface with  $K$  ample,  $(K^2) = 9, p_g = q = 0$ . *Am. J. Math.* **101**(1), 233–244 (1979)
24. G. Nebe, Finite quaternionic matrix groups. *Represent. Theory* **2**, 106–223 (1998)
25. O.T. O'Meara, *Introduction to Quadratic Forms* (Springer, Berlin, 1973)
26. R. Scharlau, Unimodular lattices over real quadratic fields. *Math. Z.* **216**, 437–452 (1994)
27. R. Scharlau, Martin Kneser's work on quadratic forms and algebraic groups, in *Quadratic Forms—Algebra, Arithmetic, and Geometry*. Contemporary Mathematics, vol. 493 (American Mathematical Society, Providence, 2009), pp. 339–357
28. G. Shimura, Arithmetic of unitary groups. *Ann. Math.* **79**, 269–409 (1964)
29. D.E. Taylor, *The Geometry of the Classical Groups*. Sigma Series in Pure Mathematics, vol. 9 (Heldermann Verlag, Berlin, 1992)

30. J. Tits, Classification of algebraic semisimple groups, in *Proceedings of the Summer Institute in Algebraic Groups and Discontinuous Groups (Boulder 1965)*. Proceedings of Symposia in Pure Mathematics, vol. 9 (American Mathematical Society, Providence, 1966), pp. 33–62
31. J. Tits, Reductive groups over local fields, in *Automorphic Forms, Representations and L-Functions (Proceedings of Symposia in Pure Mathematics)*, vol. 33 (American Mathematical Society, Providence, 1979), pp. 29–69
32. G.L. Watson, Transformations of a quadratic form which do not increase the class-number. *Proc. Lond. Math. Soc. (3)* **12**, 577–587 (1962)



# polyDB: A Database for Polytopes and Related Objects



Andreas Paffenholz

**Abstract** polyDB is a database for discrete geometric objects independent of a particular software. The database is accessible via web and an interface from the software package polymake. It contains various datasets from the area of lattice polytopes, combinatorial polytopes, matroids and tropical geometry.

In this short note we introduce the structure of the database and explain its use with a computation of the free sums and certain skew bipyramids among the class of smooth Fano polytopes in dimension up to 8.

**Keywords** Database • polyDB • Discrete geometry • Lattice polytopes • Reflexive polytopes • Smooth polytopes

**Subject Classifications** 52-04, 52B20

## 1 Introduction

In recent years availability of computational classifications of mathematical objects has proven to be an important and valuable tool to obtain new results, to check new ideas and to experiment with the objects to obtain insight into their structure and directions for further research.

We know the full list of smooth Fano polytopes (up to lattice equivalence) up to dimension 9 by an algorithm of Øbro [33], whose availability within the software package polymake has been the foundation e.g. for counter-examples to a conjecture of Batyrev and Selivanova [32] or the classification of simplicial, terminal, and reflexive polytopes with many vertices by Assarf et al. [5]. Availability of the same data in Magma [37] lead to the study of the poset of blowups by Higashitani [16] or the study of reflexive polytopes of higher index by Kasprzyk and

---

A. Paffenholz (✉)

TU Darmstadt, Dolivostr. 15, 64293 Darmstadt, Germany

e-mail: [paffenholz@opt.tu-darmstadt.de](mailto:paffenholz@opt.tu-darmstadt.de); [paffenholz@mathematik.tu-darmstadt.de](mailto:paffenholz@mathematik.tu-darmstadt.de)

NilI [24]. The classification of 0/1-polytopes up to dimension 6 by Aichholzer [2] was used in the study of permutation polytopes by Baumeister et al. [6].

We also know classifications of small oriented matroids by Miyata et al. [12], polytopes [21, 39], and reflexive polytopes up to dimension 4 by Kreuzer and Skarke [3, 26]. The symbolic data project by Gräbe et al. [13] aims to collect data from computer algebra and make it accessible in a structured and searchable form on their web page. The library *MIPLIB* by Koch et al. [25] collects discrete optimization problems for benchmarking of algorithms.

Most of these collections, however, cannot easily be used in a software package. Sometimes the data is only available in text format or, if searchable via a database, is connected to a specific software package or lacks a proper interface at all. For example, the small oriented matroids [12], polytopes [27, 39], or 0/1-polytopes [2] are available as text files, while access to the small groups library [7] is linked to GAP [36]. Altman et al. [3] have created a database for the reflexive polytopes up to dimension 4 computed by Kreuzer and Skarke [26], but it is currently not accessible at the link given in the paper.

On the other hand, the Graded Rings Database [8] project has a more general approach and provides data in a format both searchable via a web interface and accessible via a programmatic interface that can be used in software packages. It currently has a focus on data from combinatorial commutative algebra and toric geometry.

The new database `polyDB` aims to provide searchable data from a wide range of areas at a permanent location in an application independent format. It allows download in text format and access from any software package that provides an interface to the data. It is also searchable via a web interface at [db.polymake.org](http://db.polymake.org). Currently, one interface to a software package is implemented, in the software package `polymake` [4, 22]. The current collection of data is thus still inspired by the range of applications of `polymake` with data from combinatorial geometry, matroid theory, toric geometry and combinatorial topology.

In the following two sections we explain the concept of the database and introduce the interface implemented in `polymake` to access the data. The last section shows one application of the database and the interface. We will show that in dimensions up to 8 more than 80% of the smooth Fano polytopes arise from lower dimensional ones as a free sum of two lower dimensional smooth Fano polytopes or a certain skew sum construction of a smooth Fano polytope and a simplex. We give the count of polytopes decomposable in this way in Table 1. With a simple extension of the scripts one can also obtain the list of possible decompositions for each polytope.

**Table 1** Free sums, skew bipyramids and generalized smooth simplex sums among the smooth Fano polytopes

Dimension	2	3	4	5	6	7	8
Smooth Fano polytopes	5	18	124	866	7622	72,256	749,892
Free sums	1	5	28	176	1361	11,760	112,285
Skew bipyramids	1	9	57	489	4323	43,777	466,770
sg simplex-1 sums	2	13	66	556	4700	47,076	495,092
sg simplex-2 sums	1	3	31	232	2403	25,157	284,249
sg simplex-3 sums	–	1	4	52	515	6635	83,730
sg simplex-4 sums	–	–	1	5	81	961	14,598
sg simplex-5 sums	–	–	–	1	6	114	1609
sg simplex-6 sums	–	–	–	–	1	7	155
sg simplex-7 sums	–	–	–	–	–	1	8
sg simplex-8 sums	–	–	–	–	–	–	1
Total sg simplex sums	3	16	93	708	6283	61,961	657,380
Total decomposable	3	16	96	712	6346	62,331	660,792

The rows denoted by *sg simplex-n sums* for  $n$  between 1 and 8 count the simplex sums with a simplex of dimension  $n$ . The row denoted by *total sg simplex sums* gives the number of different generalized smooth simplex sums with a simplex of any dimension. The row *total decomposable* counts the number of different polytopes among the free sums and the smooth generalized simplex sums

## 2 polyDB

In this section we briefly introduce the structure of the database polyDB and the data sets already contained in it.

The database polyDB for discrete geometric objects is based on the open source NoSQL database MongoDB [31]. It has been set up at [db.polymake.org](http://db.polymake.org). The database stores its data as plain JSON documents grouped into *collections* and *databases* (To avoid confusion with this and the abstract database polyDB we will refer to this technical term introduced by MongoDB as a *collection group*). We use this to group collections from the same area of discrete geometry into a common collection group. E.g., the collection group *Objects in Tropical Geometry* currently contains two collections of such objects, the small *tropical oriented matroids* classified by Horn [17] and the *polytropes* classified by Kulas [27] and Tran [39]. polyDB stores data in a plain JSON format independent of any particular software package. See Fig. 1 for an example of an entry in the collection of smooth reflexive polytopes.

Each document contains one special entry *polyDB* (besides its *\_id*, which is required by MongoDB). Apart from this all other entries and their tags can be chosen freely depending on the data. The entry *polyDB* may specify format restrictions for the data and import or export specifications for various software packages, separated by subfields naming the software. This section may contain, e.g., information on the required version, authors of the data, and the method to load the data into the

```

{
  "_id" : "F.2D.3",
  "DIM":2,
  "FACETS" : [[1,0,1],[1,0,-1],[1,1,0],[1,-1,-1],[1,-1,0]],
  "VERTICES": [[1,-1,-1],[1,-1,1],[1,0,1],[1,1,0],[1,1,-1]],
  "F_VECTOR" : [5,5],
  "EHRHART_POLYNOMIAL_COEFF": ["1", "7/2", "7/2"],
  "H_STAR_VECTOR": [1,5,1],
  "CENTROID": ["1", "-2/21", "-2/21"],
  "N_LATTICE_POINTS":8,
  "NORMAL" : "true",
  "VERY_AMPLE" : "true",
  "LATTICE_VOLUME":7,
  "polyDB" : {
    [...]
  }
}

```

**Fig. 1** An entry in the collection of smooth Fano polytopes. Naming of the fields is in this example taken from standard properties of objects in `polymake`. However, there are no restrictions on field names

particular software package. Each collection group also has a separate collection *type\_information* that specifies the format of an entry in a collection and allows to store information applicable to all data sets in this collection, e.g., methods for import and export of the data. The web interface at [db.polymake.org](http://db.polymake.org) allows independent and searchable access to all data sets in `polyDB`.

There are currently five collections, grouped into four collection groups contained in `polyDB`. We give a brief introduction to each of the collections.

- The collection group *Lattice Polytopes* has the collection *Smooth Reflexive Polytopes* that contains low dimensional smooth reflexive polytopes based on the algorithm of Øbro [33]. Øbro used his algorithm to compute the data up to dimension 8. Later, dimension 9 was computed with an improved implementation of the algorithm by Lorenz and the author. There are 9,060,505 such polytopes.
- The collection group *Objects in Tropical Geometry* has two collections. The collection *Tropical Oriented Matroids* contains a list of 71 known non-realizable tropical oriented matroids. This data was provided by Horn [17]. The collection *Full-dimensional Polytopes in TP3* contains all 1013 polytopes in 3-dimensional tropical projective space. The collection was generated by Constantin Fischer from data of Joswig and Kulas [21] and Tran [39]. See [20] for a description.
- The collection group *Special Polytopes* has the collection *Faces of Birkhoff Polytopes* which contains all 5371 combinatorial types of faces up to dimension 8 of the Birkhoff polytope in any dimension [34].

- The collection group *Matroids* has the collection *Matroids on at most 12 elements*. This collection contains all 32,401,446 small matroids as computed by Miyata et al. [11, 12, 30].

Further collections are in preparation.

### 3 The `polymake` Interface to `polyDB`

The initiative for `polyDB` was started in 2013 by Silke Horn and the author as an *extension* for the software package `polymake` [19] with associated database. With the latest version 3.1 of `polymake` [38], released in March 2017, the interface to the database has been turned into a *bundled extension* for `polymake` that is directly delivered with the software and the database has been set up as an independent project.

However, the software package `polymake` currently provides the only interface for import of data into the database and methods to access and use it for computations. Given a search query, i.e. a list of restrictions on the properties of an object, MongoDB allows the retrieval of a single object satisfying the query, an array with all objects satisfying the query or a cursor that returns objects from the result set one after another. All three methods are also implemented in `polymake`. The implementation is based on the perl MongoDB driver [1]. With

```
polytope > db_info();
DATABASE: LatticePolytopes
This database contains various classes of lattice polytopes.

Collection: SmoothReflexive
A complete collection of smooth reflexive lattice polytopes
in dimensions up to 9, up to lattice equivalence. [...]
```

we can query which collection groups are available. The collection group and collection we want to use for our search are then specified with the keywords `db` and `collection` in any access function. The query itself is given as a perl hash. The query is not processed by `polymake` but directly handed over to MongoDB, so it allows all queries specified in the MongoDB query language. A specification of the full query language and its use from within perl can be found in the documentation of MongoDB [31] and the perl driver for it [1].

Here is an example returning an array of results.

```
polytope > $parray=db_query({"DIM"=>3, "N_FACETS"=>5},
polytope(2) > db=>"LatticePolytopes",
polytope(3) > collection=>"SmoothReflexive");
polytope > print $parray->size;
4
```

This shows that there are four polytopes in the collection *SmoothReflexive* that have dimension 3 and 5 facets. Using a loop over this array or a database cursor we can

check properties of each object returned. For example

```
polytope > $cursor=db_cursor({"DIM"=>3, "N_FACETS"=>5},
polytope (2) > db=>"LatticePolytopes",
polytope (3) > collection=>"SmoothReflexive");
polytope > while ( !$cursor->at_end() ) {
polytope (2) > $p=$cursor->next();
polytope (3) > print $p->N_LATTICE_POINTS, " ";
polytope (4) > }
34 30 31 30
```

defines a cursor over the collection *SmoothReflexive* successively returning all polytopes that satisfy the restrictions given in the query, i.e., that have five facets in dimension 3. Here it tells us that among the four polytopes found above, two have 30, one has 31 and one has 34 lattice points.

## 4 Decomposing Smooth Fano Polytopes

We illustrate the use of `polyDB` and its interface to `polymake` with a computation that uses the collection *SmoothReflexive* in the collection group *LatticePolytopes* to compute decompositions of smooth Fano polytopes in dimensions 1 to 8. With our computations we start a new statistics that counts how many of the smooth Fano polytopes can be generated from lower dimensional smooth Fano polytopes with some simple known polytope construction method that preserves both smoothness and reflexivity of the polytope. We consider three methods in this paper and determine how many of the smooth Fano polytopes in these dimensions are

- free sums of two smooth Fano polytopes
- a smooth skew bipyramid over a smooth Fano polytope as defined in [5], or
- a generalized simplex sum of a smooth Fano polytope with a smooth simplex.

This new construction method will be defined below.

All smooth skew bipyramids and many of the free sums are also generalized smooth simplex sums. We will also provide the total number of smooth Fano polytopes that can be decomposed with at least one of these constructions. The results are collected in Table 1.

We briefly explain the relevant notions. More background can, e.g., be found in the book of Ewald [10]. Let  $P \subseteq \mathbb{R}^d$  be a polytope with vertices  $v_1, \dots, v_r \in \mathbb{R}^d$ , i.e.,

$$P := \text{conv}(v_1, \dots, v_r) \tag{1}$$

is the convex hull of these points and none of the  $v_i$  can be omitted in the definition. We assume that  $P$  is full dimensional, i.e., the affine hull of  $P$  is  $\mathbb{R}^d$  (otherwise we can pass to a subspace). A polytope can equally be given as the intersection of a

finite number of half-spaces in the form

$$P = \{ x \mid Ax \leq b \} \tag{2}$$

for some  $A \in \mathbb{R}^{s \times d}$  and  $b \in \mathbb{R}^s$ . We can again assume that no inequality is redundant in this definition. In this case the rows of  $A$  are the *facet normals* of  $P$ . A *facet*  $F$  of  $P$  is the set of all  $x \in P$  that satisfy one of the inequalities in (2) with equality. A *face* of  $P$  is the common intersection in  $P$  of a subset of the facets (this may be empty). The vertices, which are the faces of dimension 0, are in the common intersection of at least  $d$  facets.

If  $0$  is strictly contained in the interior of  $P$ , then the *polar* or *dual polytope* is defined as

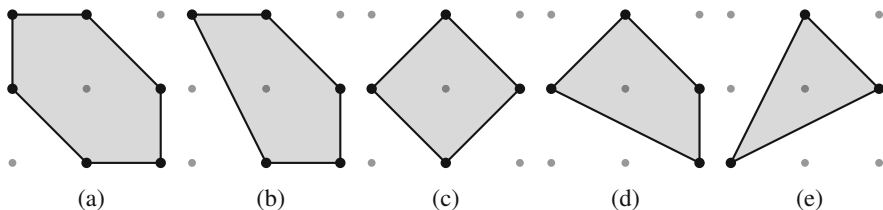
$$P^\vee := \{ v \mid \langle v, x \rangle \leq 1 \text{ for all } x \in P \}.$$

In fact, a finite subset of the inequalities in this definition suffice to define  $P^\vee$  (those corresponding to the vertices of  $P$ ), so that  $P^\vee$  is again a polytope. Further, we have  $(P^\vee)^\vee = P$ .

A lattice  $\Lambda$  is the integral span of a linearly independent set of vectors in  $\mathbb{R}^n$ . Up to a linear transformation we can assume that  $\Lambda$  is the integer lattice  $\mathbb{Z}^d \subset \mathbb{R}^n$ , and by passing to a subspace we can assume that  $n = d$ . With these assumptions a polytope  $P$  is a *lattice polytope* if all its vertices are in  $\mathbb{Z}^d$ .

In this case we can assume that both  $A$  and  $b$  are integral in (2), and that the greatest common divisor of the entries of each row of  $A$  (i.e., of the entries of each facet normal) is 1. A lattice polytope  $P$  is *reflexive* if  $P^\vee$  is again a lattice polytope. In this case  $b = \mathbf{1}$  in (2), and for both  $P$  and  $P^\vee$  the origin is the unique interior lattice point.  $P$  is *smooth* if the vertices of any facet of  $P$  are a lattice basis of  $\mathbb{Z}^d$ . In this case  $0$  is a strictly interior point of  $P$  and each facet has exactly  $d$  vertices, so  $P$  is *simplicial*. Moreover, the polar polytope is again a lattice polytope (in the dual lattice) whose vertices are the facet normals (the rows of  $A$ ), so  $P$  is also reflexive. Note that in the literature sometimes the polytopes polar to the ones defined here are called smooth.

It follows from a result of Hensley [15] and Lagarias and Ziegler [28] that there are only finitely many smooth reflexive polytopes in each dimension up to lattice equivalence (affine transformations preserving  $\mathbb{Z}^d$ ), as reflexive polytopes have exactly one interior lattice point. See Fig. 2 for the list of such polytopes in dimension 2. The complete list is contained in polyDB for  $d \leq 9$  in the collection *SmoothReflexive* of the database *LatticePolytopes*. Note however, that in the database we follow the above mentioned alternative definition and list the duals of the ones defined here. Yet, for the purpose of the following constructions it is easier to work with the definition given above, so we will use that one in the following. This requires that in the scripts we use for our computations below we polarize the polytopes obtained from the database. Sometimes this is, however, only done implicitly. This saves computation time, as it follows from the design of `polymake` that for reflexive polytopes the facets of the polytope are the vertices



**Fig. 2** The five 2-dimensional smooth Fano polytopes. (a)  $P_6$ . (b)  $P_5$ . (c)  $P_{4a}$ . (d)  $P_{4b}$ . (e)  $P_3$

of its dual. Also, as we will see below, most constructions can also be given for the duals of the polytopes.

We introduce several methods to construct a smooth Fano polytope from smaller ones. The most well known construction is the *free sum* of two polytopes  $P \subseteq \mathbb{R}^a$  and  $Q \subseteq \mathbb{R}^b$  that both contain the origin in their interior. This is the polytope

$$P \oplus Q := \text{conv}(\{(v, 0) \in \mathbb{R}^{a+b} \mid v \in P\} \cup \{(0, w) \in \mathbb{R}^{a+b} \mid w \in Q\}) .$$

We can also define this on the dual side. The *product* of polytopes  $P$  and  $Q$  is the polytope

$$P \times Q := \{ (x, y) \mid x \in P, y \in Q \} .$$

Then, if  $P$  and  $Q$  contain the origin in their interior,

$$P \oplus Q = (P^\vee \times Q^\vee)^\vee . \tag{3}$$

We will use this dual definition for the detection of free sums among the smooth Fano polytopes. See Fig. 3a for an example.

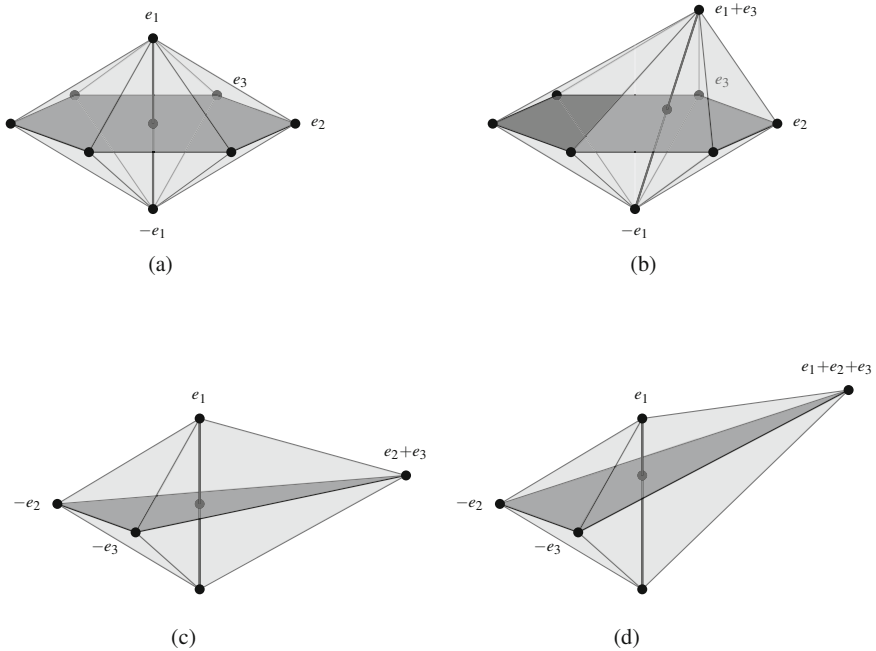
A *bipyramid* over a polytope  $P$  is the free sum of  $P$  with a segment  $S$  containing the origin in the interior. More generally, we say that  $Q$  is a *skew bipyramid* over  $P$  if  $Q$  has the same combinatorial type (the same face lattice) as a bipyramid over  $P$ . The two vertices coming from vertices of  $S$  are the two *apices* of  $Q$ .

If  $P$  is a smooth  $d$ -dimensional Fano polytope then we call the free sum with the segment  $[-1, 1]$  the *smooth bipyramid* over  $P$ . Let  $v$  be a vertex of  $P$  and  $\bar{v}$  its embedding into  $\mathbb{R}^{d+1}$  by adding a 0 at the end. Then the *smooth skew bipyramid* for vertex  $v$  as defined by Assarf et al. [5] is the polytope

$$\text{SBipy}(P, v) := \text{conv}(P \times \{0\} \cup \{-e_{d+1}, \bar{v} + e_{d+1}\}) .$$

Figure 3b shows an example of this definition. More generally, we say that  $Q$  is a *smooth generalized skew bipyramid* over  $P$  if  $Q$  is a skew bipyramid over  $P$  such that the two apices have lattice distance 1 from  $P$ . This class contains all smooth bipyramids and smooth skew bipyramids. The following proposition is an extension of Lemmas 1, 2 and 3 of [5]. The proof easily carries over into this more general setting.





**Fig. 3** Polytope constructions. (a) The free sum of a hexagon and a segment. This is at the same time also a proper bipyramid over the hexagon. (b) A skew bipyramid of a hexagon. The top apex has been shifted to  $e_1 + e_3$ . (c) A generalized simplex sum of a segment and a triangle. (d) Another generalized simplex sum of a segment with a triangle

**Proposition 4.1** *Let  $P$  and  $Q$  be smooth Fano polytopes. Then the free sum  $P \oplus Q$ , the smooth bipyramid and any smooth (generalized) skew bipyramid over  $P$  are again smooth Fano polytopes.*  $\square$

We further generalize this construction. Let  $P \subseteq \mathbb{R}^a$  be a smooth Fano polytope and  $Q \subseteq \mathbb{R}^b$  a smooth Fano simplex (this is unique up to lattice equivalence). Let  $v$  be a vertex of  $Q$ . Then  $R := P \oplus Q$  is a smooth Fano polytope and also any polytope  $R'$  obtained from  $R$  by replacing  $v$  with a lattice point  $v'$  in the hyperplane  $\mathbb{R}^a + v \subseteq \mathbb{R}^{a+b}$ , as long as  $R$  and  $R'$  have the same combinatorial type. This is again a simple extension of the proposition above. We call those polytopes *smooth generalized simplex sums*. Figure 3c, d shows two examples. Observe that any smooth (generalized skew) bipyramid is a simplex sum.

We can use `polymake` and `polyDB` to detect all free sums and smooth generalized simplex sums among the smooth Fano polytopes. Clearly, these two constructions overlap in various ways. Any proper bipyramid over all polytope  $P$  is also the free sum of a  $P$  with a segment, and polytopes may have more than one possible decomposition into a free sum. Many of the various possibilities to place the vertex  $v'$  for a generalized smooth simplex sum are lattice equivalent. Our

approach to detect all different instances is as follows: For a fixed dimension  $d$  we consider all possible splits of  $d$  as a sum of dimensions  $a$  and  $b$  and compute all free sums of smooth Fano polytopes in these two dimensions and all simplex sums of an  $a$ -dimensional smooth Fano polytope with a  $b$ -dimensional simplex. For each such polytope we run through the list of  $d$ -dimensional smooth Fano polytopes, check for lattice equivalence and store the name of the polytope we have found. We could also store the way we obtained it alongside, so that in the end we have a list of all possible splits for a given  $d$ -dimensional smooth Fano polytope.

We did the computation up to dimension 8. The results are given in Table 1. The free sums can be obtained with the small scripts given in Fig. 4. The first script `identify_smooth_polytope` takes a smooth Fano polytope, identifies it in the database and returns its name. The identification is based on the `polymake` function `lattice_isomorphic_smooth_polytopes`, that reduces the check whether two lattice polytopes are lattice isomorphic to a colored graph isomorphism problem (which is solved using `bliss` [23] or `nauty` [29]). Note that there is also the extension `LatticeNormalization` [18] to `polymake` that computes the lattice normal form of a lattice polytope (see [14] for a definition), but the reduction to colored graph isomorphism is more efficient for smooth polytopes. The simpler problem of checking combinatorial isomorphisms (i.e., graph isomorphism) can also be done with the `polymake`-function `canonical_hash` (also based on `bliss` or `nauty`). The second function `all_free_sums_in_dim` computes all possible free sums that lead to a  $d$ -dimensional smooth Fano polytope. As the database contains the polytopes dual to the ones we consider we use (3) and compute products instead of sums to avoid explicit dualization. For each product the function calls `identify_smooth_polytope` to identify it in the database. The function returns a list of all names (`_ids`) found in this way. If `splitinfo` is set to 1 it also returns all pairs of summands.

For the computation of the smooth generalized simplex sums we used the function `all_skew_simplex_sums_in_dim` available at [35]. For each combination of an  $a$ -dimensional smooth Fano polytope and a  $b$ -dimensional simplex with  $d = a + b$  we compute all possible lattice points for the shifted vertex  $v'$ , construct the polytope and again use `identify_smooth_polytope` to identify it in the database. Computation of all possible  $v'$  requires the computation of all lattice points in the hyperplane  $\mathbb{R}^a + v \subseteq \mathbb{R}^{a+b}$  that lead to a lattice polytope with the same combinatorial type as the proper free sum. This can be reduced to enumerating lattice points in the interior of a polytope, which is done in `polymake` via the interface to `Normaliz` [9]. As above the function returns a list of `ids`, and also all possible decompositions into a simplex sum if `splitinfo` is set to 1. You can save the scripts to a file in the current folder and load this into `polymake` via

```
polytope> script(<filename>);
```

```

use application "polytope";

sub identify_smooth_fano_in_polydb {
  my $p = shift;
  my $d = $p->DIM;
  my $nlp = new Int($p->N_LATTICE_POINTS);
  my $parray=db_query({"DIM"=>$d, "N_VERTICES"=>$p->N_VERTICES, "
    N_FACETS"=>$p->N_FACETS,
    "N_LATTICE_POINTS"=>$nlp, }, db=>"LatticePolytopes",
    collection=>"SmoothReflexive");
  foreach my $c ( @$parray ) {
    if ( lattice_isomorphic_smooth_polytopes($c,$p) ) { return $c
      ->name; }
  }
  die "polytope not found\n";
}

sub all_free_sums_in_dim {
  my ($d,%options) = @_;
  my $list;
  if ( $options{"splitinfo"} ) {
    $list = new Map<String,Set<Pair<String,String> > >;
  } else {
    $list = new Set<String>;
  }
  my $cur_options = { db=>"LatticePolytopes", collection=>"
    SmoothReflexive" };
  foreach my $n (1..$d/2) {
    my $curl=db_cursor({"DIM"=>$n}, $cur_options);
    while ( !$curl->at_end() ) {
      my $c1 = $curl->next();
      my $cur2=db_cursor({"DIM"=>$d-$n}, $cur_options );
      while ( !$cur2->at_end() ) {
        my $c2 = $cur2->next();
        my $name = identify_smooth_fano_in_polydb(product($c1,$c2))
          ;
        if ( $options{"splitinfo"} ) {
          my $split = new Pair<String,String>($c1->name,$c2->name);
          $list->{$name} += $split;
        } else {
          $list += $name;
        }
      }
    }
  }
  return $list;
}

```

**Fig. 4** A function to detect all free sums among the smooth Fano polytopes. This is a shortened version of the script given at [35]

Then the classification, e.g. in dimension 4, is obtained with

```

polytope > $fs = all_free_sums_in_dim(4);
polytope > print $fs->size;
28
polytope > $sb = all_skew_bipyramids_in_dim(4);
polytope > print $sb->size;
57
polytope > $s = new Set<String>;
polytope > foreach (1..4) {
polytope(2) > $st = skew_simplex_sums_in_dim(4,$_);
polytope(3) > print $st->size, " ";
polytope(4) > $s += $st;
polytope(5) > }
66 31 4 1
polytope > print $s->size;
93
polytope > print (($fs+$s)->size);
96
    
```

The scripts containing the functions `all_skew_bipyramids_in_dim` for skew bipyramids and `skew_simplex_sums_in_dim` for generalized smooth simplex sums are available from [35] and allow to store the possible decompositions. Note that the computation time for the decompositions grows quickly in the dimension. While dimension 4 runs in a few minutes on an Intel Xeon E5-4650, computations in dimension 8 took over a month.

From these computations we can, e.g., see that we have three different decompositions of the dual of the 5-dimensional polytope with index  $F.5D.0116$ . Its vertices are the rows of the matrix in Table 2a. We can decompose this into three different simplex sums. One is over dual of the 3-dimensional polytope  $P_3$  with index  $F.3D.0112$ . This is shown in Table 2b, where the shaded part corresponds to the vertices of  $P_3$ . The shifted vertex of the triangle is  $[0, 0, -2, 1, 1]$ . Note that the vertices are given as obtained by dualization from the database. Hence, the equality

**Table 2** Simplex sums leading to the dual of  $F.5D.0116$

0 0 0 0 1	0 1 1 0 0	-1 0 0 0 0	-1 0 0 0 0
0 0 1 0 -1	-1 0 0 0 0	0 0 0 -1 0	0 0 0 -1 0
0 0 -1 0 0	0 0 1 0 0	0 1 0 1 0	0 1 0 -1 0
0 0 0 -1 0	0 -1 0 0 0	0 -1 0 0 0	0 -1 0 0 0
0 0 0 0 -1	0 0 -1 0 0	0 0 0 1 0	0 0 0 1 0
-1 0 0 0 0	1 0 -1 0 0	0 0 -1 0 0	0 0 -1 0 0
0 -1 0 0 0	0 0 0 -1 0	1 0 1 2 0	1 0 1 2 0
0 1 0 0 1	0 0 0 0 -1	0 0 0 0 -1	0 0 0 0 -1
1 0 0 1 2	0 0 -2 1 1	0 0 0 -1 1	0 0 0 1 1

(a) Dual of  $F.5D.0116$

(b) Dual of  $F.3D.0112$  extended with skew triangle

(c) Dual of  $F.4D.0008$  extended with segment

(d) Dual of  $F.4D.0019$  extended with segment

is not directly visible from the vertices, as the two polytopes differ by a lattice isomorphism. We can check this with `polymake`.

```
polytope > $p3_ext = new Polytope(VERTICES=>
polytope (2) > [[1,0,1,1,0,0], [1,-1,0,0,0,0], [1,0,0,1,0,0],
polytope (3) > [1,0,-1,0,0,0], [1,0,0,-1,0,0], [1,1,0,-1,0,0],
polytope (4) > [1,0,0,0,-1,0], [1,0,0,0,0,-1],
polytope (5) > [1,0,0,-2,1,1]);
polytope > $p5 = new Polytope(VERTICES=>
polytope (2) > [[1,0,0,0,0,1], [1,0,0,1,0,-1], [1,0,0,-1,0,0],
polytope (3) > [1,0,0,0,-1,0], [1,0,0,0,0,-1], [1,-1,0,0,0,0],
polytope (4) > [1,0,-1,0,0,0], [1,0,1,0,0,1], [1,1,0,0,1,2]);
polytope > print lattice_isomorphic_smooth_polytope (
polytope (2) > polarize($p3_ext),polarize($p5));
1
```

Here, the variable `$p3_ext` contains the polytope  $P_3$  and `$p5` is  $P_5$ . As above we need to dualize for the isomorphism check. The check returns 1, which is the true-value for `polymake`.

The other two decompositions are over the 4-dimensional polytopes  $P_4^1$  and  $P_4^2$  with index F.4D.0008 and F.4D.0019. Those are shown in Table 2c and d. Again, the vertices of  $P_4^1$  and  $P_4^2$  are shaded. The shifted vertices of the 1-dimensional simplex are in the last line.

With this simple computation we have seen that over 80% of the smooth Fano polytopes can be obtained from at least one of the constructions considered here. Hence, for a structural description of all smooth Fano polytopes it suffices to look at the remaining less than 20%.

## References

1. A perl driver for MongoDB (2016). Available at [search.cpan.org/perldoc?MongoDB](http://search.cpan.org/perldoc?MongoDB)
2. O. Aichholzer, Extremal properties of 0/1-polytopes of dimension 5, in *Polytopes—Combinatorics and Computation (Oberwolfach, 1997)*, vol. 29, ed. by G. Kalai, G.M. Ziegler, DMV Seminar (Birkhäuser, Basel, 2000), pp. 111–130, [http://www.ist.tugraz.at/staff/aichholzer/research/rp/rcs/info01poly](http://www.ist.tugraz.at/staff/aichholzer/research/rp/rcs/info01poly/www.ist.tugraz.at/staff/aichholzer/research/rp/rcs/info01poly)
3. R. Altman, J. Gray, Y.H. He, V. Jejjala, B.D. Nelson, A Calabi-Yau database: threefolds constructed from the Kreuzer-Skarke list. *J. High Energy Phys.* (2) **158**, front matter+48 (2015)
4. B. Assarf, E. Gawrilow, K. Herr, M. Joswig, B. Lorenz, A. Paffenholz, T. Rehn, `Polymake` in linear and integer programming. *Math. Program. Comput.* (2014, to appear). Available at [arxiv:1408.4653](http://arxiv.org/abs/1408.4653)
5. B. Assarf, M. Joswig, A. Paffenholz, Smooth Fano polytopes with many vertices. *Discrete Comput. Geom.* **52**, 153–194 (2014)
6. B. Baumeister, C. Haase, B. Nill, A. Paffenholz, On permutation polytopes. *Adv. Math.* **222**(2), 431–452 (2009)
7. H.U. Besche, B. Eick, E. O’Brien, Small groups library (2002). Available at [www.icm.tu-bs.de/ag\\_algebra/software/small/](http://www.icm.tu-bs.de/ag_algebra/software/small/)
8. G. Brown, A. Kasprzyk, The graded rings database. Available at [www.grdb.co.uk](http://www.grdb.co.uk)

9. W. Bruns, B. Ichim, T. Römer, R. Sieg, C. Söger, Normaliz. Algorithms for rational cones and affine monoids. *J. Algebra* **324**, 1098–1113 (2010). Available at [www.normaliz.uni-osnabrueck.de](http://www.normaliz.uni-osnabrueck.de)
10. G. Ewald, *Combinatorial Convexity and Algebraic Geometry*. Graduate Texts in Mathematics, vol. 168 (Springer, New York, 1996)
11. L. Finschi, K. Fukuda, Combinatorial generation of small point configurations and hyperplane arrangements, in *Discrete and Computational Geometry*. Algorithms and Combinatorics, vol. 25 (Springer, Berlin, 2003), pp. 425–440
12. K. Fukuda, H. Miyata, S. Moriyama, Complete enumeration of small realizable oriented matroids. *Discrete Comput. Geom.* **49**(2), 359–381 (2013)
13. H.G. Gräbe, Semantic-aware fingerprints of symbolic research data, in *Mathematical Software ICMS 2016*, vol. 9725, ed. by G.M. Greuel, T. Koch, P. Paule, A. Sommese. Theoretical Computer Science and General Issues (Springer, Cham, 2016)
14. R. Grinis, A. Kasprzyk, Normal forms of convex lattice polytopes (2013). Available at [arxiv:1301.6641](https://arxiv.org/abs/1301.6641)
15. D. Hensley, Lattice vertex polytopes with interior lattice points. *Pac. J. Math.* **105**(1), 183–191 (1983)
16. A. Higashitani, Equivalence classes for smooth Fano polytopes (2015). Available at [arxiv:1503.06434](https://arxiv.org/abs/1503.06434)
17. S. Horn, Tropical oriented matroids and cubical complexes. Ph.D. thesis, 2012
18. S. Horn, A. Paffenholz, Lattice normalization: a polymake extension for lattice isomorphism checks (2014). Available at [github.com/apaffenholz/lattice\\_normalization](https://github.com/apaffenholz/lattice_normalization)
19. S. Horn, A. Paffenholz, polyDB: a database extension for polymake (2014). Available at [github.com/solros/poly\\_db](https://github.com/solros/poly_db)
20. M. Joswig, Polytopes in the tropical projective 3-torus (2016). Available at [home-pages.math.tu-berlin.de/~joswig/polymake/data/](http://home-pages.math.tu-berlin.de/~joswig/polymake/data/)
21. M. Joswig, K. Kulas, Tropical and ordinary convexity combined. *Adv. Geom.* **10**(2), 333–352 (2010)
22. M. Joswig, B. Müller, A. Paffenholz, polymake and lattice polytopes, in *21st International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2009)*, ed. by C. Krattenthaler, V. Strehl, M.N. Kauers. Discrete Mathematics and Theoretical Computer Science Proceedings, AK (Association of Discrete Mathematics and Theoretical Computer Science, Nancy, 2009), pp. 491–502
23. T. Junttila, P. Kaski, bliss: a tool for computing automorphism groups and canonical labelings of graphs (2011). Available at [www.tcs.hut.fi/Software/bliss/](http://www.tcs.hut.fi/Software/bliss/)
24. A.M. Kasprzyk, B. Nill, Reflexive polytopes of higher index and the number 12 (2011). Available at [arxiv:1107.4945](https://arxiv.org/abs/1107.4945)
25. T. Koch, T. Achterberg, E. Andersen, O. Bastert, T. Berthold, R.E. Bixby, E. Danna, G. Gamrath, A.M. Gleixner, S. Heinz, A. Lodi, H. Mittelmann, T. Ralphs, D. Salvagnin, D.E. Steffy, K. Wolter, MIPLIB 2010. *Math. Program. Comput.* **3**(2), 103–163 (2011). Available at [mpc.zib.de/index.php/MPC/article/view/56/28](http://mpc.zib.de/index.php/MPC/article/view/56/28)
26. M. Kreuzer, H. Skarke, Calabi-yau data (1997). Available at [hep.itp.tuwien.ac.at/~kreuzer/CY/](http://hep.itp.tuwien.ac.at/~kreuzer/CY/)
27. K. Kulas, Coarse types of tropical matroid polytopes (2010). Available at [arxiv:1012.3053](https://arxiv.org/abs/1012.3053)
28. J.C. Lagarias, G.M. Ziegler, Bounds for lattice polytopes containing a fixed number of interior points in a sublattice. *Can. J. Math.* **43**(5), 1022–1035 (1991). <https://doi.org/10.4153/CJM-1991-058-4>
29. B. McKay, nauty 2.5 (2013). Available at [cs.anu.edu.au/~bdm/nauty/](http://cs.anu.edu.au/~bdm/nauty/)
30. H. Miyata, S. Moriyama, K. Fukuda, Classification of oriented matroids (2013). Available at [www.imai.is.s.u-tokyo.ac.jp/~hmiyata/oriented\\_matroids/](http://www.imai.is.s.u-tokyo.ac.jp/~hmiyata/oriented_matroids/)
31. MongoDB v3.2 (2016). Available at [www.mongodb.com](http://www.mongodb.com)
32. B. Nill, A. Paffenholz, Examples of Kähler-Einstein toric Fano manifolds associated to non-symmetric reflexive polytopes. *Beitr. Algebra Geom.* **52**(2), 297–304 (2011). <https://doi.org/10.1007/s13366-011-0041-y>

33. M. Øbro, Classification of Smooth Fano Polytopes. Ph.D. thesis, University of Aarhus, 2007. Available at [pure.au.dk/portal/files/41742384/imf\\_phd\\_2008\\_moe.pdf](http://pure.au.dk/portal/files/41742384/imf_phd_2008_moe.pdf)
34. A. Paffenholz, Faces of Birkhoff polytopes. *Electron. J. Combin.* **22**(1), Paper 1.67, 36 (2015)
35. A. Paffenholz, `polymake_smooth_fano`: scripts for decompositions of Fano polytopes (2016). Available at [github.com/apaffenholz/polymake\\_smooth\\_fano](https://github.com/apaffenholz/polymake_smooth_fano)
36. The GAP Group, GAP – Groups, Algorithms, and Programming, Version 4.4.12. The GAP Group (2008). Available at [www.gap-system.org](http://www.gap-system.org)
37. The Magma Group, Magma computational algebra system (2017). Available at [magma.maths.usyd.edu.au/magma/](http://magma.maths.usyd.edu.au/magma/)
38. The `polymake`-Team, `polymake` release 3.1 (2017). Available at [github.com/polymake/polymake](https://github.com/polymake/polymake)
39. N.M. Tran, Polytopes and tropical eigenspaces: cones of linearity. *Discrete Comput. Geom.* **51**(3), 539–558 (2014)

# Construction of Neron Desingularization for Two Dimensional Rings



Gerhard Pfister and Dorin Popescu

**Abstract** Let  $u : A \rightarrow A'$  be a regular morphism of Noetherian rings and  $B$  an  $A$ -algebra of finite type. Then any  $A$ -morphism  $v : B \rightarrow A'$  factors through a smooth  $A$ -algebra  $C$ , that is  $v$  is a composite  $A$ -morphism  $B \rightarrow C \rightarrow A'$ . This theorem called General Neron Desingularization was first proved by the second author (Popescu, Nagoya Math J 100:97–126, 1985). Later different proofs were given by André (Cinq exposés sur la désingularisation. Handwritten manuscript Ecole Polytechnique Fédérale de Lausanne, 1991), Swan (Neron-Popescu desingularization. In: Kang (ed) Algebra and geometry. International Press, Cambridge, pp 135–192, 1998) and Spivakovsky (J Am Math Soc 294:381–444, 1999). All the proofs are not constructive. In Pfister and Popescu (J Symb Comput 80:570–580, 2017) the authors gave a constructive proof together with an algorithm to compute the Neron Desingularization for 1-dimensional local rings. In this paper we go one step further. We give an algorithmic proof of the General Neron Desingularization theorem for 2-dimensional local rings and morphisms with small singular locus. The main idea of the proof is to reduce the problem to the one-dimensional case. Based on this proof we give an algorithm to compute the desingularization.

**Keywords** Smooth morphisms • Regular morphisms • Neron desingularization

**2010 Mathematics Subject Classification** Primary 13B40; Secondary 14B25,13H05

---

G. Pfister (✉)

Fachbereich Mathematik, Universität Kaiserslautern, Postfach 3049, 67653 Kaiserslautern, Germany

e-mail: [pfister@mathematik.uni-kl.de](mailto:pfister@mathematik.uni-kl.de)

D. Popescu

Simion Stoilow Institute of Mathematics of the Romanian Academy, Research Unit 5, University of Bucharest, P.O. Box 1-764, Bucharest 014700, Romania

e-mail: [dorin.popescu@imar.ro](mailto:dorin.popescu@imar.ro)

© Springer International Publishing AG, part of Springer Nature 2017

G. Böckle et al. (eds.), *Algorithmic and Experimental Methods in Algebra, Geometry, and Number Theory*,

[https://doi.org/10.1007/978-3-319-70566-8\\_24](https://doi.org/10.1007/978-3-319-70566-8_24)



## 1 Introduction

The General Neron Desingularization Theorem, first proved by the second author has many important applications. One application is the generalization of Artin's famous approximation theorem [2, 8, 9].

Let us recall some definitions. A ring morphism  $u : A \rightarrow A'$  has *regular fibers* if for all prime ideals  $P \in \text{Spec } A$  the ring  $A'/PA'$  is a regular ring, i.e. its localizations are regular local rings. It has *geometrically regular fibers* if for all prime ideals  $P \in \text{Spec } A$  and all finite field extensions  $K$  of the fraction field of  $A/P$  the ring  $K \otimes_{A/P} A'/PA'$  is regular. If for all  $P \in \text{Spec } A$  the fraction field of  $A/P$  has characteristic 0 then the regular fibers of  $u$  are geometrically regular fibers. A flat morphism  $u$  is *regular* if its fibers are geometrically regular. If  $u$  is regular of finite type then  $u$  is called *smooth*. A localization of a smooth algebra is called *essentially smooth*.

**Theorem 1.1 (General Neron Desingularization, André [1], Neron [5], Popescu [7–9], Swan [12], Spivakovsky [11])** *Let  $u : A \rightarrow A'$  be a regular morphism of Noetherian rings and  $B$  an  $A$ -algebra of finite type. Then any  $A$ -morphism  $v : B \rightarrow A'$  factors through a smooth  $A$ -algebra  $C$ , that is  $v$  is a composite  $A$ -morphism  $B \rightarrow C \rightarrow A'$ .*

The proof of this theorem is not constructive. Constructive proofs for one-dimensional rings were given in Popescu and Popescu [10] and Pfister and Popescu [6]. In this paper we will treat the 2-dimensional case. The idea is to reduce the problem to the one-dimensional case. We will choose a suitable element  $a \in A$  and consider  $\bar{A} = A/aA$ ,  $\bar{B} = \bar{A} \otimes_A B$ ,  $\bar{A}' = A'/aA'$ ,  $\bar{v} = \bar{A} \otimes_A v : \bar{B} \rightarrow \bar{A}'$  to find a desingularization  $\bar{B} \rightarrow \bar{D} \rightarrow \bar{A}'$  induced by a smooth  $A$ -algebra  $D$ . This desingularization can then be lifted to a desingularization  $B \rightarrow C \rightarrow A'$ .

For the computational part we have the following assumptions: Let  $k$  be a field,  $x = (x_1, \dots, x_n)$  and  $J \subset k[x]$  be an ideal. We assume

$$A = (k[x]/J)_{\langle x \rangle} \text{ is Cohen-Macaulay of dimension 2, } A' \text{ is its completion}$$

and  $u$  the inclusion. The images of the morphism  $v : B \rightarrow A'$  need not to be in  $A$ , i.e. the input for the algorithm can only be an approximation of  $v$  by polynomials up to a given bound. The bound to obtain a desingularization of  $v$  depends also on the ring  $B$  and is usually not known in advance. If the given bound is not good enough the algorithm will fail. In this case the bound has to be enlarged and the algorithm has to be restarted with new approximations of  $v$ .

The case that the image of  $v$  is already in  $A$  is trivial because in this case we can use the smooth  $A$ -algebra  $C = A$  as desingularization.

## 2 Constructive General Neron Desingularization

Let  $u : A \rightarrow A'$  be a flat morphism of Noetherian local rings of dimension 2. Suppose that the maximal ideal  $\mathfrak{m}$  of  $A$  generates the maximal ideal of  $A'$  and the

completions of  $A, A'$  are isomorphic. Moreover suppose that  $A'$  is Henselian, and  $u$  is a regular morphism.

Let  $B = A[Y]/I, Y = (Y_1, \dots, Y_n)$ . If  $f = (f_1, \dots, f_r), r \leq n$  is a system of polynomials from  $I$  then we can define the ideal

$$\Delta_f \text{ generated by all } r \times r\text{-minors of the Jacobian matrix } (\partial f_i / \partial Y_j).$$

After Elkik [4] let  $H_{B/A}$  be the radical of the ideal  $\sum_f ((f) : I) \Delta_f B$ , where the sum is taken over all systems of polynomials  $f$  from  $I$  with  $r \leq n$ . Then for  $\mathfrak{p} \in \text{Spec } B$

$$B_{\mathfrak{p}} \text{ is essentially smooth over } A \text{ if and only if } \mathfrak{p} \not\supset H_{B/A}$$

by the Jacobian criterion for smoothness. Thus  $H_{B/A}$  measures the non smooth locus of  $B$  over  $A$ .  $B$  is *standard smooth* over  $A$  if there exists  $f$  in  $I$  as above such that

$$B = ((f) : I) \Delta_f B.$$

The aim of this paper is to give an algorithmic proof of the following theorem.

**Theorem 2.1** *Any  $A$ -morphism  $v : B \rightarrow A'$  such that  $v(H_{B/A})A'$  is  $\mathfrak{m}A'$ -primary factors through a standard smooth  $A$ -algebra  $B'$ .*

*Proof* The idea is to find a suitable element  $a \in A$  such that we can use the one-dimensional result obtained for  $\bar{A} = A/(a), \bar{B} = \bar{A} \otimes_A B, \bar{A}' = A'/(aA'), \bar{v} = \bar{A} \otimes_A v$  to find a desingularization  $\bar{D}$  induced by a standard smooth  $A$ -algebra  $D$  (Lemma 2.3). This desingularization can then be lifted using  $D$ . To simplify the proof we assume that  $A$  is Cohen-Macaulay.

We choose  $\gamma, \gamma' \in v(H_{B/A})A' \cap A$  such that  $\gamma, \gamma'$  is a regular sequence in  $A$ , let us say  $\gamma = \sum_{i=1}^q v(b_i)z_i, \gamma' = \sum_{i=1}^q v(b_i)z'_i$  for some  $b_i \in H_{B/A}$  and  $z_i, z'_i \in A'$ . Set  $B_0 = B[Z, Z']/(f, \tilde{f})$ , where  $f = -\gamma + \sum_{i=1}^q b_i z_i \in B[Z], Z = (Z_1, \dots, Z_q), \tilde{f} = -\gamma' + \sum_{i=1}^q b_i z'_i \in B[Z'], Z' = (Z'_1, \dots, Z'_q)$  and let  $v_0 : B_0 \rightarrow A'$  be the map of  $B$ -algebras given by  $Z \rightarrow z, Z' \rightarrow z'$ . Replacing  $B$  by  $B_0$  we may suppose that  $\gamma, \gamma' \in H_{B/A}$ .

We need the following lemmata.

**Lemma 2.2**

1. ([7, Lemma 3.4]) *Let  $B_1$  be the symmetric algebra  $S_B(I/I^2)$  of  $I/I^2$  over  $B$ . Then  $H_{B/A}B_1 \subset H_{B_1/A}$  and  $(\Omega_{B_1/A})_{\gamma}$  is free over  $(B_1)_{\gamma}$  for any  $\gamma \in H_{B/A}$ .*
2. ([12, Proposition 4.6]) *Suppose that  $(\Omega_{B/A})_{\gamma}$  is free over  $B_{\gamma}$ . Let  $I' = (I, Y') \subset A[Y, Y'], Y' = (Y'_1, \dots, Y'_n)$ . Then  $(I'/I'^2)_{\gamma}$  is free over  $B_{\gamma}$ .*

<sup>1</sup>For the algorithm we have to choose  $\gamma, \gamma'$  more carefully:  $\gamma \equiv \sum_{i=1}^q b_i(y')z_i$  modulo  $(\gamma', \gamma'')$ ,  $\gamma' \equiv \sum_{i=1}^q b_i(y'')z'_i$  modulo  $(\gamma', \gamma'')$  with  $z_i, z'_i \in A$ , and  $y'_i \equiv v(Y_i)$  modulo  $\mathfrak{m}^N$  in  $A, N \gg 0$ .

<sup>2</sup>Let  $M$  be a finitely represented  $B$ -module and  $B^m \xrightarrow{(a_{ij})} B^n \rightarrow M \rightarrow 0$  a presentation then  $S_B(M) = B[T_1, \dots, T_n]/J$  with  $J = (\{\sum_{i=1}^n a_{ij} T_i\}_{j=1, \dots, m})$ .

3. ([9, Corollary 5.10]) Suppose that  $(I/I^2)_\gamma$  is free over  $B_\gamma$ . Then a power of  $\gamma$  is in  $((g) : I)\Delta_g$  for some  $g = (g_1, \dots, g_r)$ ,  $r \leq n$  in  $I$ .

Using (1) of Lemma 2.2 we can reduce the proof of Theorem 2.1 to the case when  $\Omega_{B_\gamma/A}$  and  $\Omega_{B_{\gamma'}/A}$  are free over  $B_\gamma$  respectively  $B_{\gamma'}$ . Let  $B_1$  be given by (1) of Lemma 2.2. The inclusion  $B \subset B_1$  has a retraction  $w$  which maps  $I/I^2$  to zero. For the reduction we change  $B, v$  by  $B_1, vw$ .

Using (2) from Lemma 2.2 we may reduce the proof to the case when  $(I/I^2)_\gamma$  (resp.  $(I/I^2)_{\gamma'}$ ) is free over  $B_\gamma$  (resp.  $B_{\gamma'}$ ). Indeed, since  $\Omega_{B_\gamma/A}$  is free over  $B_\gamma$  we see that changing  $I$  with  $(I, Y') \subset A[Y, Y']$  we may suppose that  $(I/I^2)_\gamma$  is free over  $B_\gamma$ . Similarly, for  $\gamma'$ .

Using (3) from Lemma 2.2 we may reduce the proof to the case when a power of  $\gamma$  (resp.  $\gamma'$ ) is in  $((f) : I)\Delta_f$  (resp.  $((f') : I)\Delta_{f'}$ ) for some  $f = (f_1, \dots, f_r)$ ,  $r \leq n$  and  $f' = (f'_1, \dots, f'_{r'})$ ,  $r' \leq n$  from  $I$ .

We may now assume that a power  $d$  (resp.  $d'$ ) of  $\gamma$  (resp.  $\gamma'$ ) has the form

$$d \equiv P = \sum_{i=1}^q M_i L_i \text{ modulo } I, \quad d' \equiv P' = \sum_{i=1}^{q'} M'_i L'_i \text{ modulo } I$$

for some  $r \times r$  (resp.  $r' \times r'$ ) minors  $M_i$  (resp.  $M'_i$ ) of  $(\partial f / \partial Y)$  (resp.  $(\partial f' / \partial Y)$ ) and  $L_i \in ((f) : I)$  (resp.  $L'_i \in ((f') : I)$ ).

The Jacobian matrix  $(\partial f / \partial Y)$  (resp.  $(\partial f' / \partial Y)$ ) can be completed with  $(n - r)$  (resp.  $(n - r')$ ) rows from  $A^n$  obtaining a square  $n$  matrix  $H_i$  (resp.  $H'_i$ ) such that  $\det H_i = M_i$  (resp.  $\det H'_i = M'_i$ ). This is easy using just the integers 0, 1.

Let  $\bar{A} = A/(d^3)$ ,  $\bar{B} = \bar{A} \otimes_A B$ ,  $\bar{A}' = A'/(d^3 A')$ ,  $\bar{v} = \bar{A} \otimes_A v$ . We will now construct a standard smooth  $A$ -algebra  $D$  and an  $A$ -morphism  $\omega : D \rightarrow A'$  such that  $y = v(Y) \in \text{Im } \omega + d^3 A'$ .

**Lemma 2.3** *There exists a standard smooth  $A$ -algebra  $D$  such that  $\bar{v}$  factors through  $\bar{D} = \bar{A} \otimes_A D$ .*

*Proof* Let  $y' \in A^n$  be such that  $y = v(Y) \equiv y'$  modulo  $(d^3, d'^3)A'$ , let us say  $y - y' \equiv d'^2 \epsilon$  modulo  $d^3$  for  $\epsilon \in d'A'^n$ . Thus

$$I(y') \equiv 0 \text{ modulo } (d^3, d'^3)A'.$$

Recall that we have  $d' \equiv P'$  modulo  $I$  and so  $P'(y') \equiv d'$  modulo  $(d^3, d'^3)$  in  $A$ . Thus

$$P'(y') \equiv d' s \text{ modulo } d^3 \text{ for a certain } s \in A \text{ with } s \equiv 1 \text{ modulo } d'.$$

Let  $G'_i$  be the adjoint matrix of  $H'_i$  and  $G_i = L'_i G'_i$ . We have  $G_i H'_i = H'_i G_i = M'_i L'_i \text{Id}_n$  and so

$$P'(y') \text{Id}_n = \sum_{i=1}^{q'} G_i(y') H'_i(y').$$

But  $H'_i$  is the matrix<sup>3</sup>  $(\partial f'_k / \partial Y_j)_{k \in [r'], j \in [n]}$  completed with some  $(n - r')$  rows of 0, 1. Especially we obtain

$$(\partial f' / \partial Y)G_i = M'_i L'_i (\text{Id}_{r'} | 0). \tag{1}$$

Then  $t_i := H'_i(y')\epsilon \in d'A'^n$  satisfies

$$G_i(y')t_i = M'_i(y')L'_i(y')\epsilon$$

and so

$$\sum_{i=1}^q G_i(y')t_i = P'(y')\epsilon \equiv d's\epsilon \text{ modulo } d^3.$$

It follows that

$$s(y - y') \equiv d' \sum_{i=1}^{q'} G_i(y')t_i \text{ modulo } d^3.$$

Note that  $t_{ij} = t_{i1}$  for all  $i \in [r']$  and  $j \in [n]$  because the first  $r'$  rows of  $H'_i$  does not depend on  $i$  (they are the rows of  $(\partial f' / \partial Y)$ ).

Let

$$h = s(Y - y') - d' \sum_{i=1}^{q'} G_i(y')T_i, \tag{2}$$

where  $T_i = (T_1, \dots, T_{r'}, T_{i,r'+1}, \dots, T_{i,n})$ ,  $i \in [q']$  are new variables. We will use also  $T_{ij} = T_i$  for  $i \in [r'], j \in [n]$  because it is convenient sometimes. The kernel of the map  $\bar{\phi} : \bar{A}[Y, T] \rightarrow \bar{A}'$  given by  $Y \rightarrow y, T \rightarrow t$  contains  $h$  modulo  $d^3$ . Since

$$s(Y - y') \equiv d' \sum_{i=1}^{q'} G_i(y')T_i \text{ modulo } h$$

and

$$f'(Y) - f'(y') \equiv \sum_j (\partial f' / \partial Y_j)(y')(Y_j - y'_j) \text{ modulo higher order terms in } Y_j - y'_j$$

---

<sup>3</sup>We use the notation  $[n] = \{1, \dots, n\}$ .

by Taylor’s formula. We see that for  $p' = \max_i \deg f'_i$  we have

$$s^{p'} f'(Y) - s^{p'} f'(y') \equiv \sum_j s^{p'-1} d'(\partial f' / \partial Y_j)(y') \sum_{i=1}^{q'} G_{ij}(y') T_{ij} + d'^2 Q \text{ modulo } h \quad (3)$$

where  $Q \in T^2 A[T]^{r'}$ . We have  $f'(y') \equiv d'^2 b'$  modulo  $d^3$  for some  $b' \in d' A^{r'}$ . Then

$$g_i = s^{p'} b'_i + s^{p'} T_i + Q_i, \quad i \in [r'] \quad (4)$$

modulo  $d^3$  is in the kernel of  $\bar{\phi}$ . Indeed, we have  $s^{p'} f'_i = d'^2 g_i$  modulo  $(h, d^3)$  because of (3). Thus  $d'^2 \bar{\phi}(g) = d'^2 g(t) \in (h(y, t), f'(y)) \in d^3 A'$  and so  $g(t) \in d^3 A'$ , because  $u$  is flat and  $d'$  is regular on  $A/(d^3)$ . Set  $E = \bar{A}[Y, T]/(I, \bar{g}, h)$  and let  $\bar{\psi} : E \rightarrow \bar{A}'$  be the map induced by  $\bar{\phi}$ . Clearly,  $\bar{v}$  factors through  $\bar{\psi}$  because  $\bar{v}$  is the composed map  $\bar{B} = \bar{A}[Y]/I \rightarrow E \xrightarrow{\bar{\psi}} \bar{A}'$ .

We will see, there are  $s', s'' \in E$  such that  $E_{ss's''}$  is smooth over  $\bar{A}$  and  $\bar{\psi}$  factors through  $E_{ss's''}$ .

Note that the  $r' \times r'$ -minor  $s'$  of  $(\partial g / \partial T)$  given by the first  $r'$ -variables  $T$  is from  $s'^{p'} + (T) \subset 1 + (d', T)$  because  $Q \in (T)^2$ . Then  $V = (\bar{A}[Y, T]/(h, g))_{ss'}$  is smooth over  $\bar{A}$ . As in [6] we claim that  $I\bar{A}[Y, T] \subset (h, g)\bar{A}[Y, T]_{ss's''}$  for some  $s'' \in 1 + (d', d^3, T)A[Y, T]$ . Indeed, we have

$$P'I\bar{A}[Y, T] \subset (f')A[Y, T] \subset (h, g)\bar{A}[Y, T]_s$$

and so

$$P'(y' + s^{-1} d' G(y') T) I \subset (h, g, d^3) A[Y, T]_s.$$

Since  $P'(y' + s^{-1} d' G(y') T) \in P'(y') + d'(T)V$  we get

$$P'(y' + s^{-1} d' G(y') T) \equiv d' s'' \text{ modulo } d^3$$

for some  $s'' \in 1 + (T)A[Y, T]$ . It follows that  $s'' I \subset (((h, g) : d'), d^3)A[Y, T]_{ss'}$ . Thus  $s'' I$  is contained modulo  $d^3$  in  $(0 :_V d') = 0$  because  $d'$  is regular on  $V$ , the map  $\bar{A} \rightarrow V$  being flat. This shows our claim. It follows that  $I \subset (d^3, h, g)A[Y, T]_{ss's''}$ . Thus  $E_{ss's''} \cong V_{s''}$  is a  $\bar{B}$ -algebra which is also standard smooth over  $\bar{A}$ .

As  $u(s) \equiv 1$  modulo  $d'$  and  $\bar{\psi}(s'), \bar{\psi}(s'') \equiv 1$  modulo  $(d', d^3, t), d, d', t \in \mathfrak{m}A'$  we see that  $u(s), \bar{\psi}(s'), \bar{\psi}(s'')$  are invertible because  $A'$  is local. Thus  $\bar{\psi}$  (and so  $\bar{v}$ ) factors through the standard smooth  $\bar{A}$ -algebra  $E_{ss's''}$ , let us say by  $\bar{\omega} : E_{ss's''} \rightarrow \bar{A}'$ .

Now, let  $Y' = (Y'_1, \dots, Y'_n)$ , and  $D$  be the  $A$ -algebra isomorphic with

$$(A[Y, T]/(I, h, g))_{ss's''} \text{ by } Y' \rightarrow Y, T \rightarrow T.$$

Since  $A'$  is Henselian we may lift  $\bar{\omega}$  to a map  $(A[Y, T]/(I, h, g))_{ss's''} \rightarrow A'$  which will correspond to a map  $\omega : D \rightarrow A'$ . Then  $\bar{v}$  factors<sup>4</sup> through  $\bar{D}$ , let us say  $\bar{B} \rightarrow \bar{D} \rightarrow \bar{A}'$ , where the first map is given by  $Y \rightarrow Y'$ . This proves Lemma 2.3.

To continue with the proof of Theorem 2.1 let  $\delta$  be the  $A$ -morphism defined by

$$\delta : B \otimes_A D \cong D[Y]/ID[Y] \rightarrow A', \quad b \otimes \lambda \rightarrow v(b)\omega(\lambda).$$

*Claim*  $\delta$  factors through a special finite type  $B \otimes_A D$ -algebra  $\tilde{E}$ .

The proof will follow the proof of Lemma 2.3. Note that the map  $\bar{B} \rightarrow \bar{D}$  is given by  $Y \rightarrow Y' + d^3D$ . Thus  $I(Y') \equiv 0$  modulo  $d^3D$ . Set  $\tilde{y} = \omega(Y')$ . Since  $\bar{v}$  factors through  $\bar{\omega}$  we get

$$y - \tilde{y} = v(Y) - \tilde{y} \in d^3A^m, \quad \text{let us say } y - \tilde{y} = d^2v \text{ for } v \in dA^m.$$

Recall that  $P = \sum_i L_i \det H_i$  for  $L_i \in ((f) : I)$ . We have  $d \equiv P$  modulo  $I$  and so  $P(Y') \equiv d$  modulo  $d^3$  in  $D$  because  $I(Y') \equiv 0$  modulo  $d^3D$ . Thus  $P(Y') = d\tilde{s}$  for a certain  $\tilde{s} \in D$  with  $\tilde{s} \equiv 1$  modulo  $d$ . Let  $\tilde{G}'_i$  be the adjoint matrix of  $H_i$  and  $\tilde{G}_i = L_i \tilde{G}'_i$ . We have  $\sum_i \tilde{G}_i H_i = \sum_i H_i \tilde{G}_i = PId_n$  and so

$$d\tilde{s}Id_n = P(Y')Id_n = \sum_i \tilde{G}_i(Y')H_i(Y').$$

But  $H_i$  is the matrix  $(\partial f_i / \partial Y_j)_{i \in [r], j \in [n]}$  completed with some  $(n - r)$  rows from 0, 1. Especially we obtain

$$(\partial f / \partial Y) \sum_i \tilde{G}_i = (PId_r | 0). \tag{5}$$

Then  $\tilde{t}_i := \omega(H_i(Y'))v \in dA^m$  satisfies

$$\sum_i \tilde{G}_i(Y')\tilde{t}_i = P(Y')v = d\tilde{s}v$$

and so

$$\tilde{s}(y - \tilde{y}) = d \sum_i \omega(\tilde{G}_i(Y'))\tilde{t}_i.$$

Let

$$\tilde{h} = \tilde{s}(Y - Y') - d \sum_i \tilde{G}_i(Y')\tilde{T}_i, \tag{6}$$

---

<sup>4</sup>Note that  $v$  does not necessarily factors through  $D$ .

where  $\tilde{T} = (\tilde{T}_1, \dots, \tilde{T}_n)$  are new variables. The kernel of the map  $\tilde{\phi} : D[Y, \tilde{T}] \rightarrow A'$  given by  $Y \rightarrow y, \tilde{T} \rightarrow \tilde{t}$  contains  $\tilde{h}$ . Since

$$\tilde{s}(Y - Y') \equiv d \sum_i \tilde{G}_i(Y') \tilde{T}_i \text{ modulo } \tilde{h}$$

and

$$f(Y) - f(Y') \equiv \sum_j (\partial f / \partial Y_j)((Y')(Y_j - Y'_j)$$

modulo higher order terms in  $Y_j - Y'_j$ , by Taylor's formula we see that for  $p = \max_i \deg f_i$  we have

$$\tilde{s}^p f(Y) - \tilde{s}^p f(Y') \equiv \sum_j \tilde{s}^{p-1} d(\partial f / \partial Y_j)(Y') \sum_i \tilde{G}_{ij}(Y') \tilde{T}_{ij} + d^2 \tilde{Q} \tag{7}$$

modulo  $\tilde{h}$  where  $\tilde{Q} \in \tilde{T}^2 D[\tilde{T}]^r$ . We have  $f(Y') = d^2 \tilde{b}$  for some  $\tilde{b} \in dD^r$ . Then

$$\tilde{g}_i = \tilde{s}^p \tilde{b}_i + \tilde{s}^p \tilde{T}_i + \tilde{Q}_i, \quad i \in [r] \tag{8}$$

is in the kernel of  $\tilde{\phi}$ . Indeed, we have  $\tilde{s}^p f_i = d^2 \tilde{g}_i$  modulo  $\tilde{h}$  because of (7) and  $P(Y') = d\tilde{s}$ . Thus  $d^2 \phi(\tilde{g}) = d^2 \tilde{g}(t) \in (\tilde{h}(y, \tilde{t}), f(y)) = (0)$  and so  $\tilde{g}(\tilde{t}) = 0$ . Set  $\tilde{E} = D[Y, \tilde{T}] / (I, \tilde{g}, \tilde{h})$  and let  $\tilde{\psi} : \tilde{E} \rightarrow A'$  be the map induced by  $\tilde{\phi}$ . Clearly,  $v$  factors through  $\tilde{\psi}$  because  $v$  is the composed map

$$B \rightarrow B \otimes_A D \cong D[Y] / I \rightarrow \tilde{E} \xrightarrow{\tilde{\psi}} A'.$$

Finally we will prove that there exist  $\tilde{s}', \tilde{s}'' \in \tilde{E}$  such that  $\tilde{E}_{\tilde{s}'\tilde{s}''}$  is standard smooth over  $A$  and  $\tilde{\psi}$  factors through  $\tilde{E}_{\tilde{s}'\tilde{s}''}$ .

Note that the  $r \times r$ -minor  $\tilde{s}'$  of  $(\partial \tilde{g} / \partial \tilde{T})$  given by the first  $r$ -variables  $\tilde{T}$  is from  $\tilde{s}'^p + (\tilde{T}) \subset 1 + (d, \tilde{T})$  because  $\tilde{Q} \in (\tilde{T})^2$ . Then  $\tilde{V} = (D[Y, \tilde{T}] / (\tilde{h}, \tilde{g}))_{\tilde{s}'}$  is smooth over  $D$ . We claim that  $I \subset (\tilde{h}, \tilde{g})D[Y, \tilde{T}]_{\tilde{s}'\tilde{s}''}$  for some other  $\tilde{s}'' \in 1 + (d, \tilde{T})D[Y, \tilde{T}]$ . Indeed, we have

$$PID[Y] \subset (f)D[Y] \subset (\tilde{h}, \tilde{g})D[Y, \tilde{T}]_{\tilde{s}'}$$

and so

$$P(Y' + \tilde{s}^{-1} d \sum_i \tilde{G}_i(Y') \tilde{T}_i) I \subset (\tilde{h}, \tilde{g})D[Y, \tilde{T}]_{\tilde{s}'}$$

Since  $P(Y' + \tilde{s}^{-1} d \sum_i \tilde{G}_i(Y') \tilde{T}_i) \in P(Y') + d(\tilde{T})$  we get  $P(Y' + \tilde{s}^{-1} d \sum_i \tilde{G}_i(Y') \tilde{T}_i) = d\tilde{s}''$  for some  $\tilde{s}'' \in 1 + (\tilde{T})D[Y, \tilde{T}]$ . It follows that  $\tilde{s}'' I \subset ((\tilde{h}, \tilde{g}) : d)D[Y, \tilde{T}]_{\tilde{s}'}$ . Thus

$\tilde{s}''I \subset (0 :_{\tilde{V}} d) = 0$ , which shows our claim. It follows that  $I \subset (\tilde{h}, \tilde{g})D[Y, \tilde{T}]_{\tilde{s}\tilde{s}'\tilde{s}''}$ . Thus  $\tilde{E}_{\tilde{s}\tilde{s}'\tilde{s}''} \cong \tilde{V}_{\tilde{s}''}$  is a  $B$ -algebra which is also standard smooth over  $D$  and  $A$ .

As  $\omega(\tilde{s}) \equiv 1$  modulo  $d$  and  $\tilde{\psi}(\tilde{s}'), \tilde{\psi}(\tilde{s}'') \equiv 1$  modulo  $(d, \tilde{t}), d, \tilde{t} \in \mathfrak{m}_{A'}$  we see that  $\omega(\tilde{s}), \tilde{\psi}(\tilde{s}'), \tilde{\psi}(\tilde{s}'')$  are invertible because  $A'$  is local. Thus  $\tilde{\psi}$  (and so  $v$ ) factors through the standard smooth  $A$ -algebra  $B' = \tilde{E}_{\tilde{s}\tilde{s}'\tilde{s}''}$ . This proves Theorem 2.1.

### 3 The Algorithm

Now we want to apply Theorem 2.1 to compute the Neron desingularization. We assume  $A = (k[x]/J)_{<x>}$  is Cohen-Macaulay of dimension 2,  $A'$  is the completion of  $A$  and  $u$  the inclusion. The morphism  $v : B \rightarrow A'$  will be given by an approximation, polynomials up to a given bound. We obtain the following algorithms (which will be implemented in SINGULAR as a library [3]). The algorithm `prepareDesingularization` corresponds to Lemma 2.2 in the proof of Theorem 2.1.

---

#### Algorithm 1 `prepareDesingularization`

---

**Input:**  $A := k[x]_{(x)}/J$  given by  $J = (h_1, \dots, h_p) \subseteq k[x], x = (x_1, \dots, x_i), k$  a field  
 $B := A[Y]/I$  given by  $I = (g_1, \dots, g_l) \subseteq k[x, Y], Y = (Y_1, \dots, Y_n)$   
 and  $y' = (y'_1, \dots, y'_n) \in k[x]^n$  such that  $H_{B/A}(y')$  is zero-dimensional  
**Output:**  $B := A[Y]/I$  given by  $I = (g_1, \dots, g_l) \subseteq k[x, Y], Y = (Y_1, \dots, Y_n), y' = (y'_1, \dots, y'_n) \in k[x]^n, f, f' \in I$  and  $d, d'$  a regular sequence in  $A, d \in ((f) : I)\Delta_f$  resp.  $d' \in ((f') : I)\Delta_{f'}$ , such that  $(I/I^2)_d$  resp.  $(I/I^2)_{d'}$  are free  $B_d$  resp.  $B_{d'}$  modules.

- 1: **compute**  $H_{B/A} = (b_1, \dots, b_q)_B$  and  $H_{B/A} \cap A$
- 2: **if**  $\dim A/H_{B/A} \cap A = 0$  **then**
- 3:     **choose**  $\gamma, \gamma' \in H_{B/A} \cap A$ , a regular sequence in  $A$
- 4: **else**
- 5:     **choose**  $\gamma, \gamma' \in H_{B/A}(y')$ , a regular sequence in  $A$   
        **write**  
             $\gamma = \sum_{i=1}^q b_i(y')y'_{i+n}$  modulo  $(\gamma', \gamma'')$ ,  $\gamma' \equiv \sum_{i=1}^q b_i(y')y'_{i+n+q}$  modulo  $(\gamma', \gamma'')$  for some  $t$  and  $y'_j \in k[x]$
- 6:      $g_{l+1} := -\gamma + \sum_{i=1}^q b_i Y_{i+n}, g_{l+2} := -\gamma' + \sum_{i=1}^q b_i Y_{i+n+q},$   
         $Y := (Y_1, \dots, Y_{n+2q}); y' := (y'_1, \dots, y'_{n+2q}); I := (g_1, \dots, g_{l+2}); l := l + 2; n := n + 2q; B := A[Y]/I.$
- 7: **end if**
- 8:  $B := S_B(I/I^2), y'$  trivially extended
- 9: **write**  
         $B := A[Y]/I, n := |Y|, Y := Y, Z = (Z_1, \dots, Z_n), I := (I, Z), B := A[Y]/I, y'$  trivially extended
- 10: **compute**  $f = (f_1, \dots, f_r)$ , and  $f' = (f'_1, \dots, f'_r)$  **such that**  
        a power  $d$  of  $\gamma$ , resp.  $d'$  of  $\gamma'$  is in  $((f) : I)\Delta_f$ , resp. in  $((f') : I)\Delta_{f'}$
- 11: **return**  $B, y', f, f', d, d'$

---

The next algorithm corresponds to Lemma 2.3 in the proof of Theorem 2.1.



**Algorithm 2** reductionToDimensionOne

---

**Input:**  $A := k[x]_{(x)}/J$  given by  $J = (h_1, \dots, h_p) \subseteq k[x], x = (x_1, \dots, x_r), k$  a field  
 $B := A[Y]/I$  given by  $I = (g_1, \dots, g_t) \subseteq k[x, Y], Y = (Y_1, \dots, Y_n), y' = (y'_1, \dots, y'_n) \in k[x]^n, f' = (f'_1, \dots, f'_r), d', d \in A$  and  $\{H'_i, L'_i\}$  such that  $d' \equiv P' = \sum_{i=1}^{q'} \det(H'_i) L'_i$  modulo  $I$   
**Output:**  $D := (A[Y', T]/(I, g, h))_{ss's''}$  given by  $I, g, h, s, s', s'' \in k[x, Y', T], Y' := (Y'_1, \dots, Y'_n)$ ;

- 1: **write**  
 $P'(y') = d's$  modulo  $d^3$  for  $s \in A, s \equiv 1$  modulo  $d'$
- 2: **for**  $i = 1$  to  $q'$  **do**
- 3:     **compute**  $G'_i$  the adjoint matrix of  $H'_i$  and  $G_i = L'_i G'_i$
- 4: **end for**
- 5:  $h := s(Y - y') - d' \sum_i G_i(y') T_i, T_i = (T_1, \dots, T_{r'}, T_{i,r'+1}, \dots, T_{i,n})$
- 6:  $p' := \max_i \{\deg f'_i\}$
- 7: **write**  
 $s^{p'} f'(Y) - s^{p'} f'(y') = \sum_j s^{p'-1} d' \partial f' / \partial Y(y') \sum_i G_{ij}(y') T_{ij} + d^2 Q$  modulo  $h$
- 8: **write**  $f'(y') = d^2 b'$  modulo  $d^3$
- 9: **for**  $i = 1$  to  $r'$  **do**
- 10:      $g_i := s^{p'} b'_i + s^{p'} T_i + Q_i$
- 11: **end for**
- 12: **compute**  $s'$  the  $r'$ -minor of  $(\partial g / \partial T)$  given by the first  $r'$  variables and  $s''$  **such that**  
 $P(y' + s^{-1} d' \sum_i G_i(y') T) = d' s''$  modulo  $d^3$
- 13:  $D := (A[Y', T]/(I, g, h))_{ss's''}; Y' := (Y'_1, \dots, Y'_n); g := g(Y'); I := I(Y'); h := h(Y')$
- 14: **return**  $D$

---

**Algorithm 3** NeronDesingularization

---

**Input:**  $N \in \mathbb{Z}_{>0}$  a bound  
 $A := k[x]_{(x)}/J$  given by  $J = (h_1, \dots, h_p) \subseteq k[x], x = (x_1, \dots, x_r), k$  a field  
 $B := A[Y]/I$  given by  $I = (g_1, \dots, g_t) \subseteq k[x, Y], Y = (Y_1, \dots, Y_n) v : B \rightarrow A' \subseteq K[[x]]/JK[[x]]$  an  $A$ -morphism given by  $y' = (y'_1, \dots, y'_n) \in k[x]^n$ , approximations modulo  $(x)^N$  of  $v(Y)$ .  
**Output:** A Neron desingularization of  $v : B \rightarrow A'$  or the message “the bound is too small”

- 1:  $(B, y', f, f', d, d') := \text{prepareDesingularization}(A, B, y')$
- 2: **if**  $(d^3, d'^3) \not\subseteq (x)^N$  **then**
- 3:     **return** “the bound is too small”
- 4: **end if**
- 5: **choose**  $r$ -minors  $M_i$  (resp.  $r'$ -minors  $M'_i$ ) of  $(\partial f / \partial Y)$ , (resp.  $(\partial f' / \partial Y)$ ) and  $L_i \in ((f) : I)$ , (resp.  $L'_i \in ((f') : I)$ ) **such that**  
for  $P = \sum_i M_i L_i$  (resp.  $P' = \sum_i M'_i L'_i$ ),  $d \equiv P$  modulo  $I$  (resp.  $d' \equiv P'$  modulo  $I$ )
- 6: **complete** the Jacobian matrix  $(\partial f / \partial Y)$  (resp.  $(\partial f' / \partial Y)$ ) by  $(n - r)$  (resp.  $(n - r')$ ) rows of 0, 1 to obtain square matrices  $H_i$  (resp.  $H'_i$ ) **such that**  
 $\det H_i = M_i$  (resp.  $\det H'_i = M'_i$ )
- 7:  $D := \text{reductionToDimensionOne}(A, B, y', f', d', d, \{H'_i, L'_i\})$
- 8: **write**  $P(Y') = d\tilde{s}; \tilde{s} \equiv 1$  modulo  $d$
- 9: **compute**  $\tilde{G}'_i$  the adjoint matrix of  $H_i$  and  $\tilde{G}_i = L_i \tilde{G}'_i$
- 10:  $\tilde{h} := \tilde{s}(Y - Y') - d \sum_{i=1}^q \tilde{G}_i \tilde{T}_i, \tilde{T}_i = (\tilde{T}_1, \dots, \tilde{T}_r, \tilde{T}_{i,r+1}, \dots, \tilde{T}_{i,n})$
- 11:  $p := \max_i \{\deg \tilde{f}_i\}$
- 12: **write**  $\tilde{s}^p f(Y) - \tilde{s}^p f(Y') = \sum_j \tilde{s}^{p-1} d \partial f / \partial Y(Y') \sum_i \tilde{G}_{ij}(Y') \tilde{T}_{ij} + d^2 \tilde{Q}$  modulo  $\tilde{h}$  and
- 13: **write**  $f(Y') = d^2 \tilde{b}, \tilde{b} \in dD'$
- 14: **for**  $i = 1$  to  $r$  **do**
- 15:      $\tilde{g}_i := \tilde{s}^p \tilde{b}_i + \tilde{s}^p \tilde{T}_i + \tilde{Q}_i$
- 16: **end for**
- 17: **compute**  $\tilde{s}'$  the  $r \times r$ -minors of  $(\partial \tilde{g} / \partial \tilde{T})$  given by the first  $r$  variables of  $\tilde{T}$
- 18: **compute**  $\tilde{s}''$  **such that**  
 $P(Y' + \tilde{s}^{-1} d \sum_i \tilde{G}_i(Y') \tilde{T}) = d\tilde{s}''$
- 19: **return**  $D[Y, \tilde{T}]/(I, \tilde{g}, \tilde{h})_{\tilde{s}\tilde{s}'\tilde{s}''}$

---

## References

1. M. André, Cinq exposés sur la désingularisation. Handwritten manuscript Ecole Polytechnique Fédérale de Lausanne (1991)
2. M. Artin, Algebraic approximation of structures over complete local rings. *Publ. Math. IHES* **36**, 23–58 (1969)
3. W. Decker, G.-M. Greuel, G. Pfister, H. Schönemann, SINGULAR 4-1-0 – a computer algebra system for polynomial computations (2016). <http://www.singular.uni-kl.de>
4. R. Elkik, Solutions d'équations à coefficients dans un anneaux hensélien. *Ann. Sci. Ecole Normale Sup.* **6**, 553–604 (1973)
5. A. Néron, Modèles minimaux des variétés abéliennes sur les corps locaux et globaux. *Publ. Math. IHES* **21**, 5–128 (1964)
6. G. Pfister, D. Popescu, Constructive General Neron Desingularization for one dimensional local rings. *J. Symb. Comput.* **80**, 570–580 (2017)
7. D. Popescu, General Neron Desingularization. *Nagoya Math. J.* **100**, 97–126 (1985)
8. D. Popescu, General Neron Desingularization and approximation. *Nagoya Math. J.* **104**, 85–115 (1986)
9. D. Popescu, Artin approximation, in *Handbook of Algebra*, vol. 2, ed. by M. Hazewinkel (Elsevier, Amsterdam, 2000), pp. 321–355
10. A. Popescu, D. Popescu, A method to compute the General Neron Desingularization in the frame of one dimensional local domains, in *Singularities and Computer Algebra – Festschrift for Gert-Martin Greuel, On the Occasion of his 70th Birthday*, ed. by W. Decker, G. Pfister, M. Schulze. Springer Monograph, pp. 199–222. arXiv:AC/1508.05511
11. M. Spivakovsky, A new proof of D. Popescu's theorem on smoothing of ring homomorphisms. *J. Am. Math. Soc.* **294**, 381–444 (1999)
12. R. Swan, Neron-Popescu desingularization, in *Algebra and Geometry*, ed. by M. Kang (International Press, Cambridge, 1998), pp. 135–192

# A Framework for Computing Zeta Functions of Groups, Algebras, and Modules



Tobias Rossmann

**Abstract** We give an overview of the author's recent work on methods for explicitly computing various types of zeta functions associated with algebraic counting problems. Among the types of zeta functions that we consider are the so-called topological ones.

**Keywords** Subgroup growth • Representation growth • Zeta functions • Topological zeta functions • Unipotent groups •  $p$ -Adic integration • Newton polytopes

**Subject Classifications** 11M41, 20F69, 14M25, 20F18, 20C15, 20G30

## 1 Introduction

### 1.1 Zeta Functions in Group Theory and Related Fields

The past decades saw the development of a theory of zeta functions of groups and related algebraic structures. In this article, we consider subobject and representation zeta functions related to enumerative problems associated with nilpotent groups. For introductions to the area and surveys of developments in particular directions, we refer the reader to [15, 19, 29, 56, 57]. We will concern ourselves with zeta functions that one can attach to a suitable infinite algebraic object (e.g. a Lie algebra or a group). In a different direction, zeta functions have found striking applications in the study of infinite families of finite groups; for surveys of this active branch of asymptotic group theory, we refer to [34, 46].

---

T. Rossmann (✉)

Department of Mathematics, University of Auckland, Auckland, New Zealand  
e-mail: [tobias.rossmann@gmail.com](mailto:tobias.rossmann@gmail.com)

The subobject zeta functions of interest to us can be traced back to a number of sources. An early ancestor is given by the Dedekind zeta function of a number field, an instance of a submodule zeta function as defined below. More recently, Solomon [48] introduced zeta functions enumerating  $\mathbf{Z}G$ -lattices within a fixed  $\mathbf{Z}G$ -module for a finite group  $G$ . In a seemingly different direction, a hugely influential paper of Grunewald et al. [23] initiated the study of zeta functions arising from the enumeration of subgroups of finite index in a given finitely generated torsion-free nilpotent group (a  $\mathcal{T}$ -group, for short). In detail, given such a group  $G$ , they defined its (*global*) *subgroup zeta function* to be

$$\zeta_G^{\leq}(s) = \sum_H |G : H|^{-s},$$

where  $H$  ranges over the subgroups of finite index of  $G$ . They also established various key properties of these zeta functions such as:

- (Convergence.) Let  $h$  be the Hirsch length of  $G$ . Then  $\zeta_G^{\leq}(s)$  converges and defines an analytic function on the half-plane  $\{s \in \mathbf{C} : \operatorname{Re}(s) > h\}$ .
- (Euler product.)  $\zeta_G^{\leq}(s) = \prod_{p \text{ prime}} \zeta_{\hat{G}_p}^{\leq}(s)$ , where  $\hat{G}_p$  denotes the pro- $p$  completion of  $G$  and each *local subgroup zeta function*  $\zeta_{\hat{G}_p}^{\leq}(s)$  is defined by enumerating open subgroups of  $\hat{G}_p$  according to their indices.
- (Rationality.) Each  $\zeta_{\hat{G}_p}^{\leq}(s)$  is rational in  $p^{-s}$  over  $\mathbf{Q}$ .

For  $G = \mathbf{Z}$ , we recover the Riemann zeta function  $\zeta(s) = \zeta_{\mathbf{Z}}(s) = \sum_{n=1}^{\infty} n^{-s}$  and the classical Euler product  $\zeta(s) = \prod_p (1 - p^{-s})^{-1}$ . This simple illustration notwithstanding, while the first two of the above points are elementary, the rationality of local subgroup zeta functions is a deep theorem.

By only considering normal subgroups of finite index of  $G$ , the *normal subgroup zeta function*  $\zeta_G^{\triangleleft}(s)$  of  $G$  is obtained; it satisfies the evident analogues of the properties stated above. The subalgebra and submodule zeta functions defined in Sect. 2.1 essentially constitute generalisations of the local and global (normal) subgroup zeta functions associated with nilpotent groups. Indeed, as explained in [23], the Mal'cev correspondence allows us to linearise the enumeration of subgroups by replacing the nilpotent group in question by a suitable nilpotent Lie  $\mathbf{Z}$ -algebra (at the cost of having to discard finitely many Euler factors).

Apart from subobject zeta functions, we also consider representation zeta functions. These are Dirichlet series enumerating certain finite-dimensional irreducible representations of a suitable group up to adequate notions of equivalence. Representation zeta functions were introduced by Witten [59] in the context of complex Lie groups. Jaikin-Zapirain [26] made fundamental contributions to the study of representation zeta functions of compact  $p$ -adic analytic groups. Within infinite group theory, a substantial amount of recent work has been devoted to

representation zeta functions of groups arising from semisimple algebraic groups; see e.g. work of Larsen and Lubotzky [32] and Avni et al. [2]. In a seemingly different direction, Hrushovski and Martin [24] (v1, 2006) introduced representation zeta functions of  $\mathcal{T}$ -groups; these are the representation zeta functions that we shall consider.

As we will explain in Sect. 3, the subobject and representation zeta functions considered here share a crucial common feature: in each case, a single global object (e.g. a  $\mathcal{T}$ -group) gives rise to a family of associated local zeta functions indexed by primes (or places of a number field) and, after excluding finitely many exceptions, these local zeta functions can all be described in terms of a single “formula”. We express this by saying that there exists such a “formula” for the *generic* local zeta functions in question. In a surprising number of interesting cases (including most cases that have been successfully computed so far), the local zeta functions under consideration are in fact given by a rational function  $W(p, p^{-s})$  in  $p$  and  $p^{-s}$  over  $\mathbf{Q}$  (again after possibly excluding finitely many primes). This phenomenon is referred to as (*almost*) *uniformity* in [19, §1.2.4]. In such uniform cases, we interpret the natural task of computing the generic local zeta functions under consideration as computing the (uniquely determined) rational function  $W$ . For instance, if  $H(\mathbf{Z})$  is the discrete Heisenberg group, then, by [23, Prop. 8.1],

$$\zeta_{H(\mathbf{Z})}^{\leq}(s) = \zeta(s)\zeta(s-1)\zeta(2s-2)\zeta(2s-3)\zeta(3s-3)^{-1} = \prod_{p \text{ prime}} W(p, p^{-s}), \quad (1)$$

where  $W(X, Y) = (1 - X^3Y^3)/((1 - Y)(1 - XY)(1 - X^2Y^2)(1 - X^3Y^2))$ .

### 1.2 Computations: Limitations and Previous Work

Theoretical results on the subobject and representation zeta functions considered here frequently rely on impractical or even non-constructive methods. In particular, in one of the central papers in the area, du Sautoy and Grunewald [16] showed that generic local subobject zeta functions are in principle “computable” (in the sense that one can compute certain formulae for them, see Sect. 3)—provided that one happens to know an embedded resolution of singularities of some (usually highly singular) hypersurface inside some affine space (of dimension  $\geq 6$  in all cases of interest); Voll [55, §3.4] obtained a similar result for representation zeta functions of nilpotent groups.

Apart from striking theoretical applications, the methods developed by du Sautoy and Grunewald [16] and Voll [55] (and Stasinski and Voll [49]) have also been successfully used to compute zeta functions in small examples (see, in particular, the computation of du Sautoy and Taylor [18] of the subalgebra zeta function of  $\mathfrak{sl}_2(\mathbf{Z})$ ; for related computations, see [14, 25, 30]). However, when it comes to explicit computations, the practical scope of these techniques is usually rather limited.

A substantial number of subobject zeta functions (primarily of nilpotent groups and Lie algebras) were computed by Woodward [60]. He relied on a combination of human guidance and computer calculations. Unfortunately, due to a lack of documentation, his findings are hard to reproduce. A number of ad hoc computations of representation zeta functions of nilpotent groups have been carried out by Ezzat [22], Snocken [47], and Stasinski and Voll [50].

### 1.3 Topological Zeta Functions

A common catchphrase in the area is that topological zeta functions are obtained from local ones (such as the  $\zeta_{G_p}^{\leq}(s)$  from above) by passing to a limit “ $p \rightarrow 1$ ”. Indeed, Denef and Loeser [13] introduced topological zeta functions of polynomials by justifying that such a limit can be applied to Igusa’s local zeta function (see [11, 36] for introductions). Despite their arithmetic ancestry (for Igusa’s local zeta function enumerates solutions to congruences), research on topological zeta functions has been primarily motivated by questions from singularity theory. In recent years, topological zeta functions of polynomials have mostly been studied within the realm of motivic zeta functions.

Using such a “motivic” point of view, topological subobject zeta functions were introduced by du Sautoy and Loeser [17]; these zeta functions are related to, but different from, Evseev’s “reduced zeta functions” [21]. Apart from giving a definition of these zeta functions, they also computed a few small examples. Further examples were determined by the author [38, 39] who also began an investigation of topological representation zeta functions [40]. The topological subobject and representation zeta functions studied by the author seem to exhibit a number of features distinct from the well-studied case of topological zeta functions of polynomials; we will discuss some of these features in Sect. 8.

### 1.4 Computations: A Framework

The author’s articles [38–41] provide a practical framework for explicitly computing numerous types of (generic) local and topological zeta functions in “fortunate” cases related to geometric genericity conditions. The main purpose of the present article is to provide a self-contained and unified introduction that takes into account theoretical developments that occurred over the course of the project.

In summary, the author's methods for computing topological [38–40] or generic local zeta functions [41] all proceed along the following lines:

1. (Translation.) Express the associated generic local zeta functions in terms of  $p$ -adic integrals defined in terms of certain global “data”.
2. (The simplify-balance-reduce loop.) After discarding finitely many primes, attempt to write the integrals from the first step as sums of integrals of the same shape but defined in terms of “regular” (i.e. sufficiently generic) data.
3. (Evaluation.) Assuming the second step succeeds, explicitly compute “formulae” for the generic local or topological zeta functions associated with the integrals attached to the regular data from the second step.
4. (Final summation.) Add the formulae from the third step.

The first step is based on known results. For the computation of subobject zeta functions, we use the formalism of “cone integrals” of du Sautoy and Grunewald [16]. For representation zeta functions associated with unipotent groups, we rely on the formulation in terms of  $p$ -adic integrals due to Stasinski and Voll [49] (which extends Voll's formalism from [55]); see also related work of Avni et al. [2] on representation zeta functions of arithmetic groups.

The second step is concerned with manipulations of  $p$ -adic integrals represented in terms of the “toric data” from [39] or the “representation data” from [40]. Either type of datum consists of algebraic ingredients (Laurent polynomials) and convex-geometric data (“half-open cones”). Regularity is an algebro-geometric genericity condition which allows us to invoke the machinery developed in [38] in order to compute the  $p$ -adic integral in question (or the associated topological zeta function). Being “balanced” is a much weaker property and it is always possible to write the integral associated with an arbitrary toric/representation datum as a sum of integrals associated with balanced data—this corresponds to the middle part of the name of the second step. In fortunate case (related to the notion of non-degeneracy from [38]), applying the balancing procedure to our initial datum from the first step will produce a family of regular data. The purpose of the reduction step is to modify balanced but singular (i.e. not regular) data, the goal being to derive regular data. This may or may not succeed for a given example and it is the main reason why the author's methods may fail for specific examples. While the final summation step is mathematically trivial, it is often computationally daunting.

## 1.5 Overview

In Sect. 2, we recall definitions of the global and local zeta functions that we consider. We then discuss the existence of “formulae” for generic local zeta functions in Sect. 3. Topological zeta functions are the subject of Sect. 4. Prior to presenting our computational framework and its computer implementation in Sect. 6, we collect

some background material from convex geometry in Sect. 5. As a demonstration of the practical usefulness of the author’s methods, a number of applications are discussed in Sect. 7. Finally, Sect. 8 is devoted to two particularly interesting conjectures that arose from the author’s computations.

## 2 Global and Local Zeta Functions

### 2.1 Formal Subalgebra, Ideal, and Submodule Zeta Functions

Let  $R$  be a commutative ring and let  $\mathbf{A}$  be an  $R$ -algebra, i.e. an  $R$ -module endowed with a multiplication  $\mathbf{A} \otimes_R \mathbf{A} \rightarrow \mathbf{A}$  (which need not be associative or Lie). A *subalgebra* of  $\mathbf{A}$  is an  $R$ -submodule which is stable under the given multiplication. As usual, by the *index*  $|\mathbf{A} : \mathbf{U}|$  of an  $R$ -submodule  $\mathbf{U} \leq \mathbf{A}$ , we mean the cardinality of the  $R$ -module quotient  $\mathbf{A}/\mathbf{U}$ . Let  $a_n^{\leq}(\mathbf{A})$  denote the number of subalgebras of  $\mathbf{A}$  of index  $n$ . Assuming that these numbers are all finite, we define the *subalgebra zeta function* of  $\mathbf{A}$  to be the formal Dirichlet series

$$\zeta_{\mathbf{A}}^{\leq}(s) = \sum_{n=1}^{\infty} a_n^{\leq}(\mathbf{A})n^{-s}.$$

If we only consider (2-sided  $R$ -)ideals of  $\mathbf{A}$ , then we obtain the *ideal zeta function*  $\zeta_{\mathbf{A}}^{\triangleleft}(s)$  of  $\mathbf{A}$ . These notions are all natural generalisations of the subring and ideal zeta functions introduced by Grunewald et al. [23].

Let  $\mathbf{M}$  be an  $R$ -module and let  $\Omega$  be a set of endomorphisms of  $\mathbf{M}$ . Let  $a_n(\Omega \curvearrowright \mathbf{M})$  denote the number of submodules  $\mathbf{U}$  of  $\mathbf{M}$  with  $|\mathbf{M} : \mathbf{U}| = n$  and such that  $\mathbf{U}$  is invariant under each element of  $\Omega$ . Assuming that each  $a_n(\Omega \curvearrowright \mathbf{M})$  is finite, we define the *submodule zeta function* of  $\Omega$  acting on  $\mathbf{M}$  to be

$$\zeta_{\Omega \curvearrowright \mathbf{M}}^{\leq}(s) = \sum_{n=1}^{\infty} a_n(\Omega \curvearrowright \mathbf{M})n^{-s}.$$

These zeta functions generalise those of Solomon [48]. It is frequently useful to note that  $\zeta_{\Omega \curvearrowright \mathbf{M}}(s)$  only depends on the unitary associative subalgebra of  $\text{End}(\mathbf{M})$  generated by  $\Omega$ . Moreover, as pointed out in [38], submodule zeta functions as defined here generalise the ideal zeta functions from above.

Generalising further, we could take into account a given  $R$ -module decomposition of an  $R$ -algebra  $\mathbf{A}$  or an  $R$ -module  $\mathbf{M}$  and consider associated graded counting problems as in [41, §3]; apart from the author’s work, such graded zeta functions have recently been studied by Lee and Voll [33]. For the sake of simplicity, while many results and ideas apply in this greater generality, in the following, we only consider subalgebra and submodule zeta functions of the form  $\zeta_{\mathbf{A}}^{\leq}(s)$  and  $\zeta_{\Omega \curvearrowright \mathbf{M}}(s)$  and we refer to these as *subject zeta functions*.



## 2.2 Number Fields and Euler Products

The subalgebra and submodule zeta functions defined in Sect. 2.1 are formal Dirichlet series. Further assumptions are needed for these to give rise to analytic functions.

We first set up some notation that will be used for the remainder of this article. Let  $k$  be a number field with ring of integers  $\mathfrak{o}$ . Let  $\mathcal{V}_k$  be the set of non-Archimedean places of  $k$ ; we identify  $\mathcal{V}_\mathbb{Q}$  with the set of prime numbers. For  $v \in \mathcal{V}_k$ , let  $\mathfrak{p}_v \in \text{Spec}(\mathfrak{o})$  correspond to  $v$ , let  $k_v$  denote the  $v$ -adic completion of  $k$ , and let  $\mathfrak{o}_v$  be the valuation ring of  $k_v$ . We write  $\mathfrak{K}_v$  for the residue field of  $k_v$  and  $q_v$  for its size.

For an  $\mathfrak{o}$ -object (e.g. an  $\mathfrak{o}$ -module)  $\mathbf{X}$ , we write  $X_v$  for the associated  $\mathfrak{o}_v$ -object obtained after base change (e.g.  $\mathbf{X} \otimes_{\mathfrak{o}} \mathfrak{o}_v$ ). Let  $\mathbf{A}$  be an  $\mathfrak{o}$ -algebra, let  $\mathbf{M}$  be an  $\mathfrak{o}$ -module, and let  $\Omega \subseteq \text{End}(\mathbf{M})$ . We assume that  $\mathbf{A}$  and  $\mathbf{M}$  are both free of rank  $d$  as  $\mathfrak{o}$ -modules. It is well-known (cf. [23, Prop. 1]) that  $\zeta_{\mathbf{A}}^{\leq}(s)$  and  $\zeta_{\Omega \curvearrowright \mathbf{M}}(s)$  both converge for  $\text{Re}(s) > d$ . Furthermore, we obtain Euler products

$$\zeta_{\mathbf{A}}^{\leq}(s) = \prod_{v \in \mathcal{V}_k} \zeta_{\mathbf{A}_v}^{\leq}(s), \quad \zeta_{\Omega \curvearrowright \mathbf{M}}(s) = \prod_{v \in \mathcal{V}_k} \zeta_{\Omega \curvearrowright \mathbf{M}_v}(s);$$

see [38, Lem. 2.3]. In [16], du Sautoy and Grunewald showed that  $\zeta_{\mathbf{A}}^{\leq}(s)$  and  $\zeta_{\Omega \curvearrowright \mathbf{M}}(s)$  have rational abscissae of convergence and admit meromorphic continuation to some larger half-planes than their initial half-planes of convergence. Furthermore, using their techniques (or the model-theoretic arguments from [23]), each  $\zeta_{\mathbf{A}_v}^{\leq}(s)$  and  $\zeta_{\Omega \curvearrowright \mathbf{M}_v}(s)$  is found to be rational in  $q_v^{-s}$ .

## 2.3 Representation Zeta Functions of Unipotent Groups

Given a topological group  $G$ , let  $\tilde{r}_n(G)$  denote the number of its continuous irreducible  $n$ -dimensional complex representations, counted up to equivalence and tensoring with continuous 1-dimensional representations (“twisting”). The motivation for allowing 1-dimensional “twists” comes from the case of nilpotent groups: while an infinite (discrete)  $\mathcal{T}$ -group  $G$  has infinitely many homomorphisms to  $\text{GL}_1(\mathbb{C})$ , Lubotzky and Magid [35] showed that each  $\tilde{r}_n(G)$  is finite. Following Hrushovski and Martin [24] (v1), if each  $\tilde{r}_n(G)$  is finite, we define the (*twist*) *representation zeta function* of  $G$  to be the formal Dirichlet series

$$\tilde{\zeta}_G^{\text{irr}}(s) = \sum_{n=1}^{\infty} \tilde{r}_n(G) n^{-s}.$$

Let  $G$  be a  $\mathcal{T}$ -group. Then  $\zeta_G^{\text{irr}}(s)$  converges in some complex half-plane, see [49, Lem. 2.1]. Moreover, crucial properties of  $\zeta_G^{\text{irr}}(s)$  such as its abscissa of convergence only depend on the commensurability class of  $G$ , see [20, Cor. B]. It is well-known that commensurability classes of  $\mathcal{T}$ -groups are in natural bijection with isomorphism classes of unipotent algebraic groups over  $\mathbf{Q}$ . Following Stasinski and Voll [49], we consider representation zeta functions of  $\mathcal{T}$ -groups associated with unipotent algebraic groups over number fields.

We first recall some facts on unipotent algebraic groups. Let  $U_d$  be the subgroup scheme of  $\text{GL}_d$  consisting of upper unitriangular matrices. An algebraic group  $\mathbf{G}$  over the number field  $k$  is *unipotent* if and only if it embeds into some  $U_d \otimes k$ ; for other characterisations of unipotence, see [10, Ch. IV]. Let  $\mathbf{G}$  be a unipotent algebraic group over  $k$ . After choosing an embedding of  $\mathbf{G}$  into some  $U_d \otimes k$ , we obtain an associated  $\mathfrak{o}$ -form  $\mathbf{G}$  of  $\mathbf{G}$  as a group scheme by taking the scheme-theoretic closure of  $\mathbf{G}$  within  $U_d \otimes \mathfrak{o}$ . We regard the  $\mathcal{T}$ -group  $\mathbf{G}(\mathfrak{o})$  as a discrete topological group and for  $v \in \mathcal{V}_k$ , we naturally regard  $\mathbf{G}(\mathfrak{o}_v)$  as a pro- $p_v$  group, where  $p_v$  is the rational prime contained in  $\mathfrak{p}_v$ . By [49, Prop. 2.2],  $\zeta_{\mathbf{G}(\mathfrak{o})}^{\text{irr}}(s) = \prod_{v \in \mathcal{V}_k} \zeta_{\mathbf{G}(\mathfrak{o}_v)}^{\text{irr}}(s)$ . Duong and Voll [20] and Hrushovski et al. [24] have shown that  $\zeta_{\mathbf{G}(\mathfrak{o}_v)}^{\text{irr}}(s)$  is rational in  $q_v^{-s}$  for almost all  $v \in \mathcal{V}_k$  and that  $\zeta_{\mathbf{G}(\mathfrak{o})}^{\text{irr}}(s)$  has rational abscissa of convergence. Duong and Voll also showed that, as in the enumeration of subobjects in Sect. 2.2,  $\zeta_{\mathbf{G}(\mathfrak{o})}^{\text{irr}}(s)$  admits meromorphic continuation to the left of its abscissa of convergence.

### 3 Computability of Generic Local Zeta Functions

We now explain in which sense generic local subobject and representation zeta functions are, in principle, computable. Let  $\mathbf{Z} = (\mathbf{Z}_v(s))_{v \in \mathcal{V}_k}$  be a family of local zeta functions defined in one of the following ways:

- $\mathbf{Z}_v(s) = (1 - q_v^{-1})^d \cdot \zeta_{\mathbf{A}_v}^{\leq}(s)$ , where  $\mathbf{A}$  is an  $\mathfrak{o}$ -algebra whose underlying  $\mathfrak{o}$ -module is free of rank  $d$ .
- $\mathbf{Z}_v(s) = (1 - q_v^{-1})^d \cdot \zeta_{\Omega \curvearrowright \mathbf{M}_v}(s)$ , where  $\mathbf{M}$  is a free  $\mathfrak{o}$ -module of rank  $d$  and  $\Omega \subseteq \text{End}(\mathbf{M})$ .
- $\mathbf{Z}_v(s) = \zeta_{\mathbf{G}(\mathfrak{o}_v)}^{\text{irr}}(s)$ , where  $\mathbf{G} \leq U_d \otimes \mathfrak{o}$  is the natural  $\mathfrak{o}$ -form of  $\mathbf{G} \leq U_d \otimes k$ .

The role of the factors  $(1 - q_v^{-1})^d$  will be explained in Sect. 4. The global zeta function associated with  $\mathbf{Z}$  is in general a subtle analytic object which we shall not consider further. Instead, we focus on the already quite difficult local picture.

In the study of local zeta functions  $\mathbf{Z}_v(s)$  attached to a global object, the exclusion of finite sets of exceptional places is often unavoidable. For example, while the subalgebra zeta function of  $\mathfrak{sl}_2(\mathbf{Z}_p)$  is given by a simple formula which is valid for all odd primes  $p$ , the case  $p = 2$  is exceptional; see [18]. Fortunately, interesting properties of global zeta functions often remain unaffected when finitely

many places are dropped; this is, for instance, the case for the global abscissae of convergence of subobject zeta functions, cf. [42, Lem. 5.3, Rem. 5.4]. Henceforth, we focus on the *generic* local zeta functions  $Z_v(s)$  obtained after discarding  $Z_w(s)$  for finitely many  $w \in \mathcal{V}_k$ .

As we mentioned above,  $Z_v(s)$  is rational in  $q_v^{-s}$  for almost every  $v \in \mathcal{V}_k$ . The task of “computing”  $Z_v(s)$  then means to determine  $W_v(Y) \in \mathbf{Q}(Y)$  with  $Z_v(s) = W_v(q_v^{-s})$ . The non-trivial fact that it is even possible to do this algorithmically is a consequence of the proof of the following deep theorem.

**Theorem 3.1** *Let  $\mathbf{Z} = (Z_v(s))_{v \in \mathcal{V}_k}$  be a family of local subalgebra, submodule, or representation zeta functions as above. There are  $k$ -varieties  $V_1, \dots, V_r$  and rational functions  $W_1(X, Y), \dots, W_r(X, Y) \in \mathbf{Q}(X, Y)$  such that for almost all  $v \in \mathcal{V}_k$ ,*

$$Z_v(s) = \sum_{i=1}^r \#\bar{V}_i(\mathfrak{K}_v) \cdot W_i(q_v, q_v^{-s}), \tag{2}$$

where  $\bar{V}_i$  denotes the reduction modulo  $\mathfrak{p}_v$  of a fixed  $\mathfrak{o}$ -model of  $V_i$ .

*Proof* For subobject zeta functions, this is due to du Sautoy and Grunewald [16] (cf. [38, Ex. 5.11(iii)]). For representation zeta functions associated with unipotent groups, it was proved by Stasinski and Voll [49, Pf of Thm A] (building upon previous work of Voll [55, §3.4]).

*Remark 3.2* A seemingly stronger version of Theorem 3.1 is given by [41, Thm 4.1]. This strengthened version takes into account not only the variation of the place  $v$  but also allows local base extensions in a suitable manner. However, by [43], the validity of (2) under variation of  $v$  (excluding finitely many exceptions) already implies the validity of its analogues after local base extensions. This observation allows us to rephrase some of our previous results more concisely in the present article.

While the proofs of Theorem 3.1 in the sources cited above are constructive, they all rely on some form of resolution of singularities for  $k$ -varieties; for non-constructive model-theoretic approaches, see e.g. [24, 37]. Even though algorithms for constructive resolution of singularities are known (see [7, 54]), these are typically impractical in the present context. Nonetheless, we obtain an algorithm which computes  $Z_v(s)$ , for each  $v \in \mathcal{V}_k$  outside of some finite set, as a rational function in  $q_v^{-s}$ . It is tempting to regard the explicit construction of (2) as the simultaneous computation of all generic local zeta functions  $Z_v(s)$  at once. This point of view, however, is not entirely satisfactory. For instance, it is unclear how to decide if two formulae of the form (2) define the same rational function for almost all  $v \in \mathcal{V}_k$ .

For many examples of interest, the phenomenon of “uniformity” mentioned in the introduction allows us to bypass such problems. We say that  $\mathbf{Z}$  is *uniform* if there exists  $W(X, Y) \in \mathbf{Q}(X, Y)$  such that  $Z_v(s) = W(q_v, q_v^{-s})$  for almost all  $v \in \mathcal{V}_k$ . While the author is not aware of any method for testing uniformity of  $\mathbf{Z}$ , if it is indeed uniform, our goal is to compute  $W(X, Y)$ .

Before we proceed further with our work towards this goal, the author would like to emphasise two points. First, he is not aware of a better general notion of computing generic local zeta functions than to construct a formula (2) (or a motivic analogue as in [17]). Secondly, he is not aware of a method for carrying out such a construction which is both general and practical. These two points explain why the author’s practical methods for computing generic local zeta functions, described in Sect. 6 below, are not general, i.e. they will not succeed in all cases.

## 4 Topological Zeta Functions

We now introduce the protagonist of [38–40]: topological zeta functions. These functions are defined analogously to topological zeta functions of polynomials, as introduced by Denef and Loeser [13]; topological subobject zeta functions were first defined by du Sautoy and Loeser [17].

### 4.1 An Informal “Definition”

Informally, we obtain the topological zeta function associated with a family  $Z = (Z_v(s))_{v \in \mathcal{V}_k}$  as in Theorem 3.1 by taking the limit “ $q_v \rightarrow 1$ ”, obtained as the constant term in the binomial expansion of a “generic”  $Z_v(s)$  as a series in  $q_v - 1$ . For example, by [49, Thm B], if  $H = U_3$  is the Heisenberg group scheme, then for each  $v \in \mathcal{V}_k$ ,

$$\widetilde{\zeta}_{H(\sigma_v)}^{\text{irr}}(s) = \frac{1 - q_v^{-s}}{1 - q_v^{1-s}} \tag{3}$$

and by symbolically expanding

$$q_v^{a-bs} = (1 + (q_v - 1))^{a-bs} = \sum_{\ell=0}^{\infty} \binom{a-bs}{\ell} (q_v - 1)^\ell,$$

we obtain  $\widetilde{\zeta}_{H(\sigma_v)}^{\text{irr}}(s) = \frac{s}{s-1} + \mathcal{O}(q_v - 1)$  whence the topological representation zeta function of  $H$  is  $\zeta_{H,\text{top}}(s) = s/(s - 1)$ . By [43, §3], this informal “definition” of topological zeta functions is rigorous in uniform cases such as (3). However, the author is not aware of a definition of topological zeta functions which is at the same time elementary, general, rigorous, and short. A pragmatic motivation for studying topological zeta functions is that they turn out to be the type of mathematical invariant which, while hard to define, can often be computed and studied effectively. Moreover, as observed by Denef and Loeser [13, Thm 2.2], by the very nature of the limit “ $q_v \rightarrow 1$ ” used to define them, topological zeta functions preserve interesting analytic properties of their local relatives.

### 4.2 A Rigorous Definition

Nowadays, topological zeta functions are most commonly studied in the context of motivic zeta functions and integrals. In contrast, the following exposition is based on the author’s axiomatisation [38, §5] of the original “arithmetic” definition of the topological zeta function of a polynomial by Denef and Loeser [13].

A rigorous notion of the limit “ $q_v \rightarrow 1$ ” is based on a formula (2). Specifically, we define such a limit separately for the terms “ $\#\bar{V}_i(\mathfrak{R}_v)$ ” and “ $W_i(q_v, q_v^{-s})$ ” and then combine them in the evident way.

First, we formalise taking a limit “ $q_v \rightarrow 1$ ” of  $W(q_v, q_v^{-s})$ . For  $e \in \mathbf{Q}[s]$ , write  $X^e := \sum_{\ell=0}^{\infty} \binom{e}{\ell} (X-1)^\ell \in \mathbf{Q}[s][[X-1]]$ . The map  $f(X, Y) \mapsto f(X, X^{-s})$  yields an embedding of  $\mathbf{Q}(X, Y)$  into  $\mathbf{Q}(s)((X-1))$ . In general,  $W(X, X^{-s})$  need not be a power series in  $X-1$  for  $W(X, Y) \in \mathbf{Q}(X, Y)$ . We will restrict attention to certain rational functions for which it is:

**Definition 4.1**

1. Let  $\mathbf{M}[X, Y] \subseteq \mathbf{Q}(X, Y)$  be the  $\mathbf{Q}$ -algebra consisting of those rational functions  $W(X, Y) \in \mathbf{Q}(X, Y)$  with  $W(X, X^{-s}) \in \mathbf{Q}(s)[[X-1]]$  and such that  $W(X, Y) = f(X, Y)/((1-X^{a_1}Y^{b_1}) \cdots (1-X^{a_r}Y^{b_r}))$  for non-zero  $(a_1, b_1), \dots, (a_r, b_r) \in \mathbf{Z}^2$  and a suitable  $f(X, Y) \in \mathbf{Q}[X^{\pm 1}, Y^{\pm 1}]$ .
2. Write  $\lfloor W(s) \rfloor$  for the image of  $W(X, Y) \in \mathbf{M}[X, Y]$  under “formal reduction modulo  $X-1$ ”, i.e. under the map  $f(X, Y) \mapsto f(X, X^{-s}) \bmod (X-1)$ .

The factors  $(1-q_v^{-1})^d$  in the definition of  $Z_v(s)$  in Sect. 3 were included to ensure the validity of the following:

**Lemma 4.2** *We may assume that  $W_1(X, Y), \dots, W_r(X, Y) \in \mathbf{M}[X, Y]$  in Theorem 3.1.*

*Proof* Combine [38, Thm 5.16] and [40, Lem. 3.4].

It remains to define a limit “ $q_v \rightarrow 1$ ” of  $\#\bar{V}_i(\mathfrak{R}_v)$  in (2). For background and further details on the following, we refer to [45, §4]. For  $v \in \mathcal{V}_k$ , fix an algebraic closure  $\bar{\mathfrak{K}}_v$  of  $\mathfrak{K}_v$  and denote by  $\mathfrak{R}_v^{(f)}$  the extension of  $\mathfrak{R}_v$  of degree  $f$  within  $\bar{\mathfrak{K}}_v$ . Let  $V$  be a  $k$ -variety. As above, we fix an  $\mathfrak{o}$ -model,  $\mathbf{V}$  say, of  $V$  and given  $v \in \mathcal{V}_k$ , we let  $\bar{V}$  denote the reduction modulo  $\mathfrak{p}_v$  of  $\mathbf{V}$ . It follows from Grothendieck’s trace formula and comparison theorems for  $\ell$ -adic cohomology that for almost all  $v \in \mathcal{V}_k$ , there are finitely many non-zero complex numbers  $\alpha_{ij}$  ( $i, j \geq 0$ ) such that for all  $f \in \mathbf{N}$ ,  $\#\bar{V}(\mathfrak{R}_v^{(f)}) = \sum_{i,j} (-1)^i \alpha_{ij}^f$  and, moreover,  $\#\bar{V}(\mathfrak{R}_v^{(0)}) := \sum_{i,j} (-1)^i \alpha_{ij}^0 = \chi(V(\mathbf{C}))$ ; here, the topological Euler characteristic  $\chi(V(\mathbf{C}))$  is taken with respect to an arbitrary embedding of  $k$  into  $\mathbf{C}$ . Numerous results in [45] justify defining  $\#\bar{V}(\mathfrak{R}_v^{(0)})$  as  $\chi(V(\mathbf{C}))$ . For example, by [41, Lem. 7.1] (an application of Chebotarev’s density theorem similar to arguments from [45]), if  $f(X) \in \mathbf{Z}[X]$  satisfies  $\#\bar{V}(\mathfrak{R}_v) = f(q_v)$  for almost all  $v \in \mathcal{V}_k$ , then  $\chi(V(\mathbf{C})) = f(1)$ .

In summary, our candidate for the topological zeta function associated with a family  $Z = (Z_v(s))_{v \in \mathcal{V}_k}$  as in Theorem 3.1 is  $\sum_{i=1}^r \chi(V_i(\mathbf{C})) \cdot [W_i(s)] \in \mathbf{Q}(s)$ . It remains to show that this rational function does not depend on the choice of the particular formula (2). This is the content of the following theorem.

**Theorem 4.3** *For  $v \in \mathcal{V}_k$ , let  $Z_v(s)$  be an analytic function on some complex right half-plane. Let  $V_1, \dots, V_r$  be  $k$ -varieties, let  $W_1(X, Y), \dots, W_r(X, Y) \in \mathbf{M}[X, Y]$ , and suppose that for almost all  $v \in \mathcal{V}_k$ ,  $Z_v(s) = \sum_{i=1}^r \# \bar{V}_i(\mathfrak{K}_v) \cdot W_i(q_v, q_v^{-s})$ . Then the following rational function is independent of the  $V_i$  and the  $W_i(X, Y)$ :*

$$Z_{\text{top}}(s) := \sum_{i=1}^r \chi(V_i(\mathbf{C})) \cdot [W_i(s)] \in \mathbf{Q}(s).$$

*Proof* Combine [38, Thm 5.12] and [43, Thm 3.2].

Theorem 4.3 generalises an insight of Denef and Loeser [13, (2.4)] at the heart of their original definition of topological zeta functions of polynomials.

**Definition 4.4** ([38, Def. 5.17]; [40, Def. 3.5]) In the setting of Theorem 3.1, we define the *topological subalgebra, submodule, or representation zeta function*  $\zeta_{\mathbf{A}, \text{top}}^{\leq}(s)$ ,  $\zeta_{\Omega \cap \mathbf{M}, \text{top}}(s)$ , or  $\zeta_{\mathbf{G}, \text{top}}^{\text{irr}}(s)$ , respectively, to be  $Z_{\text{top}}(s) \in \mathbf{Q}(s)$ , where  $Z$  is defined as in Sect. 3.

Up to a simple shift, our definition of topological subalgebra zeta functions is consistent with that of du Sautoy and Loeser [17, §8].

*Example 4.5* Let  $\mathfrak{h}$  be the Heisenberg Lie  $\mathbf{Z}$ -algebra. The subalgebra zeta function of  $\mathfrak{h}$  coincides with the subgroup zeta function of the discrete Heisenberg group in (1). Hence, for each prime  $p$ ,  $\zeta_{\mathfrak{h} \otimes \mathbf{Z}_p}^{\leq}(s) = W(p, p^{-s})$ , where  $W(X, Y)$  is given after (1). Thus, the topological subalgebra zeta function of  $\mathfrak{h}$  is the constant term of  $(1 - X^{-1})^3 W(X, X^{-s})$  as a series in  $X - 1$ , i.e.

$$\zeta_{\mathfrak{h}, \text{top}}^{\leq}(s) = \frac{3s - 3}{s(s - 1)(2s - 2)(2s - 3)} = \frac{3}{2s(s - 1)(2s - 3)}.$$

Observe that the real poles of  $\zeta_{\mathfrak{h} \otimes \mathbf{Z}_p}^{\leq}(s)$  and  $\zeta_{\mathfrak{h}, \text{top}}^{\leq}(s)$  coincide. While this is not a general phenomenon, Denef and Loeser [13, Thm 2.2] showed that poles of topological zeta functions always give rise to poles of suitable associated local zeta functions. In view of Igusa’s Monodromy Conjecture (see [11, §2.3]), this connection between poles of local and topological zeta functions provides one of the key motivations for studying the latter.

Example 4.5 is misleading in the simplicity of the formula for the topological zeta function and its derivation from knowledge of the associated local zeta functions. Indeed, one of the key features of the method for computing topological zeta functions in Sect. 6 is that it does not rely on computations of local zeta functions.

## 5 Tools from Convex Geometry

We briefly recall basic notions from convex geometry needed in the following.

### 5.1 Cones and Generating Functions

For details on most of the following, see e.g. [4]. A (linear) *half-space* in  $\mathbf{R}^n$  is a set of the form  $\{\omega \in \mathbf{R}^n : \langle \alpha, \omega \rangle \geq 0\}$ , where  $\alpha \in \mathbf{R}^n$  is non-zero and  $\langle \cdot, \cdot \rangle$  denotes the usual inner product. If  $\alpha$  can be chosen to have rational entries, then the half-space is *rational*. By a *cone* in  $\mathbf{R}^n$ , we mean a finite intersection of linear half-spaces; note that cones are (convex) polyhedra. If these half-spaces can all be taken to be rational, then we say that the cone is rational. By a *half-open* cone, we mean a set of the form  $\mathcal{C}_0 = \mathcal{C} \setminus (\mathcal{C}_1 \cup \dots \cup \mathcal{C}_r)$ , where  $\mathcal{C}$  is a cone and each  $\mathcal{C}_i$  is a face of  $\mathcal{C}$  (i.e. the intersection of  $\mathcal{C}$  with a supporting hyperplane). If the  $\mathcal{C}_i$  can be chosen to be precisely the faces of  $\mathcal{C}$  other than  $\mathcal{C}$  itself, then  $\mathcal{C}_0$  is a *relatively open cone*. If  $\mathcal{C}$  can be chosen to be rational, then we say that  $\mathcal{C}_0$  is rational. We say that  $\mathcal{C}_0$  is *pointed* if its closure does not contain a non-zero linear subspace. Supposing that  $\mathcal{C}_0$  is rational and pointed, it is well-known that the generating function  $\sum_{\omega \in \mathcal{C}_0 \cap \mathbf{Z}^n} \lambda^\omega \in \mathbf{Q}[[\lambda_1, \dots, \lambda_n]]$  enumerating (integer) lattice points in  $\mathcal{C}_0$  is given by a rational function in  $\mathbf{Q}(\lambda_1, \dots, \lambda_n)$ . The standard proof of rationality proceeds by triangulating the closure of  $\mathcal{C}_0$  followed by an application of the inclusion-exclusion principle. This argument does not, in general, lead to a practical algorithm. A more sophisticated method for computing generating functions is “Barvinok’s algorithm” as described by Barvinok and Woods [5]. The implementation of this algorithm as part of LattE [3] plays a vital role in the author’s software *Zeta*, to be described below.

Half-open cones are convenient for theoretical purposes. However, they appear scarcely in the literature and software usually does not support them directly. Fortunately, as explained in [39, §8.4], we may perform all computations required by the method described below using suitable polyhedra (non-canonically) attached to the half-open cones in question.

### 5.2 Newton Polytopes and Initial Forms

Most of the following is well-known but the term “balanced” is non-standard; for references, see [38, §4.1].

Let  $f \in k[X^{\pm 1}] := k[X_1^{\pm 1}, \dots, X_n^{\pm 1}]$ . Write  $f = \sum_{\alpha \in \mathbf{Z}^n} c_\alpha X^\alpha$ , where  $c_\alpha \in k$  and  $c_\alpha = 0$  for almost all  $\alpha \in \mathbf{Z}^n$ . Let  $\text{supp}(f) := \{\alpha \in \mathbf{Z}^n : c_\alpha \neq 0\}$  and define the *Newton polytope*  $\text{New}(f)$  of  $f$  to be the convex hull of  $\text{supp}(f)$  in  $\mathbf{R}^n$ . Suppose that  $f \neq 0$  so that  $\text{New}(f) \neq \emptyset$ . For  $\omega \in \mathbf{R}^n$ , let  $m(f, \omega) := \min_{\alpha \in \text{supp}(f)} \langle \alpha, \omega \rangle$ . We define

the *initial form* of  $f$  in the direction  $\omega$  to be

$$\text{in}_\omega(f) := \sum_{\substack{\alpha \in \text{supp}(f), \\ \langle \alpha, \omega \rangle = n(f, \omega)}} c_\alpha X^\alpha.$$

**Definition 5.1** ([39, Def. 5.1(i)]) Let  $\emptyset \neq \mathcal{M} \subseteq \mathbf{R}^n$  and let  $0 \neq f \in k[X^{\pm 1}]$ . We say that  $f$  is  $\mathcal{M}$ -balanced if  $\omega \mapsto \text{in}_\omega(f)$  is constant on  $\mathcal{M}$ .

Define an equivalence relation  $\sim_f$  on  $\mathbf{R}^n$  by letting  $\omega \sim_f \omega'$  if and only if  $\text{in}_\omega(f) = \text{in}_{\omega'}(f)$ . Thus,  $f$  is  $\mathcal{M}$ -balanced if and only if  $\mathcal{M}$  is contained in one of the equivalence classes of  $\sim_f$ . We will now recall descriptions of these classes in terms of the Newton polytope of  $f$ .

Given a non-empty polytope  $\mathcal{P} \subseteq \mathbf{R}^n$  and  $\omega \in \mathbf{R}^n$ , let  $\text{face}_\omega(\mathcal{P})$  be the face of  $\mathcal{P}$  consisting of those  $\alpha \in \mathcal{P}$  which minimise  $\langle \alpha, \omega \rangle$  over  $\mathcal{P}$ . The (relatively open) *normal cone* of a face  $\tau \subseteq \mathcal{P}$  is  $N_\tau(\mathcal{P}) := \{\omega \in \mathbf{R}^n : \text{face}_\omega(\mathcal{P}) = \tau\}$ . The equivalence classes of  $\sim_f$  from above are precisely the normal cones  $N_\tau(\text{New}(f))$  for faces  $\tau \subseteq \text{New}(f)$ . In particular, the finite set  $\{\text{in}_\omega(f) : \omega \in \mathbf{R}^n\}$  is in natural bijection with the set of faces of  $\text{New}(f)$ . The following is now obvious.

**Lemma 5.2** ([39, Lem. 5.3]) *Let  $\emptyset \neq \mathcal{M} \subseteq \mathbf{R}^n$  and  $0 \neq f \in k[X^{\pm 1}]$ . Then  $f$  is  $\mathcal{M}$ -balanced if and only if there exists a face  $\tau \subseteq \text{New}(f)$  with  $\mathcal{M} \subseteq N_\tau(\text{New}(f))$ .*

Now suppose that  $f = (f_1, \dots, f_r)$  for non-zero  $f_1, \dots, f_r \in k[X^{\pm 1}]$ . One can show (cf. [38, §3.3]) that the equivalence classes of  $\sim_f$  defined by letting  $\omega \sim_f \omega'$  if and only if  $\text{in}_\omega(f_i) = \text{in}_{\omega'}(f_i)$  for  $i = 1, \dots, r$  are precisely the normal cones associated with faces of  $\text{New}(f_1 \cdots f_r)$ .

## 6 A Framework for Computing Zeta Functions

In this section, we provide a unified summary of the author’s methods for computing generic local and topological zeta functions. For the sake of a more streamlined exposition, we only spell out the case of subalgebra zeta functions.

We begin by recalling the translation step (Sect. 6.1) which reduces the computation of local zeta functions to that of computing  $p$ -adic integrals. As we will explain in Sect. 6.2, these integrals can be encoded in terms of objects that we call “toric data”. In Sect. 6.3, we introduce the key notions of “balanced” and “regular” toric data. The “simplify-balance-reduce loop” at the heart of our method is discussed in Sect. 6.4. Assuming its successful completion, we face different tasks depending on whether we seek to compute (generic) local or topological zeta functions; these tasks are discussed in Sect. 6.5.



### 6.1 *p*-Adic Integration

Let  $\mathbf{A}$  be an  $\mathfrak{o}$ -algebra which is free of rank  $d$  as an  $\mathfrak{o}$ -module. By choosing a basis, we identify  $\mathbf{A}$  and  $\mathfrak{o}^d$  as  $\mathfrak{o}$ -modules. This allows us to parameterise submodules of  $\mathbf{A}$  using the row spans of upper-triangular  $d \times d$  matrices. Building upon work of Grunewald et al. [23, §3], du Sautoy and Grunewald [16, Thm 5.5] observed that those submodules of  $\mathbf{A}$  which are subalgebras can be characterised in terms of polynomial divisibility conditions in the entries of matrices. We formalise this as in [38, Rem. 2.7(ii)].

Let  $R := \mathfrak{o}[\mathbf{X}] := \mathfrak{o}[X_{ij} : 1 \leq i \leq j \leq d]$  and  $C := [\delta_{i \leq j} \cdot X_{ij}] \in \text{Tr}_d(R)$ , where  $\delta_{i \leq j} = 1$  if  $i \leq j$  and  $\delta_{i \leq j} = 0$  otherwise. We identify  $R^d = \mathbf{A} \otimes_{\mathfrak{o}} R$  and in particular regard  $R^d$  as an  $R$ -algebra. Let  $C_1, \dots, C_d$  be the rows of  $C$ . Let  $\text{adj}(C) \in \text{Tr}_d(R)$  be the adjugate matrix of  $C$ ; hence, over  $k(\mathbf{X})$ ,  $\text{adj}(C) = \det(C)C^{-1}$ .

Henceforth, for  $v \in \mathcal{V}_k$ , let  $\mu_v$  denote the additive Haar measure on  $k_v$  with  $\mu_v(\mathfrak{o}_v) = 1$ ; we use the same symbol for the product measure on  $k_v^d$  and  $\text{Tr}_d(k_v)$ . Moreover, we let  $|\cdot|_v$  denote the usual  $v$ -adic absolute value with  $|\pi|_v = q_v^{-1}$  for  $\pi \in \mathfrak{p}_v \setminus \mathfrak{p}_v^2$ . Finally, we write  $\|A\|_v := \sup(|a|_v : a \in A)$ . The following expresses each  $\zeta_{\mathbf{A}_v}^{\leq}(s)$  as a ‘‘cone integral’’ in the sense of du Sautoy and Grunewald [16].

**Theorem 6.1** ([16, Thm 5.5]; cf. [23, Prop. 3.1]) *Let  $f \subseteq \mathfrak{o}[\mathbf{X}^{\pm 1}]$  consist of the non-zero entries of all tuples of the form  $\det(C)^{-1}(C_m C_n) \text{adj}(C)$  for  $1 \leq m, n \leq d$ . Then for each  $v \in \mathcal{V}_k$ ,*

$$\zeta_{\mathbf{A}_v}^{\leq}(s) = (1 - q_v^{-1})^{-d} \int_{\{\mathbf{x} \in \text{Tr}_d(\mathfrak{o}_v) : \|f(\mathbf{x})\|_v \leq 1\}} \prod_{i=1}^d |x_{ii}|_v^{s-i} d\mu_v(\mathbf{x}). \tag{4}$$

We remark that local submodule zeta functions can be similarly expressed in terms of  $p$ -adic integrals of the same shape as (4).

Let  $\mathbf{G} \leq U_d \otimes k$  be a unipotent algebraic group over  $k$  with associated  $\mathfrak{o}$ -form  $\mathbf{G} \leq U_d \otimes \mathfrak{o}$ . Stasinski and Voll [49, §2.2.3] expressed  $\zeta_{\mathbf{G}(\mathfrak{o}_v)}^{\text{irr}}(s)$ , for almost all  $v \in \mathcal{V}_k$ , in terms of a  $p$ -adic integral defined using a fixed set of globally defined polynomials. While the author’s framework is flexible enough to accommodate these integrals (see [38, Def. 4.6] and [40, §5.1]), for the sake of simplicity, in the following, we only consider integrals of the form (4).

### 6.2 Toric Data and Associated Integrals

Henceforth, in addition to  $k$ , we fix an ‘‘ambient space’’ of dimension  $n$ ; in the setting of Theorem 6.1, our ambient space will be  $\text{Tr}_d$  so that  $n = d(d + 1)/2$ .

**Definition 6.2** ([39, Def. 3.1]) A *toric datum* is a pair  $\mathcal{T} = (\mathcal{C}_0; \mathbf{f})$ , where  $\mathcal{C}_0 \subseteq \mathbf{R}_{\geq 0}^n$  is a half-open cone (see Sect. 5.1) and  $\mathbf{f} = (f_1, \dots, f_r)$  is a finite family of non-zero Laurent polynomials  $f_i \in k[\mathbf{X}^{\pm 1}] := k[X_1^{\pm 1}, \dots, X_n^{\pm 1}]$ .

Henceforth, we tacitly assume that  $\mathcal{C}_0 \neq \emptyset$ . We now explain how a toric datum  $\mathcal{T} = (\mathcal{C}_0; \mathbf{f})$  gives rise to  $p$ -adic integrals. For  $v \in \mathcal{V}_k$  and  $\mathbf{x} \in k_v^n$ , write  $v(\mathbf{x}) = (v(x_1), \dots, v(x_n))$ ; an elementary but crucial observation is that if  $x_1 \cdots x_n \neq 0$  and  $\alpha \in \mathbf{Z}^n$ , then  $v(\mathbf{x}^\alpha) = \langle \alpha, v(\mathbf{x}) \rangle$ . Define  $\mathcal{C}_0(\mathfrak{o}_v) := \{\mathbf{x} \in \mathfrak{o}_v^n : v(\mathbf{x}) \in \mathcal{C}_0\}$ .

**Definition 6.3** Let  $\mathcal{T} = (\mathcal{C}_0; \mathbf{f})$  be a toric datum,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)$  for  $\beta_1, \dots, \beta_m \in \mathbf{N}_0^n$ ,  $v \in \mathcal{V}_k$ , and let  $s_1, \dots, s_m$  be complex variables. Define

$$Z_v^{\mathcal{T}, \boldsymbol{\beta}}(s_1, \dots, s_m) := \int_{\{\mathbf{x} \in \mathcal{C}_0(\mathfrak{o}_v) : \|f(\mathbf{x})\|_v \leq 1\}} |\mathbf{x}^{\beta_1}|_v^{s_1} \cdots |\mathbf{x}^{\beta_m}|_v^{s_m} d\mu_v(\mathbf{x}). \tag{5}$$

Thus, the integral in (4) is a univariate specialisation of (5) (with  $\mathcal{C}_0 = \mathbf{R}_{\geq 0}^n$ ).

### 6.3 Balanced and Regular Toric Data

We will now explain how under a suitable regularity hypothesis for a toric datum  $\mathcal{T}$ , we may construct an explicit (multivariate analogue of) formula (2) for the integrals  $Z_v^{\mathcal{T}, \boldsymbol{\beta}}(s_1, \dots, s_m)$ . Write  $\mathbf{T}^n := \text{Spec}(\mathbf{Z}[X_1^{\pm 1}, \dots, X_n^{\pm 1}])$  and identify  $\mathbf{T}^n(R) = (R^\times)^n$  for any commutative ring  $R$ . Let  $\bar{k}$  be an algebraic closure of  $k$ .

**Definition 6.4** ([39, Def. 5.1(ii), Def. 5.5]) Let  $\mathcal{T} = (\mathcal{C}_0; \mathbf{f})$  be a toric datum with  $\mathbf{f} = (f_1, \dots, f_r)$  as above.

- $\mathcal{T}$  is *balanced* if  $f_i$  is  $\mathcal{C}_0$ -balanced (see Definition 5.1) for  $i = 1, \dots, r$ .
- $\mathcal{T}$  is *regular* if it is balanced and the following holds:  
 for each  $J \subseteq \{1, \dots, r\}$  and  $\mathbf{u} \in \mathbf{T}^n(\bar{k})$  with  $f_j(\mathbf{u}) = 0$  for all  $j \in J$ , the rank of

$$\left[ \frac{\partial \text{in}_\omega(f_j)(\mathbf{u})}{\partial X_i} \right]_{\substack{i=1, \dots, n; \\ j \in J}}$$

is  $\#J$ , where  $\omega \in \mathcal{C}_0$  is arbitrary (the particular choice of  $\omega$  being irrelevant).

Using Lemma 5.2 and the comments following it, we may test if a toric datum  $\mathcal{T}$  is balanced. As explained in [39, §5.2], regularity can then be tested using Gröbner bases techniques.

**Theorem 6.5** ([39, Thm 5.8]) Let  $\mathcal{T} = (\mathcal{C}_0; \mathbf{f})$  be a regular toric datum, where  $\mathbf{f} = (f_1, \dots, f_r)$ . Let  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)$  be as in Definition 6.3. For  $J \subseteq \{1, \dots, r\}$ , let  $V_J^\circ \subseteq \mathbf{T}^n \otimes k$  be the subvariety defined by  $f_j = 0$  for  $j \in J$  and  $f_i \neq 0$  for  $i \notin J$ .

Then there are explicit  $W_J \in \mathbf{Q}(X, Y_1, \dots, Y_m)$  such that for almost all  $v \in \mathcal{V}_k$ ,

$$Z_v^{\mathcal{T}, \beta}(s_1, \dots, s_m) = q_v^{-n} \sum_{J \subseteq \{1, \dots, r\}} \# \bar{V}_J^\circ(\mathcal{R}_v) \cdot (q_v - 1)^{\#J} \cdot W_J(q_v, q_v^{-s_1}, \dots, q_v^{-s_m}). \tag{6}$$

The  $W_J$  in Theorem 6.5 are given explicitly in the sense that they arise via (explicit) monomial substitutions from generating functions enumerating lattice points inside certain half-open cones  $\mathcal{C}_0^J \subseteq \mathcal{C}_0 \times \mathbf{R}^{\#J}$ ; see [39, §5.5] for details.

Theorem 6.5 is an algorithmically-minded consequence of [38, Thm 4.10]. The latter theorem provides formulae such as (6) for  $p$ -adic integrals of a quite general shape under suitable “non-degeneracy” conditions (closely related to the above concept of regularity for toric data). Such notions of non-degeneracy have their origin in work of Khovanskii [27, 28] and others [6, 31, 52] in toric geometry. They also found numerous applications in the study of Igusa’s local zeta function, a close relative of the zeta functions studied here. Indeed, [38, Thm 4.10] was inspired by (and generalises) a result of Denef and Hoornaert [12, Thm 4.2]; another source of inspiration is given by work of Veys and Zúñiga-Galindo [53, §4]. For a more detailed comparison between the author’s approach and previous work in the literature, we refer to [38, §4.4].

Much like (2), the formalism for attaching topological zeta functions to families of local ones in Sect. 4.2 admits a natural multivariate version; see [38, §5]. However, as a technical inconvenience, we cannot pass directly from (6) to the associated topological zeta function since the Laurent series  $(X - 1)^{\#J} W_J(X, X^{-s_1}, \dots, X^{-s_m})$  in  $X - 1$  over  $\mathbf{Q}(s_1, \dots, s_m)$  typically fail to be power series in  $X - 1$ . Fortunately, as explained in [39, §6.4], it turns out that we may rewrite (6) (altering both the varieties and the rational functions involved in the process) in a way that allows us to pass to the associated topological zeta function analogously to Theorem 4.3. We note that passing from multivariate local zeta functions to topological ones is compatible with suitable univariate specialisations such as the ones used here; see [38, Rem. 5.15].

### 6.4 The Simplify-Balance-Reduce Loop

We now discuss the heart of our method. Starting with an  $\mathfrak{o}$ -algebra  $\mathbf{A}$ , we seek to construct a formula (2) for its generic local subalgebra zeta functions. As we have seen, these zeta functions are expressible in terms of  $p$ -adic integrals attached to an initial toric datum  $\mathcal{T}^0 = (\mathbf{R}_{\geq 0}^n; \mathbf{f}^0)$ , where  $\mathbf{f}^0$  is a set of Laurent polynomials such as the set  $\mathbf{f}$  in Theorem 6.1. (The integrand encoded by  $\beta$  in Definition 6.3 is all but insignificant and will be ignored in the following.) Our method is based on several operations applied to toric data as part of a loop.

### 6.4.1 Overview

At all times of our loop, we maintain a finite collection,  $\mathcal{T}$  say, of toric data such that for almost all  $v \in \mathcal{V}_k$ , the integral (5) associated with our initial toric datum  $\mathcal{T}^0$  (essentially the subalgebra zeta function of  $\mathbf{A}_v$ ) is given by the sum of the integrals corresponding to the elements of  $\mathcal{T}$ ; similarly, the topological zeta function associated with  $\mathcal{T}^0$  (or, equivalently, the topological subalgebra zeta function of  $\mathbf{A}$ ) will be expressed as a sum of the topological zeta functions attached to the elements of  $\mathcal{T}$ . Initially,  $\mathcal{T}$  only consists of  $\mathcal{T}^0$ . We repeatedly process those elements of  $\mathcal{T}$  that we have not already found to be regular. More precisely, if any such element,  $\mathcal{T}$  say, fails to meet certain criteria, then we derive new toric data  $\mathcal{T}_1, \dots, \mathcal{T}_N$ , say, from  $\mathcal{T}$ , remove  $\mathcal{T}$  from  $\mathcal{T}$ , insert  $\mathcal{T}_1, \dots, \mathcal{T}_N$  into  $\mathcal{T}$ , and resume processing the elements of  $\mathcal{T}$ . We now give details on how exactly we process a given toric datum  $\mathcal{T} = (\mathcal{C}_0; \mathbf{f}) \in \mathcal{T}$ .

### 6.4.2 Simplification

First, we “simplify”  $\mathcal{T}$ . The key observation is that we may replace  $\mathcal{T}$  by any other toric datum if almost all of the associated  $p$ -adic integrals remain unchanged. Apart from obvious operations such as removing duplicates or constants from  $\mathbf{f}$ , we are e.g. also free to replace  $\mathbf{f}$  by another finite generating set of the same  $k[\mathbf{X}]$ -submodule of  $k[\mathbf{X}^{\pm 1}]$ . Moreover, we may remove all Laurent monomials from  $\mathbf{f}$  for an integrality condition “ $|\mathbf{x}^\alpha|_v \leq 1$ ” is equivalent to the constraint “ $\langle \alpha, v(\mathbf{x}) \rangle \geq 0$ ” on  $v(\mathbf{x})$  which can be encoded by shrinking  $\mathcal{C}_0$  accordingly. The precise operations that we carry out are explained in [39, §§7.1–7.2].

### 6.4.3 Balancing

Suppose that  $\mathcal{T}$  has been simplified but that it is not balanced. By considering the non-empty intersections of  $\mathcal{C}_0$  with the normal cones of  $\text{New}(\prod \mathbf{f})$  (see Sect. 5.2), we obtain a partition  $\mathcal{C}_0 = \bigcup_{i=1}^N \mathcal{C}_0^i$  such that each  $\mathcal{T}_i := (\mathcal{C}_0^i; \mathbf{f})$  is balanced. We then remove  $\mathcal{T}$  from  $\mathcal{T}$  and insert  $\mathcal{T}_1, \dots, \mathcal{T}_N$ .

### 6.4.4 Reduction

It remains to consider the case that  $\mathcal{T}$  is *singular*, i.e. balanced but not regular. The author is unaware of a practically useful method for dealing with these cases in general. Instead, the “reduction step” from [39, §7.3] is an attempt to repair certain specific types of singularity which the author frequently encountered in examples of interest. This method may not lead to immediate improvements and to ensure termination, we impose a bound on the number of subsequent reduction steps. If this number is exceeded, we let our method fail.

Instead of reiterating [39, §7.3], we illustrate the reduction step by discussing the special case that gave rise to the general form. Namely, suppose that  $\mathcal{T} = (\mathcal{C}_0; \mathbf{f})$  is balanced, where  $\mathbf{f} = (f_1, \dots, f_r)$  and  $r \geq 2$ . Choose  $\omega \in \mathcal{C}_0$ . Further suppose that there are  $\alpha_1, \alpha_2 \in \mathbf{Z}^n$  and  $g \in k[X^{\pm 1}]$  such that  $\text{in}_\omega(f_i) = X^{\alpha_i}g$  for  $i = 1, 2$ ; write  $h_i = f_i - X^{\alpha_i}g$ . We assume that  $g$  consists of more than one term (i.e.  $\#\text{supp}(g) \geq 2$ ) whence  $\mathcal{T}$  is singular. We decompose  $\mathcal{C}_0$  into half-open cones  $\mathcal{C}_0^{\leq}$  and  $\mathcal{C}_0^>$  defined by

$$\begin{aligned} \mathcal{C}_0^{\leq} &:= \{\lambda \in \mathcal{C}_0 : \langle \alpha_1, \lambda \rangle \leq \langle \alpha_2, \lambda \rangle\} \text{ and} \\ \mathcal{C}_0^> &:= \{\lambda \in \mathcal{C}_0 : \langle \alpha_1, \lambda \rangle > \langle \alpha_2, \lambda \rangle\}. \end{aligned}$$

Instead of  $\mathcal{T}$ , we may then consider the two toric data  $\mathcal{T}^{\leq} := (\mathcal{C}_0^{\leq}; \mathbf{f})$  and  $\mathcal{T}^> := (\mathcal{C}_0^>; \mathbf{f})$ . We only consider  $\mathcal{T}^{\leq}$  in the following, the case of  $\mathcal{T}^>$  being analogous. We also assume that  $\mathcal{C}_0^{\leq}$  is non-empty. If  $\mathbf{x} \in \mathfrak{o}_v^n$  with  $v(\mathbf{x}) \in \mathcal{C}_0^{\leq}$ , then  $v(\mathbf{x}^{\alpha_2 - \alpha_1}) \geq 0$ . It follows that  $\mathbf{Z}_v^{\mathcal{T}^{\leq}, \beta}(s_1, \dots, s_m)$  remains unaffected if we “remove” one reason for the singularity of  $\mathcal{T}^{\leq}$ , namely the summand “ $X^{\alpha_2}g$ ” of  $f_2$ , by replacing  $f_2$  by  $f'_2 := f_2 - X^{\alpha_2 - \alpha_1}f_1 = h_2 - X^{\alpha_2 - \alpha_1}h_1$ . The resulting toric datum may no longer be balanced. We therefore process it using the steps discussed so far in the hope that eventually, all singularities will be successfully removed.

### 6.5 Processing the Pieces

Assuming successful termination of the “simplify-balance-reduce loop”, we obtain a formula (2) for  $\zeta_{\mathbf{A}_v}^{\leq}(s)$  (and almost every  $v \in \mathcal{V}_k$ ) by applying Theorem 6.5 to each regular toric datum that we constructed. In this formula, the  $V_i$  are given as subvarieties of tori  $\mathbf{T}^{m_i} \otimes k$  and the  $W_i(X, Y)$  are “described” combinatorially (but not yet computed) in terms of generating functions enumerating lattice points inside certain half-open cones. Our next step is to carry out further computations involving the  $V_i$  and  $W_i(X, Y)$ . These computations will depend on whether we seek to compute topological or generic local zeta functions.

#### 6.5.1 Topological Computations

We first consider the computation of  $\zeta_{\mathbf{A}, \text{top}}^{\leq}(s) \in \mathbf{Q}(s)$ . As we mentioned above, by rewriting the formulae obtained using Theorem 6.5 as in [39, §6.4], we may assume that  $W_i(X, Y) \in \mathbf{M}[X, Y]$  for each  $i$  in (2); the  $V_i$  will still be given as subvarieties (closed ones even) of tori  $\mathbf{T}^{m_i} \otimes k$ . We are thus left with three steps:

- (T1) compute each  $\chi(V_i(\mathbf{C}))$ ,
- (T2) compute each  $[W_i(s)]$ , and
- (T3) compute  $\sum_{i=1}^r \chi(V_i(\mathbf{C})) \cdot [W_i(s)]$  as a fraction of polynomials from  $\mathbf{Q}[s]$ .

Regarding (T1), there are general-purpose algorithms for computing Euler characteristics of varieties; see, in particular, work of Aluffi [1]. We do not make any use of these techniques in practice. Instead, we rely on the following two ingredients. First, the Bernstein-Khovanskii-Kushnirenko (BKK) Theorem [28, §3, Thm 2] provides a convex-geometric formula for the topological Euler characteristic of (the complex analytic space associated with) a closed subvariety  $f_1 = \dots = f_m = 0$  of  $\mathbf{T}^n \otimes \mathbf{C}$  if  $(f_1, \dots, f_m)$  is non-degenerate in the sense of [27, §2]. Khovanskii’s notion of non-degeneracy is closely related to our concept of regularity; see [38, §4.2]. In particular, if the reduction step from Sect. 6.4 should not be needed during our computations, then the BKK Theorem can be applied to all varieties that we encounter; cf. [39, Rem. 6.15(ii)]. Secondly, we employ a recursive procedure which seeks to compute topological Euler characteristics associated with closed subvarieties of  $\mathbf{T}^n \otimes k$  by decomposing these varieties using subvarieties of lower-dimensional tori. While this procedure is not guaranteed to work in all cases, it has proven to be very useful in practice. Details are given in [39, §6.6] (with some further explanations in [41, §5]).

For (T2), in case  $W_i(X, Y)$  is obtained using the method from above, the computation of  $[W_i(s)]$  is described in [39, §6.5]. An important observation (already used implicitly by Denef and Loeser [13, §5]) is that while  $W_i(X, Y)$  arises from a generating function enumerating lattice points inside a half-open cone,  $\mathcal{D}_0$  say,  $[W_i(s)]$  can be written as a sum of rational functions indexed by the cones of maximal dimension in a triangulation of the closure of  $\mathcal{D}_0$ .

Finally, step (T3) remains. As described in [39, §8.3], we can easily keep track of a common denominator of all  $[W_i(s)]$  which allows us to recover  $\zeta_{\mathbf{A}, \text{top}}^{\leq}(s)$  using evaluation at random points and polynomial interpolation. This concludes our method for computing topological subalgebra zeta functions.

### 6.5.2 Generic Local Computations

Given an associated formula (2) obtained as above, we consider the computation of the generic local subalgebra zeta functions  $\zeta_{\mathbf{A}_v}^{\leq}(s)$ . We make an assumption which is even stronger than uniformity as defined in Sect. 3. Namely, we assume that for  $i = 1, \dots, r$ , there exists  $c_i(X) \in \mathbf{Z}[X]$  such that  $\#\bar{V}_i(\mathfrak{K}_v) = c_i(q_v)$  for almost all  $v \in \mathcal{V}_k$ . The author would like to note that he is not aware of any method for deciding if this assumption is satisfied (or for computing the  $c_i(X)$  if it is); for computations in possibly non-uniform settings, see [41, §§5–6,8].

Inspired by steps (T1)–(T3) from above, we proceed as follows:

- (L1) attempt to construct each  $c_i(X) \in \mathbf{Z}[X]$  (failure being an option),
- (L2) compute each  $W_i(X, Y)$  as a sum of bivariate rational functions, and
- (L3) compute  $W(X, Y) \in \mathbf{Q}(X, Y)$  with  $Z_{\mathbf{A}_v}^{\leq}(s) = W(q_v, q_v^{-s})$  for almost all  $v \in \mathcal{V}_k$ .

For (L1), we extend ideas from the computation of Euler characteristics in (T1). We sketch the key ingredients; for details, see [41, §5]. Let  $f_1, \dots, f_m \in k[X^{\pm 1}] = k[X_1^{\pm 1}, \dots, X_n^{\pm 1}]$  be non-zero. Let  $V \subseteq \mathbf{T}^n \otimes k$  be defined by  $f_1 = \dots = f_m = 0$ . We seek to find  $c(X) \in \mathbf{Z}[X]$  such that  $\#\bar{V}(\mathfrak{K}_v) = c(q_v)$  for almost all  $v \in \mathcal{V}_k$ . This is trivial for  $n = 0$ . For  $n = 1$ , the Euclidean algorithm allows us to assume that  $m = 1$ . We then check if the roots of  $f_1$  lie in  $k$  (in which case, we take  $c(X)$  to be the number of distinct roots) and abort if it does not. We may thus assume  $n > 1$ . Similarly to the simplification step in Sect. 6.4, we use the fact that we are free to replace the  $f_i$  by any collection of Laurent polynomials which generates the same ideal of  $k[X^{\pm 1}]$ . As one potential reduction of dimensions, we then construct an isomorphism  $V \approx_k U \times_k (\mathbf{T}^{n-d} \otimes k)$ , where  $U$  is a closed subvariety of  $\mathbf{T}^d \otimes k$  and  $d = \dim(\text{New}(f_1 \cdots f_m))$ ; see [38, §6.1] and [39, §6.3]. Other potential reductions of dimensions are obtained by trying to solve each  $f_i = 0$  for one of the variables as in [41, Lem. 5.1–5.2].

For (L2), we use algorithms due to Barvinok and others [5] for computing and manipulating generating functions associated with polyhedra. Using these methods, each  $W_i(X, Y)$  will be expressed as a sum of rational functions. For (L3), similarly to (T3), we write the final sum of (2) over a common denominator. However, due to the frequently large degree of said denominator (in  $X$  and  $Y$ ), at least a naive variation of the approach based on polynomial interpolation from (T3) is often impractical. Instead, after grouping together rational functions based on heuristics (partially inspired by ideas of Woodward [60, §2.5]), we add all numerators over our common denominator. This is usually by far the most computationally involved step of all.

## 6.6 Zeta

The author’s software package **Zeta** [44] for Sage [51] implements his methods for computing generic local and topological subalgebra, submodule, and representation zeta functions; moreover, **Zeta** offers basic support for Igusa-type zeta functions associated with polynomials and polynomial mappings (as in [53] but using the author’s notion of non-degeneracy instead of [53, Def. 4.1]). Apart from functionality built into Sage, **Zeta** makes critical use of Singular [9] (polynomial arithmetic, Gröbner bases), Normaliz [8] (triangulations), and LattE [3] (generating functions associated with polyhedra).

## 7 Highlights of Computations Using Zeta

The topological subalgebra and representation zeta functions as defined in Sect. 4.2 are all invariant under base change in the sense that they only depend on the  $\mathbf{C}$ -isomorphism class of  $\mathbf{A} \otimes_{\circ} \mathbf{C}$  and  $\mathbf{G} \otimes_{\circ} \mathbf{C}$ , respectively; see [38, Prop. 5.19]

and [40, Prop. 4.3]. Apart from the  $\mathbf{C}$ -isomorphism class of a single 5-dimensional algebra, dubbed  $\text{Fil}_4$  by Woodward, the topological subalgebra zeta functions of nilpotent Lie algebras of dimension  $\leq 5$  can all be derived (via [43, §3]) from previous  $p$ -adic calculations. The algebra  $\text{Fil}_4$  has a basis  $(e_1, \dots, e_5)$  with  $[e_1, e_2] = e_3$ ,  $[e_1, e_3] = e_4$ ,  $[e_1, e_4] = e_5$ ,  $[e_2, e_3] = e_5$  and such that all remaining commutators of basis elements (except for those implied by anti-commutativity) are zero. Based on computations using *Zeta*, the following was first announced in [38, §7.3]:

**Theorem 7.1 ([39, §9.1])**

$$\begin{aligned} \zeta_{\text{Fil}_4, \text{top}}(s) = & (392031360s^9 - 5741480808s^8 + 37286908278s^7 - \\ & 140917681751s^6 + 341501393670s^5 - 550262853249s^4 + \\ & 589429290044s^3 - 404678115300s^2 + 161557332768s - \\ & 28569052512) / (3(15s - 26)(7s - 12)(7s - 13)(6s - 11)^3 \\ & (5s - 8)(5s - 9)(4s - 7)^2(3s - 4)(2s - 3)(s - 1)s). \end{aligned}$$

The seemingly bizarre numbers in the numerator are consistent with the four conjectures stated in [38, §8], two of which we will discuss below. The generic local subalgebra zeta functions associated with  $\text{Fil}_4$  remain unknown.

Prior to the following,  $\mathfrak{sl}_2(\mathbf{Q})$  was the only example of an insoluble Lie algebra whose associated generic local subalgebra zeta functions had been computed.

**Theorem 7.2 ([41, Thm 9.1])** *For almost all primes  $p$ ,  $\zeta_{\mathfrak{gl}_2(\mathbf{Z}_p)}^{\leq}(s) = W(p, p^{-s})$ , where*

$$\begin{aligned} W(X, Y) = & (-X^8Y^{10} - X^8Y^9 - X^7Y^9 - 2X^7Y^8 + X^7Y^7 - X^6Y^8 - X^6Y^7 + 2X^6Y^6 \\ & - 2X^5Y^7 + 2X^5Y^5 - 3X^4Y^6 + 3X^4Y^4 - 2X^3Y^5 + 2X^3Y^3 - 2X^2Y^4 \\ & + X^2Y^3 + X^2Y^2 - XY^3 + 2XY^2 + XY + Y + 1) / ((1 - X^7Y^6) \\ & (1 - X^3Y^3)(1 - X^2Y^2)^2(1 - Y)). \end{aligned}$$

Noting that  $\mathfrak{gl}_2(\mathbf{Z}_p) \approx \mathfrak{sl}_2(\mathbf{Z}_p) \oplus \mathbf{Z}_p$  for  $p \neq 2$ , this formula in particular illustrates the generally wild effect of direct sums on subalgebra zeta functions. We note that Theorem 7.2 is consistent with results of Voll [55, Thm A] and Evseev [21, Thm 3.3].

Other computations of particular interest are that of  $\zeta_{U_d(\mathbf{Z}_p) \curvearrowright \mathbf{Z}_p^d}(s)$  for  $d \leq 5$  and almost all primes  $p$  (see [41, §9.4]); the formula for  $d = 5$  fills about three pages. These computations are consistent with functional equations recently established by Voll [58, Thm 5.5] as well as with [42, Prop. 6.1] (which implies that the abscissa of convergence of  $\zeta_{U_d(\mathbf{Z}) \curvearrowright \mathbf{Z}^d}(s)$  is 1 for any  $d \geq 1$ ).



Regarding representation zeta functions, extending previous work of others, the author (with the help of **Zeta**) finished the determination of the generic local representation zeta functions of unipotent algebraic groups of dimension at most 6 over number fields (see [41, §8]); we note that there are infinitely many such groups of dimension 6. The representation zeta functions of  $U_d(\mathbf{Z}_p)$  are only known for  $d \leq 5$ ; the case  $d = 5$  was settled, for almost all  $p$ , using **Zeta** (see [41, Thm 8.4]).

For comments on limitations of the author’s method, see [39, §8.2] and [41, §6.4]. In particular, to the author’s knowledge, not a single explicit example of a (local or topological) subalgebra or ideal zeta function associated with a nilpotent Lie algebra of class at least 5 is known. It seems likely that new theoretical insights will be needed to compute such examples. Regarding practical limitations, **Zeta** can express the generic local subalgebra zeta functions associated with  $\text{Fil}_4$  in terms of a sum of bivariate rational functions (thus, in particular, proving uniformity in the sense of Sect. 3). However, due to the number and complexity of these rational functions, the author has so far been unable to calculate their sum as a (reduced) fraction of polynomials. The author feels cautiously optimistic that further developments of computational techniques will eventually overcome such obstacles.

## 8 Conjectures

### 8.1 Local and Topological Zeta Functions at Zero

Every non-trivial local subobject zeta function known to the author has a pole at zero. No explanation of this phenomenon seems to have been provided. Under nilpotency assumptions, much more seems to be true.

*Conjecture 8.1 ([38, Conj. IV])* Let  $\mathbf{A}$  be a nilpotent  $\mathfrak{o}$ -algebra (associative or Lie, say). Let the underlying  $\mathfrak{o}$ -module of  $\mathbf{A}$  be free of rank  $d$ . Then for all  $v \in \mathcal{V}_k$ ,

$$\zeta_{\mathbf{A}_v}^{\leq}(s) \cdot (1 - q_v^{-s}) \cdots (1 - q_v^{d-1-s}) \Big|_{s=0} = 1.$$

Conjecture 8.1 was first observed by the author in a “topological form” which asserts that  $\zeta_{\mathbf{A}, \text{top}}^{\leq}(s)$  has a simple pole at zero with residue  $(-1)^{d-1}/(d-1)!$ . (For an example, consider the formula in Theorem 7.1.) Numerous examples illustrate that Conjecture 8.1 and its topological form may or may not be satisfied for non-nilpotent examples. The author’s “semi-simplification conjecture” [42, Conj. E] disposes of nilpotency assumptions and predicts the exact behaviour of generic local submodule zeta functions  $\zeta_{\Omega \curvearrowright M_v}(s)$  in terms of the Wedderburn decomposition of the associative unital algebra generated by  $\Omega$ . (The special case  $\Omega = \{\omega\}$  of the semi-simplification conjecture follows from [42, Thm 5.1].) It remains an interesting problem to even state a generalisation of Conjecture 8.1 for possibly non-nilpotent algebras.

## 8.2 Topological Zeta Functions at Infinity

In contrast to the behaviour at zero in Sect. 8.1, the author is not aware of a useful local analogue of the following.

*Conjecture 8.2 (“Degree conjecture”; [38, Conj. I])* Let  $\mathbf{A}$  be an  $\sigma$ -algebra whose underlying  $\sigma$ -module is free of rank  $d$ . Then  $\zeta_{\mathbf{A},\text{top}}^{\leq}(s)$  has degree  $-d$  as a rational function in  $s$ .

For example, the topological zeta function in Theorem 7.1 has degree  $-5$ , as predicted by Conjecture 8.2. As explained in [38, §8.1], the degree of a topological zeta function carries valuable information about the associated local zeta functions. We note that [55, Thm A] implies that for almost all  $v \in \mathcal{V}_k$ , the degree of  $\zeta_{\mathbf{A}_v}^{\leq}(s)$  as a rational function in  $q_v^{-s}$  is  $-d$  (cf. [58, §1.3]). A refinement of Conjecture 8.2 asserts that  $s^d \zeta_{\mathbf{A},\text{top}}^{\leq}(s) \Big|_{s=\infty}$  is a *positive* rational number. Finding an interpretation (even conjectural) of this number remains an interesting open problem.

In contrast to the mysterious case of subobject zeta functions, the author found topological representation zeta functions associated with unipotent groups to always have degree 0; see [40, Cor. 4.7].

**Acknowledgements** The author thanks Christopher Voll and the anonymous referee for their comments and the Alexander von Humboldt-Foundation for support during the preparation of this article. The work described here was funded by the DFG Priority Programme “Algorithmic and Experimental Methods in Algebra, Geometry and Number Theory” (SPP 1489).

## References

1. P. Aluffi, Characteristic classes of singular varieties, in *Topics in Cohomological Studies of Algebraic Varieties* (Birkhauser, Basel, 2005), pp. 1–32
2. N. Avni, B. Klopsch, U. Onn, C. Voll, Representation zeta functions of compact  $p$ -adic analytic groups and arithmetic groups. *Duke Math. J.* **162**(1), 111–197 (2013)
3. V. Baldoni, N. Berline, J.A. De Loera, B. Dutra, M. Köppe, S. Moreinis, G. Pinto, M. Vergne, J. Wu, *A User’s Guide for LattE Integrale v1.7.3* (2015). Software package. LattE is available at <http://www.math.ucdavis.edu/~latte/>
4. A. Barvinok, *Integer Points in Polyhedra*. Zurich Lectures in Advanced Mathematics (European Mathematical Society, Zürich, 2008)
5. A. Barvinok, K. Woods, Short rational generating functions for lattice point problems. *J. Am. Math. Soc.* **16**(4), 957–979 (2003) (electronic)
6. D.N. Bernstein, A.G. Kushnirenko, A.G. Khovanskii, Newton polyhedra (Russian). *Usp. Mat. Nauk* **31**(3(189)), 201–202 (1976)
7. E. Bierstone, P.D. Milman, Canonical desingularization in characteristic zero by blowing up the maximum strata of a local invariant. *Invent. Math.* **128**(2), 207–302 (1997)
8. W. Bruns, B. Ichim, T. Römer, C. Söger, *Normaliz 3.1.2. Algorithms for Rational Cones and Affine Monoids* (2016). Available from <http://www.math.uos.de/normaliz/>
9. W. Decker, G.-M. Greuel, G. Pfister, H. Schönemann, *Singular 4-1-0—A Computer Algebra System for Polynomial Computations* (2016). Available from <http://www.singular.uni-kl.de/>
10. M. Demazure, P. Gabriel, *Groupes algébriques. Tome I: Géométrie algébrique, généralités, groupes commutatifs*. Avec un appendice it Corps de classes local par Michiel Hazewinkel (Masson & Cie/North-Holland Publishing Co., Éditeur Paris/Amsterdam, 1970)

11. J. Denef, Report on Igusa's local zeta function. Séminaire Bourbaki, vol. 1990/1991. Astérisque **201–203** (1991). Exp. No. 741, 359–386 (1992)
12. J. Denef, K. Hoornaert, Newton polyhedra and Igusa's local zeta function. *J. Number Theory* **89**(1), 31–64 (2001)
13. J. Denef, F. Loeser, Caractéristiques d'Euler-Poincaré, fonctions zêta locales et modifications analytiques. *J. Am. Math. Soc.* **5**(4), 705–720 (1992)
14. M.P.F. du Sautoy, The zeta function of  $\mathfrak{sl}_2(\mathbb{Z})$ . *Forum Math.* **12**(2), 197–221 (2000)
15. M.P.F. du Sautoy, Zeta functions of groups: the quest for order versus the flight from ennui, in *Groups St. Andrews 2001 in Oxford*, vol. I (Cambridge University Press, Cambridge, 2003), pp. 150–189
16. M.P.F. du Sautoy, F.J. Grunewald, Analytic properties of zeta functions and subgroup growth. *Ann. Math. (2)* **152**(3), 793–833 (2000)
17. M.P.F. du Sautoy, F. Loeser, Motivic zeta functions of infinite-dimensional Lie algebras. *Sel. Math. N. Ser.* **10**(2), 253–303 (2004)
18. M.P.F. du Sautoy, G. Taylor, The zeta function of  $\mathfrak{sl}_2$  and resolution of singularities. *Math. Proc. Camb. Philos. Soc.* **132**(1), 57–73 (2002)
19. M.P.F. du Sautoy, L. Woodward, *Zeta Functions of Groups and Rings*. Lecture Notes in Mathematics, vol. 1925 (Springer, Berlin, 2008)
20. D.H. Dung, C. Voll, Uniform analytic properties of representation zeta functions of finitely generated nilpotent groups. *Trans. Am. Math. Soc.* **369**(9), 6327–6349 (2017)
21. A. Evseev, Reduced zeta functions of Lie algebras. *J. Reine Angew. Math.* **633**, 197–211 (2009)
22. S. Ezzat, Representation growth of finitely generated torsion-free nilpotent groups: methods and examples, Ph.D. thesis, 2012. See <http://hdl.handle.net/10092/7235>
23. F.J. Grunewald, D. Segal, G.C. Smith, Subgroups of finite index in nilpotent groups. *Invent. Math.* **93**(1), 185–223 (1988)
24. E. Hrushovski, B. Martin, S. Rideau, R. Cluckers, Definable equivalence relations and zeta functions of groups. *J. Eur. Math. Soc.* (2017, to appear). arXiv:math/0701011
25. I. Ilani, Zeta functions related to the group  $SL_2(\mathbb{Z}_p)$ . *Isr. J. Math.* **109**, 157–172 (1999)
26. A. Jaikin-Zapirain, Zeta function of representations of compact  $p$ -adic analytic groups. *J. Am. Math. Soc.* **19**(1), 91–118 (2006) (electronic)
27. A.G. Khovanskii, Newton polyhedra, and toroidal varieties. *Funkcional. Anal. i Priložen* **11**(4), 56–64, 96 (1977)
28. A.G. Khovanskii, Newton polyhedra, and the genus of complete intersections. *Funktsional. Anal. i Prilozhen* **12**(1), 51–61 (1978)
29. B. Klopsch, Representation growth and representation zeta functions of groups. *Note Mat.* **33**(1), 107–120 (2013)
30. B. Klopsch, C. Voll, Zeta functions of three-dimensional  $p$ -adic Lie algebras. *Math. Z.* **263**(1), 195–210 (2009)
31. A.G. Kushnirenko, Polyèdres de Newton et nombres de Milnor. *Invent. Math.* **32**(1), 1–31 (1976)
32. M. Larsen, A. Lubotzky, Representation growth of linear groups. *J. Eur. Math. Soc.* **10**(2), 351–390 (2008)
33. S. Lee, C. Voll, Enumerating graded ideals in graded rings associated to free nilpotent Lie rings (2016). Preprint. arXiv:1606.04515
34. M.W. Liebeck, Probabilistic and asymptotic aspects of finite simple groups, in *Probabilistic Group Theory Combinatorics, and Computing* (Springer, London, 2013), pp. 1–34
35. A. Lubotzky, A.R. Magid, Varieties of representations of finitely generated groups. *Mem. Am. Math. Soc.* **58**(336), xi+117 (1985)
36. J. Nicaise, An introduction to  $p$ -adic and motivic zeta functions and the monodromy conjecture, in *Algebraic and Analytic Aspects of Zeta Functions and L-Functions* (World Scientific, Singapore, 2010), pp. 141–166
37. J. Pas, Uniform  $p$ -adic cell decomposition and local zeta functions. *J. Reine Angew. Math.* **399**, 137–172 (1989)

38. T. Rossmann, Computing topological zeta functions of groups, algebras, and modules, I. Proc. Lond. Math. Soc. (3) **110**(5), 1099–1134 (2015)
39. T. Rossmann, Computing topological zeta functions of groups, algebras, and modules, II. J. Algebra **444**, 567–605 (2015)
40. T. Rossmann, Topological representation zeta functions of unipotent groups. J. Algebra **448**, 210–237 (2016)
41. T. Rossmann, Computing local zeta functions of groups, algebras, and modules. Trans. Am. Math. Soc. (2017). <https://doi.org/10.1090/tran/7361>
42. T. Rossmann, Enumerating submodules invariant under an endomorphism. Math. Ann. **368**(1–2), 391–417 (2017)
43. T. Rossmann, Stability results for local zeta functions of groups algebras, and modules. Math. Proc. Camb. Philos. Soc. 1–10 (2017). <https://doi.org/10.1017/S0305004117000585>
44. T. Rossmann, Zeta, Version 0.3.2 (2017). See <http://www.math.uni-bielefeld.de/~rossmann/Zeta/>
45. J.-P. Serre, *Lectures on  $N_X(p)$* . Chapman & Hall/CRC Research Notes in Mathematics, vol. 11 (CRC Press, Boca Raton, FL, 2012)
46. A. Shalev, Applications of some zeta functions in group theory, in *Zeta Functions in Algebra and Geometry* (American Mathematical Society, Providence, RI, 2012), pp. 331–344
47. R. Snocken, Zeta functions of groups and rings, Ph.D. thesis, 2012. See <http://eprints.soton.ac.uk/id/eprint/372833>
48. L. Solomon, Zeta functions and integral representation theory. Adv. Math. **26**(3), 306–326 (1977)
49. A. Stasinski, C. Voll, Representation zeta functions of nilpotent groups and generating functions for Weyl groups of type B. Am. J. Math. **136**(2), 501–550 (2014)
50. A. Stasinski, C. Voll, Representation zeta functions of some nilpotent groups associated to prehomogeneous vector spaces. Forum Math. **29**(3), 717–734 (2017)
51. W.A. Stein et al., *Sage Mathematics Software (Version 7.4)*. The Sage Development Team (2016). Available from <http://www.sagemath.org/>
52. A.N. Varchenko, Zeta-function of monodromy and Newton’s diagram. Invent. Math. **37**(3), 253–262 (1976)
53. W. Veys, W.A. Zúñiga-Galindo, Zeta functions for analytic mappings, log-principalization of ideals, and Newton polyhedra. Trans. Am. Math. Soc. **360**(4), 2205–2227 (2008)
54. O. Villamayor, Constructiveness of Hironaka’s resolution. Ann. Sci. Éc. Norm. Supér. (4) **22**(1), 1–32 (1989)
55. C. Voll, Functional equations for zeta functions of groups and rings. Ann. Math. (2) **172**(2), 1181–1218 (2010)
56. C. Voll, A newcomer’s guide to zeta functions of groups and rings, in *Lectures on Profinite Topics in Group Theory* (Cambridge University Press, Cambridge, 2011), pp. 99–144
57. C. Voll, Zeta functions of groups and rings—recent developments, in *Groups St Andrews 2013* (University of St Andrews, St Andrews, 2015), pp. 469–492
58. C. Voll, Local functional equations for submodule zeta functions associated to nilpotent algebras of endomorphisms. Int. Math. Res. Not. (2017). <https://doi.org/10.1093/imrn/rmx186>
59. E. Witten, On quantum gauge theories in two dimensions. Commun. Math. Phys. **141**(1), 153–209 (1991)
60. L. Woodward, Zeta functions of groups: computer calculations and functional equations, Ph.D. thesis, 2005

# On Decomposition Numbers of Diagram Algebras



Armin Shalile

**Abstract** In this paper, we survey an algorithm which determines the decomposition numbers of the partition algebra, Brauer algebra and walled Brauer algebra over a field of characteristic 0. The algorithm is based on the action of a set of distinguished elements of the algebra, the so-called Jucys-Murphy elements. We also outline the proof which is remarkably uniform.

**Keywords** Diagram algebras • Cellular algebras • Decomposition matrices

**Subject Classifications** 20C30, 20G05

## 1 Introduction

Diagram algebras denotes a class of algebras which among others encompass Brauer, Temperley-Lieb and partition algebras. They often arise in the context of generalizations of Schur-Weyl duality, a correspondence which classically relates the representation theory of the symmetric and the general linear group [28]. For example Brauer algebras play the role of the symmetric group in this correspondence when the general linear group is replaced by the orthogonal or symplectic group. Hence diagram algebras are a very important tool but they are also interesting in their own right. They display a rich and interesting structure, for example they usually admit a cellular structure [10] and are even cellularly stratified [13].

In this paper, we survey the determination of decomposition numbers of three important diagram algebras over certain fields using a family of commuting elements, the so called Jucys-Murphy elements. The diagram algebras we consider are the partition algebra, the Brauer algebra and the walled Brauer algebra. Each

---

A. Shalile (✉)

Institute for Algebra and Number Theory, University of Stuttgart, Pfaffenwaldring 57,  
70569 Stuttgart, Germany

e-mail: [shalile@mathematik.uni-stuttgart.de](mailto:shalile@mathematik.uni-stuttgart.de); [a.shalile@gmail.com](mailto:a.shalile@gmail.com)

of these algebras has a quotient isomorphic to a symmetric group algebra and the fields we consider are such that this quotient is semisimple. Various approaches to this problem, mostly for the easier characteristic 0 case, exist, see [5, 20, 21]. But also first attempts at the more difficult case of positive characteristic case have been made, see [17] for the partition algebra and [14] and [15] where Schur algebras of Brauer algebras are studied thus providing further insights into the decomposition number problem. The striking feature of the approach presented here is that the description and to a large extent also the proofs are very similar for each of these algebras. This indicates that the approach might also be adapted to other diagram algebras and that there might be a more general way to prove it for a large class of diagram algebras simultaneously. Because of the description in terms of Jucys-Murphy elements, the approach is also in principle compatible with the affine setting, for example for affine cellular algebras [18].

The similarity of the representation theory of different diagram algebras has been frequently exploited, a particularly fruitful strategy being to lift results from the symmetric group. For example in the case of the Brauer algebra, ordinary character theory [27], modular character theory [29], Murphy bases [8, 25], permutation and Young modules [12] as well as a description of the center in terms of conjugacy class sums [30] have been studied. Some of this theory was even studied for various diagram algebras at the same time [13, 23, 26].

Throughout this paper, we work over a field  $F$  of characteristic 0 bearing in mind, however, that the approach works with some modifications in the case of large enough finite characteristic. Here, large enough means larger than the degree (defined below) of the underlying diagram algebra. Let  $A$  be either the partition algebra, Brauer or walled Brauer algebra. Each of these algebras is cellular by [10] and therefore comes equipped with a set of distinguished modules, called cell modules and denoted by  $\Delta(\lambda)$  for some label  $\lambda$ . The decomposition matrix records the composition factors for each such cell module. Just as the symmetric groups, the algebras in question over different degrees form a nested family of algebras and in particular algebras of higher degrees contain algebras of lower degrees. The main technique which we will employ is to exploit refined versions of restriction of cell modules to family members of smaller degrees. This yields a combinatorial object called tableau which essentially are paths in a tree with vertices labelled by cell modules and an edge between cell modules if one of the cell modules occurs as a summand in the restriction of the other cell module. The key for decomposition numbers is that there is a set of distinguished elements called Jucys-Murphy (JM) elements for each such algebra which induce colourings of the edges of the graph which determines decomposition numbers: A module  $L(\mu)$  is a composition factor of  $\Delta(\lambda)$  if and only if there are tableaux for  $\Delta(\lambda)$  and  $\Delta(\mu)$  with the same colouring at each level. Some restriction for  $\lambda$  and  $\mu$  are necessary here (see Theorem 4.7 for an exact statement) but nevertheless this methods yields all decomposition numbers.

This paper is structured as follows. After the definition of the diagram algebras in question, we collect some well-known properties in Sect. 3 and define tableaux, JM elements and the “action” of JM elements on tableaux, see Sect. 4. This will already suffice to state the main result, Theorem 4.7. Before we give a sketch of the proof of

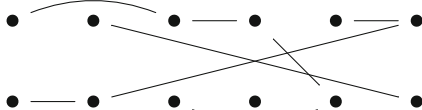
the theorem in Sect. 6, we explain a convenient way to check the conditions stated in the main theorem in Sect. 5.

## 2 Definition of the Algebras in Question

Throughout this paper, let  $F$  be a field of characteristic 0 and fix a parameter  $\delta \in F$ .

### 2.1 Definition of Partition Algebras

Partition algebras were introduced by Paul Martin [19] for the study of the Potts model in statistical mechanics and independently by Jones [16]. The partition algebra  $P_r(\delta)$  of degree  $r \in \mathbb{N}$  with parameter  $\delta$  has a basis consisting of set partitions of the set  $\{1, 1', 2, 2', \dots, r, r'\}$  which are visualized diagrammatically as partition diagrams. A partition diagram is a  $2 \times r$  array of dots, which are labelled  $1, \dots, r$  in the top row and  $1', \dots, r'$  in the bottom row. We connect dots by edges in such a way that the connected components are precisely the blocks of a partition and a minimal number of edges is used. Notice that this is not unique.



For example, the diagram represents the set partition  $\{\{1, 3, 4, 3', 5'\}, \{2, 6'\}, \{4'\}, \{5, 6, 1', 2'\}\}$ .

We can define a multiplication on partitions diagrammatically by a process called concatenation. To concatenate two partition diagrams  $a$  and  $b$ , we write  $a$  on top of  $b$  and identify adjacent rows. The concatenation  $a \circ b$  is the diagram obtained from this construction by deleting all connected components which are not connected to the top or bottom row and premultiplying the resulting diagram with a power of the parameter where the exponent of the parameter counts the number of connected components deleted. An example of this process is given in Fig. 1.

The partition algebra  $P_r(\delta)$  has a subalgebra  $P_{r-1/2}(\delta)$  which is spanned by all diagrams where  $r$  and  $r'$  belong to the same block of the partition. The algebra

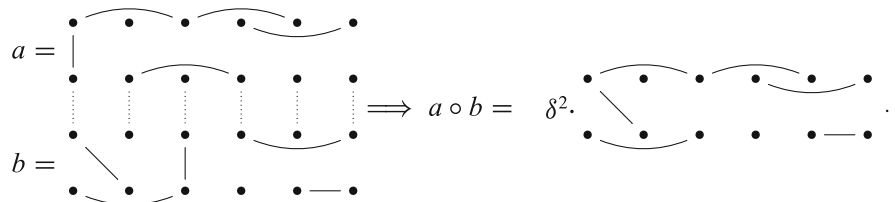


Fig. 1 An example of concatenation

$P_{r-1/2}(\delta)$  in turn contains the partition algebra  $P_{r-1}(\delta)$  as a subalgebra where we identify  $P_{r-1}(\delta)$  with the subalgebra spanned by all partitions in which  $\{r, r'\}$  is a block. Thus we get a chain of subalgebras

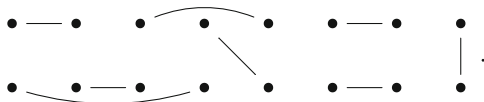
$$P_{1/2}(\delta) \subset P_1(\delta) \subset P_{3/2}(\delta) \subset P_2(\delta) \subset \dots \subset P_r(\delta).$$

The half integer degree partition algebras are in fact Morita equivalent to full integer degree partition algebras:

**Theorem 2.1 (Proposition 5 of [20])** *There is an equivalence  $T$  between the module categories of  $P_{r+1/2}(\delta)$  and  $P_r(\delta - 1)$ .*

### 2.2 Definition of Brauer Algebras

Brauer algebras were introduced by Richard Brauer in order to generalize Schur-Weyl duality to the orthogonal and symplectic groups [2]. The Brauer algebra is defined in a similar fashion as the partition algebra, except that we only allow a subset of the partition diagrams, namely Brauer diagrams. By a Brauer diagram on  $2r$  ( $r \in \mathbb{N}$ ) dots we mean a  $2 \times r$ -array of dots such that each dot is connected by an edge to exactly one other dot distinct from itself. An example of a diagram on 16 dots is



**Definition 2.2** The Brauer algebra is the subalgebra of the partition algebra spanned by Brauer diagrams.

### 2.3 Definition of Walled Brauer Algebras

The walled Brauer algebra was introduced in [1] for the study of Schur-Weyl duality for the so called mixed tensor space, see [1] for details. It is a subalgebra of the Brauer algebra spanned by walled Brauer diagrams. Let  $r, s \in \mathbb{N}$ . A walled Brauer diagram on  $2(r + s)$  dots is a Brauer diagram on  $2(r + s)$  dots with a vertical wall between columns  $r$  and  $r + 1$  such that

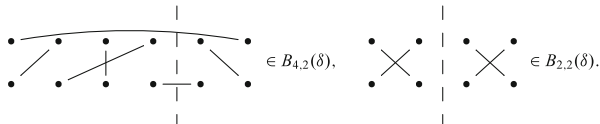
1. No through arc may intersect the wall.
2. All horizontal arcs must intersect the wall.

In this definition, a horizontal arc is an edge connecting dots in the same row and a through arc is an edge which connects dots in the top and bottom row. The



subalgebra of  $B_r(\delta)$  spanned by all walled Brauer diagrams on  $2(r + s)$  dots is called the walled Brauer algebra and denoted  $B_{r,s}(\delta)$ .

*Example 2.3* The following are examples of walled Brauer diagrams where the wall is indicated by a dotted line:



### 3 Simple Modules and Cellularity

All algebras treated here are cellular algebras in the sense of Graham and Lehrer [10].

**Theorem 3.1** ([6, 10, 33]) *The Brauer, walled Brauer and partition algebras are cellular for all choices of  $r, s$  and  $\delta$ .*

The concept of a cellular algebra is very useful as it reduces many questions about the algebra to problems in linear algebra (which does not mean that they become trivial or easy). In particular, cellular algebras come equipped with a natural class of modules, called cell modules. The simple modules occur as simple heads of a subset of the cell modules and the radical of these cell modules can be explicitly described by a bilinear form. In this way, the dimension of simple modules can be obtained by computing the ranks of Gram matrices associated with the bilinear form. There is, however, no general method to compute decomposition numbers for cellular algebras.

In order to describe the cellular structure in more detail, we first need to recall some definitions: A partition of a natural number  $n$  is a tuple  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$  such that  $\lambda_i \in \mathbb{N}$ ,  $\lambda_i \geq \lambda_{i+1}$  for  $i = 1, \dots, k - 1$  and  $\sum_{i=1}^k \lambda_i = n$ . Each  $\lambda_i$  is called a part of  $\lambda$  and the natural number  $k$  is called the number of parts of  $\lambda$ . We occasionally view the partition  $\lambda$  as  $(\lambda_1, \lambda_2, \dots, \lambda_k, 0, 0, \dots)$ . We consider  $\lambda = \emptyset$  the unique partition of 0.  $\lambda$  is called  $p$ -singular if there is  $j \leq k - p + 1$  such that  $\lambda_j = \lambda_{j+1} = \lambda_{j+p-1}$  and  $p$ -regular otherwise. For  $p = 0$  all partitions are considered  $p$ -regular.

**Theorem 3.2** ([6, 10, 33]) *The labelling set  $\Lambda$  of the cell modules is given as follows:*

1. In case of the partition algebra  $P_r(\delta)$ ,  $\Lambda$  is the set of partitions of non-negative integers  $l \leq r$ .
2. In case of the half partition algebra  $P_{r+1/2}(\delta)$ ,  $\Lambda$  is also the set of partitions of non-negative integers  $l \leq r$ .
3. In case of the Brauer algebra  $B_r(\delta)$ ,  $\Lambda$  is given by the partitions of  $r - 2t$  where  $t \in \mathbb{N}_0$  with  $r - 2t \geq 0$ .

4. In case of the walled Brauer algebra  $B_{r,s}(\delta)$ ,  $\Lambda$  is given by the set of all ordered tuples  $(\lambda, \lambda')$  with  $\lambda$  a partition of  $r - l \leq r$  and  $\lambda'$  a partition of  $s - l \leq s$  for some non-negative integer  $l$ .

The following is well known and implies among other things that the algebras are even quasi-hereditary in the cases we consider:

**Theorem 3.3** ([6, 10, 33]) *When  $F$  is a field of characteristic 0, then, for each of the algebras of the previous theorem,  $\Lambda$  coincides with the set of simple modules except (a) in the case of the Brauer algebra  $B_r(0)$  with  $r$  even and (b) in the case of the partition algebra with  $\delta = 0$ , in both of which cases the empty partition has to be excluded.*

We will abuse notation slightly and denote both the cell modules and their simple heads by  $\Delta_r(\lambda)$  and  $L_r(\lambda)$ , respectively, for  $\lambda \in \Lambda$ , keeping in mind that in Case (a) and (b) of the previous theorem, there is no simple module labelled by the empty partition. Furthermore, we will omit indices and simply write  $\Delta(\lambda)$  and  $L(\lambda)$ , respectively, when the parameters are clear from the context. In case of the walled Brauer algebra, we write  $\Delta(\lambda, \lambda')$  for the cell and  $L(\lambda, \lambda')$  for the simple module labelled by  $(\lambda, \lambda') \in \Lambda$ . Sometime we will abuse notation and omit the  $\lambda'$  to unify statements.

One aim of this paper is to understand the composition factors of the cell modules  $\Delta(\lambda)$ . We denote by  $d_{\lambda\mu} = [\Delta(\lambda) : L(\mu)]$  the composition multiplicity of the simple module  $L(\mu)$  in the cell module  $\Delta(\lambda)$ . The decomposition matrix  $D$  is the matrix with rows and columns labelled by  $\Lambda$  and entry  $(\lambda, \mu)$  equal to  $d_{\lambda\mu}$ . The following theorem follows from [13]:

**Theorem 3.4 (Proposition 6.1 and Corollary 6.2 in [13])** *Let  $D_A$  denote the decomposition matrix of the cellular algebra  $A$ . There is an ordering of cell and simple modules such that*

1.  $D_{P_d(\delta)} = \begin{pmatrix} D_{FS_r} & 0 \\ * & D_{P_{d-1}(\delta)} \end{pmatrix}$  where  $d \in \{r, r + 1/2\}$ .
2.  $D_{B_r(\delta)} = \begin{pmatrix} D_{FS_r} & 0 \\ * & D_{B_{r-2}(\delta)} \end{pmatrix}$ .
3.  $D_{B_{r,s}(\delta)} = \begin{pmatrix} D_{F(S_r \times S_s)} & 0 \\ * & D_{B_{r-1,s-1}(\delta)} \end{pmatrix}$ .

**Remark 3.5** It is well known that each of the algebras we consider contain an ideal such that the algebra modulo this ideal is isomorphic to the symmetric group algebra  $FS_r$  or  $F(S_r \times S_s)$ , respectively. The simple module labelled by a partition  $\lambda$  of  $r$  (and  $\lambda'$  a partition of  $s$  in the case of the walled Brauer algebra) are then inflated simple modules (also called Specht modules  $S(\lambda)$ ) for the symmetric group algebra. We sometimes refer to these partitions (or pairs of partitions) as maximal partitions.

One important corollary of this theorem is that in order to understand all decomposition numbers, it is enough to understand the composition factors corresponding to partitions of maximal size (in the sense of the previous remark).

We will usually identify a partition with its Young diagram. Thus, we will speak of addable and removable boxes of a partition: These are boxes of the Young diagram of  $\lambda$  such that adding or removing this box will still yield a partition. If the box is  $\epsilon$ , then we denote the newly created partition by  $\lambda + \epsilon$  and  $\lambda - \epsilon$ , respectively. We denote by  $\text{add } \lambda$  and  $\text{rem } \lambda$  the set of addable and removable boxes of  $\lambda$ . The content of a box  $\epsilon$  which is in row  $i$  and column  $j$  of the Young diagram is defined to be  $c(\epsilon) = j - i$ .

Adding and removing boxes of a Young diagram are intimately connected with induction and restriction of cell modules, which are defined as follows:

**Definition 3.6** Let  $A_d$  be equal to  $P_r(\delta)$ ,  $P_{r+1/2}(\delta)$ ,  $B_r(\delta)$  or  $B_{r,s}(\delta)$  and  $A_{d-1}$  be equal to  $P_{r-1/2}(\delta)$ ,  $P_r(\delta)$ ,  $B_{r-1}(\delta)$  or  $B_{r-1,s}$  (or  $B_{r,s-1}$  if  $r = 0$ ), respectively. Then  $\text{res}_d : A_d\text{-mod} \rightarrow A_{d-1}\text{-mod}$  denotes the restriction functor of  $A_d$ -modules to  $A_{d-1}$ -modules and  $\text{ind}_d : A_d\text{-mod} \rightarrow A_{d+1}\text{-mod}$  denotes the induction functor of  $A_d$ -modules to  $A_{d+1}$ -modules, that is, given an  $A_d$ -module  $M$ , we set  $\text{ind}_d M = A_{d+1} \otimes_{A_d} M$ .

The induction and restriction of cell modules is quite well understood:

**Theorem 3.7 ([4, 7, 31])** *Suppose  $F$  is a field of characteristic 0. Then the following sequences are exact*

1. *In the case of the partition algebra  $P_r(\delta)$ ,*

$$0 \longrightarrow \bigoplus_{\epsilon \in \text{rem}(\lambda)} \Delta_{r-1/2}(\lambda - \epsilon) \longrightarrow \text{res} \Delta_r(\lambda) \longrightarrow \Delta_{r-1/2}(\lambda) \longrightarrow 0.$$

2. *In the case of the half partition algebra  $P_{r+1/2}(\delta)$ ,*

$$0 \longrightarrow \Delta_r(\lambda) \longrightarrow \text{res} \Delta_{r+1/2}(\lambda) \longrightarrow \bigoplus_{\epsilon \in \text{add}(\lambda)} \Delta_r(\lambda + \epsilon) \longrightarrow 0.$$

3. *In the case of the Brauer algebra,*

$$0 \longrightarrow \bigoplus_{\epsilon \in \text{rem}(\lambda)} \Delta(\lambda - \epsilon) \longrightarrow \text{res} \Delta(\lambda) \longrightarrow \bigoplus_{\epsilon' \in \text{add}(\lambda)} \Delta(\lambda + \epsilon') \longrightarrow 0.$$

4. *In the case of the walled Brauer algebra  $B_{r,s}(\delta)$  with  $r > 0$ ,*

$$0 \longrightarrow \bigoplus_{\epsilon \in \text{rem}(\lambda)} \Delta(\lambda - \epsilon, \lambda') \longrightarrow \text{res} \Delta(\lambda, \lambda') \longrightarrow \bigoplus_{\epsilon' \in \text{add}(\lambda')} \Delta(\lambda, \lambda' + \epsilon') \longrightarrow 0.$$

5. *In the case of the walled Brauer algebra  $B_{0,s}(\delta)$ ,*

$$\text{res} \Delta_{0,s}(\emptyset, \lambda') = \bigoplus_{\epsilon \in \text{rem}(\lambda')} \Delta_{0,s-1}(\emptyset, \lambda' - \epsilon).$$

Here we adopt the convention that  $\Delta(\theta) = 0$  whenever  $\theta \notin \Lambda$ . In particular, the decomposition is multiplicity free in each case.

*Remark 3.8* Similar statements for induction are also known and follow essentially from adjointness of induction and restriction.

### 4 Jucys-Murphy Elements and Gelfand-Zetlin Basis

The multiplicity freeness of restriction in the previous section guarantees the existence of a natural basis for cell modules, sometimes called Gelfand-Zetlin basis, which we will now define. In the semisimple case, a restriction of a cell module  $\Delta(\lambda)$  to the family member in the next lowest degree yields a direct sum decomposition of cell modules  $\bigoplus \Delta(\lambda_i)$  in which no cell module occurs twice. Repeating this with the cell modules  $\Delta(\lambda_i)$  in the decomposition, we receive yet another multiplicity free decomposition. This defines a tree sometimes also called Bratelli diagram which has levels corresponding to degrees of the diagram algebras and vertices at level  $k$  labelled by simple modules of the family member in degree  $k$  and an edge from a vertex  $\lambda_i$  in level  $k$  to a vertex  $\lambda_j$  at the next lowest level precisely when  $\Delta(\lambda_j)$  occurs as a direct summand in the restriction of  $\Delta(\lambda_i)$ . The paths from top to bottom of such a graph are called tableaux. This iterated restriction ultimately yields a unique direct sum decomposition into 1-dimensional subspaces because if we decrease the degree within the family, we eventually end up with a field. The paths in the Bratelli diagram therefore label this basis in a natural way which is unique up to scalar multiples. The construction of the basis heavily relies on the semisimplicity of the algebra in question. Nevertheless we will use the vectors  $\{v_t\}$  of this basis and especially the tableaux  $t$  labelling it in the non-semisimple case as well. This can be done by exploiting the fact that the algebras treated here are generically semisimple and we can pass to the field  $F$  we are working with by specializing the parameter. The problem is that, in general, there will be denominators which become zero when specializing the parameter which can be avoided by multiplying  $v_t$  with the least common multiple of the denominators occurring. Notice that the Gelfand-Zetlin basis is also only defined up to scalar multiples. However, it should be noted that the vectors  $\{v_t\}$  potentially become linearly dependent after specialization.

With our knowledge of restriction, we can also define the notion of a tableau directly as follows:

**Definition 4.1** A tableau  $t$  of a partition  $\lambda$  (or  $(\lambda, \lambda')$  in case of  $B_{r,s}(\delta)$ ) is a sequence of partitions (or pairs of partitions)

1.  $(t^{1/2} = \emptyset, t^1, t^{3/2}, \dots, t^d = \lambda)$  for  $P_d(\delta)$  with  $d \in \{r, r + 1/2\}$ ,
2.  $(t^0 = \emptyset, t^1, \dots, t^r = \lambda)$  for  $B_r(\delta)$ ,
3.  $(t^0 = (\emptyset, \emptyset), t^1 = (t^1_L, t^1_R), t^2 = (t^2_L, t^2_R), \dots, t^{r+s} = (\lambda, \lambda'))$  for  $B_{r,s}(\delta)$ ,

such that for all  $k \in \mathbb{N}$  (or  $k \in \frac{1}{2}\mathbb{N}$  in case of  $P_r(\delta)$  and  $P_{r+1/2}(\delta)$ )

1.  $t^k = \begin{cases} t^{k-1/2} \text{ or } t^{k-1/2} + \epsilon & \text{if } k \in \mathbb{N} \\ t^{k-1/2} \text{ or } t^{k-1/2} - \epsilon & \text{if } k \notin \mathbb{N} \end{cases}$ ,
2.  $t^k = t^{k-1} + \epsilon$  or  $t^k = t^{k-1} - \epsilon$ ,
3.  $t^k = \begin{cases} (t_L^{k-1}, t_R^{k-1} + \epsilon) & \text{if } k \leq s \\ (t_L^{k-1} + \epsilon, t_R^{k-1}) \text{ or } (t_L^{k-1}, t_R^{k-1} - \epsilon) & \text{if } k > s \end{cases}$ ,

respectively. Here,  $\epsilon$  is an addable or removable box, respectively, of the appropriate partition. We will sometimes refer to the transition from  $t^{k-1}$  to  $t^k$  by adding or removing a box at the  $k$ th step of  $t$ .

The set of tableaux of a given partition  $\lambda \in \Lambda$  is denoted by  $\text{Tab}(\lambda)$  (the degree will always be clear from the context).

*Example 4.2* Examples of tableaux are:

1. For the partition algebra of degree  $d = 4 + 1/2$ ,

$$u = \left( \emptyset, \square, \square, \square, \square, \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}, \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array}, \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array}, \begin{array}{|c|c|c|c|c|} \hline \square & \square & \square & \square & \square \\ \hline \square & \square & \square & \square & \square \\ \hline \end{array} \right),$$

$$t = \left( \emptyset, \square, \square, \square, \square, \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}, \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}, \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \right).$$

2. For the Brauer algebra with  $r = 4$ ,

$$u = \left( \emptyset, \square, \square, \square, \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}, \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array} \right),$$

$$t = \left( \emptyset, \square, \square, \square, \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \right).$$

3. For the walled Brauer algebra with  $(r, s) = (3, 2)$ ,

$$u = \left( (\emptyset, \emptyset), (\emptyset, \square), \left( \emptyset, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \right), \left( \square, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \right), \left( \square, \square, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \right), \left( \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \right), \right),$$

$$t = \left( (\emptyset, \emptyset), (\emptyset, \square), \left( \emptyset, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \right), \left( \square, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \right), \left( \square, \square, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \right), \left( \square, \square, \square \right), \right).$$

The missing ingredient for the determination of decomposition numbers are a distinguished family of elements of the algebras we are considering, the so called Jucys-Murphy elements or sometimes just called Murphy elements. This is a family of elements  $\{L_k\}_{k \in I}$  where the indexing set  $I$  is  $\mathbb{N}$  in the Brauer and walled Brauer algebra case and  $1/2\mathbb{N}$  in the partition algebra case. Their exact definition will

not be important and we refer the reader to the papers [11, Section 3], [24] and [4, Equations (2.19) and (2.20)]. However, they have some properties which are important for us:

**Theorem 4.3** ([4, 11, 24])

1. The JM elements  $\{L_k\}_{k \in I}$  commute with each other.
2. The JM element  $L_k$  commutes with  $P_{k-1/2}(\delta)$ ,  $B_{k-1}(\delta)$  and  $B_{0,k-1}(\delta)$  (if  $k < s$ ) or  $B_{k-s,s}(\delta)$  (if  $k \geq s$ ).
3. The elements of the Gelfand-Zetlin basis are eigenvectors for all  $L_k$ .

The last property will precisely give rise to the colouring of the tableaux which was mentioned in the introduction: Each tableau  $t$  can be coloured at step  $k$  by the eigenvalue  $L_k(t)$  by which the JM element  $L_k$  acts on the Gelfand-Zetlin basis element labelled by this tableau. The eigenvalues  $L_k(t)$  have been determined:

**Theorem 4.4** ([4, 11, 24])

1. For the partition algebra, we have

$$L_k(t) = \begin{cases} c(\epsilon) & \text{if } t^k = t^{k-1/2} + \epsilon, k \in \mathbb{N} \\ \delta - |t^k| & \text{if } t^k = t^{k-1/2}, k \in \mathbb{N} \\ \delta - c(\epsilon) & \text{if } t^k = t^{k-1/2} - \epsilon, k \notin \mathbb{N} \\ |t^k| & \text{if } t^k = t^{k-1/2}, k \notin \mathbb{N} \end{cases}.$$

2. For the Brauer algebra, we have

$$L_k(t) = \begin{cases} c(\epsilon) & \text{if } t^k = t^{k-1} + \epsilon \\ 1 - \delta - c(\epsilon) & \text{if } t^k = t^{k-1} - \epsilon \end{cases}.$$

3. For the walled Brauer algebra, we have

$$L_k(t) = \begin{cases} c(\epsilon) & \text{if a box was added at the } k\text{th step of } t \\ -\delta - c(\epsilon) & \text{if a box was removed at the } k\text{th step of } t \end{cases}.$$

**Definition 4.5** Let  $\lambda \in \Lambda$ . The weight  $\text{wt}(t)$  of a  $\lambda$ -tableau  $t$  is defined to be the tuple

$(L_{1/2}(t), L_1(t), L_{3/2}(t), \dots, L_d(t))$  (for  $P_d(\delta)$  with  $d = r$  or  $d = r + 1/2$ ) or  $(L_0(t), L_1(t), \dots, L_d(t))$  (in the Brauer/walled Brauer case with  $d = r$  or  $d = r + s$ ). We say that  $\lambda, \mu \in \Lambda$  have a common JM weight if there is a  $\mu$ -tableau  $u$  and a  $\lambda$ -tableau  $t$  with  $\text{wt}(u) = \text{wt}(t)$ .

*Example 4.6* The following examples show how to represent a tableau together with its weight by a coloured graph:

1. For the partition algebra:

$$\begin{aligned}
 u : \emptyset \xrightarrow{0} \square \xrightarrow{1} \square \xrightarrow{-1} \square \square \xrightarrow{-2} \square \square \xrightarrow{-1} \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \xrightarrow{-3} \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \xrightarrow{-2} \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array} \xrightarrow{-4} \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array}, \\
 t : \emptyset \xrightarrow{0} \square \xrightarrow{1} \square \xrightarrow{-1} \square \square \xrightarrow{-2} \square \square \xrightarrow{-1} \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \xrightarrow{-3} \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \xrightarrow{\delta-3} \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \xrightarrow{\delta-1} \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array}.
 \end{aligned}$$

If we consider the case  $\delta = 5$ , then the tableau  $u$  and  $t$  have the same weight. Thus,  $(3, 1)$  and  $(1, 1)$  have a common JM weight. Furthermore, one can prove that  $[\Delta(1, 1) : L(3, 1)] \neq 0$ .

2. For the Brauer algebra:

$$\begin{aligned}
 u : \left( \emptyset \xrightarrow{0} \square \xrightarrow{1} \square \square \xrightarrow{-1} \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \xrightarrow{-2} \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array} \right), \\
 t : \left( \emptyset \xrightarrow{0} \square \xrightarrow{1} \square \square \xrightarrow{-1} \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \xrightarrow{1-\delta+1} \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \right).
 \end{aligned}$$

From this we can read off that  $\text{wt}(t) = (1, -1, 2)$  and  $\text{wt}(u) = (1, -1, 2 - \delta)$ . Thus, if  $\delta = 0$ , then  $(3, 1)$  and  $(1, 1)$  again have a common JM weight. At the same time, it is known that  $[\Delta(1, 1) : L(3, 1)] \neq 0$  with the chosen value of  $\delta$ .

3. For the walled Brauer algebra:

$$\begin{aligned}
 u : (\emptyset, \emptyset) \xrightarrow{0} (\emptyset, \square) \xrightarrow{-1} \left( \emptyset, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \right) \xrightarrow{0} \left( \square, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \right) \xrightarrow{-1} \left( \square \square, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \right) \xrightarrow{-1} \left( \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \right), \\
 t : (\emptyset, \emptyset) \xrightarrow{0} (\emptyset, \square) \xrightarrow{-1} \left( \emptyset, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \right) \xrightarrow{0} \left( \square, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \right) \xrightarrow{-1} \left( \square \square, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \right) \xrightarrow{-\delta+1} (\square \square, \square).
 \end{aligned}$$

Again for  $\delta = 2$  the tableau  $u$  and  $t$  have the same weight. Thus,  $((2, 1), (1, 1))$  and  $((2), (1))$  have a common JM weight and at the same time  $[\Delta((2), (1)) : L((2, 1), (1, 1))] \neq 0$ .

The fact that having a common JM weight is related to a non-zero decomposition number is not a coincidence and this is precisely the main theorem:

**Theorem 4.7** *If  $\mu$  is a partition of  $r$ ,  $\mu'$  is a partition of  $s$  and  $\lambda \in \Lambda$  (or  $(\lambda, \lambda') \in \Lambda$ ), then  $L(\mu)$  (or  $L(\mu, \mu')$ ) is a composition factor of  $\Delta(\lambda)$  (or  $\Delta(\lambda, \lambda')$ ) if and only if there is a  $\mu$ -tableau (or  $(\mu, \mu')$ -tableau)  $u$  and a  $\lambda$ -tableau (or  $(\lambda, \lambda')$ -tableau)  $t$  with  $\text{wt}(u) = \text{wt}(t)$  and, in the case of the Brauer algebra only, the tableau  $u$  and  $t$  have to be strongly balanced (see Section 5 of [32] for the definition of strongly balanced).*

*Remark 4.8* As we remarked earlier, the statement above actually suffices to determine the entire decomposition matrix by Theorem 3.4 and the fact that decomposition numbers are always zero or one in the cases we consider, see Corollary 4.10.

The fact that the previous statement can be formulated in such a unified way is probably not a coincidence. We will see later that to a large extent (modulo some combinatorial input), even the proof of the statement is quite uniform. Notice that one quite easily obtains one direction of our main theorem which is proved in Corollary 2.2 in [32] in a more general context.

**Proposition 4.9** *Let  $\lambda, \mu \in \Lambda$  and suppose  $L(\mu)$  is a composition factor of  $\Delta(\lambda)$ . Then  $\lambda$  and  $\mu$  have a common JM weight. If  $\mu$  is a partition of  $r$  (and  $\mu'$  a partition of  $s$  in the walled Brauer case), then for every  $\mu$ -tableau  $u$  there is a  $\lambda$ -tableau  $t$  with  $\text{wt}(u) = \text{wt}(t)$ .*

For the complete determination of decomposition numbers, we also have to show that decomposition multiplicities must always be either 0 or 1:

**Corollary 4.10** *Let  $\lambda, \mu \in \Lambda$  with  $\mu$  a partition of  $r$  (and  $\mu'$  a partition of  $s$ ). Then:*

- (a) *If  $\lambda$  and  $\mu$  have a common JM weight, then for every  $\mu$ -tableau  $u$  there is precisely one  $\lambda$ -tableau  $t$  with  $\text{wt}(u) = \text{wt}(t)$ .*
- (b)  $[\Delta(\lambda) : L(\mu)] \in \{0, 1\}$ .

*Proof* The first part was proved for the partition algebra in Corollary 2.12 of [31] and for the Brauer algebra in Lemma 3.2 of [32]. The proof for the walled Brauer algebra is almost identical. We need to simply replace  $\lambda$  by  $(\lambda, \lambda')$ ,  $\mu$  by  $(\mu, \mu')$  and  $1 - \delta$  by  $-\delta$  in the proof. The proof only uses arrow diagrams (which will be introduced in the next section) and the action of adding/removing boxes with the same induced JM eigenvalue is the same for the two algebras, as we will see in Fig. 2. The only difference is that there is no left corner for the walled Brauer algebra since the arrow diagram is doubly infinite in this case. This further restricts the number of cases we have to consider in the proof making it actually shorter.

The second part is immediately implied by the first since if  $[\Delta(\lambda) : L(\mu)] > 1$ , then  $\Delta(\lambda)$  would contain at least two copies of the simple module  $L(\mu) = \Delta(\mu)$ . Therefore, the generalized eigenspace corresponding to  $\text{wt}(t)$  contains at least two linearly independent elements implying that for every  $\mu$ -tableau  $u$ , there would be at least two different  $\lambda$ -tableau of the same weight, see also the proof of Corollary 3.5 in [32].  $\square$

## 5 Checking for a Common JM Weight

Theorem 4.7 gives quite a nice description of decomposition numbers, however, checking the conditions of the theorem is rather computationally involved. It requires us in principle to compute all possible tableaux, all colourings and then



to look for matching weights. It turns out that there is a much more efficient way to do so. It is based on developing a diagrammatic way to represent the main operations involved in the theorem, namely adding and removing boxes to partitions and pairs of partitions and at the same time keeping control of the action of the JM elements.

In order to get an idea how to do this, consider the partition  $(6, 5, 3, 1)$ . Let us draw the Young diagram and write the content into every box:

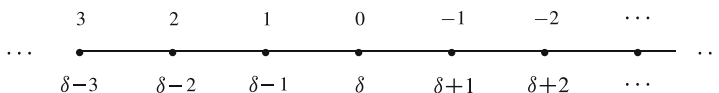
0	1	2	3	4	5
-	1	0	1	2	3
-	2	-	1	0	
-	3				

If we are just given the set of contents of the last boxes in each row (in our case  $\{5, 3, 0, -3\}$ ), then we can recover the partition from this data. Moreover, removing for example the box filled with a 5 amounts to changing the content 5 in our set by a 4 so that we get  $\{4, 3, 0, -3\}$ . Similarly, adding a box next to the box labelled 5 results in the set  $\{6, 3, 0, -3\}$ . So adding and removing boxes can be implemented in this simple model. Notice that there can never be two boxes in the last column of a row with the same content and this constraint coincides precisely with addable or removable boxes. Thus, a model of our partition could be to have beads on the usual number line and adding boxes amounts to moving beads up and removing boxes amount to moving beads down. A bead is then addable, if there is a free space to the right of it.

To obtain a combinatorial model of partitions adapted to each of our algebras, we now need to consider when moving two beads amounts to the same JM eigenvalue. For example in the Brauer case, adding a box of content  $i$  amounts to the same JM eigenvalue as removing a box of content  $1 - \delta - i$ . To keep track of such moves it is convenient to fold the number line over in such a way that adding a box of content  $i$  and removing a box of content  $1 - \delta - i$  both amount to moving a bead from the same column in the same direction, see examples below. This naturally leads to the arrow diagrams which were first defined by Brundan and Stroppel in [3] and are defined as follows:

**Definition 5.1** Given a partition  $\lambda = (\lambda_1, \lambda_2, \dots) \in \Lambda$ , (where we include all zero entries) we define its diagram or arrow diagram  $d(\lambda)$  as follows:

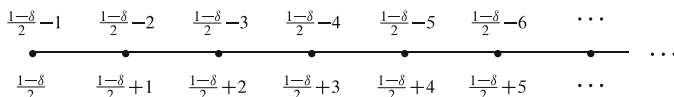
1. For the partition algebra, the arrow diagram consists of a doubly infinite line with positions above and below the line labelled as follows:



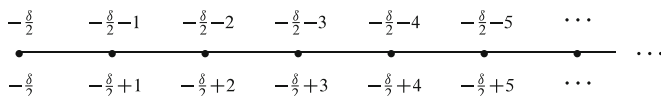
Thus, opposite labels always sum to  $\delta$ . For each part  $\lambda_i$  of  $\lambda$ , we draw a  $\nabla$  above the line at the label  $\lambda_i - i$  (that is, at the content of the last box of each row). Furthermore, we draw a single  $\wedge$  below the line at the label  $|\lambda|$ .

- In the Brauer algebra case, we have a half infinite line with positions labelled above and below the line as follows:

Case 1:  $\delta$  odd.

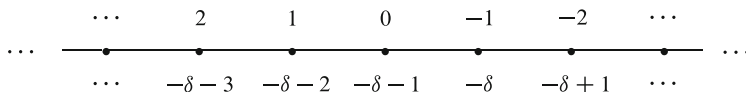


Case 2:  $\delta$  even.



For each part  $\lambda_i$  of  $\lambda$  we draw at the position labelled  $\lambda_i - i$  a  $\nabla$  if that position is above the line and a  $\wedge$  if the position is below the line. If  $\delta$  is even and  $\lambda_i - i = -\frac{\delta}{2}$ , then we only draw a  $\nabla$  at the corresponding position above the line.

- For the walled Brauer algebra, we draw a doubly infinite line with positions labelled above and below the line as follows:



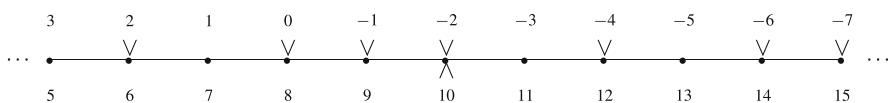
If  $\lambda = (\lambda_1, \lambda_2, \dots)$  with  $\lambda_i \geq 0$ , then we draw for each part  $\lambda_i$  of  $\lambda$  a  $\nabla$  at the position above the line labelled  $\lambda_i - i$ . Similarly, if  $\lambda' = (\lambda'_1, \lambda'_2, \dots)$  with  $\lambda'_i \geq 0$ , then we draw for each  $\lambda'_i$  a  $\wedge$  below the line at position  $\lambda'_i - i$ .

*Remark 5.2* Notice that  $\lambda$  and  $d(\lambda)$  determine each other and we will often identify them. Furthermore, adding a box to  $\lambda$  in some row corresponds to moving the arrow  $\nabla$  corresponding to that row up by 1 in the arrow diagram, that is to a position with a higher label. Also, a box is addable if and only if the next higher position is empty in the arrow diagram. Similar statements hold for removable boxes. We will see in Fig. 2 that arrow diagrams are additionally particularly well behaved with respect to the action of the JM elements.

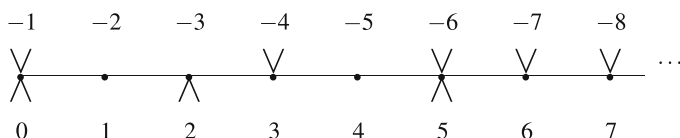
*Example 5.3* The definitions are best understood by examples.

- In the partition algebra case, we consider the partition  $\lambda = (3, 2, 2, 2, 1, 0, 0, \dots)$  with  $\delta = 8$ . The contents of the last boxes in each row are  $(2, 0, -1, -2, -4,$

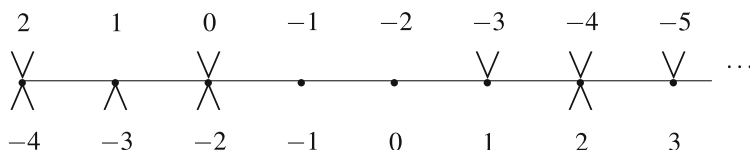
$-6, -7, \dots$ ) and  $|\lambda| = 10$ . Therefore, we get the following arrow diagram  $d(\lambda)$ :



2. In the Brauer case with  $\lambda = (6, 4, 3, 3, 1, 0, \dots)$  and  $\delta = 1$ , we get the following diagram:

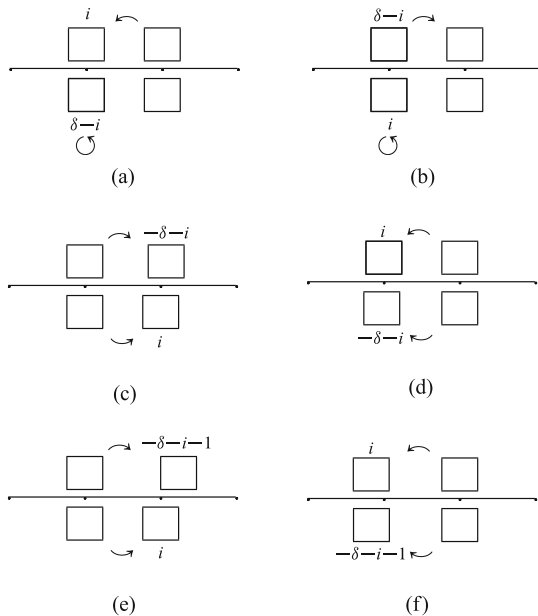


3. In the walled Brauer case, with  $(\lambda, \lambda') = ((3, 1, 0, \dots), (3, 0, \dots))$  and  $\delta = 1$ , we get



We will frequently use shorthand notation for arrow diagrams. For example,  $\lambda$  as in Example 5.3(1) will be represented as  $\dots \circ \overset{2}{\vee} \circ \vee \vee \times \circ \vee \circ \vee \dots$  where  $\vee$  stands for a column which only contains a  $\vee$ , and  $\wedge/\times/\circ$  represent columns containing only a  $\wedge$ , both a  $\vee$  and a  $\wedge$ , or no arrows at all, respectively. The label at the top/bottom of a symbol is precisely the label above/below the line in that column and specifying one will determine all other labels. Sometimes we represent  $\vee/\wedge/\times/\circ$  in extended notation by  $\begin{matrix} \vee/\circ & \vee/\circ \\ \circ/\wedge & \circ/\wedge \end{matrix}$ .

We will now connect the study of arrow diagrams with tableaux and the action of JM elements. Our aim is to relate the property of having a common JM weight to the study of composition factors of the cell modules  $\Delta(\lambda)$ . It turns out that having a common JM weight is quite restrictive. If we view a tableau as a construction plan for the corresponding partition, then the JM weight only leaves few possibilities for this construction. If the  $k$ th entry of the JM weight is  $i$ , that is, the JM element  $L_k$  acts by  $i$  on the Gelfand-Zetlin basis vector corresponding to the given tableau, then there are only at most two possibilities which are depicted in Fig. 2. Here the boxes stand for potential positions of  $\wedge$  and  $\vee$  and the arrows indicate which way the arrows may be moved. For example in the partition algebra, if  $k$  is not a natural number, then we may either add a box of content  $i$  at the  $k$ th step or leave the partition unchanged if the size of the partition in question is  $\delta - i$ . If  $k$  is a natural number, then we may either remove a box of content  $\delta - i$  or leave the partition unchanged if its size is  $i$ . The Brauer and walled Brauer algebra cases are also depicted, where we have to distinguish the cases when the label  $i$  is above or below the line.



**Fig. 2** Possible ways to move arrows with JM eigenvalue  $i$ . (a)  $P_d(\delta)$ , half to full degree. (b)  $P_d(\delta)$ , full to half degree. (c)  $B_r(\delta)$ , Case I. (d)  $B_r(\delta)$ , Case II. (e)  $B_{r,s}(\delta)$ , Case I. (f)  $B_{r,s}(\delta)$ , Case II

Using these rules we may characterize partitions with a common JM weight. The following proposition holds for all algebras treated above except the half integer partition algebras (where one first has to apply the equivalence of Theorem 2.1, see Proposition 2.9 of [31] for details):

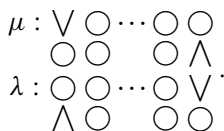
**Proposition 5.4** *Suppose  $\lambda, \theta \in \Lambda$  have a common JM weight. Then for all  $k$  the  $k$ th columns of  $d(\lambda)$  and  $d(\theta)$  contain the same number of arrows (that is, counting  $\wedge$  and  $\vee$ ).*

*Proof* This is proved for the partition algebra in Proposition 2.9 of [31] and for the Brauer algebra in Lemma 3.1 of [32]. For the walled Brauer algebra, the proof is virtually identical as in the Brauer case and also follows immediately from Fig. 2.

In case of the partition algebra, we can easily state an exact criterion when two tableaux have a common JM weight. In the Brauer and walled Brauer algebra case, one first constructs from the ordinary arrow diagram  $d(\mu)$  a so-called cup-curl diagram for the Brauer and a cup-cap diagram for the walled Brauer algebra, denoted  $c(\mu)$  and superposes this onto the arrow diagram of  $d(\lambda)$ . Since we do not need the exact criterion in this article, refer the reader to [5] for the definition of these diagrams. Then one obtains the following result where we again leave out the half integer partition algebra case for simplicity:

**Theorem 5.5** *Suppose  $\lambda, \mu \in \Lambda$  with  $\lambda \neq \mu$  and  $\mu$  is a partition of  $r$  (and  $\mu'$  is a partition of  $s$ ). Then  $\lambda$  and  $\mu$  have a common JM weight if and only if*

1. *in the partition algebra case, the arrow diagrams  $d(\lambda)$  and  $d(\mu)$  are the same except for one configuration of  $l$  neighbouring columns of the following form:*



2. *in the case of the Brauer algebra, the superposition of the cup-curl diagram  $c(\mu)$  on top of  $d(\lambda)$  is oriented.*
3. *in the case of the walled Brauer algebra, the superposition of the cup-cap diagram  $c(\mu)$  on top of  $d(\lambda)$  is oriented.*

*Proof* The proof is not difficult but requires us to check a lot of cases. We outline the idea here. To show that given two partitions with a common JM weight, we get a diagram of the required form, we use an inductive proof. We start with a diagram of the required form and show that under moves with the same induced JM eigenvalue, we get another diagram of the same form. We can control these moves very well, see Fig. 2, but we have to basically check all possible forms which two adjacent columns of a diagram can take.

For the other direction, we interpret the cup-curl or cup-cap diagram as a construction plan for the tableau which give the common JM weight. The broad idea is to consider the arrow diagrams of  $\lambda$  and  $\mu$  and see whether it is possible to transform them through same column moves into the same diagram. We refer the reader to the proof of Theorem 6.4 in [32] for details. Notice that the walled Brauer algebra case is a subcase of the Brauer algebra case because the proof again only uses graphical calculus.

## 6 On the Proof of Theorem 4.7

We have to show that if  $\lambda$  and  $\mu$  have a common JM weight with  $\mu$  of maximal size, then  $L(\mu)$  is a composition factor of  $\Delta(\lambda)$ . We will achieve this by an induction on the length of a tableau. If we have a  $\mu$ -tableau  $u$  and a  $\lambda$ -tableau  $t$  of the same weight and length  $k$ , say, then certainly the partitions (or pairs of partitions)  $u^{k-1}$  and  $t^{k-1}$  have a common JM weight. The idea is to lift the induced homomorphism by using induction functors. In order to keep more control in the induction process, we first “factor” the induction functor, as we will see in the next definition which is completely analogous to 6.4 of [22]. For that definition, let  $d \in \{r, r + 1/2, (r, s)\}$  (depending on whether we are in the full partition algebra/Brauer algebra, half partition algebra or walled Brauer algebra case) and  $d'$  be equal to either  $d - 1/2$  (partition algebra),  $d - 1$  (Brauer algebra),  $d - (1, 0)$  (walled Brauer algebra with

$r > 0$ ) or  $d = (0, 1)$  (walled Brauer algebra with  $r = 0$ ). Furthermore, let  $c_n$  be the sum of the JM elements up to the element  $L_n$ . The element  $c_n$  is central in  $A_n$  for each of the algebras in question, see [9, Theorem 3.10], [24, Corollary 2.4] and [4, Lemma 2.1]. Therefore, the generalized eigenspaces of the action of  $c_n$  on any  $A_n$ -module  $M$  induce an  $A_n$ -module decomposition of  $M$ .

**Definition 6.1** Let  $M$  be an  $A_{d'}$ -module whose generalized eigenspace decomposition with respect to  $c_{d'}$  is trivial. For  $i \in \mathbb{Z}$ , define  $i\text{-ind}_{d'} M$  to be the projection onto the generalized eigenspace of the action of the JM-element  $L_d$  on  $\text{ind}_{d'} M$  with eigenvalue  $i$ . Similarly, for an  $A_d$ -module  $M$  whose  $c_d$  generalized eigenspace decomposition is trivial, we set  $i\text{-res}_d M$  to be the generalized eigenspace of the action of the JM-element  $L_{d'}$  on  $\text{res}_d M$  with eigenvalue  $i$ . Extend this definition to arbitrary  $A_{d'}$ -modules and  $A_d$ -modules, respectively, by first decomposing them into generalized eigenspaces with respect to  $c_{d'}$  and  $c_d$ , respectively, and then applying  $i\text{-ind}$  and  $i\text{-res}$ , respectively, to each summand. We usually omit  $d$  and  $d'$  if they are clear from the context.

*Remark 6.2*

1. Notice that the definition of the  $i$ -induction map is well-defined since it is equivalent to first decomposing a module into generalized eigenspaces with respect to the central element  $c_{d'}$ , then inducing to  $A_d$  and finally decomposing into generalized eigenspaces with respect to the central element  $c_d$  since  $c_d - c_{d'} = L_d$ . A similar remark holds for  $i$ -restriction.
2. Since the cell modules for the algebra  $A_n$  are generically simple, the central element  $c_n$  acts by a scalar multiple of the identity on them and, in particular, their generalized eigenspace decomposition with respect to  $c_n$  is always trivial.

These maps have the usual adjointness properties by exactly the same argument as in 6.4 of [22].

**Proposition 6.3** *The maps*

$$i\text{-ind} : A_{d'}\text{-mod} \rightarrow A_d\text{-mod} \quad i\text{-res} : A_d\text{-mod} \rightarrow A_{d'}\text{-mod}$$

*are functorial and  $i\text{-ind}$  is a left-adjoint to  $i\text{-res}$ .*

For cell modules, the following restriction rules can be deduced from Theorem 3.7 by projection onto the corresponding JM weight space:

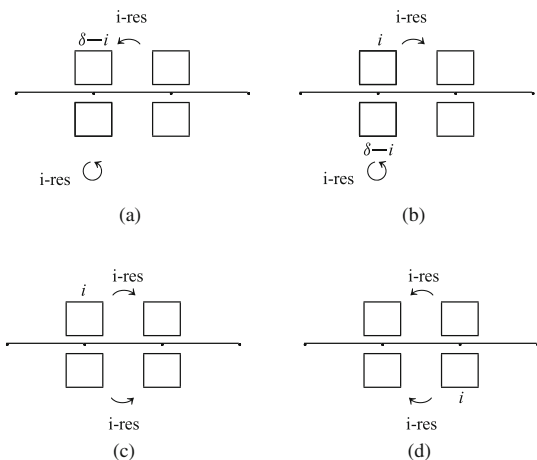
**Theorem 6.4** *Suppose  $\lambda \in \Lambda$  with  $d$  and  $d'$  as before. Then there are exact sequences as follows (where in each case we set  $\Delta(\lambda \pm \epsilon) = 0$  if no addable/removable box  $\epsilon$  with the required content exists)*

1. *Partition algebra:*

$$(a) \quad 0 \longrightarrow \Delta_r(\lambda) \longrightarrow i\text{-res} \Delta_{r+1/2}(\lambda) \longrightarrow \Delta_r(\lambda + \epsilon) \longrightarrow 0,$$

*where  $\epsilon$  is an addable box of content  $\delta - i$  and the term on the left is zero unless  $|\lambda| = i$ .*

**Fig. 3** Possible ways to move arrows when *i*-restricting. (a) Half to full degree. (b) Full to half degree. (c) (Walled) Brauer algebra I. (d) (Walled) Brauer algebra II



(b)  $0 \longrightarrow \Delta_{r-1/2}(\lambda - \epsilon) \longrightarrow \text{i-res } \Delta_r(\lambda) \longrightarrow \Delta_{r-1/2}(\lambda) \longrightarrow 0$ ,  
 where  $\epsilon$  is a removable box of content  $i$  and the term on the right is zero unless  $|\lambda| = \delta - i$ .

2. Brauer algebra:

$$0 \longrightarrow \Delta_{r-1}(\lambda - \epsilon) \longrightarrow \text{i-res } \Delta_r(\lambda) \longrightarrow \Delta_{r-1}(\lambda + \epsilon') \longrightarrow 0$$

where  $\epsilon$  is a removable box of  $\lambda$  of content  $i$  and  $\epsilon'$  is an addable box of  $\lambda$  of content  $1 - \delta - i$ .

3. Walled Brauer algebra:

a. If  $r > 0$ , then

$$0 \longrightarrow \Delta_{r-1,s}(\lambda - \epsilon, \lambda') \longrightarrow \text{i-res } \Delta_{r,s}(\lambda, \lambda') \longrightarrow \Delta_{r-1,s}(\lambda, \lambda' + \epsilon') \longrightarrow 0$$

b.  $\text{i-res } \Delta_{0,s}(\emptyset, \lambda') \cong \Delta_{0,s-1}(\emptyset, \lambda' - \epsilon)$ .

where  $\epsilon$  is a removable box of  $\lambda$  of content  $i$  and  $\epsilon'$  is an addable box of  $\lambda$  of content  $-\delta - i$ .

Remark 6.5

1. Similar statements exist for *i*-induction.
2. Notice that there is always at most one addable and at most one removable box of a given content when working over characteristic 0.

Figure 3 summarizes the theorem in a similar way as Fig. 2. Notice that the Brauer and walled Brauer case are identical since we did not indicate the label opposite to *i*.

We will also need restriction rules for simple modules. Unfortunately, we can only describe the socle in full generality which will, however, be sufficient for our purposes. In the following theorem, we abuse notation for simplicity and write  $\Delta(\lambda)$  instead of  $\Delta(\lambda, \lambda')$ . Furthermore, the symbol  $?$  is fixed for each situation and either stands for  $\vee$  or  $\bigcirc$ .

**Theorem 6.6** *Consider the partition, Brauer or walled Brauer algebra over a field of characteristic 0. Then*

1. If  $i\text{-res } \Delta(\lambda) = 0$ , then  $i\text{-res } L(\lambda) = 0$ .
2. If  $i\text{-res } \Delta(\lambda) = \Delta(\lambda')$  and  $i\text{-ind } \Delta(\lambda') = \Delta(\lambda)$ , then  $i\text{-res } L(\lambda) = L(\lambda')$ .
3. Partition algebra:

a. Let  $\lambda^+ = \dots \overset{i}{\vee} \wedge \dots \in \Lambda$  and  $\lambda^- = \dots \overset{i}{\wedge} \vee \dots \in \Lambda$ , then

$$i\text{-res } L(\lambda^+) = L(\lambda^-) \text{ and } i\text{-res } L(\lambda^-) = 0.$$

b. If  $\lambda = \dots \overset{?}{\bigcirc} \overset{i}{\vee} \bigcirc \dots \in \Lambda$ , then the socle of  $i\text{-res } L(\lambda)$  is isomorphic to

$$L(\dots \overset{?}{\bigcirc} \overset{i}{\vee} \bigcirc \dots) \oplus L(\dots \overset{?}{\bigcirc} \bigcirc \overset{i}{\vee} \dots).$$

4. Brauer algebra:

a. Suppose  $\lambda' \in \Lambda'$  (where  $\Lambda'$  is labelling set for simples for  $B_{r-1}(\delta)$ ) has an addable box  $\epsilon$  of content  $i$  and a removable box  $\epsilon'$  of content  $1 - \delta - i$ , where these boxes need not be distinct. Set  $\lambda^+ = \lambda' + \epsilon \in \Lambda$  and  $\lambda^- = \lambda' - \epsilon' \in \Lambda$ . Then

$$i\text{-res } L(\lambda^+) = L(\lambda') \text{ and } i\text{-res } L(\lambda^-) = \begin{cases} L(\lambda') & \text{if } \epsilon + \epsilon' = (1, 1) \\ 0 & \text{otherwise} \end{cases}.$$

b. Suppose we are not in case (1),(2) or (4)(a). Then  $\lambda$  has a removable box  $\epsilon'$  of content  $i$  and an addable box  $\epsilon$  of content  $1 - \delta - i$  such that the socle of  $i\text{-res } L(\lambda)$  is isomorphic to

$$L(\lambda - \epsilon') \oplus L(\lambda + \epsilon).$$

5. Walled Brauer algebra  $B_{r,s}(\delta)$  with  $r > 0$ :

a. Suppose  $\lambda' = (\lambda'_L, \lambda'_R) \in \Lambda'$  (where  $\Lambda'$  is labelling set for simples for  $B_{r-1,s}(\delta)$ ) and  $\lambda'_L$  has an addable box  $\epsilon$  of content  $i$  and  $\lambda'_R$  a removable box  $\epsilon'$  of content  $-\delta - i$ . Set  $\lambda^+ = (\lambda'_L + \epsilon, \lambda'_R) \in \Lambda$  and  $\lambda^- = (\lambda'_L, \lambda'_R - \epsilon') \in \Lambda$ .



Then

$$i\text{-res } L(\lambda^+) = L(\lambda') \text{ and } i\text{-res } L(\lambda^-) = 0.$$

- b. Suppose we are not in case (1),(2) or (5)(a) and  $\lambda = (\lambda_L, \lambda_R) \in \Lambda$ . Then  $\lambda_L$  has a removable box  $\epsilon'$  of content  $i$  and  $\lambda_R$  has an addable box  $\epsilon$  of content  $-\delta - i$  such that the socle of  $i\text{-res } L(\lambda)$  is isomorphic to

$$L(\lambda_L - \epsilon', \lambda_R) \oplus L(\lambda_L, \lambda_R + \epsilon).$$

*Remark 6.7* We did not include the case of the walled Brauer algebra  $B_{r,s}(\delta)$  with  $r = 0$ , since  $B_{0,s}(\delta) \cong FS_s$  and the restriction rules are precisely the ordinary branching rules for cell modules.

*Proof* This was proved for the partition algebra in [31, Theorem 3.3] and for the Brauer algebra in [32, Theorem 4.5]. The proof of the walled Brauer algebra is again almost exactly the same as the Brauer algebra case since the proof mostly uses general properties of induction and restriction functors as well as graphical calculus of arrow diagrams both of which are identical for both algebras. We simply have to change  $1 - \delta$  to  $-\delta$ , replace partitions by pairs of partitions and be careful whether the boxes which are added and removed are done so to the left or to the right partition in the pair. Furthermore, the corner cases all vanish and in particular, we do not need to consider the case when the added and removed boxes form the partition  $(1, 1)$  as  $L(\lambda^+)$  will always be a composition factor of  $\Delta(\lambda^-)$ .

We are finally in a position to sketch the proof of Theorem 4.7:

*Proof* We have to show that if there are tableaux  $t \in \text{Tab}(\lambda)$  and  $u \in \text{Tab}(\mu)$  with  $\text{wt}(u) = \text{wt}(t)$  and  $\mu$  is a maximal partition, then  $[\Delta(\lambda) : L(\mu)] \neq 0$ . The proof will be by induction on the degree of the algebras. The result is trivially true for  $r = 0$  (and  $s = 0$ ) or  $r = 1/2$  in the partition algebra case. Assume it holds in smaller degrees.

We will denote by  $\lambda'$  and  $\mu'$  the second-last entry of  $t$  and  $u$ , respectively (the last entry of course being  $\lambda$  and  $\mu$  themselves). It is possible to show that without loss of generality, we may choose  $u$  and  $t$  in such a way that  $i\text{-res } \Delta(\lambda) \cong \Delta(\lambda')$  with  $i = c(\epsilon)$  where  $\epsilon$  is the box added at the last step of  $u$ , see Lemma 4.7 of [32], the walled Brauer case being the same except that we do not even have to worry about the critical corner.

If we just leave out the last partition in  $u$  and  $t$ , we obtain  $t' \in \text{Tab}(\lambda')$  and  $u' \in \text{Tab}(\mu')$  with  $\text{wt}(u') = \text{wt}(t')$  and  $\mu'$  is of maximal size, so that by induction assumption we have  $[\Delta(\lambda') : L(\mu')] \neq 0$ . Therefore,  $\text{Hom}(\Delta(\mu'), \Delta(\lambda')/M) \neq 0$  for some submodule  $M' \leq \Delta(\lambda')$ .

Thus, there is a submodule  $U \leq \Delta(\lambda')$  with unique simple top  $L(\mu') = \Delta(\mu')$  (for example let  $U$  be a submodule of  $\Delta(\lambda')$  of minimal length which contains  $\Delta(\mu')$ )

as a quotient).

$$0 \neq \text{Hom}(U, \Delta(\lambda')) = \text{Hom}(U, \text{i-res } \Delta(\lambda)) = \text{Hom}(\text{i-ind } U, \Delta(\lambda)).$$

We will examine the head of  $\text{i-ind } U$  more closely.

Suppose the simple module  $L(\tau)$  occurs in the head of  $\text{i-ind } U$  and is a composition factor of  $\Delta(\lambda)$ . Then

$$0 \neq \text{Hom}(\text{i-ind } U, L(\tau)) = \text{Hom}(U, \text{i-res } L(\tau)).$$

Since  $L(\tau)$  is a composition factor of  $\Delta(\lambda)$ ,  $\tau$  and  $\lambda$  have a common JM-weight. Therefore,  $d(\tau)$  and  $d(\lambda)$  have the same number of arrows in each column by Proposition 5.4. Since  $\text{i-res } \Delta(\lambda) \cong \Delta(\lambda')$ ,  $\lambda$  and hence  $\tau$  cannot be of the form in Case (3)(b), (4)(b) or (5)(b), respectively, of Theorem 6.6. In particular,  $\text{i-res } L(\tau)$  is either equal to  $L(\tau')$ , for some partition  $\tau'$  differing from  $\tau$  by one box, or zero. The latter case cannot happen, since the Hom-space is non-zero. Since  $U$  has unique simple top  $L(\mu')$ , we can conclude that  $\text{i-res } L(\tau) = L(\mu')$  and therefore also  $\tau' = \mu'$ .

To prove the theorem, it will suffice to show that  $\tau$  must be equal to  $\mu$ . But  $\mu'$  was obtained from  $\mu$  by moving an arrow of label  $i$  to the label  $i - 1$  and we can therefore show the result by simply considering the possibilities for the column  $c$  of  $d(\mu') = d(\tau')$  containing the label  $i - 1$ . This is done in the proof of Theorem 4.1 of [31] for the partition algebra and Theorem 5.3 of [32] for the Brauer algebra. The case of the walled Brauer algebra is again identical to the Brauer algebra case since it is an exercise in arrow diagram calculus which is identical for both algebras (and even simpler for the walled Brauer algebra, as there is no corner case to be considered). Notice that in Case II of the proof of Theorem 5.3 in [32], we always have  $\text{i-res } L(\mu^-) = 0$  in the walled Brauer case, so there is no need to distinguish whether  $\epsilon + \epsilon' = (1, 1)$  or not and the proof is much shorter.

## References

1. G. Benkart, M. Chakrabarti, T. Halverson, R. Leduc, C. Lee, J. Stroomer, Tensor product representations of general linear groups and their connections with Brauer algebras. *J. Algebra* **166**, 529–567 (1994)
2. R. Brauer, On algebras which are connected with the semisimple continuous groups. *Ann. Math.* **38**, 857–872 (1937)
3. J. Brundan, C. Stroppel, Highest weight categories arising from Khovanov’s diagram algebra I: cellularity. *Mosc. Math. J.* **11**, 685–722 (2011)
4. J. Brundan, C. Stroppel, Gradings on walled Brauer algebras and Khovanov’s arc algebra. *Adv. Math.* **231**, 709–773 (2012)
5. A. Cox, M. De Visscher, Diagrammatic Kazhdan-Lusztig theory for the (walled) Brauer algebra. *J. Algebra* **340**, 151–181 (2011)
6. A. Cox, P. Martin, A. Parker, C. Xi, Representation theory of towers of recollement: theory, notes, and examples. *J. Algebra* **302**, 340–360 (2006)

7. W.F. Doran, D.B. Wales, P.J. Hanlon, On the semisimplicity of the Brauer centralizer algebras. *J. Algebra* **211**, 647–685 (1999)
8. J. Enyang, Cellular bases for the Brauer and Birman-Murakami-Wenzl algebras. *J. Algebra* **281**, 413–449 (2004)
9. J. Enyang, Jucys-Murphy elements and a presentation for partition algebras. *J. Algebraic Combin.* **37**, 401–454 (2013)
10. J.J. Graham, G.I. Lehrer, Cellular algebras. *Invent. Math.* **123**, 1–34 (1996)
11. T. Halverson, A. Ram, Partition algebras. *Eur. J. Comb.* **26**, 869–921 (2005)
12. R. Hartmann, R. Paget, Young modules and filtration multiplicities for Brauer algebras. *Math. Z.* **254**, 333–357 (2006)
13. R. Hartmann, A. Henke, S. Koenig, R. Paget, Cohomological stratification of diagram algebras. *Math. Ann.* **347**, 765–804 (2010). <https://doi.org/10.1007/s00208-009-0458-x>
14. A. Henke, S. Koenig, Schur algebras of Brauer algebras I. *Math. Z.* **272**, 729–759 (2012)
15. A. Henke, S. Koenig, Schur algebras of Brauer algebras, II. *Math. Z.* **276**, 1077–1099 (2014)
16. V.F.R. Jones, The Potts model and the symmetric group, in *Subfactors (Kyuzeso, 1993)* (World Scientific Publishing, River Edge, NJ, 1994), pp. 259–267
17. O. King, Decomposition numbers of partition algebras in non-dividing characteristic. Preprint (2013)
18. S. Koenig, C. Xi, Affine cellular algebras. *Adv. Math.* **229**, 139–182 (2012)
19. P. Martin, Temperley-Lieb algebras for nonplanar statistical mechanics—the partition algebra construction. *J. Knot Theor. Ramif.* **3**, 51–82 (1994)
20. P. Martin, The partition algebra and the Potts model transfer matrix spectrum in high dimensions. *J. Phys. A* **33**, 3669–3695 (2000)
21. P. Martin, The decomposition matrices of the Brauer algebra over the complex field (2009). arXiv:0908.1500
22. A. Mathas, *Iwahori-Hecke Algebras and Schur Algebras of the Symmetric Group*. University Lecture Series, vol. 15 (American Mathematical Society, Providence, RI, 1999)
23. A. Mathas, Seminormal forms and Gram determinants for cellular algebras. *J. Reine Angew. Math.* **619**, 141–173 (2008)
24. M. Nazarov, Young’s orthogonal form for Brauer’s centralizer algebra. *J. Algebra* **182**, 664–693 (1996)
25. D. Nguyen, A cellular basis of the q-brauer algebra related with murphy bases of the Hecke algebras (2013). arXiv:1302.4272
26. I. Paul, Structure theory for cellularly stratified diagram algebras, Doctoral thesis, 2016
27. A. Ram, Characters of Brauer’s centralizer algebras. *Pacific J. Math.* **169**, 173–200 (1995)
28. I. Schur, *Über die rationalen Darstellungen der allgemeinen linearen Gruppe* (Sitzungsberichte Akad., Berlin, 1927), pp. 58–75
29. A. Shalile, Conjugation in Brauer algebras and applications to character theory. *J. Pure Appl. Algebra* **215**, 2694–2714 (2011)
30. A. Shalile, On the center of the Brauer algebra. *Algebr. Represent. Theory* **16**, 65–100 (2013)
31. A. Shalile, On the modular representation theory of the partition algebra. *J. Algebraic Combin.* **42**, 245–282 (2015)
32. A. Shalile, Decomposition numbers of Brauer algebras in non-dividing characteristic. *J. Algebra* **423**, 963–1009 (2015)
33. C. Xi, Partition algebras are cellular. *Compos. Math.* **119**, 99–109 (1999)

# Koblitz's Conjecture for Abelian Varieties



Ute Spreckels and Andreas Stein

**Abstract** Consider a principally polarized abelian variety  $A$  of dimension  $d$  defined over a number field  $F$ . If  $\mathfrak{p}$  is a prime ideal in  $F$  such that  $A$  has good reduction at  $\mathfrak{p}$ , let  $N_{\mathfrak{p}}$  be the order of  $A \bmod \mathfrak{p}$ . We have formulae for the density  $p_{\ell}$  of primes  $\mathfrak{p}$  such that  $N_{\mathfrak{p}}$  is divisible by a fixed prime number  $\ell$  in two cases:  $A$  is a CM abelian variety and the CM-field is contained in  $F$ , or  $A$  has trivial endomorphism ring and its dimension is 2, 6 or odd. In both cases, we can prove that  $C_A = \prod_{\ell} \frac{1-p_{\ell}}{1-1/\ell}$  is a positive constant. We conjecture that the number of primes  $\mathfrak{p}$  with norm up to  $n$  such that  $N_{\mathfrak{p}}$  is prime is given by the formula  $C_A \frac{n}{d \log(n)^2}$ , generalizing a formula by N. Koblitz, conjectured in 1988 for elliptic curves. Numerical evidence that supports this conjectural formula is provided.

**Keywords** Abelian varieties over finite fields • Galois representations • General symplectic group over a finite field • Serre's open image theorem • Torsion points of abelian varieties

**Subject Classifications** 11G10, 11N05, 11F80, 11G20

## 1 Introduction

Let  $A$  be a principally polarized abelian variety of dimension  $d$  defined over a number field  $F$  with absolute Galois group  $G_F$ . Denote by  $\mathfrak{p}$  a prime ideal of  $F$  of inertia degree 1 such that  $A$  has good reduction at  $\mathfrak{p}$ . Let  $p$  be the order of the residue field of  $\mathfrak{p}$ . Consider  $N_{\mathfrak{p}} = \# \bar{A}(\mathbb{F}_p)$  to be a random variable depending on  $\mathfrak{p}$  where  $\bar{A} = A \bmod \mathfrak{p}$ . A random integer in the size of  $p^d \approx \# \bar{A}(\mathbb{F}_p)$  is prime with

---

U. Spreckels • A. Stein (✉)

Institut für Mathematik, Carl von Ossietzky Universität Oldenburg, 26111 Oldenburg, Germany  
e-mail: [ute.spreckels@uni-oldenburg.de](mailto:ute.spreckels@uni-oldenburg.de); [andreas.stein@uni-oldenburg.de](mailto:andreas.stein@uni-oldenburg.de)

probability  $(d \log(p))^{-1}$  by the prime number theorem. Hence if  $N_{\mathfrak{p}}$  were random, there would be about

$$\sum_{p \leq n} \frac{1}{d \log(n)} \approx \frac{n}{d \log(n)^2} \tag{1}$$

primes  $\mathfrak{p}$  with norm less than  $n$  such that  $N_{\mathfrak{p}}$  is prime. However, the orders  $N_{\mathfrak{p}}$  do not behave like random integers in respect to primality.

For every prime number  $\ell$ , the action of Galois automorphisms on torsion points of  $A$  induces a representation of  $G_F$  on  $\text{Aut}(A[\ell]) \cong \mathbb{F}_{\ell}^{2d}$ . Since  $A$  is principally polarized, the Galois action respects the symplectic Weil pairing on  $A$ , thus there is a representation

$$\rho_{\ell} : G_F \rightarrow \text{GSp}_{2d}(\mathbb{F}_{\ell}).$$

The fixed field of the kernel of  $\rho_{\ell}$  is called the  $\ell$ -division field of  $A$ . Its Galois group is denoted by  $G_{\ell}$ . Let  $\mathfrak{p} \mid p$  be a prime ideal of  $F$  such that  $A$  has good reduction at  $p$ . It is known that the order of the group of  $\mathbb{F}_{\mathfrak{p}}$ -rational points on  $A \bmod \mathfrak{p}$  is divisible by  $\ell$  if and only if the Frobenius in  $G_{\ell}$ , represented in  $\text{GSp}_{2d}(\mathbb{F}_{\ell})$ , has eigenvalue 1 [11, Lem. 2.1]. If we denote by  $p_{\ell}$  the density of prime ideals  $\mathfrak{p}$  of  $F$  such that  $\ell \mid \#A(\mathbb{F}_{\mathfrak{p}})$ , then Chebotarev’s density theorem yields that  $p_{\ell}$  equals the proportion of Galois automorphisms in  $G_{\ell}$  that have eigenvalue 1.

As first realized by Koblitz [4],  $p_{\ell}$  can be used to improve the estimate (1). Let

$$C_A = \prod_{\ell \text{ prime}} \frac{1 - p_{\ell}}{1 - 1/\ell}.$$

Based on (1) we conjecture that there are about

$$C_A \sum_{p \leq n} \frac{1}{d \log(n)} \approx C_A \frac{n}{d \log(n)^2} \tag{2}$$

primes  $\mathfrak{p}$  with norm less than  $n$  such that  $N_{\mathfrak{p}}$  is prime. We expect the conjecture to hold true in case that the events  $\ell \mid N_{\mathfrak{p}}$ , where  $\ell$  is any prime number, are mutually independent. The elliptic curve version of this conjecture can be found in Koblitz’s paper [4], a dimension 2 version in [11, Conj. 4.1] and the generalized conjecture in [6, Conj. 1.3]. Koblitz presented examples of elliptic curves over  $\mathbb{Q}$  supporting this conjecture for  $d = 1$ . Zywinia noticed that the Koblitz conjecture is false if one does not ensure that the events  $\ell \mid N_{\mathfrak{p}}$  are mutually independent [12]. In case that dependencies occur, the constant  $C_A$  cannot be the product of probabilities as above. Zywinia presented an example of an elliptic curve where Koblitz’s original conjecture actually fails and thus suggested refinements for  $C_A$  [12, Section 2.1]. Besides, he extended Koblitz’s conjecture to elliptic curves defined over number fields [12, Conj. 1.2]. We do not know of any counterexample of (2) for  $d > 1$ .

It is an interesting problem to find abelian varieties such that the events  $\ell \mid N_p$  are dependent. However, we do not consider this situation. From now on, we therefore assume that the events  $\ell \mid N_p$  are independent and work with  $C_A$  as defined above.

In case that  $A$  has CM, the Galois group of the  $\ell$ -division field can be embedded in a maximal torus in the general symplectic group  $\mathrm{GSp}_{2d}(\mathbb{F}_\ell)$ . We obtained a formula for  $p_\ell$  by a matrix counting technique in [6]. This led us to the following theorem.

**Theorem 1.1** *Let  $A$  be a principally polarized abelian variety of arbitrary dimension  $d$  defined over a number field  $F$  such that  $A$  has CM by a CM-field  $K \subseteq F$ . For a prime number  $\ell$ , let  $G_\ell$  be the Galois group of the  $\ell$ -division field of  $A$  and assume that  $G_\ell$  is isomorphic to a maximal torus of  $\mathrm{GSp}_{2d}(\mathbb{F}_\ell)$  for almost all prime numbers  $\ell$ . Then  $C_A = \prod_{\ell \text{ prime}} \frac{1-p_\ell}{1-1/\ell}$  is a positive constant.*

*Proof* See [6], Main Theorem 1.2. □

The formula for  $C_A$  that can be found in [6] does not depend on  $A$  itself but only on its CM-field. Hence it is reasonable not to consider a single abelian variety but every abelian variety that has CM by a fixed CM-field at once to test the conjectural formula (2). The conjecture provides us with a prediction of how many of the orders  $N_p$  are prime. To check this prediction we thus have to determine the number of points of abelian varieties over various finite prime fields. We were able to avoid long-lasting computations by exploiting the fact that there is a bijection

$$\mathcal{A}_{p,K} \leftrightarrow \{[w] \mid w \in \mathcal{O}_K, w\bar{w} = p, \mathbb{Q}(w) = K\}$$

sending  $[A]$  to the conjugacy class of the Frobenius endomorphism of  $A$  (by a theorem of Honda and Tate, see also [6, Theorem 5.2]). Here  $p$  is a prime number and  $\mathcal{A}_{p,K}$  is the set of isogeny classes of simple abelian varieties over  $\mathbb{F}_p$  which have CM by the fixed CM-field  $K$ . The algebraic integers  $w$  in the set on the right-hand side are so-called Weil  $p$ -integers. Let  $P_w$  denote the characteristic polynomial of a Weil  $p$ -integer  $w$ . If  $A$  corresponds to  $w$  by the above bijection, then the number of points of  $A/\mathbb{F}_p$  equals  $P_w(1)$ . Thus for our purpose it suffices to compute the set of Weil  $p$ -integers for several  $p$  (in [6] we considered  $p$  up to  $10^7$ ) and evaluate their characteristic polynomials at 1 to obtain the orders. We then may check how many of these orders are prime and compare the result with the number predicted by the conjecture.

Tables with numerical results can be found in both [11] (dimension 2) and [6] (dimension 3). The experimental data support the conjecture.

In this paper we are interested in the constant  $C_A$  and the conjecture (2) in the case where Serre’s open image theorem applies, i.e. for abelian varieties  $A$  with  $\mathrm{End}(A) = \mathbb{Z}$  and  $d = 2, 6$  or odd. The main result is the following.

**Theorem 1.2** *Let  $A$  be a principally polarized abelian variety of arbitrary dimension  $d$  defined over a number field  $F$  such that  $\mathrm{End}(A) = \mathbb{Z}$  and  $d = 2, 6$  or odd. Then  $C_A = \prod_{\ell \text{ prime}} \frac{1-p_\ell}{1-1/\ell}$  is a positive constant.*

By Serre’s open image theorem  $G_\ell$  is isomorphic to  $\mathrm{GSp}_{2d}(F_\ell)$  for  $\ell \gg 0$ . Thus the ratio of matrices with eigenvalue 1 in  $\mathrm{GSp}_{2d}(\mathbb{F}_\ell)$  equals  $p_\ell$ . We prove the convergence of  $C_A$  with intricate computations using results of [10].

We conclude the paper by presenting numerical evidence to support the conjectural formula (2). We counted the number of prime orders  $N_p$  with  $\mathrm{Norm}(\mathfrak{p}) \leq n$  for the Jacobians of the genus 2 curves 587.a.587.1, 743.a.743.1, 971.a.971.1 and 1051.a.1051.1 from the L-Functions and Modular Forms Database [7] ( $n = 4000$ ). We also treated a genus 3 curve that stems from [13] ( $n = 4300$ ).

## 2 Abelian Varieties with Trivial Endomorphism Ring

By Serre’s open image theorem [5, Theorem 7.3 and subsequent corollary], if  $A$  is an abelian variety with trivial endomorphism ring and dimension 2 or 6 or odd, then the representation  $\rho_\ell$  defined in the introduction is surjective for almost all  $\ell$ . Thus  $G_\ell$  is isomorphic to  $\mathrm{GSp}_{2d}(\mathbb{F}_\ell)$  for almost all  $\ell$  in this case. Recall that

$$\mathrm{GSp}_{2d}(\mathbb{F}_\ell) = \{M \in \mathrm{GL}_{2d}(\mathbb{F}_\ell) \mid \exists \lambda \in \mathbb{F}_\ell^\times : M'JM = \lambda J\}$$

where  $J = \begin{pmatrix} 0 & I_d \\ -I_d & 0 \end{pmatrix}$ .

**Proposition 2.1** *The order of the general symplectic group is*

$$\#\mathrm{GSp}_{2d}(\mathbb{F}_\ell) = \ell^{d^2}(\ell - 1) \prod_{i=1}^d (\ell^{2i} - 1) = \ell^{2d^2+d+1} - \ell^{2d^2+d} + \mathcal{O}(\ell^{2d^2+d-1}).$$

*Proof* It is easy to see that there is an isomorphism

$$\mathrm{Sp}_{2d}(\mathbb{F}_\ell) \times \left\{ \begin{pmatrix} \lambda I_d & 0 \\ 0 & I_d \end{pmatrix} \mid \lambda \in \mathbb{F}_\ell^\times \right\} \cong \mathrm{GSp}_{2d}(\mathbb{F}_\ell)$$

and the order of  $\mathrm{Sp}_{2d}(\mathbb{F}_\ell)$  is  $\ell^{d^2} \prod_{i=1}^d (\ell^{2i} - 1)$ . □

Let  $e_\ell(d)$  be the number of matrices with eigenvalue 1 in  $\mathrm{GSp}_{2d}(\mathbb{F}_\ell)$ . A formula for  $e_\ell(d)$  was computed in various context e.g. [1, 3] and in [10]. We will rely on the latter as this work was done explicitly with the idea to apply to our setting. We want to prove that

$$e_\ell(d) = \ell^{2d^2+d} - 3\ell^{2d^2+d-2} + \mathcal{O}(\ell^{2d^2+d-3}).$$

Let  $m, s \in \mathbb{N}_0$  with  $m \leq 2d$  and  $2s \leq m$ . Define

$$T_\ell(d, m, 0) = (\ell - 1)\ell^{d^2 - \frac{m(m-1)}{2}} \prod_{j=1}^d (\ell^{2j} - 1) \prod_{j=1}^m (\ell^j - 1)^{-1}, \tag{3}$$

and for  $s > 0$ ,

$$T_\ell(d, m, s) = \ell^{d^2 - (m-s)^2 - \frac{(m-2s+1)(m-2s)}{2}} \prod_{j=s+1}^d (\ell^{2j} - 1) \prod_{j=1}^{m-2s} (\ell^j - 1)^{-1}. \tag{4}$$

**Theorem 2.2 ([10], Thm. 3.1)** *The number of matrices with eigenvalue 1 in  $\text{GSp}_{2d}(\mathbb{F}_\ell)$  is*

$$e_\ell(d) = \sum_{i=0}^{2d-1} (-1)^i \ell^{\frac{i(i+1)}{2}} \sum_{s=\max(0, i+1-d)}^{\lfloor \frac{i+1}{2} \rfloor} T_\ell(d, i+1, s)$$

*Proof* This is an immediate consequence of [10, Thm. 3.1 and Cor. 2.1] and the last formula on [10, p. 9] (which in turn stems from [9]), applied with  $m = 1$ .  $\square$

To examine the convergence of  $C_A$  we will concentrate on the highest powers of  $\ell$  in  $e_\ell(d)$ . We need the following statements.

**Lemma 2.3**

1. *Let  $d, m, s \in \mathbb{N}$  with  $m \leq 2d$ ,  $s \leq d$  and  $2s \leq m$ . We have*

$$T_\ell(d, m, s) = T_\ell(d - 1, m, s)\ell^{2d-1}(\ell^{2d} - 1).$$

2. *Let  $d \in \mathbb{N}$  with  $d \geq 2$ . We have*

$$T_\ell(d, 2d - 1, d - 1) = \ell^{2d} + \ell^{2d-1} + \dots + \ell, \\ T_\ell(d, 2d, d) = 1.$$

*Proof* The claims follow immediately from the definition of  $T$ . For instance, for  $s = 0$ , by (3) we have

$$T_\ell(d, m, 0) = (\ell - 1)\ell^{d^2 - \frac{(m-1)m}{2}} \prod_{j=1}^d (\ell^{2j} - 1) \prod_{j=1}^m (\ell^j - 1)^{-1} \\ = (\ell - 1)\ell^{(d-1)^2 + 2d-1 - \frac{(m-1)m}{2}} (\ell^{2d} - 1) \prod_{j=1}^{d-1} (\ell^{2j} - 1) \prod_{j=1}^m (\ell^j - 1)^{-1} \\ = T_\ell(d - 1, m, 0)\ell^{2d-1}(\ell^{2d} - 1),$$



and for  $s > 0$

$$\begin{aligned}
 T_\ell(d, m, s) &= \ell^{d^2 - (m-s)^2 - \frac{(m+1-2s)(m-2s)}{2}} \prod_{j=s+1}^d (\ell^{2j} - 1) \prod_{j=1}^{m-2s} (\ell^j - 1)^{-1} \\
 &= T_\ell(d-1, m, s) \ell^{2d-1} (\ell^{2d} - 1),
 \end{aligned}$$

so the same identity holds in both cases.

The second assertion also is an immediate consequence of (4). □

We are now ready to prove the following.

**Corollary 2.4** *Assume  $d \geq 2$ . The number of matrices with eigenvalue 1 in  $\text{GSp}_{2d}(\mathbb{F}_\ell)$  is*

$$e_\ell(d) = \ell^{2d^2+d} - 3\ell^{2d^2+d-2} + O(\ell^{2d^2+d-3}).$$

*Proof* We will prove the claim using induction over  $d$ . For  $d = 2$ , we obtain from Theorem 2.2 and the definitions (3) and (4),

$$e_\ell(2) = \ell^{10} - 3\ell^8 + \ell^7 - \ell^6 + \ell^5 + 4\ell^4.$$

Assume the claim holds for  $d - 1$  (where  $d \geq 3$ ), i.e. we have

$$\begin{aligned}
 e_\ell(d-1) &= \ell^{2(d-1)^2+d-1} - 3\ell^{2(d-1)^2+d-3} + O(\ell^{2(d-1)^2+d-4}) \\
 &= \ell^{2d^2-3d+1} - 3\ell^{2d^2-3d-1} + O(\ell^{2(d-1)^2+d-4})
 \end{aligned}$$

Using Lemma 2.3 several times, we can express  $e_\ell(d)$  in terms of  $e_\ell(d - 1)$ :

$$\begin{aligned}
 e_\ell(d) &= \sum_{i=0}^{2(d-1)-1} (-1)^i \ell^{\frac{i(i+1)}{2}} \sum_{s=\max(0, i+1-d)}^{\lfloor \frac{i+1}{2} \rfloor} T_\ell(d, i+1, s) \\
 &\quad + \ell^{\frac{(2d-2)(2d-1)}{2}} T_\ell(d, 2d-1, d-1) - \ell^{\frac{(2d-1)2d}{2}} T_\ell(d, 2d, d) \\
 &= \sum_{i=0}^{2(d-1)-1} (-1)^i \ell^{\frac{i(i+1)}{2}} \sum_{s=\max(0, i+1-(d-1))}^{\lfloor \frac{i+1}{2} \rfloor} T_\ell(d-1, i+1, s) \cdot \ell^{2d-1} (\ell^{2d} - 1) \\
 &\quad + \sum_{i=d-1}^{2(d-1)-1} (-1)^i \ell^{\frac{i(i+1)}{2}} T_\ell(d, i+1, i+1-d) \\
 &\quad + \ell^{2d^2-3d+1} (\ell^{2d} + \ell^{2d-1} + \dots + \ell) - \ell^{2d^2-d}
 \end{aligned}$$

$$\begin{aligned}
 &= e_\ell(d-1)\ell^{2d-1}(\ell^{2d}-1) \\
 &\quad + \sum_{i=d-1}^{2d-3} (-1)^i \ell^{\frac{i(i+1)}{2}} T_\ell(d, i+1, i+1-d) \\
 &\quad + \ell^{2d^2-d+1} + \ell^{2d^2-d-1} + \dots + \ell^{2d^2-3d+2}
 \end{aligned} \tag{5}$$

The induction hypothesis implies that

$$\begin{aligned}
 e_\ell(d-1)\ell^{2d-1}(\ell^{2d}-1) &= (\ell^{2d^2-3d+1} - 3\ell^{2d^2-3d-1} + \dots)\ell^{2d-1}(\ell^{2d}-1) \\
 &= \ell^{2d^2+d} - 3\ell^{2d^2+d-2} + \dots
 \end{aligned} \tag{6}$$

The exponents of  $\ell$  in the terms in the last line of Eq. (5) are lower than  $2d^2 + d - 2$ . Hence we only have to consider the sum

$$\sum_{i=d-1}^{2d-3} (-1)^i \ell^{\frac{i(i+1)}{2}} T_\ell(d, i+1, i+1-d). \tag{7}$$

The first term of the sum which we obtain for  $i = d - 1$  contains  $T_\ell(d, d, 0)$ . We use (3) to obtain

$$T_\ell(d, d, 0) = (\ell - 1)\ell^{d^2 - \frac{d(d-1)}{2}} \prod_{j=1}^d (\ell^{2j} - 1) \prod_{j=1}^d (\ell^j - 1)^{-1}.$$

Thus we find that the highest exponent of  $\ell$  in the first term of (7) is

$$\frac{d(d-1)}{2} + 1 + d^2 - \frac{d(d-1)}{2} + \frac{d(d+1)}{2} = \frac{d(3d+1)}{2} + 1$$

and for all  $d \geq 3$  this is lower than  $2d^2 + d - 2$ .

For the terms of (7) with  $d \leq i \leq 2d - 3$  we are left to examine  $T_\ell(d, m, s)$  with  $s = m - d > 0$ ,  $m = i + 1$  and  $d + 1 \leq m \leq 2d - 2$ . For this we use (4) in order to find the highest exponent of  $\ell$  in  $T_\ell(d, m, s)$ :

$$\begin{aligned}
 T_\ell(d, m, s) &= \ell^{d^2 - (m-s)^2 - \frac{(m-2s+1)(m-2s)}{2}} \prod_{j=s+1}^d (\ell^{2j} - 1) \prod_{j=1}^{m-2s} (\ell^j - 1)^{-1} \\
 &= \ell^{d^2 - m^2 + 2ms - s^2 - \frac{(m-2s+1)(m-2s)}{2} + d^2 + d - s^2 - s - \frac{(m-2s+1)(m-2s)}{2}} + \dots \\
 &= \ell^{2d^2 + d - 2m^2 + (6s-1)m - 6s^2 + s} + \dots
 \end{aligned}$$

Replacing  $s$  by  $m - d$  we find that the highest exponent of  $\ell$  in the terms of (7) for  $d + 1 \leq m = i + 1 \leq 2d - 2$  is

$$\frac{m(m-1)}{2} + 6md - 4d^2 - 2m^2.$$

The maximum occurs for  $m = 2d - 2$ , i.e.  $i = 2d - 3$ . Thus the highest exponent of  $\ell$  in the terms of (7) for  $d \leq i \leq 2d - 3$  is  $2d^2 - d - 5$  and this is smaller than  $2d^2 + d - 2$  for every  $d$ . Hence if the claim is true for  $d - 1$ , it is true for  $d$  as well.  $\square$

We are now ready to proof our main result (Theorem 1.2 in the introduction). We have to show that  $C_A = \prod_{\ell} a_{\ell}$  is a positive constant.

*Proof (of Theorem 1.2)* There are only finitely many  $\ell$  such that

$$p_{\ell} \neq \frac{e_{\ell}}{\#\mathrm{GSp}_d(\mathbb{F}_{\ell})},$$

hence if  $\ell$  is big enough, we may compute  $p_{\ell}$  using Proposition 2.1 and Corollary 2.4. We find that there is a function  $f(\ell)$  with  $\lim_{\ell \rightarrow \infty} f(\ell) = 1$  such that for big  $\ell$ ,

$$\begin{aligned} p_{\ell} &= \frac{\ell^{2d^2+d} - 3\ell^{2d^2+d-2} + O(\ell^{2d^2+d-3})}{\ell^{2d^2+d+1} - \ell^{2d^2+d} + \Theta(\ell^{2d^2+d-1})} \\ &= \frac{1}{\ell^2} \left( \ell + \frac{\ell^{2d^2+d-1} + O(\ell^{2d^2+d-2})}{\ell^{2d^2+d-1} - \ell^{2d^2+d-2} + \Theta(\ell^{2d^2+d-3})} \right) \\ &= \frac{\ell + f(\ell)}{\ell^2} \end{aligned}$$

and hence

$$a_{\ell} = \frac{1 - p_{\ell}}{1 - \frac{1}{\ell}} = \frac{\ell^2 - \ell - f(\ell)}{\ell^2 - \ell} = 1 - \frac{f(\ell)}{\ell^2 - \ell}.$$

If the infinite sum

$$S = \sum_{\ell} \frac{|f(\ell)|}{\ell^2 - \ell}$$

converges, the product  $C_A = \prod_{\ell} a_{\ell}$  converges as well by [8], Theorem 1. Since  $f(\ell) \rightarrow 1$ , there is an  $\ell_0$  such that  $|f(\ell)| \leq 2$  for all  $\ell \geq \ell_0$ . Let  $S_n$  be the partial

sum of  $S$  taken over all  $\ell \leq n$ . The sequence  $(S_n)$  is monotonous and bounded since

$$\sum_{\ell \geq \ell_0} \frac{|f(\ell)|}{\ell^2 - \ell} \leq 2 \sum_{\ell \geq \ell_0} \frac{1}{\ell^2 - \ell} \leq 2 \sum_{n \geq 2} \frac{1}{n^2 - n} = 2 \sum_{n \geq 2} \left( \frac{1}{n-1} - \frac{1}{n} \right) = 2.$$

Hence  $S$  and  $C_A$  converge. □

We computed approximations of  $C_A$  for some small dimensions  $d$ . We took the product over all primes  $\ell$  up to  $10^9$  (i.e. we assumed that  $G_\ell \cong \text{GSp}_{2d}(\mathbb{F}_\ell)$  for all  $\ell \leq 10^9$ ) and got the following values.

$$\begin{aligned} d = 2 : \quad C_A &\approx 0.6946382884 \\ d = 3 : \quad C_A &\approx 0.6885179362 \\ d = 5 : \quad C_A &\approx 0.6885714948 \end{aligned} \tag{8}$$

We conclude by presenting some examples supporting the conjectural formula (2). We consider some examples with  $d = 2$  and one examples with  $d = 3$ .

All genus 2 curves and their properties are taken from the L-Functions and Modular Forms Database (LMFDB) [7]. Their Jacobians are abelian varieties of dimension 2. The Jacobians have trivial torsion unit subgroup and endomorphism ring  $\mathbb{Z}$ .

*Example 2.5* Consider the Jacobian  $J_1$  of the curve  $D_1$  defined by

$$y^2 + x^3y + xy + y + x^2 + x = 0$$

(curve 587.a.587.1). Bad reduction occurs at 587. We use the computer algebra system Magma [2] to obtain the group order  $N_p = \#J_1(\mathbb{F}_p)$  for all primes  $p \in \{2, \dots, n\} \setminus \{587\}$  where  $n \in \{1000, 2000, 3000, 4000\}$  and count how many of these orders are prime. We denote by  $A_n(J_1)$  the number of prime group orders  $N_p$  with  $p \leq n$ . We do not check whether the representation  $\rho_\ell$ , induced by the action of  $G_\mathbb{Q}$  on the  $\ell$ -torsion points of  $J_1$  is surjective for all primes (recall that the open image theorem only ensures this holds for *almost* all primes), but simply assume that the constant  $C_{J_1}$  is (approximately) 0.6946382884 as computed before. Under this assumption, this constant depends only on the dimension of  $J_1$ , so we write  $C_2$  instead. In Fig. 1 we list the results and the predicted values.

*Example 2.6* Consider the Jacobian  $J_2$  of the curve  $D_2$  defined by

$$y^2 + x^3y + xy + y + x^4 - x^2 = 0$$

(curve 743.a.743.1). Bad reduction occurs at 743. We denote by  $A_n(J_2)$  the number of prime group orders  $N_p$  where  $p \leq n$  is a prime with good reduction. See Fig. 1 for the results.

$n$	$C_2 \frac{n}{2\log(n)^2}$	$C_2 \sum_{p \leq n} \frac{1}{2\log(p)}$	$A_n(J_1)$	$A_n(J_2)$	$A_n(J_3)$	$A_n(J_4)$
1000	7.3	11.1	15	15	5	8
2000	12.0	17.6	20	23	7	12
3000	16.3	23.2	25	28	8	15
4000	20.2	28.3	30	30	13	18

**Fig. 1** Results of Examples 2.5–2.8,  $C_2 = 0.6946382884$

*Example 2.7* Consider the Jacobian  $J_3$  of the curve  $D_3$  defined by

$$y^2 + y - x^5 + 2x^3 - x = 0$$

(curve 971.a.971.1). Bad reduction occurs at 971. We denote by  $A_n(J_3)$  the number of prime group orders  $N_p$  where  $p \leq n$  is a prime with good reduction. See Fig. 1 for the results.

*Example 2.8* Consider the Jacobian  $J_4$  of the curve  $D_4$  defined by

$$y^2 + y - x^5 + x^4 - x^2 + x = 0$$

(curve 1051.a.1051.1). Bad reduction occurs at 1051. We denote by  $A_n(J_4)$  the number of prime group orders  $N_p$  where  $p \leq n$  is a prime with good reduction. See Fig. 1 for the results.

One would expect the second prediction,  $C_2 \sum_{p \leq n} \frac{1}{2\log(p)}$ , to be more accurate than the first prediction,  $C_2 \frac{n}{2\log(n)^2}$ . This is confirmed by the first two examples: Although we considered a rather small number of primes  $p$  (since it quite time-consuming to compute  $N_p$ ), the predictions for  $J_1$  and  $J_2$  are quite good. For  $J_4$  the number of prime orders we counted goes well with the first prediction, however they are a bit small compared to the second, more accurate prediction. For  $J_3$ , the numbers of prime orders we observed are only half of the predicted values. Either the number of primes we considered was limited to obtain reliable results in this examples or the constants  $C_{J_3}$  and  $C_{J_4}$  differ considerably from the constant  $C_2 = 0.6946382884$  we computed in the last section. Recall that when we computed  $C_2$  we assumed that  $G_\ell \cong \text{GSp}_{2d}(\mathbb{F}_\ell)$  for all  $\ell$  up to  $10^9$ . If this assumption does not hold for some small  $\ell$ ,  $C_2$  is not a good approximation for the correct constants  $C_{J_3}$  and  $C_{J_4}$ .

*Example 2.9* Let  $J$  be the Jacobian of the non-hyperelliptic curve  $D$  defined by

$$D : X^3Y - X^2Y^2 + XY^3 - Y^4 + Y^3Z + X^2Z^2 - XYZ^2 - Y^2Z^2 - XZ^3 - YZ^3 = 0.$$

$D$  and  $J$  are defined over  $\mathbb{Q}$ .  $D$  has genus 3, thus  $J$  has dimension 3. Let  $G_{\mathbb{Q}}$  be the absolute Galois group of  $\mathbb{Q}$ . Zywinia [13] has shown that the representation induced by the action of  $G_{\mathbb{Q}}$  on  $\ell$ -torsion points of  $J$ ,

$$\rho_{\ell} : G_{\mathbb{Q}} \rightarrow \mathrm{GSp}_6(\mathbb{F}_{\ell}),$$

is surjective for all primes  $\ell$ . Hence we may use the approximation for  $C_J$  computed in (8).

We used Magma [2] to compute the L-polynomial of  $D \bmod p$  and evaluate it at 1 for all primes numbers  $p \in \{2, \dots, 4300\} \setminus S$ , where  $S = \{7, 11, 83\}$  is the set of those primes where  $D$  and  $J$  have bad reduction. The result is the order  $N_p$  of the Jacobian  $J_p = J \bmod p$ . We found 21 primes such that  $N_p$  is prime. They are listed in Figs. 2 and 3 shows that the conjectured values and the actual numbers of  $p \leq n$  such that  $N_p$  is prime go well together.

$p$	$N_p$ prime	$p$	$N_p$ prime
5	307	2393	13741348627
557	171219319	2447	14776842569
613	242503939	2459	14560673957
637	325438297	2477	15536256587
863	671644577	2503	15936831311
1103	1333585301	2741	20088600289
1297	2192725411	2843	23227460699
1433	2940030209	2887	24174730291
1567	3915860191	3389	38453606281
1597	4109726147	3929	60013471627
2087	9044075741		

Fig. 2 Prime orders

$n$	$C_J \frac{n}{3 \log(n)^2}$	$C_J \sum_{p \leq n} \frac{1}{3 \log(p)}$	# $N_p$ prime with $p \leq n$
1000	4.8	7.6	5
2000	7.9	11.8	10
3000	10.7	15.6	19
4300	14.1	20.0	21

Fig. 3 Comparison of predicted and actual values

## References

1. J.D. Achter, J. Holden, Notes on an analogue of the Fontaine-Mazur conjecture. *J. Théor. Nombr. Bordx.* **15**, 627–637 (2003). <https://doi.org/10.5802/jtnb.416>
2. W. Bosma, J. Cannon, C. Playoust, The Magma algebra system. I. The user language. *J. Symb. Comput.* **24**(3–4), 235–265 (1997)
3. J. Fulman, A probabilistic approach to conjugacy classes in the finite symplectic orthogonal groups. *J. Algebra* **234**, 207–224 (2000). <https://doi.org/10.1006/jabr.2000.8455>
4. N. Koblitz, Primality of the number of points on an elliptic curve over a finite field. *Pac. J. Math.* **131**(1), 157–165 (1988). <https://doi.org/10.2140/pjm.1988.131.157>
5. J.-P. Serre, Lettre à Marie-France Vignéras du 10/2/1986, in *Œuvres, Collected Papers*, vol. IV (Springer, Berlin, 2000), pp. 38–55
6. U. Spreckels, On the order of CM abelian varieties over finite prime fields. *Finite Fields Appl.* **45**, 386–405 (2017). <https://doi.org/10.1016/j.ffa.2017.01.004>
7. The LMFDB Collaboration, The l-functions and modular forms database (2016), <http://www.lmfdb.org>. Accessed 27 Sept 2016
8. W.F. Trench, Conditional convergence of infinite products. *Am. Math. Mon.* **106**, 646–651 (1999). <https://doi.org/10.2307/2589494>
9. L.C. Washington, Some remarks on Cohen-Lenstra heuristics. *Math. Comp.* **47**(176), 741–474 (1986). <https://doi.org/10.2307/2008187>
10. A. Weng, On the order of abelian varieties with trivial endomorphism ring reduced modulo a prime. *Finite Fields Appl.* **28**, 115–122 (2014). <https://doi.org/10.1016/j.ffa.2014.01.002>
11. A. Weng, On the order of abelian surfaces of CM-type over finite prime fields. *Quaest. Math.* **38**(6), 771–787 (2015). <https://doi.org/10.2989/16073606.2014.981720>
12. D. Zywina, A refinement of Koblitz’s conjecture. *Int. J. Number Theor.* **7**(3), 739–769 (2011). <https://doi.org/10.1142/S1793042111004411>
13. D. Zywina, An explicit Jacobian of dimension 3 with maximal Galois action (2015), <http://www.math.cornell.edu/~zywina/papers/GSp6example.pdf>. Accessed 15 Jul 2015

# Chabauty Without the Mordell-Weil Group



Michael Stoll

**Abstract** Based on ideas from recent joint work with Bjorn Poonen, we describe an algorithm that can in certain cases determine the set of rational points on a curve  $C$ , given only the  $p$ -Selmer group  $S$  of its Jacobian (or some other abelian variety  $C$  maps to) and the image of the  $p$ -Selmer set of  $C$  in  $S$ . The method is more likely to succeed when the genus is large, which is when it is usually rather difficult to obtain generators of a finite-index subgroup of the Mordell-Weil group, which one would need to apply Chabauty's method in the usual way. We give some applications, for example to generalized Fermat equations of the form  $x^5 + y^5 = z^p$ .

**Keywords** Rational points on curves • Chabauty's method • Selmer group

**Subject Classifications** 11G30, 14G05, 14G25, 14H25, 11Y50, 11D41

## 1 Introduction

When one is faced with the task of determining the set of rational points on a (say) hyperelliptic curve  $C: y^2 = f(x)$ , then the usual way to proceed is in the following steps. For the following discussion, we assume that  $f$  has odd degree, which implies that there is a rational point at infinity on  $C$ , which eliminates possible shortcuts that can be used to show that a curve does not have any rational points. We denote the Jacobian variety of  $C$  by  $J$ .

### 1. Search for rational points on $C$ .

This can be done reasonably efficiently for  $x$ -coordinates whose numerator and denominator are at most  $10^5$ , say. Rational points on curves of genus  $\geq 2$  are expected to be fairly small (in relation to the coefficients of the defining

---

M. Stoll (✉)

Mathematisches Institut, Universität Bayreuth, 95440 Bayreuth, Germany  
e-mail: [Michael.Stoll@uni-bayreuth.de](mailto:Michael.Stoll@uni-bayreuth.de)



equation), so the result very likely is  $C(\mathbb{Q})$ . It remains to show that we have not overlooked any points.

2. Compute the 2-Selmer group  $\text{Sel}_2 J$  [14].

The ‘global’ part of this computation requires arithmetic information related to class group and unit group data for the number fields generated by the roots of  $f$ . If the degrees of the irreducible factors of  $f$  are not too large (and the coefficients are of moderate size), then this computation is feasible in many cases, possibly assuming the Generalized Riemann Hypothesis (GRH in the following) to speed up the class group computation. The ‘local’ part of the computation is fairly easy for the infinite place and the odd finite places, but it can be quite involved to find a basis of  $J(\mathbb{Q}_2)/2J(\mathbb{Q}_2)$ .

To proceed further, we need the resulting bound  $r$  for the rank of  $J(\mathbb{Q})$ ,

$$r = \dim_{\mathbb{F}_2} \text{Sel}_2 J - \dim_{\mathbb{F}_2} J(\mathbb{Q})[2] ,$$

to be strictly less than the genus  $g$  of  $C$ . By work of Bhargava and Gross [2] it is known that the Selmer group is small on average, independent of the genus, so when  $g$  is not very small, this condition is likely to be satisfied.

3. Find  $r$  independent points in  $J(\mathbb{Q})$ .

We can use the points on  $C$  we have found in Step 1 to get some points in  $J(\mathbb{Q})$ . However, it can be quite hard to find further points if the points we get from the curve generate a subgroup of rank  $< r$ . There are two potential problems. The first is of theoretical nature: the rank of  $J(\mathbb{Q})$  can be strictly smaller than  $r$ , in which case it is obviously impossible to find  $r$  independent points. Standard conjectures imply that the difference between  $r$  and the rank is even, so we will not be in this situation when we are missing just one point. In any case, if we suspect our bound is not tight, we can try to use visualization techniques [4] to improve the bound. The second problem is practical: some of the generators of  $J(\mathbb{Q})$  can have fairly large height and are therefore likely to fall outside our search space. When the genus  $g$  is moderately large, then we also have the very basic problem that the dimension of our search space is large.

To proceed further, we need to know generators of a finite-index subgroup  $G$  of  $J(\mathbb{Q})$ .

4. Fix some (preferably small) prime  $p$  (preferably of good reduction) and use the knowledge of  $G$  to compute a basis of the space  $V$  of  $\mathbb{Q}_p$ -defined regular differentials on  $C$  that kill the Mordell-Weil group  $J(\mathbb{Q})$  under the Chabauty-Coleman pairing (see for example [15]).

This requires evaluating a number of  $p$ -adic abelian integrals on  $C$ , which (in the case of good reduction with  $p$  odd) can be done by an algorithm due to Bradshaw and Kedlaya and made practical by Balakrishnan [1]. Alternatively, one can use the group structure to reduce to the computation of ‘tiny’ integrals, which can be evaluated using power series.

5. Find the common zeros of the functions  $P \mapsto \int_{\infty}^P \omega$  on  $C(\mathbb{Q}_p)$ , where  $\omega$  runs through a basis of  $V$ .

The rational points are among this set. If there are additional zeros, then they can usually be excluded by an application of the Mordell-Weil sieve [6].

The most serious stumbling block is Step 3, in particular when the genus  $g$  is of ‘medium’ size (say between 5 and 15). In this case Step 2 is feasible, but we are likely to run into problems when trying to find sufficiently many independent points in the Mordell-Weil group.

In this paper we propose an approach that circumvents this problem. Its great advantage is that it uses only the 2-Selmer group and data that can be obtained by a purely 2-adic computation. Its disadvantage is that it may fail: for it to work, several conditions have to be satisfied, which, however, are likely to hold, in particular when the genus gets large.

Generally speaking, the method tries to use the ideas of [11] (where it is shown that many curves as above have the point at infinity as their only rational point) to deal with given concrete curves. Section 2 gives a slightly more flexible version of one of the relevant results of this paper. In Sect. 4, we formulate the algorithm for hyperelliptic curves of odd degree that is based on this key result. The method will apply in other situations as well (whenever we are able to compute a suitable Selmer group), and we plan to work this out in more detail in a follow-up paper for the case of general hyperelliptic curves and also for the setting of ‘Elliptic Curve Chabauty’, where one wants to find the set of  $k$ -points  $P$  on an elliptic curve  $E$  defined over a number field  $k$  such that  $f(P) \in \mathbb{P}^1(\mathbb{Q})$ , where  $f: E \rightarrow \mathbb{P}^1$  is a non-constant  $k$ -morphism. One application in the latter setting is given at the end of this paper. The approach has also already been applied in [8] to complete the resolution of the Generalized Fermat Equation  $x^2 + y^3 = z^{11}$ .

One ingredient of the algorithm is the computation of ‘halves’ of points in the group  $J(\mathbb{Q}_2)$ . In Sect. 5 we give a general procedure for doing this in  $J(k)$ , when  $J$  is the Jacobian of an odd degree hyperelliptic curve and  $k$  is any field not of characteristic 2. In Sect. 6, we demonstrate the usefulness of our approach by showing that the only integral solutions of  $y^2 - y = x^{21} - x$  are the obvious ones.

In Sect. 7, we show how our method leads to a fairly simple criterion that implies the validity of Fermat’s Last Theorem for a given prime exponent. This does not lead to any new results, of course, but it gives a nice illustration of the power of the method. In Sect. 8, we then apply our approach to the curves  $5y^2 = 4x^p + 1$ . Carrying out the computations, we can show that the only rational points on these curves are the three obvious ones, namely  $\infty, (1, 1)$  and  $(1, -1)$ , when  $p$  is a prime  $\leq 53$  (assuming GRH for  $p \geq 23$ ). A result due to Dahmen and Siksek [7] then implies that the only coprime integer solutions of the Generalized Fermat Equation

$$x^5 + y^5 = z^p$$

are the trivial ones (where  $xyz = 0$ ).

As already mentioned, we end this paper with another type of example, which uses the method in the context of ‘Elliptic Curve Chabauty’ to show that a certain hyperelliptic curve of genus 4 over  $\mathbb{Q}$  has only the obvious pair of rational points.

The Mordell-Weil rank is 4 in this case, so no variant of Chabauty’s method applies directly to the curve.

## 2 The Algorithm

In this section we formulate and prove a variant of [11, Prop. 6.2]. We then use it to give an algorithm that can show that the set of known rational points in some subset  $X$  of the  $p$ -adic points of a curve already consists of all rational points contained in  $X$ , using as input only the  $p$ -Selmer group of the Jacobian of the curve. The idea behind this goes back to McCallum’s paper [10].

Let  $k$  be a number field, let  $C/k$  be a nice (meaning smooth, projective and geometrically irreducible) curve of genus  $g \geq 2$  and let  $A/k$  be an abelian variety, together with a map  $i: C \rightarrow A$  such that  $A$  is generated by the image of  $C$  (for example,  $A$  could be the Jacobian of  $C$  and  $i$  the embedding given by taking some  $k$ -rational point  $P_0 \in C(k)$  as basepoint). Fix a prime number  $p$ . We write  $\text{Sel}_p A$  for the  $p$ -Selmer group of  $A$ . Recall that this is defined as the kernel of the diagonal homomorphism in the commuting diagram with exact rows

$$\begin{array}{ccccccc}
 0 & \longrightarrow & \frac{A(k)}{pA(k)} & \xrightarrow{\delta} & H^1(k, A[p]) & \longrightarrow & H^1(k, A)[p] \longrightarrow 0 \\
 & & \downarrow & & \downarrow & \searrow & \downarrow \\
 0 & \longrightarrow & \prod_v \frac{A(k_v)}{pA(k_v)} & \xrightarrow{\delta} & \prod_v H^1(k_v, A[p]) & \longrightarrow & \prod_v H^1(k_v, A)[p] \longrightarrow 0
 \end{array}$$

that is induced by applying Galois cohomology to the short exact sequence

$$0 \longrightarrow A[p] \longrightarrow A \xrightarrow{ip} A \longrightarrow 0$$

of Galois modules over  $k$  and over all completions  $k_v$  of  $k$ , so the products in the second row run over all places  $v$  of  $k$ . The vertical maps are induced by  $k \hookrightarrow k_v$ . In particular, for each place  $v$  there is a canonical map  $\text{Sel}_p(A) \rightarrow A(k_v)/pA(k_v)$ .

We write  $k_p = k \otimes_{\mathbb{Q}} \mathbb{Q}_p$ ; this is the product of the various completions of  $k$  at places above  $p$ . The set  $C(k_p)$  and the group  $A(k_p)$  can similarly be understood as products of the sets or groups of  $k_v$ -points, for the various  $v \mid p$ . The inclusion  $k \hookrightarrow k_p$  induces natural maps  $C(k) \hookrightarrow C(k_p)$  and  $A(k) \hookrightarrow A(k_p)$ . Let  $X \subseteq C(k_p)$  be a subset (for example, the points in a product of  $v$ -adic residue disks). We then have the following commutative diagram of maps.

$$\begin{array}{ccccccc}
 C(k) \cap X \hookrightarrow & C(k) & \xrightarrow{i} & A(k) & \xrightarrow{\pi} & \frac{A(k)}{pA(k)} & \xrightarrow{\delta} \text{Sel}_p A \\
 \downarrow & \downarrow & & \downarrow & & \downarrow & \swarrow \sigma \\
 X \hookrightarrow & C(k_p) & \xrightarrow{i} & A(k_p) & \xrightarrow{\pi_p} & \frac{A(k_p)}{pA(k_p)} & 
 \end{array}$$

We introduce some more notation. For  $P \in A(k_p)$ , we set

$$q(P) := \{ \pi_p(Q) : Q \in A(k_p), \exists n \geq 0 : p^n Q = P \} \subseteq \frac{A(k_p)}{pA(k_p)}, \tag{1}$$

and for a subset  $S \subseteq A(k_p)$ , we set  $q(S) = \bigcup_{P \in S} q(P)$ . We further define

$$\tau(P) := \sup\{n : n \geq 0, P \in p^n A(k_p)\} \in \mathbb{Z}_{\geq 0} \cup \{\infty\}.$$

Note that  $\tau(P) = \infty$  is equivalent to  $P$  having finite order prime to  $p$ ; on the complement of the finite set consisting of such  $P$ , the quantities  $\tau$  and  $q$  are locally constant.

With a view toward further applications, we first state a more general version of our result, which we will then specialize (see Theorem 2.6 below). We remark that  $C$  could also be a variety of higher dimension here.

**Theorem 2.1** *In the situation described above, fix some subgroup  $\Gamma \subseteq A(k)$  and assume that*

- (1)  $\ker \sigma \subseteq \delta(\pi(\Gamma))$ , and that
- (2)  $q(i(X) + \Gamma) \cap \text{im}(\sigma) \subseteq \pi_p(\Gamma)$ .

Then  $i(X \cap C(k)) \subseteq \bar{\Gamma} := \{Q \in A(k) : \exists n \geq 1 : nQ \in \Gamma\}$ .

*Proof* Let  $P \in X \cap C(k)$ . We show by induction on  $n$  that for each  $n \geq 0$ , there are  $T_n \in \Gamma$  and  $Q_n \in A(k)$  such that  $i(P) = T_n + p^n Q_n$ . This is clear for  $n = 0$  (take  $T_0 = 0$  and  $Q_0 = i(P)$ ). Now assume that  $T_n$  and  $Q_n$  exist. Note that  $\pi_p(Q_n) \in q(i(P) - T_n)$ , so

$$\pi_p(Q_n) = \sigma(\delta(\pi(Q_n))) \in q(i(X) + \Gamma) \cap \text{im}(\sigma);$$

by assumption (2) this implies  $\sigma(\delta(\pi(Q_n))) \in \pi_p(\Gamma) = \sigma(\delta(\pi(\Gamma)))$ . This shows that  $Q_n \in \Gamma + \ker(\sigma \circ \delta \circ \pi)$ . By assumption (1) and since  $\delta$  is injective, we have

$$\ker(\sigma \circ \delta \circ \pi) = \pi^{-1}(\delta^{-1}(\ker \sigma)) \subseteq \pi^{-1}(\pi(\Gamma)) = \Gamma + \ker \pi = \Gamma + pA(k),$$

which implies that  $Q_n \in \Gamma + pA(k)$ . So there are  $T' \in \Gamma$  and  $Q_{n+1} \in A(k)$  such that  $Q_n = T' + pQ_{n+1}$ . We set  $T_{n+1} = T_n + p^n T' \in \Gamma$ ; then

$$i(P) = T_n + p^n Q_n = T_n + p^n(T' + pQ_{n+1}) = T_{n+1} + p^{n+1}Q_{n+1} .$$

Now consider the quotient map  $\psi: A(k) \twoheadrightarrow A(k)/\bar{\Gamma}$ . Since  $\bar{\Gamma}$  is saturated in the finitely generated group  $A(k)$ , the quotient group is torsion free and hence free. Observe that for every  $n \geq 0$ ,

$$\psi(i(P)) = \psi(T_n + p^n Q_n) = \psi(T_n) + p^n \psi(Q_n) = p^n \psi(Q_n) \in p^n(A(k)/\bar{\Gamma}) ,$$

which implies that  $\psi(i(P)) = 0$  and so  $i(P) \in \bar{\Gamma}$ . □

The point of formulating the statement in this way (as compared to [11]) is that we avoid the use of  $p$ -adic abelian logarithms, which would require us to compute  $p$ -adic abelian integrals, usually with  $p = 2$  and in a situation when the curve has bad reduction at 2. Instead, we need to be able to compute  $q(P)$  for a given point  $P$ , which comes down to finding its  $p$ -division points. At least in some cases of interest, this approach seems to be computationally preferable.

*Remark 2.2* Instead of considering multiplication by  $p$ , we could use an endomorphism  $\psi$  of  $A$  that is an isogeny of degree a power of  $p$  and such that some power of  $\psi$  is divisible by  $p$  in the endomorphism ring of  $A$ . We then consider  $A(k)/\psi A(k)$ ,  $A(k_p)/\psi A(k_p)$  and the  $\psi$ -Selmer group  $\text{Sel}_\psi A$ . Note that when  $\psi: A \rightarrow A'$  is any isogeny whose kernel has order a power of  $p$  and with dual isogeny  $\hat{\psi}$ , then we can consider  $A \times A'$  with the endomorphism  $\tilde{\psi}: (P, P') \mapsto (\hat{\psi}(P'), \psi(P))$ , which satisfies  $\tilde{\psi}^2 = \deg \psi = p^m$ , together with the morphism  $\tilde{\iota}: X \rightarrow A \times A', P \mapsto (i(P), 0)$ . Taking  $\Gamma \times \{0\}$  in place of  $\Gamma$  and writing the relevant maps as

$$A(k) \xrightarrow{\pi} \frac{A(k)}{\hat{\psi}(A'(k))} \xrightarrow{\subset \delta} \text{Sel}_{\hat{\psi}} A' \xrightarrow{\sigma} \frac{A(k_p)}{\hat{\psi}(A'(k_p))}$$

and

$$A'(k) \xrightarrow{\pi'} \frac{A'(k)}{\psi(A(k))} \xrightarrow{\subset \delta'} \text{Sel}_\psi A \xrightarrow{\sigma'} \frac{A'(k_p)}{\psi(A(k_p))} ,$$

the second condition in Theorem 2.1 translates into

$$q_A(i(X) + \Gamma) \cap \text{im}(\sigma) \subseteq \sigma(\delta(\pi(\Gamma)))$$

and

$$q_{A'}(\hat{\psi}^{-1}(i(X)) + A'(k)_{\text{tors}}) \cap \text{im}(\sigma') \subseteq \sigma'(\delta'(\pi'(A'(k)_{\text{tors}}))) .$$

*Remark 2.3* The set  $X \cap i^{-1}(\bar{\Gamma})$  that contains  $C(k) \cap X$  when Theorem 2.1 applies can in many cases be determined by the usual Chabauty-Coleman techniques; see for example [15]. Of course, if  $\Gamma$  is finite (and so  $\bar{\Gamma} = A(k)_{\text{tors}}$  is finite as well), which is usually the case in applications, then determining  $X \cap i^{-1}(\bar{\Gamma})$  is essentially trivial.

We give some indication of how one can compute a set such as  $q(P + \Gamma)$ , where  $P \in A(k_p)$ . We assume that, given  $P \in A(k_p)$ , we can find all  $Q \in A(k_p)$  such that  $pQ = P$ .

**Lemma 2.4** *With the notations used in Theorem 2.1, fix a complete set of representatives  $R \subseteq \Gamma$  for  $\Gamma/p\Gamma$ . Let  $P \in A(k_p)$  and set*

$$\mathcal{Q} = \{Q \in A(k_p) : \exists T \in R: pQ = P + T\} .$$

*Define an equivalence relation on  $\mathcal{Q}$  via  $Q \sim Q' \iff Q - Q' \in \Gamma$ , and let  $\mathcal{Q}'$  be a complete set of representatives for  $\mathcal{Q}/\sim$ . Then*

$$q(P + \Gamma) = \{\pi_p(P + T) : T \in R\} \cup \bigcup_{Q \in \mathcal{Q}'} q(Q + \Gamma) .$$

*Proof* Since  $Q + \Gamma = Q' + \Gamma$  whenever  $Q \sim Q'$ , it is sufficient to prove the equality with  $\mathcal{Q}$  in place of  $\mathcal{Q}'$ . We first show that the set on the right is contained in the set on the left. This is clear for the elements  $\pi_p(P + T)$ , taking  $n = 0$  in (1). So let now  $Q \in \mathcal{Q}$  and  $\xi \in q(Q + \Gamma)$ . Then there are  $n \geq 0$ ,  $T' \in \Gamma$  and  $Q' \in A(k_p)$  such that  $p^n Q' = Q + T'$  and  $\pi_p(Q') = \xi$ . There is also  $T \in \Gamma$  such that  $pQ = P + T$ . We then have

$$p^{n+1}Q' = p(Q + T') = P + (T + pT') \in P + \Gamma$$

and so  $\xi = \pi_p(Q') \in q(P + \Gamma)$ .

Now we show the reverse inclusion. Let  $\xi \in q(P + \Gamma)$ , so there are  $n \geq 0$ ,  $T' \in \Gamma$ ,  $Q' \in A(k_p)$  such that  $p^n Q' = P + T'$  and  $\pi_p(Q') = \xi$ . There is also some  $T \in R$  such that  $T - T' = pT''$  with  $T'' \in \Gamma$ . If  $n = 0$ , then  $\xi = \pi_p(P + T') = \pi_p(P + T)$ . If  $n > 0$ , we can write

$$P + T = (P + T') + pT'' = p(p^{n-1}Q' + T'') = pQ$$

with  $Q = p^{n-1}Q' + T'' \in \mathcal{Q}$ , and  $\xi = \pi_p(Q') \in q(Q - T'') \subseteq q(Q + \Gamma)$ . □

Whenever  $\sup \tau(P + \Gamma) < \infty$ , the recursion implied by Lemma 2.4 will terminate, and so the lemma translates into an algorithm for computing  $q(P + \Gamma)$ . We make this condition more explicit.

**Lemma 2.5** *Write  $\text{cl}(\Gamma)$  for the topological closure of  $\Gamma$  in  $A(k_p)$ . Let  $P \in A(k_p)$ . Then  $\sup \tau(P + \Gamma) = \infty$  if and only if there is a point  $T \in A(k_p)$  of finite order prime to  $p$  such that  $P \in T + \text{cl}(\Gamma)$ .*

*Proof* Let  $A(k_p)_1$  be the kernel of reduction (i.e., the product of the kernels of reduction of the various  $A(k_v)$  with  $v$  a place above  $p$ ) and let  $m$  denote the exponent of the finite group  $A(k_p)/A(k_p)_1$ . Then for all  $P \in A(k_p)$ ,  $p^n mP$  tends to the origin as  $n \rightarrow \infty$ . If  $\sup \tau(P + \Gamma) = \infty$ , then there are arbitrarily large  $n$  such there exist  $\gamma_n \in \Gamma$  and  $Q_n \in A(k_p)$  with  $P = \gamma_n + p^n Q_n$ , so  $mP - m\gamma_n$  tends to the origin as  $n$  gets large. Then  $P - \gamma_n$  must be close to a point of order  $m$ ; by restricting to a sub-sequence, we find that  $P - \gamma_n$  approaches a point  $T \in A(k_p)[m]$ . Since  $T$  is close to  $P - \gamma_n = p^n Q$  for arbitrarily large  $n$ ,  $T$  must be infinitely divisible by  $p$ , so the order of  $T$  is prime to  $p$ . We clearly have  $P \in T + \text{cl}(\Gamma)$ .

For the converse, it suffices to consider  $P$  in the closure of  $\Gamma$ , since the points of finite order prime to  $p$  in  $A(k_p)$  are infinitely  $p$ -divisible. Since any point sufficiently close to the origin is highly  $p$ -divisible, this implies that for each  $n \geq 0$  we can find  $\gamma_n \in \Gamma$  and  $Q_n \in A(k_p)$  such that  $P - \gamma_n = p^n Q_n$ . This implies that  $\sup \tau(P + \Gamma) = \infty$ . □

We specialize Theorem 2.1 to the case that  $k = \mathbb{Q}$  and  $i$  embeds the curve into its Jacobian. Let  $\mathcal{C}$  be a proper regular model of  $C$  over  $\mathbb{Z}_p$ . Then the reduction map sends  $C(\mathbb{Q}_p) = \mathcal{C}(\mathbb{Z}_p)$  to the set of smooth  $\mathbb{F}_p$ -points on the special fiber of  $\mathcal{C}$ . The preimage  $D$  of a smooth  $\mathbb{F}_p$ -point on the special fiber of  $\mathcal{C}$  under the reduction map is called a *residue disk* in  $C(\mathbb{Q}_p)$ ; see [11]. It follows from Hensel’s Lemma that there is an analytic map  $\varphi$  from the open  $p$ -adic unit disk to  $C$  such that  $D = \varphi(p\mathbb{Z}_p)$ . If  $p = 2$ , then we call the subsets  $\varphi(4\mathbb{Z}_2)$  and  $\varphi(2 + 4\mathbb{Z}_2)$  *half residue disks*.

**Theorem 2.6** *Let  $C$  be a nice curve over  $\mathbb{Q}$ , with Jacobian  $J$ . Let  $P_0 \in C(\mathbb{Q})$  and take  $X \subseteq C(\mathbb{Q}_p)$  to be contained in a residue disk or, when  $p = 2$  and  $J(\mathbb{Q})[2] \neq 0$ , in a half residue disk, and to contain  $P_0$ . Let  $i: C \rightarrow J$  be the embedding sending  $P_0$  to zero. With the notation introduced above, assume that*

- (1)  $\ker \sigma \subseteq \delta(\pi(J(\mathbb{Q})_{\text{tors}}))$ , and that
- (2)  $q(i(X) + J(\mathbb{Q})_{\text{tors}}) \cap \text{im}(\sigma) \subseteq \pi_p(J(\mathbb{Q})_{\text{tors}})$ .

Then  $C(\mathbb{Q}) \cap X = \{P_0\}$ .

*Proof* We apply Theorem 2.1 with  $k = \mathbb{Q}$ ,  $C$  our curve,  $A = J$ ,  $i$  as given in the statement and  $\Gamma = \bar{\Gamma} = J(\mathbb{Q})_{\text{tors}}$ . This tells us that  $i(C(\mathbb{Q}) \cap X) \subseteq J(\mathbb{Q})_{\text{tors}}$ . If  $p > 2$ , or  $p = 2$  and  $J(\mathbb{Q})[2] = 0$ , then the only rational torsion point in the kernel of reduction of  $J(\mathbb{Q}_p)$  is the origin, which implies that there cannot be two distinct points in  $X$  both mapping to torsion under  $i$ . If  $p = 2$  and  $J(\mathbb{Q})[2] \neq 0$ , then the corresponding statement is true if  $X$  is a half residue disk, which means that  $i(X)$  is contained in  $K_2$ , the second kernel of reduction; see Sect. 3 below. In both cases, we find that there is at most one rational point in  $X$ ; since  $P_0$  is one such point, it must be the only one. □

This leads to the following algorithm. It either returns **FAIL** or it returns the set of rational points on the curve  $C$ . We refer to [16] for the definition of the  $p$ -Selmer set  $\text{Sel}_p(C)$  of the curve  $C$ . Given an embedding  $i$  of  $C$  into its Jacobian  $J$ , it can be interpreted as the subset of  $\text{Sel}_p(J)$  consisting of elements that locally come from points on the curve.

**Algorithm 2.7**

**Input:** A nice curve  $C$ , defined over  $\mathbb{Q}$ , with Jacobian  $J$ .

A point  $P_0 \in C(\mathbb{Q})$ , defining an embedding  $i: C \rightarrow J$ .

A prime number  $p$ .

**Output:** The set of rational points on  $C$ , or **FAIL**.

1. Compute the  $p$ -Selmer group  $\text{Sel}_p J$  and the  $p$ -Selmer set  $\text{Sel}_p C$ ;  $i$  induces a map  $i_*: \text{Sel}_p C \hookrightarrow \text{Sel}_p J$ .
2. Search for rational points on  $C$  and collect them in a set  $C(\mathbb{Q})_{\text{known}}$ .
3. Let  $\sigma: \text{Sel}_p J \rightarrow J(\mathbb{Q}_p)/pJ(\mathbb{Q}_p)$  be the canonical map. If  $\ker \sigma \not\subseteq \delta(\pi(J(\mathbb{Q})_{\text{tors}}))$ , then return **FAIL**.
4. Let  $R$  be the image of  $J(\mathbb{Q})_{\text{tors}}$  in  $J(\mathbb{Q}_p)/pJ(\mathbb{Q}_p)$ .
5. Let  $\mathcal{X}$  be a partition of  $C(\mathbb{Q}_p)$  into residue disks whose image in  $J(\mathbb{Q}_p)/pJ(\mathbb{Q}_p)$  consists of one element and that are contained in half residue disks when  $p = 2$  and  $J(\mathbb{Q})[2] \neq 0$ .
6. For each  $X \in \mathcal{X}$  do the following:
  - a. If  $X \cap C(\mathbb{Q})_{\text{known}} = \emptyset$ :  
If  $\pi_p(X) \subseteq \text{im}(\sigma \circ i_*)$ , then return **FAIL**;  
otherwise continue with the next  $X$ .
  - b. Pick some  $P_1 \in C(\mathbb{Q})_{\text{known}} \cap X$ .
  - c. Compute  $Y = \bigcup_{P \in X, T \in J(\mathbb{Q})_{\text{tors}}} q([P - P_1] + T) \subseteq J(\mathbb{Q}_p)/pJ(\mathbb{Q}_p)$ .
  - d. If  $Y \cap \text{im}(\sigma) \not\subseteq R$ , then return **FAIL**.
7. Return  $C(\mathbb{Q})_{\text{known}}$ .

**Proposition 2.8** *The algorithm is correct: if it does not return FAIL, then it returns the set of rational points on  $C$ .*

*Proof* First note that Step 3 verifies the first assumption of Theorem 2.6; it returns **FAIL** when the assumption does not hold. It is also clear that if the algorithm does not return **FAIL**, then the set it returns is a subset of  $C(\mathbb{Q})$ . We show the reverse inclusion. So let  $P \in C(\mathbb{Q})$  be some rational point. There will be some  $X \in \mathcal{X}$  such that  $P \in X$ . Then  $\pi_p(X)$  is contained in  $\text{im}(\sigma \circ i_*)$ , so since the algorithm did not return **FAIL**, by Step 6a. it follows that  $X \cap C(\mathbb{Q})_{\text{known}} \neq \emptyset$ ; let  $P_1 \in X \cap C(\mathbb{Q})_{\text{known}}$  as in Step 6b. Now by Step 6d. the second assumption of Theorem 2.6 is satisfied, taking the embedding with base-point  $P_1$ . So the theorem applies, and it shows that there is only one rational point in  $X$ , so  $P = P_1 \in C(\mathbb{Q})_{\text{known}}$ . □

*Remark 2.9* We note that in Step 6, the set  $X$  can be further partitioned if necessary. If there are several points in  $C(\mathbb{Q})_{\text{known}}$  that end up in the same set  $X$ , then the second assumption of Theorem 2.6 cannot be satisfied. But it is still possible that the theorem can be applied to smaller disks that separate the points. (If the points are too close  $p$ -adically, this will not work, though. In this case, one could try to use  $\Gamma + J(\mathbb{Q})_{\text{tors}}$  in the more general version of the theorem, where  $\Gamma$  is the subgroup generated by the difference of the two points.)

There are also cases when it helps to combine several sets  $X$  into one. One such situation is when there are points in  $C(\mathbb{Q}_p)$  that differ by a torsion point of order



prime to  $p$  and such that only one of the corresponding sets  $X$  contains a (known) rational point.

A particularly useful case is when  $C$  is hyperelliptic,  $A = J$  is the Jacobian of  $C$ , and we consider  $p = 2$ . There is an algorithm that computes 2-Selmer group  $\text{Sel}_2 J$ , which is feasible in many cases, compare [14]. We discuss this further in Sect. 4 below in the case when the curve has a rational Weierstrass point at infinity.

Another useful case (using a slightly more general setting) is related to “Elliptic curve Chabauty”. Here  $A$  is the Weil restriction of an elliptic curve  $E$  over some number field  $k$  such that there is a non-constant morphism  $C \rightarrow E$  defined over  $k$ . We give an example of this in Sect. 9.

### 3 Computing the Image Under $q$ of a Disk

In this section, we discuss in some detail how to find the image under  $q$  of (the image in  $J$  of) a residue disk of  $C(k_p)$ . The basic idea is that  $q$  is locally constant on the curve even near points where  $\tau$  becomes infinite (a variant of this was already used in [11]). To get a practical algorithm out of this idea, we have to produce an explicit neighborhood on which  $q$  is constant. We will do this first away from the points where  $\tau$  becomes infinite and then also on residue disks centered at a point where  $\tau$  becomes infinite.

Since objects over  $k_p$  are products of objects over the various completions  $k_v$  at places  $v$  above  $p$ , we will now work over a fixed such completion. We fix a non-constant morphism  $i: C \rightarrow J$ , where  $J$  can be any abelian variety that is spanned by  $i(C)$ . To ease notation, we denote the map  $J(k_v) \rightarrow J(k_v)/pJ(k_v)$  by  $\pi$  instead of  $\pi_v$ .

We assume that we can compute  $q(P)$  for any given point  $P \in J(k_v)$  that is not (too close to) a point of finite order prime to  $p$ . When  $p = 2$  and  $C$  is hyperelliptic of odd degree and  $J$  is the Jacobian, this can be done by using the halving algorithm of Sect. 5 below: we compute the image of  $P$  in  $L_2^\square$  and record it; if the image is trivial, then we compute all halves of  $P$  and apply the same procedure to them. Since by assumption  $P$  is not infinitely 2-divisible, the recursion will eventually stop with an empty set of points still to be considered.

The following is essentially immediate from the definitions.

**Lemma 3.1** *Let  $P_1, P_2 \in J(k_v)$  and assume that  $P_1 \equiv P_2 \not\equiv 0 \pmod{p^{m+1}J(k_v)}$ . Then  $\tau(P_1) = \tau(P_2)$  and  $q(P_1) = q(P_2)$ .*

*Proof* The assumptions imply that  $P_1, P_2 \notin p^{m+1}J(k_v)$ , so whenever there are  $Q \in J(k_v)$  and  $n \geq 0$  such that  $p^n Q = P_1$  or  $P_2$ , then  $n \leq m$ . Let  $P' \in J(k_v)$  such that  $P_2 = P_1 + p^{m+1}P'$ . Then  $p^n Q = P_1$  implies  $p^n(Q + p^{m+1-n}P') = P_2$ , so that  $\tau(P_2) \geq \tau(P_1)$ , and by symmetry, we obtain equality.

Let  $\xi \in q(P_1)$ ; then  $\xi = \pi(Q)$  for some  $Q$  such that  $p^n Q = P_1$  as above. Then  $n \leq m$  and so  $\xi = \pi(Q) = \pi(Q + p^{m+1-n}P') \in q(P_2)$  as well. This shows that  $q(P_1) \subseteq q(P_2)$ ; the reverse inclusion follows again by symmetry.  $\square$

We write  $\mathcal{O}_v$  for the ring of integers in  $k_v$  and  $\varpi$  for a uniformizer. We abuse notation and write  $v: k_v^\times \rightarrow \mathbb{Z}$  for the additive valuation, normalized such that  $v(\varpi) = 1$ . Then  $e = v(p)$  is the absolute ramification index of  $K_v$ . We fix a proper regular model  $\mathcal{C}$  of  $C$  over  $\mathcal{O}_v$ . Let  $\mathcal{J}$  be the Néron model of  $J$  over  $\mathcal{O}_v$ . For  $n \geq 1$ , we denote by

$$K_n := \ker(J(k_v) = \mathcal{J}(\mathcal{O}_v) \rightarrow \mathcal{J}(\mathcal{O}_v/\varpi^n \mathcal{O}_v))$$

the ‘higher kernels of reduction’;  $K_n$  is also the group of  $\varpi^n \mathcal{O}_v$ -points of the formal group associated to  $\mathcal{J}$ .

We now fix a residue disk  $D \subseteq C(K_v)$  with respect to  $\mathcal{C}$ ; we will denote an analytic parameterization  $D_0 \rightarrow D$  by  $\varphi$ , where  $D_0$  is the open unit disk. Since  $i$  induces a morphism from the smooth part of  $\mathcal{C}$  to  $\mathcal{J}$ , it follows that

$$t, t' \in \varpi \mathcal{O}_v, \quad v(t - t') \geq m \implies i(\varphi(t)) - i(\varphi(t')) \in K_m. \tag{2}$$

The formal logarithm converges on  $K_1$  and gives a homomorphism  $K_1 \rightarrow k_v^{\dim J}$ . Restricted to  $K_m$  with  $m > e/(p - 1)$ , the formal exponential provides an inverse, so that the formal logarithm gives an isomorphism  $K_m \rightarrow (\varpi^m \mathcal{O}_v)^{\dim J}$ . It follows that  $pK_m = K_{m+e}$ ; in particular,

$$K_{ne+m} = p^n K_m \subseteq p^n J(k_v) \quad \text{for all } n \geq 0. \tag{3}$$

This implies together with (2) that for  $m$  as above and  $n \geq 0$ ,

$$t, t' \in \varpi \mathcal{O}_v, \quad v(t - t') \geq ne + m \implies i(\varphi(t)) \equiv i(\varphi(t')) \pmod{p^n J(k_v)}. \tag{4}$$

In the following we write  $\mu$  for  $\lfloor e/(p - 1) \rfloor + 1$ ; this is the smallest choice of  $m$  in the considerations above. If  $k_v = \mathbb{Q}_p$  (or, more generally, an unramified extension of  $\mathbb{Q}_p$ , so that  $e = 1$ ), then  $\mu = 1$  when  $p$  is odd, and  $\mu = 2$  when  $p = 2$ .

**Corollary 3.2** *Consider  $\varphi: D_0 \rightarrow D \subseteq C(k_v)$  as above, and let  $t_0 \in \varpi \mathcal{O}_v$  be such that  $\tau(i(\varphi(t_0))) \leq n$ . Then for all  $t$  with  $v(t - t_0) \geq e(n + 1) + \mu$ , we have*

$$\tau(i(\varphi(t))) = \tau(i(\varphi(t_0))) \quad \text{and} \quad q(i(\varphi(t))) = q(i(\varphi(t_0))).$$

*More generally, let  $\Gamma \subseteq J(k_v)$  be a subgroup. If  $\max \tau(i(\varphi(t_0)) + \Gamma) \leq n$ , then for all  $t$  with  $v(t - t_0) \geq e(n + 1) + \mu$ , we have*

$$\max \tau(i(\varphi(t)) + \Gamma) = \max \tau(i(\varphi(t_0)) + \Gamma)$$

and

$$q(i(\varphi(t)) + \Gamma) = q(i(\varphi(t_0)) + \Gamma).$$

*Proof* By (4), we have  $i(\varphi(t)) \equiv i(\varphi(t_0)) \pmod{p^{n+1}J(k_v)}$ . The first claim now follows from Lemma 3.1. The second claim follows from the first by considering  $i(\varphi(t)) + \gamma$  for each  $\gamma \in \Gamma$  separately, and applying the first claim to the shifted embedding  $P \mapsto i(P) + \gamma$ .  $\square$

If the image of the disk  $D$  in  $J$  does not contain a point of finite order prime to  $p$ , then  $\tau$  will be bounded on  $D$ . Corollary 3.2 then provides a partition of  $D$  into finitely many sub-disks such that  $q \circ i$  is constant on each of them. In this way, we can compute  $q(i(D))$ . In a similar way, this allows us to compute  $q(i(D) + \Gamma)$  if  $i(D)$  does not meet  $\text{cl}(\Gamma) + J(k_v)[p']$ , where  $G[p']$  denotes the subgroup of an abelian group  $G$  consisting of elements of finite order prime to  $p$ ; compare Lemma 2.5.

We now consider the case when  $D$  contains a point  $P_0$  such that  $i(P_0) \in J(k_v)[p']$ . In this situation, the result above will not produce a finite partition into sub-disks, so we need to have an explicit estimate for the size of the pointed disk around  $P_0$  on which  $q \circ i$  is constant. Without loss of generality,  $i(P_0) = 0$ . We also assume that  $\varphi(0) = P_0$ , so that  $i(\varphi(0)) = 0 \in J$ .

In the following, we write  $n_{\text{tors}}$  for the smallest  $n \geq 0$  such that  $J(k_v)[p^\infty] \subseteq J[p^n]$ . In other words,  $p^{n_{\text{tors}}}$  is the exponent of the  $p$ -power torsion subgroup  $J(k_v)[p^\infty]$ .

**Lemma 3.3** *Let  $P \in J(k_v)$ .*

- (1) *If  $n_{\text{tors}} = 0$ , then  $\tau(pP) = \tau(P) + 1$  and  $q(P) \subseteq q(pP) \subseteq q(P) \cup \{0\}$ .*
- (2)  *$\tau(P) > n_{\text{tors}}$ , then  $\tau(pP) = \tau(P) + 1$  and  $q(pP) = q(P)$ .*

*Proof* Since  $p^n Q = P$  implies  $p^{n+1} Q = pP$ , the inclusion  $q(P) \subseteq q(pP)$  is clear, as is the inequality  $\tau(pP) \geq \tau(P) + 1$ , for arbitrary  $P$ .

First assume that  $n_{\text{tors}} = 0$ . Consider  $\xi \in q(pP)$ , so there are  $Q \in J(k_v)$  and  $n \geq 0$  such that  $p^n Q = pP$  and  $\pi(Q) = \xi$ . If  $n = 0$ , then  $\xi = \pi(Q) = \pi(pP) = 0$ . If  $n \geq 1$ , then we must have  $p^{n-1} Q = P$  (there is no nontrivial  $p$ -torsion), so  $\xi = \pi(Q) \in q(P)$ . Taking  $n = \tau(pP)$  shows that  $\tau(P) \geq \tau(pP) - 1$ .

Now assume that  $\tau(P) > n_{\text{tors}}$  and write  $P = p^{n_{\text{tors}}+1} P_0$  for  $P_0 \in J(k_v)$ . We first show that  $\pi(J(k_v)[p^\infty]) \subseteq q(P)$ . For this, let  $T \in J(k_v)[p^\infty] = J(k_v)[p^{n_{\text{tors}}}]$ . Then  $p^{n_{\text{tors}}}(T + pP_0) = P$ , so  $\pi(T) = \pi(T + pP_0) \in q(P)$  by (1).

To show that  $q(pP) \subseteq q(P)$ , let  $\xi \in q(pP)$ , so there are some  $Q \in J(k_v)$  and  $n \geq 0$  with  $\pi(Q) = \xi$  such that  $p^n Q = pP = p^{n_{\text{tors}}+2} P_0$ . If  $n \leq n_{\text{tors}} + 1$ , then it follows that  $Q = p^{n_{\text{tors}}+2-n} P_0 + T$  with  $T \in J(k_v)[p^\infty]$ , so  $\xi = \pi(Q) = \pi(T) \in q(P)$  by the argument above. If  $n \geq n_{\text{tors}} + 2$ , then  $p^{n-n_{\text{tors}}-2} Q = P_0 + T$  with  $T \in J(k_v)[p^{n_{\text{tors}}}]$ , and therefore  $p^{n-1} Q = p^{n_{\text{tors}}+1} P_0 = P$ , so  $\xi = \pi(Q) \in q(P)$ . Carrying out this argument with  $n = \tau(pP)$  and a suitable  $Q$ , we also get that  $\tau(P) \geq \tau(pP) - 1$ .  $\square$

For  $m \geq 1$  we define

$$N(m) = 1 + \min \left\{ \left\lfloor \frac{km - v(k) - \mu}{e} \right\rfloor : k \geq 2 \right\}.$$

Then  $N(m)e \geq 2m - a$  for some constant  $a$ .

**Lemma 3.4** *Assume that  $v(t) = m \geq 1$ . Then*

$$p \cdot i(\varphi(t)) \equiv i(\varphi(pt)) \pmod{p^{N(m)}J(k_v)} .$$

*Proof* In terms of formal group coordinates, we can write

$$\log_J i(\varphi(t)) = c_1 t + \frac{c_2}{2} t^2 + \frac{c_3}{3} t^3 + \dots$$

with  $c_1, c_2, c_3, \dots \in \mathcal{O}_v^{\dim J}$ . We find that

$$\begin{aligned} & \log_J (pi(\varphi(t)) - i(\varphi(pt))) \\ &= p \log_J i(\varphi(t)) - \log_J i(\varphi(pt)) \\ &= c_2 \frac{p-p^2}{2} t^2 + c_3 \frac{p-p^3}{3} t^3 + c_4 \frac{p-p^4}{4} t^4 + \dots \in (\varpi^{N(m)e+\mu} \mathcal{O}_v)^{\dim J} , \end{aligned}$$

by the definition of  $N(m)$ . We have that

$$p \cdot i(\varphi(t)) - i(\varphi(pt)) \in pK_m + K_{m+e} = K_{m+e} \subseteq K_\mu ,$$

so we are in the domain of the isomorphism induced by the formal logarithm, which allows us to conclude that  $pi(\varphi(t)) - i(\varphi(pt)) \in K_{N(m)e+\mu}$ . The claim then follows from (3). □

**Corollary 3.5** *If we have  $p = 2$  and  $e = 1$  (which is the case when  $k_v = \mathbb{Q}_2$ ) in the situation of Lemma 3.4, then*

$$2i(\varphi(t)) \equiv i(\varphi(2t)) \pmod{2^{2m-2}J(\mathbb{Q}_2)} .$$

*If in addition  $C$  is hyperelliptic,  $\varphi(0)$  is a Weierstrass point and  $\varphi(-t) = \iota(\varphi(t))$ , where  $\iota$  is the hyperelliptic involution, then*

$$2i(\varphi(t)) \equiv i(\varphi(2t)) \pmod{2^{3m-1}J(\mathbb{Q}_2)} .$$

*Proof* If  $p = 2$  and  $e = 1$ , then  $\mu = 2$  and so  $N(m) = 2m - 2$  in Lemma 3.4 (the minimum is attained for  $k = 2$ ).

Under the additional assumptions on  $C$  and  $\varphi$ , it follows that  $\log_J \circ i \circ \varphi$  is odd, so that  $c_{2n} = 0$  for all  $n \geq 1$  in the proof of Lemma 3.4. We then obtain the better bound in the same way as in that proof, noting that we can restrict to odd  $k$  (which have  $v(k) = 0$ ). □

For our fixed  $\varphi$  and  $i$ , we define, for  $m \geq 1$ ,

$$n_m := \max\{\tau(i(\varphi(t))) : t \in \varpi \mathcal{O}_v, v(t) = m\} .$$

**Lemma 3.6** *There is some  $b \in \mathbb{Z}$  such that  $n_m e \leq m + b$  for all  $m \geq 1$ .*

*Proof* First note that  $i(\varphi(t)) \in K_{m+a} \setminus K_{m+a+1}$  for some fixed  $a$  when  $m$  is sufficiently large, where  $a$  is the valuation of  $c_1$  in the proof of Lemma 3.4 above.

Next, let  $a'$  denote the  $p$ -adic valuation of the exponent of the (finite) quotient group  $J(k_v)/K_\mu$ . Then for  $n \geq \mu$  and  $P \in K_n \setminus K_{n+1}$ , we have  $\tau(P)e \leq n - \mu + a'e$ . To see this, write  $P = p^{\tau(P)}Q$  for some  $Q \in J(k_v)$ ; assume  $\tau(P)e > n - \mu + a'e$ . Then  $p^{a'}Q$  maps to an element of order prime to  $p$  in  $J(k_v)/K_\mu$ , and since  $P = p^{\tau(P)-a'}(p^{a'}Q) \in K_\mu$ , it follows that  $p^{a'}Q \in K_\mu$  (its class in  $J(k_v)/K_\mu$  has order prime to  $p$  and a power of  $p$  at the same time, so it must be zero). This in turn implies, using (3),

$$P = p^{\tau(P)}Q = p^{\tau(P)-a'} \cdot (p^{a'}Q) \in K_{\mu+(\tau(P)-a')e} \subseteq K_{n+1} ,$$

a contradiction. So  $\tau(P)e \leq n - \mu + a'e$  as claimed.

Finally, combining these arguments, we see that  $n_me \leq m + (a + a'e - \mu)$  for large  $m$ , which implies the claim. □

**Lemma 3.7** *Let  $m_0 = 1$  if  $n_{\text{tors}} = 0$  and  $m_0 = n_{\text{tors}}e + \mu + e$  otherwise. There is some  $m \geq m_0$  such that  $N(m) \geq n_m + 1$ . For any such  $m$ , we have that*

$$q(i(\varphi(\{t : m \leq v(t) < \infty\}))) = q(i(\varphi(\{t : v(t) = m\}))) \cup \{0\} .$$

*Proof* By Lemma 3.6,  $n_me \leq m + b$  for some  $b$ ; on the other hand,  $N(m)e \geq 2m - a$  for some  $a$ , so whenever  $m \geq a + b + e$ , the inequality  $N(m) \geq n_m + 1$  holds. Fix such an  $m$  that also satisfies  $m \geq m_0$ . We now show that if  $v(t) = m$ , then

$$\tau(i(\varphi(p^n t))) = \tau(i(\varphi(t))) + n \quad \text{and} \quad q(i(\varphi(p^n t))) \subseteq q(i(\varphi(t))) \cup \{0\}$$

for all  $n \geq 0$ , which implies the claim (note that  $0 \in q(i(P))$  if  $P$  is sufficiently close to  $P_0$ ). Note that  $m \geq n_{\text{tors}}e + \mu + e$  implies  $n_m \geq n_{\text{tors}} + 1$  by (4) (taking  $t' = 0$ ). We proceed by induction on  $n$ , the case  $n = 0$  being trivial. So consider  $n \geq 1$ . By the inductive assumption, we have

$$\tau(i(\varphi(p^{n-1} t))) = \tau(i(\varphi(t))) + n - 1 \leq n_m + n - 1$$

and

$$q(i(\varphi(p^{n-1} t))) \subseteq q(i(\varphi(t))) \cup \{0\} .$$

By Lemma 3.4, this implies  $p \cdot i(\varphi(p^{n-1} t)) \equiv i(\varphi(p^n t)) \pmod{p^{N(m+(n-1)e)}J(k_v)}$ , and since

$$N(m + (n - 1)e) \geq N(m) + 2n - 2 \geq n_m + n \geq \tau(i(\varphi(p^{n-1} t))) + 1$$

and  $n_m \geq n_{\text{tors}} + 1$  in case  $n_{\text{tors}} > 0$ , by Lemmas 3.1 and 3.3 it follows that

$$\tau(i(\varphi(p^n t))) = \tau(p \cdot i(\varphi(p^{n-1} t))) = \tau(i(\varphi(p^{n-1} t))) + 1 = \tau(i(\varphi(t))) + n$$

and

$$q(i(\varphi(p^n t))) = q(p \cdot i(\varphi(p^{n-1} t))) \subseteq q(i(\varphi(p^{n-1} t))) \cup \{0\} \subseteq q(i(\varphi(t))) \cup \{0\} .$$

□

**Corollary 3.8** *If  $p = 2$  and  $e = 1$  in the situation of Lemma 3.7, then we take  $m_0 = 1$  if  $n_{\text{tors}} = 0$  and  $m_0 = n_{\text{tors}} + 3$  otherwise. There is then some  $m \geq m_0$  such that  $2m - 3 \geq n_m$ . For any such  $m$ , we have that*

$$q(i(\varphi(\{t : m \leq v(t) < \infty\}))) = q(i(\varphi(\{t : v(t) = m\}))) \cup \{0\} .$$

*If the curve is hyperelliptic,  $P_0 = \varphi(0)$  is a Weierstrass point and  $\varphi(-t) = \iota(\varphi(t))$ , where  $\iota$  is the hyperelliptic involution, then the condition above can be replaced by  $3m - 2 \geq n_m$ .*

*Proof* This follows again from  $\mu = 2$  and  $N(m) \geq 2m - 2$ . The improved statement under the additional assumptions follows in the same way as for Corollary 3.5. □

This now allows us to find  $q(i(D))$  when  $0 \in i(D)$ . First we use Corollary 3.2 to determine  $q(i(\varphi(\{t : 1 \leq v(t) \leq m_0 - 1\})))$ . Then for  $m = m_0, m_0 + 1, \dots$ , we find in a similar way  $n_m$  and  $q(i(\varphi(\{t : v(t) = m\})))$ . As soon as  $n_m + 1 \leq N(m)$ , we can stop the computation; we then have

$$q(i(D \setminus \{P_0\})) = q(i(\varphi(\{t : 1 \leq v(t) \leq m\}))) \cup \{0\} .$$

We state a special case for later use.

**Corollary 3.9** *Assume that  $C$  is hyperelliptic, of good reduction mod 2, and satisfies  $J(\mathbb{Q}_2)[2] = 0$  and  $J(\mathbb{F}_2)[2] = 0$ . Let  $P_0 \in C(\mathbb{Q}_2)$ , choose a parameterization  $\varphi$  of a residue disk  $D$  centered at  $P_0$  and let  $i_{P_0}$  denote the embedding of  $C$  into  $J$  sending  $P_0$  to 0. Then*

- (1)  $q(i_{P_0}(D)) = q(i_{P_0}(\varphi(2\mathbb{Z}_2^\times \cup 4\mathbb{Z}_2^\times))) \cup \{0\}$ , and
- (2) *if  $P_0$  is a Weierstrass point and  $\varphi$  satisfies  $\varphi(-t) = \iota(\varphi(t))$ , then*  
 $q(i_{P_0}(D)) = q(i_{P_0}(\varphi(2\mathbb{Z}_2^\times))) \cup \{0\}$ .

*Proof* Since  $k_v = \mathbb{Q}_2$ , we are in the case  $p = 2$  and  $e = 1$ . The assumptions on 2-torsion over  $\mathbb{Q}_2$  and over  $\mathbb{F}_2$  imply that  $n_{\text{tors}} = 0$ , which in turn implies that when  $m \geq 1$  and  $P \in K_m \setminus K_{m+1}$ , we have  $\tau(P) \in \{m - 2, m - 1\}$ , compare [11, Lemma 10.1] and its proof. Also,  $K_1$  has odd index in  $J(\mathbb{Q}_2)$ . We can therefore take  $b = -1$  in Lemma 3.6. Then  $m = 2$  is a suitable value in Corollary 3.8. When  $P_0$  is a Weierstrass point, then by Corollary 3.8 again even  $m = 1$  is sufficient. □

We now give a version of Lemma 3.7 that applies when we work with a subgroup  $\Gamma$  that does not consist of torsion points only. We restrict here to the

case  $k_v = \mathbb{Q}_2$ ; a general statement can be obtained and proved along the same lines, with changes similar to the statement and proof of Lemma 3.7.

We let  $\Gamma \subseteq J(\mathbb{Q}_2)$  be a subgroup such that  $\Gamma \cap 2J(\mathbb{Q}_2) = 2\Gamma$  and such that  $\text{cl}(\Gamma)$  is not of finite index in  $J(\mathbb{Q}_2)$ . We define

$$n_{m,\Gamma} := \sup\{\tau(i(\varphi(t)) + \gamma) : \gamma \in \Gamma, t \in 2\mathbb{Z}_2, v(t) = m\} .$$

**Lemma 3.10** *Let  $m_0 = 2$  if  $n_{\text{tors}} = 0$  and  $m_0 = n_{\text{tors}} + 3$  otherwise. Assume that there is  $m \geq m_0$  such that  $2m - 3 \geq n_{m,\Gamma}$ . For any such  $m$ , we have that*

$$q(i(\varphi(\{t : m \leq v(t) < \infty\})) + \Gamma) = q(i(\varphi(\{t : v(t) = m\})) + \Gamma) \cup q(\Gamma) .$$

*If the curve is hyperelliptic,  $P_0 = \varphi(0)$  is a Weierstrass point and  $\varphi(-t) = \iota(\varphi(t))$ , where  $\iota$  is the hyperelliptic involution, then the condition above can be replaced by  $3m - 2 \geq n_{m,\Gamma}$ .*

By standard Chabauty-Coleman, the intersection of  $i(D)$  with  $\text{cl}(\Gamma)$  is finite. So for  $m$  sufficiently large,  $i(\varphi(2^m\mathbb{Z}_2))$  will meet  $\text{cl}(\Gamma)$  only in  $P_0$ , hence  $n_{m,\Gamma} < \infty$ . So we can hope to find an  $m$  as in the lemma. It is conceivable, however, that the image of the curve meets  $\text{cl}(\Gamma)$  at  $i(P_0)$  with higher multiplicity, in which case  $n_{m,\Gamma}$  may grow too fast with  $m$ .

*Proof* We show again inductively that if  $v(t) = m$ , then

$$\max \tau(i(\varphi(2^n t)) + \Gamma) = \max \tau(i(\varphi(t)) + \Gamma) + n$$

and

$$q(i(\varphi(2^n t)) + \Gamma) \subseteq q(i(\varphi(t)) + \Gamma) \cup q(\Gamma)$$

for all  $n \geq 0$  (note that  $q(\Gamma) \subseteq q(i(P) + \Gamma)$  if  $P$  is sufficiently close to  $P_0$ ). The case  $n = 0$  is trivial. So consider  $n \geq 1$ . By the inductive assumption, we have

$$\max \tau(i(\varphi(2^{n-1} t)) + \Gamma) = \max \tau(i(\varphi(t)) + \Gamma) + n - 1 \leq n_{m,\Gamma} + n - 1$$

and

$$q(i(\varphi(2^{n-1} t)) + \Gamma) \subseteq q(i(\varphi(t)) + \Gamma) \cup q(\Gamma) .$$

By Corollary 3.5, we have  $2i(\varphi(2^{n-1} t)) \equiv i(\varphi(2^n t)) \pmod{2^{2m+2n-4}J(\mathbb{Q}_2)}$ . So for every  $\gamma \in \Gamma$ , we have

$$2(i(\varphi(2^{n-1} t)) + \gamma) \equiv i(\varphi(2^n t)) + 2\gamma \pmod{2^{2m+2n-4}J(\mathbb{Q}_2)} .$$

Since

$$2m + 2n - 4 \geq n_{m,\Gamma} + n \geq \tau(i(\varphi(2^{n-1} t)) + \gamma) + 1$$

and  $n_{m,\Gamma} \geq n_{\text{tors}} + 1$  in case  $n_{\text{tors}} > 0$ , by Lemmas 3.1 and 3.3 it follows that

$$\tau(i(\varphi(2^n t)) + 2\gamma) = \tau(2(i(\varphi(2^{n-1} t)) + \gamma)) = \tau(i(\varphi(2^{n-1} t)) + \gamma) + 1$$

and

$$\begin{aligned} q(i(\varphi(2^n t)) + 2\gamma) &= q(2(i(\varphi(2^{n-1} t)) + \gamma)) \\ &\subseteq q(i(\varphi(2^{n-1} t)) + \gamma) \cup \{0\} \subseteq q(i(\varphi(t)) + \Gamma) \cup q(\Gamma) . \end{aligned}$$

Now consider  $\gamma \in \Gamma \setminus 2\Gamma$ . Since  $i(\varphi(2^n t)) \in 2^{n+m-2}J(\mathbb{Q}_2)$  and  $n + m - 2 \geq 1$ , we get

$$\tau(i(\varphi(2^n t)) + \gamma) = 0 \quad \text{and} \quad q(i(\varphi(2^n t)) + \gamma) = \{\pi(\gamma)\} \subseteq q(\Gamma) .$$

(We use here that  $\gamma \notin 2J(\mathbb{Q}_2)$ .) Together, these relations imply that

$$\max \tau(i(\varphi(2^n t)) + \Gamma) = \max \tau(i(\varphi(t)) + \Gamma) + n$$

and

$$q(i(\varphi(2^n t)) + \Gamma) \subseteq q(i(\varphi(t)) + \Gamma) \cup q(\Gamma)$$

as claimed.

The improved statement under the additional assumptions follows again in the same way. □

## 4 Determining the Set of Rational Points on Odd Degree Hyperelliptic Curves

In this section, we specialize the algorithm formulated in Sect. 2 to hyperelliptic curves of odd degree over  $\mathbb{Q}$ . So let

$$C: y^2 = f(x)$$

be a hyperelliptic curve, given by a squarefree polynomial  $f \in \mathbb{Z}[x]$  of odd degree  $2g + 1$  (then  $g$  is the genus of  $C$ ). We understand  $C$  to be the smooth projective model of the affine curve given by the equation; then  $C$  is a nice curve. We write  $J$  for the Jacobian of  $C$ . For a point  $P_0 \in C(\mathbb{Q})$  (or  $C(\mathbb{Q}_2)$ ), we let  $i_{P_0}: C \rightarrow J$  denote the embedding that sends  $P_0$  to the origin of  $J$ .

To carry out one of the relevant steps, we have to compute  $q(P)$  for points  $P \in J(\mathbb{Q}_2)$  (where  $q(P)$  is defined as above with  $p = 2$ ). The basic strategy for this was



explained in Sect. 3. To implement it, we need to be able to divide by 2 in  $J(\mathbb{Q}_2)$ . We consider this problem in Sect. 5 below.

We recall the algorithm for computing the 2-Selmer group of  $J$ , compare [12, 14]. Let  $C$  be given by the affine equation  $y^2 = f(x)$  with  $f \in \mathbb{Z}[x]$  squarefree and of odd degree  $2g + 1$ , where  $g$  is the genus of  $C$ . Let  $L = \mathbb{Q}[x]/\langle f \rangle$  be the associated étale algebra and write  $\theta$  for the image of  $x$  in  $L$ . If  $A$  is any commutative ring, then we write  $A^\square$  for the group  $A^\times / (A^\times)^2$  of square classes in the multiplicative group  $A^\times$  of  $A$ .

For any field extension  $k$  of  $\mathbb{Q}$ , there is an isomorphism

$$H^1(k, J[2]) \xrightarrow{\cong} \ker(N_{(L \otimes_{\mathbb{Q}} k)/k}: (L \otimes_{\mathbb{Q}} k)^\square \rightarrow k^\square) \tag{5}$$

realizing the Galois cohomology group on the left in a concrete way, and there is the ‘Cassels map’ or ‘ $x - T$ ’ map

$$\mu_k: J(k) \longrightarrow J(k)/2J(k) \hookrightarrow (L \otimes_{\mathbb{Q}} k)^\square$$

that is induced by evaluating  $x - \theta$  (multiplicatively) on divisors whose support is disjoint from the set of Weierstrass points of  $C$ . The image of  $\mu_k$  is contained in the kernel of the norm map above;  $\mu_k$  is the composition of the connecting map  $\delta_k: J(k) \rightarrow H^1(k, J[2])$  induced by the exact sequence of Galois modules

$$0 \longrightarrow J[2] \longrightarrow J(\bar{k}) \xrightarrow{\cdot 2} J(\bar{k}) \longrightarrow 0$$

with the isomorphism (5). We write  $\mu = \mu_{\mathbb{Q}}$ , and for  $v$  a place of  $\mathbb{Q}$ , we write  $L_v = L \otimes_{\mathbb{Q}} \mathbb{Q}_v$  (with  $\mathbb{Q}_\infty = \mathbb{R}$  as usual) and set  $\mu_v = \mu_{\mathbb{Q}_v}$ .

Let  $\Sigma$  be the set of places of  $\mathbb{Q}$  consisting of 2 and the finite places  $v$  such that the Tamagawa number of  $J$  at  $v$  is even. The subgroup  $L(\Sigma, 2)$  of  $L^\square$  consists of the elements represented by  $\alpha \in L^\times$  such that the fractional ideal generated by  $\alpha$  has the form  $I_1^2 I_2$  with  $I_2$  supported on the primes above primes in  $\Sigma$ . Then the isomorphic image of  $\text{Sel}_2 J$  in  $L^\square$ , which we will identify with  $\text{Sel}_2 J$ , is given by

$$\text{Sel}_2 J = \{ \xi \in L(\Sigma, 2) : N_{L/\mathbb{Q}}(\xi) = \square, \forall v \in \Sigma \cup \{ \infty \}: \rho_v(\xi) \in \text{im}(\mu_v) \},$$

where  $\rho_v: L^\square \rightarrow L_v^\square$  is the canonical map. There is also the 2-Selmer set of  $C$ , given by

$$\text{Sel}_2 C = \{ \xi \in L(\Sigma, 2) : N_{L/\mathbb{Q}}(\xi) = \square, \forall v: \rho_v(\xi) \in \mu_v(i_\infty(C(\mathbb{Q}_v))) \}.$$

It is a subset of the 2-Selmer group. The set of places  $v$  in the condition can be restricted to the set  $\Sigma \cup \{ \infty \}$  together with all ‘small’ primes, where ‘small’ in practice can be rather large; see [5].

Algorithm 2, combined with the representation of  $J(k)/2J(k)$  as a subgroup of  $(L \otimes_{\mathbb{Q}} k)^\square$ , then leads to the following.

---

**Algorithm 4.1**

---

**Input:** A polynomial  $f \in \mathbb{Z}[x]$ , squarefree and of odd degree  $2g + 1$ .

**Output:** The set of rational points on  $C$ ;  $y^2 = f(x)$ , or FAIL.

1. Let  $J$  denote the Jacobian of  $C$ . Set  $L = \mathbb{Q}[x]/\langle f \rangle$ .
  2. Compute  $\text{Sel}_2 J$  and  $\text{Sel}_2 C$  as a subgroup and a subset of  $L^\square$ .
  3. Let  $L_2 = L \otimes_{\mathbb{Q}} \mathbb{Q}_2$ ; let  $r: L^\square \rightarrow L_2^\square$  be the map induced by  $\mathbb{Q} \rightarrow \mathbb{Q}_2$ .  
If  $\ker r \cap \text{Sel}_2 J \not\subseteq \delta(\pi(J(\mathbb{Q})[2^\infty]))$ , then return FAIL.
  4. Search for rational points on  $C$  and collect them in a set  $C(\mathbb{Q})_{\text{known}}$ .
  5. Let  $\mathcal{X}$  be a partition of  $C(\mathbb{Q}_2)$  into residue disks whose image in  $L_2^\square$  consists of one element and that are contained in half residue disks when  $J(\mathbb{Q})[2] \neq 0$ .
  6. Let  $R$  denote the image of  $J(\mathbb{Q})[2^\infty]$  in  $L_2^\square$ .
  7. For each  $X \in \mathcal{X}$  do the following:
    - a. If  $X \cap C(\mathbb{Q})_{\text{known}} = \emptyset$ :  
If  $\mu_2(X) \subseteq \text{Sel}_2 C$ , then return FAIL;  
otherwise continue with the next  $X$ .
    - b. Pick some  $P_0 \in C(\mathbb{Q})_{\text{known}} \cap X$ .
    - c. Compute  $Y = \mu_2(q(i_{P_0}(X) + J(\mathbb{Q})[2^\infty])) \subseteq L_2^\square$ .
    - d. If  $Y \cap r(\text{Sel}_2 J) \not\subseteq R$ , then return FAIL.
  8. Return  $C(\mathbb{Q})_{\text{known}}$ .
- 

That the algorithm is correct is a special case of Proposition 2.8, taking into account that torsion points of odd order are infinitely 2-divisible, which allows us to replace  $J(\mathbb{Q})_{\text{tors}}$  with  $J(\mathbb{Q})[2^\infty]$  at the places where the latter occurs.

Remark 2.9 applies in the same way as to the general algorithm.

*Remark 4.2* We note that the (image of the) Selmer group in  $L^\square$  that is used in the algorithm can be replaced by any subgroup  $S$  of  $L^\square$  that contains it (and similarly for the Selmer set). For example, we can take

$$S = \{ \xi \in L(\Sigma, 2) : N_{L/\mathbb{Q}}(\xi) = \square, \forall v \in \Sigma \cup \{ \infty \} \setminus \{ 2 \}; \text{res}_v(\xi) \in \text{im}(\mu_v) \} ,$$

where  $\Sigma$  is the set of ‘bad primes’ for 2-descent on  $J$ . This leaves out the 2-adic Selmer condition. Taking it into account requires the computation of  $\mu_2(J(\mathbb{Q}_2))$ , which is usually the most time-consuming step in the local part of the computation of  $\text{Sel}_2 J$ . We can do without it, since using  $S$  in the algorithm is actually equivalent to using  $\text{Sel}_2 J$ . To see this, first consider Step 3. Since all elements in the kernel of  $r$  satisfy the 2-adic Selmer condition trivially, it follows that  $\ker r \cap S = \ker r \cap \text{Sel}_2 J$ , so that the outcome of Step 3 is the same in both cases. Now consider Step 7a. This does not involve  $\text{Sel}_2 J$ , so its outcome is trivially the same in both cases. Finally consider Step 7d. If  $Y \cap r(S) \not\subseteq R$ , then there is some  $s \in S$  such that  $r(s) \notin R$  and  $r(s) \in Y$ . But everything in  $Y$  is of the form  $\mu_2(Q)$  for some  $Q \in J(\mathbb{Q}_2)$ , so  $Y \subseteq \text{im}(\mu_2)$ , which means that  $s$  satisfies the 2-adic Selmer condition. This shows that  $s \in \text{Sel}_2 J$  and then implies that  $Y \cap r(\text{Sel}_2 J) \ni r(s) \notin R$ , so that the outcome

of this step is again the same in both cases. The preceding arguments show that the algorithm fails on  $S$  if and only if it fails on  $\text{Sel}_2 J$ . Finally, it is clear that the result will be the same, namely  $C(\mathbb{Q})_{\text{known}}$ , in both cases when the algorithm does not output FAIL.

If  $\Sigma \subseteq \{2, p\}$  with  $p \not\equiv \pm 1 \pmod 8$ , then we can also leave out the condition  $N_{L/\mathbb{Q}}(\xi) = \square$ , since then  $\mathbb{Q}(\Sigma, 2)$  injects into  $\mathbb{Q}_2^\square$ , so the norm condition is implied by the image under  $r$  being in  $Y$ .

Of course, we can also use a subset of  $L^\square$  that is possibly larger than  $\text{Sel}_2 C$  instead of the 2-Selmer set. In fact, this is what we have to do in practice, since the computation of the exact 2-Selmer set usually requires taking into account the local conditions for all primes up to some bound that is exponential in the genus of  $C$ ; compare [5].

If we assume that  $C(\mathbb{Q})_{\text{known}}$  meets every set in  $\mathcal{X}$ , then the other conditions required to avoid failure of the algorithm are likely to be satisfied. This follows from work of Bhargava and Gross [2], which we use in a similar way as in [11]: the ‘probability’ that the map  $\text{Sel}_2 J \rightarrow J(\mathbb{Q}_2)/2J(\mathbb{Q}_2)$  is injective is at least  $1 - 2^{1-g-\dim_{\mathbb{F}_2} J(\mathbb{Q}_2)/2}$ , and the ‘probability’ that the image has intersection with  $Y$  contained in  $R$  is at least  $1 - (\#(Y/R) - 1)2^{1-g}$ . Since by the results of [11]  $Y$  is usually small and by [17] the size of  $Y$  modulo  $R$  is uniformly bounded by some constant times  $g^2$ , there is a very good chance that both conditions are satisfied when  $g$  is large.

## 5 Halving Points on Odd Degree Hyperelliptic Jacobians

In this section we describe an algorithm that computes one ‘half’ or all ‘halves’ of a point  $P \in 2J(k)$ , where  $J$  is the Jacobian of a hyperelliptic curve  $C$  of odd degree over the field  $k$ . We assume that  $\text{char}(k) \neq 2$ , so that  $C$  can be given by an equation  $y^2 = f(x)$  with  $f \in k[x]$  squarefree and of odd degree  $2g + 1$ .

Recall that each point in  $J(k)$  is uniquely represented in the form  $[D - d\infty]$ , where  $D$  is an effective divisor in general position defined over  $k$  and  $d = \deg D \leq g$ . An effective divisor  $D$  is said to be *in general position* if its support does not contain  $\infty$  and  $D \not\geq P + \iota(P)$  for any point  $P \in C$ , where  $\iota: C \rightarrow C$  is the hyperelliptic involution.

Any effective divisor  $D$  in general position can be described by its *Mumford representation*  $(a, b)$ . Here  $a \in k[x]$  is a monic polynomial of degree  $d = \deg D$  whose roots are the  $x$ -coordinates of the points in the support of  $D$ , with appropriate multiplicity (so that  $a$  corresponds to the image of  $D$  under the hyperelliptic quotient map to  $\mathbb{P}^1$ ), and  $b \in k[x]$  is another polynomial such that  $b(\xi) = \eta$  for any point  $P = (\xi, \eta)$  in the support of  $D$  and satisfying  $a \mid f - b^2$ . This polynomial  $b$  is uniquely determined modulo  $a$ ; in particular, we obtain a unique representation if we require  $\deg(b) < d$ . However, it is sometimes useful to allow additional flexibility, so we will not always insist on this normalization. In fact, we may also want to allow

polynomials  $a$  of larger degree (this leads to even more non-unique representations, but can be useful in certain situations).

We will use the notation  $(a, b)$  to denote the divisor  $D$ , and we will write  $[a, b] = [(a, b) - d\infty]$  for the point on  $J$  corresponding to it.

Let  $c$  be the leading coefficient of  $f$ . Then in terms of the Mumford representation, the descent map  $\mu: J(k) \rightarrow L^\square$  is given by

$$[a, b] \mapsto (-c)^{\deg(a)} a(\theta) \cdot (L^\times)^2$$

if  $a$  and  $f$  are coprime. In the general case, write  $a_1$  and  $f_1$  for  $a$  and  $f$  divided by their (monic) gcd; then

$$\mu([a, b]) = \tilde{\mu}(a) := (-c)^{\deg(a)} (a(\theta) - a_1(\theta)f_1(\theta)) \cdot (L^\times)^2 ;$$

compare [12].

Since the kernel of  $\mu$  is  $2J(k)$ , this gives us a way of deciding if a point  $P \in J(k)$  is divisible by 2 in  $J(k)$ : this is equivalent to the existence of a polynomial  $s \in k[x]$  such that

$$s(\theta)^2 = (-c)^{\deg(a)} (a(\theta) - a_1(\theta)f_1(\theta)) ;$$

equivalently,

$$s^2 \equiv (-c)^{\deg(a)} (a - a_1f_1) \pmod{f} .$$

We will now state a result that shows how to compute a point  $Q \in J(k)$  such that  $2Q = P$ , given such a polynomial  $s$ .

Note that when  $a = a_1^2 a_2$ , then  $P = [a, b]$  is divisible by 2 if and only if  $P_2 = [a_2, b]$  is, and each point  $Q$  such that  $2Q = P$  has the form  $Q = Q_1 + Q_2$  where  $Q_2$  satisfies  $2Q_2 = P_2$  and  $Q_1 = [a_1, b]$ . So we can assume that  $a$  is squarefree.

**Proposition 5.1** *Let  $a \in k[x]$  be monic and squarefree, of degree  $\leq 2g + 1$ . Let  $d$  denote  $\gcd(a, f)$ , so that  $a = da_1$  and  $f = df_1$  as above. Suppose we have  $b, s \in k[x]$  with*

$$f \equiv b^2 \pmod{a} \quad \text{and} \quad (-c)^{\deg(a)} (a - a_1f_1) \equiv s^2 \pmod{f} ,$$

so that  $[a, b] \in 2J(k)$ . For polynomials  $u, v$  and  $w$ , consider the following system of congruences:

$$vd \equiv ws \pmod{f_1}, \quad vd \equiv ub \pmod{a_1}, \quad uf_1 \equiv ws \pmod{d}. \tag{6}$$

Then this system has a nontrivial solution  $(u, v, w)$  with  $w$  monic such that

$$\deg(u) < \deg(a)/2, \quad \deg(v) \leq g + \deg(a)/2 - \deg(d) \quad \text{and} \quad \deg(w) \leq g. \tag{7}$$

Each such solution satisfies the relation

$$u^2 f_1 = dv^2 - (-c)^{\deg(a)} a_1 w^2. \tag{8}$$

Now assume that  $(u, v, w)$  is a solution such that  $w$  has minimal degree. Let  $d_1 = \gcd(u, w)$ ; then  $d_1$  divides  $fa_1$  and  $v$ . Write  $d_1 = d_f d_a$  with  $d_f = \gcd(d_1, f)$  and  $d_a = \gcd(d_1, a_1)$ . Set  $w_1 = w/d_1$ ,  $u_1 = u/d_1$ ,  $v_1 = v/d_1$  and let  $r \in k[x]$  be such that

$$ru_1 \equiv -v_1 d \pmod{w_1 d_a} \quad \text{and} \quad r \equiv 0 \pmod{d_f}.$$

Then  $Q = [w, r]$  satisfies  $P = 2Q$ .

If  $Q$  and  $Q'$  are computed starting from  $s$  and  $s'$  such that  $s' \not\equiv \pm s \pmod{f}$ , then  $Q$  and  $Q'$  are distinct.

*Proof* First note that, since  $f$  is squarefree, we have that  $d$  and  $f_1$  are coprime. Also,  $d$  and  $a_1$  are coprime, since a divisor in general position contains no ramification point with multiplicity 2 or more. So  $f_1, a_1$  and  $d$  are coprime in pairs and squarefree. The fact that  $a$  divides  $f - b^2$  implies that  $d$  divides  $b$  and that  $d$  is also the gcd of  $a$  and  $b$ .

The first claim is that the system of congruences has a nontrivial solution when the degrees of the polynomials are bounded as stated. To see this, note that the conditions are linear in (the coefficients of)  $u, v$  and  $w$ , and that the total number of coefficients of  $u, v$  and  $w$  is

$$\begin{aligned} & [\deg(a)/2] + (g + \lfloor \deg(a)/2 \rfloor - \deg(d) + 1) + (g + 1) \\ & = 2g + \deg(a) - \deg(d) + 2 = \deg(f_1) + \deg(a_1) + \deg(d) + 1. \end{aligned}$$

On the other hand, the number of linear constraints is  $\deg(f_1) + \deg(a_1) + \deg(d)$ . So there are more variables than constraints, hence nontrivial solutions exist.

We claim that  $w$  cannot be zero in such a solution. Otherwise, the first congruence would imply that  $f_1$  divides  $v$  (since  $f_1, a_1$  and  $d$  are coprime in pairs), which for degree reasons (recall that  $\deg(a) \leq 2g + 1$ ) is only possible when  $v = 0$ . In a similar way, the second congruence would then imply that  $a_1$  divides  $u$  (since  $a_1$  is coprime to  $b$ ), whereas the third congruence implies that  $d$  divides  $u$ , so  $a$  divides  $u$ , which is only possible when  $u = 0$ . But then our solution is trivial, a contradiction. So  $w \neq 0$ , and without loss of generality,  $w$  can be taken to be monic.

We show that every solution as above satisfies relation (8). Namely, by the first congruence and since  $s^2 \equiv (-c)^{\deg(a)} a \pmod{f_1}$ ,

$$d^2 v^2 = (dv)^2 \equiv (sw)^2 = s^2 w^2 \equiv (-c)^{\deg(a)} a w^2 = (-c)^{\deg(a)} d a_1 w^2 \pmod{f_1},$$

so (since  $d$  and  $f_1$  are coprime), the relation holds mod  $f_1$ . Next, by the second congruence,

$$d^2 v^2 = (dv)^2 \equiv (bu)^2 = b^2 u^2 \equiv fu^2 = d f_1 u^2 \pmod{a_1},$$

so (since  $d$  and  $a_1$  are coprime), the relation holds mod  $a_1$ . Finally, by the last congruence,

$$u^2 f_1^2 = (u f_1)^2 \equiv (s w)^2 = s^2 w^2 \equiv -(-c)^{\deg(a)} a_1 f_1 w^2 \pmod{d},$$

so (since  $d$  and  $f_1$  are coprime again), the relation holds also mod  $d$ . It follows that it holds mod  $f_1 a_1 d$ . Since the degrees of all terms are strictly less than the degree of  $f_1 a_1 d$ , equality follows, and (8) is verified.

We note that the fact shown above that a nontrivial solution has  $w \neq 0$  implies that  $w$  determines the solution uniquely. It follows that there is in fact a *unique* solution with  $w$  monic and  $\deg(w)$  minimal.

Since  $d$  is squarefree, (8) implies that the gcd  $d_1$  of  $w$  and  $u$  also divides  $v$ . We can therefore divide all three by this gcd, obtaining  $u_1, v_1$  and  $w_1$ ; they satisfy

$$u_1^2 f_1 = d v_1^2 - (-c)^{\deg(a)} a_1 w_1^2.$$

If some irreducible factor  $p$  of  $d_1$  does not divide  $f a_1$ , then  $(u/p, v/p, w/p)$  also satisfy the system of congruences, contradicting the minimality of  $\deg(w)$ . Now assume that  $p^2$  divides  $d_1$  for some irreducible polynomial  $p$ . Then  $p$  divides  $f_1, a_1$  or  $d$ , say  $p \mid a_1$  (the other cases are analogous). Since  $a_1$  is squarefree, the congruence  $vd \equiv ub \pmod{a_1}$  implies  $(v/p)d \equiv (u/p)b \pmod{a_1}$ , and so again  $(u/p, v/p, w/p)$  satisfy the system of congruences, contradiction. So  $d_1$  is squarefree and must therefore divide  $a_1 f_1 d$ . In particular, we can write  $d_1 = d_f d_a$  as claimed.

Note that  $(u_1 b)^2 \equiv u_1^2 f \equiv (v_1 d)^2 \pmod{a_1}$ , so  $a_1$  divides  $(u_1 b - v_1 d)(u_1 b + v_1 d)$ . We claim that  $d_a = \gcd(u_1 b + v_1 d, a_1)$ . For this, consider an irreducible factor  $p$  of  $d_a$ . If  $p$  divides  $u_1 b - v_1 d$ , then  $(u/p, v/p, w/p)$  is a solution, a contradiction. So  $p$  must divide  $u_1 b + v_1 d$ . Conversely, if  $p$  is any irreducible factor of  $a_1$  that divides  $u_1 b + v_1 d$ , then (noticing that  $b$  is invertible mod  $a_1$ ) for  $p$  to divide  $ub - vd$ , it must necessarily divide  $u$  and  $v$ , so  $p \mid d_a$ .

$u_1$  is invertible mod  $w_1$ , but also mod  $d_a$  (since  $u_1$  and  $v_1$  are coprime as well— $a_1$  is squarefree—and  $d_a$  is coprime with  $f_1$  and  $d$ ). Furthermore,  $d_f$  is coprime with  $w_1$  (and of course also with  $d_a$ ), for essentially the same reason. Therefore a polynomial  $r$  exists such that  $u_1 r \equiv -v_1 d \pmod{w_1 d_a}$  and  $r \equiv 0 \pmod{d_f}$ .

Now we consider the function

$$\phi = u(x)y - v(x)d(x) = d_f(x)d_a(x) (u_1(x)y - v_1(x)d(x))$$

on  $C$ . Its divisor of zeros is

$$\begin{aligned} & 2(d_f, 0) + ((d_a, b) + (d_a, -b)) + ((d, 0) + (d_a, -b) + (a_1/d_a, b) + 2(w_1, -r)) \\ &= (a_1, b) + (d, 0) + 2((d_f, 0) + (d_a, -b) + (w_1, -r)) \\ &= (a, b) + 2(w, -r). \end{aligned}$$

To see this, note that the norm in  $k[x]$  of the last factor of  $\phi$  is  $u_1^2 f - v_1^2 d^2 = (-c)^{\deg(a)} da_1 w_1^2$  and that  $u_1 b \equiv v_1 d \pmod{a_1/d_a}$  and  $u_1 b \equiv -v_1 d \pmod{d_a}$  (and so also  $r \equiv b \pmod{d_a}$ ). Setting  $Q = [w, r]$ , we therefore obtain  $2Q = P$ .

We now show that  $Q$  determines  $s \pmod f$  up to sign. Given  $Q = [w, r]$  such that  $2Q = P$ , there is a unique function (up to scaling) on  $C$  whose divisor is  $(a, b) + 2(w, -r) - n\infty$  (where  $n = \deg(a) + 2 \deg(w)$ ); this function must then be  $\phi$ , which gives us  $u$  and  $v$  up to scaling; the relation  $u^2 f_1 = dv^2 - (-c)^{\deg(a)} aw^2$  then fixes them up to a common sign. Write  $d_f = d_{f_1} d_d$  with  $d_{f_1} = \gcd(d_f, f_1)$  and  $d_d = \gcd(d_f, d)$ . In a similar way as above for  $d_a$ , one shows that  $d_{f_1} = \gcd(w_1 s + v_1 d, f_1)$  and  $d_d = \gcd(u_1 f_1 + w_1 s, d)$ . Since  $w_1$  is coprime with  $f$ , this determines  $s \pmod f$  via the congruences

$$\begin{aligned} w_1 s &\equiv v_1 d \pmod{f_1/d_{f_1}}, & w_1 s &\equiv -v_1 d \pmod{d_{f_1}}, \\ w_1 s &\equiv u_1 f_1 \pmod{d/d_d}, & w_1 s &\equiv -u_1 f_1 \pmod{d_d}. \end{aligned}$$

A common sign change of  $u$  and  $v$  (which is the only ambiguity here) results in a sign change of  $s$ . □

We can try to use the algorithm implied by Proposition 5.1 over a  $p$ -adic field. It will possibly run into precision problems when some of the roots of  $a$  get close to roots of  $f$  (but with the resultant of  $a$  and  $f$  still being nonzero, albeit  $p$ -adically small) or when the resulting point is represented by a divisor of lower degree or such that some points are close to the point at infinity. In practice, however, these problems occur fairly rarely. A possible remedy in such a case is to replace  $(a, b)$  by another representation  $(a', b')$  such that  $[a', b'] = [a, b]$  and  $\deg(a) > g$ . Writing  $f - b^2 = ac$ , we have  $[c - 2hb - h^2 a, -b - ha] = [a, b]$  for all polynomials  $h$ . Taking  $h$  to be constant already allows us to replace  $a$  by a polynomial  $a'$  that is coprime with  $f$  (and probably we can also arrange  $a'$  to be squarefree) and satisfies  $\deg(a') \leq g + 1$  if  $\deg(a) = g$ . Another possibility is to consider points in a residue disk given by suitable Laurent series, perform the computation on the Laurent series and then specialize.

*Remark 5.2* In the context of computing  $q(P)$ , the following observation can be useful. Given  $P = [a, b]$  with  $\deg(a) \leq g + 1$  and  $T = [h, 0] \in J(k)[2]$  with  $h \mid f$  and  $\deg(h) \leq g$ , we can use the method described in Proposition 5.1 to compute halves of  $P + T$  without first computing a representation of the sum. For simplicity assume  $\gcd(a, f) = 1$  (this can be arranged, see above). Then  $P + T = [ah, b'h]$  where  $b'h \equiv b \pmod a$ . There will be  $s_1$  and  $s_2$  such that  $s_1^2 \equiv (-c)^{\deg(a)} ah \pmod{f/h}$  and  $s_2^2 \equiv -(-c)^{\deg(a)} a(f/h) \pmod h$ . We obtain the congruences

$$vh \equiv ws_1 \pmod{f/h}, \quad vh \equiv ub \pmod a, \quad u(f/h) \equiv ws_2 \pmod h$$

with the bounds  $\deg(u) < (\deg(a) + \deg(h))/2$ ,  $\deg(v) \leq g + (\deg(a) - \deg(h))/2$  and  $\deg(w) \leq g$ .

In a similar way, we can divide  $P + P'$  by 2: let  $P = [a, b]$ ,  $P' = [a', b']$  and assume that  $\deg(a) + \deg(a') \leq 2g + 1$  and that  $a, a'$  and  $f$  are coprime in pairs. Given a polynomial  $s$  such that  $s^2 \equiv (-c)^{\deg(a)+\deg(a')}aa' \pmod{f}$ , the system to be solved is

$$v \equiv ws \pmod{f}, \quad v \equiv ub \pmod{a}, \quad v \equiv ub' \pmod{a'}$$

where we require  $\deg(u) < (\deg(a)+\deg(a'))/2$ ,  $\deg(v) \leq g + (\deg(a)+\deg(a'))/2$  and  $\deg(w) \leq g$ .

We mention one implication that can be helpful in applications.

**Corollary 5.3** *Let  $[a', b]$  be the Mumford representation of a point  $P \in J(k)$ , write  $a' = a_0^2a$  with  $a$  squarefree and monic and fix a polynomial  $s$  such that*

$$s^2 \equiv (-c)^{\deg(a)}(a - a_1f_1) \pmod{f}$$

as above. Let  $(u, v, w)$  be the solution with  $w$  monic and of smallest degree of the system (6) with the restrictions in (7), and let  $Q \in J(k)$  be the associated point such that  $2Q = P$ . Then  $\mu(Q) = \tilde{\mu}(a_0)\tilde{\mu}(w)$ .

*Proof* This is because according to Proposition 5.1, we have  $Q = [a_0, b] + [w, r]$  for some  $r \in k[x]$ . □

**Corollary 5.4** *In the situation of Corollary 5.3, we have the following special cases.*

- (1) *If  $P = [(\xi, \eta) - \infty] \in 2J(k)$  with  $\eta \neq 0$ , fix  $s \in k[x]$  such that  $s^2 \equiv c(\xi - x) \pmod{f}$ . Let  $w$  be the monic polynomial of smallest degree such that the residue of smallest degree of  $ws$  modulo  $f$  has degree  $\leq g$ . Then the point  $Q \in J(k)$  with  $2Q = P$  that is associated to  $s$  satisfies*

$$\mu(Q) = \tilde{\mu}(w) .$$

- (2) *If  $P = [(\xi_1, \eta_1) - (\xi_2, \eta_2)] \in 2J(k)$  with  $\xi_1 \neq \xi_2$  and  $\eta_j \neq 0$  for  $j \in \{1, 2\}$ , fix  $s \in k[x]$  such that  $s^2 \equiv (x - \xi_1)(x - \xi_2) \pmod{f}$ . Let  $w$  be the monic polynomial of smallest degree such that the residue  $v$  of smallest degree of  $ws$  modulo  $f$  has degree  $\leq g + 1$  and satisfies  $\eta_2v(\xi_1) + \eta_1v(\xi_2) = 0$ . Then the point  $Q \in J(k)$  with  $2Q = P$  that is associated to  $s$  satisfies*

$$\mu(Q) = \tilde{\mu}(w) .$$

*Proof* This follows directly from Corollary 5.3, using that  $d = 1$  (in the notation of Proposition 5.1) in both cases and that  $u$  has to be constant. In the first case, the congruence  $v \equiv ub \pmod{a}$  is redundant, and the system reduces to just  $v \equiv ws \pmod{f}$ . In the second case, the congruence  $v \equiv ub \pmod{a}$  is equivalent to the condition  $\eta_2v(\xi_1) + \eta_1v(\xi_2) = 0$ . □



## 6 A Concrete Example

In this section we use the approach described above to show the following result.

**Theorem 6.1** *Assuming GRH, the only integral solutions of the equation*

$$y^2 - y = x^{21} - x$$

have  $x \in \{-1, 0, 1\}$ .

We remark that  $l = 21$  is the smallest odd exponent such that our method can be successfully applied to determine the set of integral points on the curve given by  $y^2 - y = x^l - x$ . One can check that for  $l \in \{5, 7, 9, 11, 13, 17\}$  the 2-Selmer rank of the Jacobian is  $\geq g = (l - 1)/2$ , and for  $l \in \{15, 19\}$ , the map from  $\text{Sel}_2 J$  to  $J(\mathbb{Q}_2)/2J(\mathbb{Q}_2)$  is not injective.

We also note that all these curves have a pair of rational points with  $x = 1/4$ ; these points are of the form  $\varphi(2u)$  for a parameterization  $\varphi$  of the residue disk at infinity, where  $u \in \mathbb{Z}_2^\times$ . For such a point  $P$ ,  $[P - \infty]$  has nontrivial image in  $J(\mathbb{Q}_2)/2J(\mathbb{Q}_2)$ , and this image is contained in the image of the Selmer group. On the other hand, by Corollary 3.9, the value of  $q$  on the residue disk of  $\infty$  is given by the values at points of the form  $\varphi(2u)$ , so  $q(i_\infty(D))$  will meet the image of the Selmer group non-trivially for every disk  $D$  around infinity, no matter how small. This implies that our approach cannot be used to show that  $\infty$  is the only rational point 2-adically close to  $\infty$ . This is why we restrict to integral points in the statement of Theorem 6.1. The result is in fact stronger: it covers all rational solutions whose  $x$ -coordinate has odd denominator.

In principle, one could try to deal with the residue disk at infinity using  $\Gamma = \langle \gamma \rangle$  where  $\gamma = [(\frac{1}{4}, \frac{1}{2} + \frac{1}{221}) - \infty]$ , since the three (known) rational points in the disk map into this group. Unfortunately, it turns out that  $q(i_\infty(P_4))$  meets the image of the Selmer group outside the image of  $\Gamma$ , which prevents us from applying Theorem 2.1. Here  $P_4 = \varphi(4)$  denotes a point with  $x$ -coordinate  $1/4^2$  (we can use a parameterization of the disk at infinity whose  $x$ -coordinate is given by  $t^{-2}$ ):  $i_\infty(P_4) + 6\gamma = 2^3Q$  with  $\pi_2(Q) \in \sigma(\text{Sel}_2 J) \setminus \pi_2(\Gamma)$ .

*Proof* Let  $C$  denote the curve defined by the equation  $y^2 - y = x^{21} - x$ , and let  $J$  be its Jacobian. Note that  $C$  is isomorphic to the curve given by  $y^2 = 4x^{21} - 4x + 1$ ; write  $f = 4x^{21} - 4x + 1$  and let  $L = \mathbb{Q}[x]/\langle f \rangle$ . We compute a group  $S \subset L^\square$  containing  $\text{Sel}_2 J$  using the algorithm described in [14]. The discriminant of  $f$  is  $-2^{40}$  times the product of six distinct odd primes. This implies that 2 is the only ‘bad’ prime for 2-descent, so that the image of the Selmer group is contained in  $L(\{2\}, 2)$ . Since  $L$  is totally ramified at 2, we can reduce this to  $S = L(\emptyset, 2)$  (if  $\xi$  represents an element of  $L(\{2\}, 2)$  and  $N_{L/\mathbb{Q}}(\xi)$  is a square, then the ideal generated by  $\xi$  must be a square). The class group of  $L$  turns out to be trivial, so that  $L(\emptyset, 2) = \mathcal{O}_L^\square$ , but we do not need this fact. We do need to compute  $L(\emptyset, 2)$  and explicit generators of it, though. This is where we use GRH to make the computation feasible in reasonable time. We check that the map  $S \rightarrow L_2^\square$  is injective.

The curve has good reduction mod 2, and  $J(\mathbb{F}_2)$  and  $J(\mathbb{Q}_2)$  both have no elements of order 2. Up to the action of the hyperelliptic involution, there are two residue disks with 2-adically integral  $x$ -coordinates; we can center them at the rational points  $(0, 0)$  and  $(1, 0)$ , respectively. By [14, Lemma 6.3], it follows that the image in  $L_2^\square$  of a point  $P \in C(\mathbb{Q}_2)$  with  $x(P) \in \mathbb{Z}_2$  depends only on  $x \pmod 4$ . We check that the image in  $L_2^\square$  of the points with  $x(P) \equiv 2 \pmod 4$  is not in the image of  $S$ . This shows that any (2-adically) integral point  $P \in C(\mathbb{Q})$  must have  $x(P) \equiv -1, 0$  or  $1 \pmod 4$ . We consider each of the corresponding (pairs of) half residue disks separately. Let  $P_0$  be one of the points  $(-1, 0)$ ,  $(0, 0)$  or  $(1, 0)$  on  $C$  and let  $D$  be the disk around  $P_0$  consisting of points  $P$  with  $x(P) \equiv x(P_0) \pmod 4$  and  $y(P) \equiv 0 \pmod 2$ . By Corollary 3.9 (note that the disk  $D$  corresponds to  $m \geq 2$  in terms of the maximal residue disk around  $P_0$ ), we have

$$q(i_{P_0}(D)) = q(i_{P_0}(\varphi(4\mathbb{Z}_2^\times))) ,$$

where  $\varphi$  is a parameterization of the residue disk containing  $P_0$  such that  $\varphi(0) = P_0$  and  $D = \varphi(4\mathbb{Z}_2)$ . By Lemma 3.1 and since  $\tau(i_{P_0}(\varphi(4u))) = 1$  for some  $u \in \mathbb{Z}_2^\times$  (as becomes apparent in the course of the computation), it is sufficient to consider  $\varphi(4)$  and  $\varphi(-4)$ . So we compute the (unique) half of  $i_{P_0}(P)$  for each point  $P \in D$  such that  $x(P) = x(P_0) \pm 4$ ; we find that its image in  $L_2^\square$  is nontrivial (and does not depend on the sign) and is not contained in the image of  $S$ . By Theorem 2.6 this now implies that  $D \cap C(\mathbb{Q}) = \{P_0\}$ , for each of the three points. So we obtain the result that

$$C(\mathbb{Q}) \cap C(\mathbb{Z}_2) = \{(-1, 0), (-1, 1), (0, 0), (0, 1), (1, 0), (1, 1)\}$$

as claimed. □

## 7 An Application to Fermat’s Last Theorem

In this section we apply the criterion that is given by the algorithm in Sect. 4 to a certain family of hyperelliptic curves that are related to Fermat curves. This leads to a criterion for Fermat’s Last Theorem to hold for a given prime  $p$ . Of course, FLT has been proved in general by Wiles [18, 21], so this will not produce a new result. On the other hand, it shows that the method does work in practice. In the next section, we will deal with a similar family of curves that are related to certain generalized Fermat equations; our method applies again and does indeed solve some new cases of generalized Fermat equations.

Consider

$$C_l: y^2 = f(x) := 4x^l + 1$$

with  $l = 2g + 1$ . This curve has good reduction at 2, since it is isomorphic to the curve  $y^2 + y = x^l$ . The reduction has three  $\mathbb{F}_2$ -points, so there are three residue classes in  $C_l(\mathbb{Q}_2)$ . We also note that  $C_l$  has the three obvious rational points  $\infty$ ,  $(0, 1)$  and  $(0, -1)$  and that  $[(0, \pm 1) - \infty] \in J_l(\mathbb{Q})$ , where  $J_l$  denotes the Jacobian of  $C_l$ , is a point of odd order  $l$ . We note that  $J_l(\mathbb{Q}_2)$  and  $J_l(\mathbb{F}_2)$  contain no points of order 2.

**Corollary 7.1** *Let  $\varphi: D_0 \rightarrow D \subseteq C_l(\mathbb{Q}_2)$  be a parameterization of one of the three residue disks of  $C_l(\mathbb{Q}_2)$ , with  $\varphi(0)$  being  $\infty$  or  $(0, \pm 1)$ . Then*

$$q(i_\infty(D)) = \begin{cases} q(i_\infty(\varphi(2\mathbb{Z}_2^\times \cup 4\mathbb{Z}_2^\times))) \cup \{0\} & \text{if } \varphi(0) = (0, \pm 1); \\ q(i_\infty(\varphi(2\mathbb{Z}_2^\times))) \cup \{0\} & \text{if } \varphi(0) = \infty. \end{cases}$$

*Proof* This is simply Corollary 3.9 specialized to the case at hand. □

We now want to find  $q(i_\infty(C_l(\mathbb{Q}_2)))$  in terms of its image in  $L_2^\square$  as in Algorithm 4. To do this, we need a basis for the latter group. We first note that  $f$  is irreducible over  $\mathbb{Q}_2$ , so  $L_2$  is a field. Let  $\lambda = 2^{1/l}$ , then  $L_2 = \mathbb{Q}_2(\lambda)$  is totally and tamely ramified and  $\theta = -\lambda^{-2}$  is a root of  $f$ . Clearly,  $2 = \lambda^l$ . Note that an element of the form  $1 + 4\alpha\lambda = 1 + \alpha\lambda^{2l+1}$  with  $\alpha \in \mathcal{O}_{L_2}$  is always a square in  $L_2$  (the power series for  $\sqrt{1+x}$  converges when the valuation of  $x$  exceeds that of 4). Furthermore,

$$(1 + \lambda^n)^2 = 1 + \lambda^{2n} + \lambda^{n+l} = (1 + \lambda^{2n})(1 + \lambda^{n+l} + \dots),$$

which allows us to eliminate factors of the form  $1 + \lambda^{2n}$  for  $n \leq l - 1$  when working modulo squares. In this way, we find that the following elements represent an  $\mathbb{F}_2$ -basis for  $L_2^\square$ :

$$\lambda, 1 + \lambda, 1 + \lambda^3, \dots, 1 + \lambda^{2n+1}, \dots, 1 + \lambda^{2l-3}, 1 + \lambda^{2l-1}, 1 + \lambda^{2l}.$$

**Lemma 7.2** *The image of  $q(i_\infty(C_l(\mathbb{Q}_2)))$  in  $L_2^\square$  consists of the classes of*

$$1, \quad 1 + \lambda^{l+2}, \quad 1 + \lambda^{2l-1}, \quad \prod_{k \geq 1} (1 + \lambda^{l+2^k}).$$

We let  $Z$  denote the set consisting of the three nontrivial classes in this image.

*Proof* We first consider the residue disk  $D_\infty$  around  $\infty$ . By Corollary 7.1, it is sufficient to find  $\mu_2(q(i_\infty(\varphi(t))))$  for  $t = 2u$  with  $u \in \mathbb{Z}_2^\times$ . One choice of  $\varphi$  is

$$\varphi(t) = (t^{-2}, 2t^{-l}(1 + 2^{-3}t^{2l} - 2^{-7}t^{4l} \pm \dots)).$$

Then  $\mu_2(i_\infty(\varphi(2u)))$  is the class of  $(2u)^{-2} + \lambda^{-2}$  in  $L_2^\square$ . We have

$$(2u)^{-2} + \lambda^{-2} = (2u)^{-2}(1 + u^2\lambda^{2l-2}) \sim 1 + \lambda^{2l-2} \sim 1 + \lambda^{2l-1},$$

where  $\sim$  denotes equivalence mod squares, by the relation

$$1 \sim (1 + \lambda^{l-1})^2 = 1 + \lambda^{2l-2} + \lambda^{2l-1} \sim (1 + \lambda^{2l-2})(1 + \lambda^{2l-1}).$$

We conclude that (specifying elements of  $L_2^\square$  using representatives in  $L_2^\times$ )

$$\mu_2(q(i_\infty(D_\infty))) = \{1, 1 + \lambda^{2l-1}\}.$$

(Compare [11, Lemma 10.2], which says that the image of the residue disk at infinity under the  $\rho$  log map has just one element.)

Now we consider the residue disk  $D_{(0,1)}$  around  $(0, 1)$ . If  $P = (\xi, \eta) \in C_l(\mathbb{Q}_2)$  has integral  $x$ -coordinate, then we must have  $\xi \in 2\mathbb{Z}_2$  (otherwise the right hand side is  $5 \pmod 8$  and therefore not a square). We can parameterize  $D_{(0,1)}$  by

$$\varphi(t) = \left( t, \sqrt{1 + 4t^l} = 1 + 2t^l - 2t^{2l} + \dots \right).$$

Then  $\mu_2(i_\infty(\varphi(2u)))$  is the class of

$$2u + \lambda^{-2} = \lambda^{-2}(1 + \lambda^{l+2}u) \sim 1 + \lambda^{l+2};$$

the latter relation holds when  $u$  is a unit. By Corollary 7.1, we also need to find the image under  $q$  of points given by  $t \in 4\mathbb{Z}_2^\times$ , so  $t = 4u$  with  $u \in \mathbb{Z}_2^\times$ . In this case (recall that  $\theta = -\lambda^{-2}$ )

$$4u - \theta = (2\theta^{g+1})^2(1 - 4u/\theta) = s(\theta)^2$$

where  $s \in \mathbb{Q}_2[x]$  is a polynomial of degree  $\leq l - 1$  such that

$$\begin{aligned} s(\theta) &= 2\theta^{g+1} \sqrt{1 - 4u/\theta} \\ &= 2(\theta^{g+1} - 2u\theta^g - 2u^2\theta^{g-1} - 4u^3\theta^{g-2} - \dots). \end{aligned}$$

The coefficients of

$$\sqrt{1 - 4x} = \sum_{n=0}^\infty 2^{2n} (-1)^n \binom{1/2}{n} x^n = 1 - \sum_{n=1}^\infty 2^n \frac{1 \cdot 3 \cdot 5 \cdots (2n - 3)}{n!} x^n$$

(except for the constant term) all have 2-adic valuation at least 1 (since  $v_2(n!) \leq n - 1$ ) and  $\theta^{-1} = -4\theta^{2g}$ , so

$$\frac{1}{2}s(x) \equiv x^{g+1} - 2ux^g - 2u^2x^{g-1} - \dots - c_{g+1}u^{g+1} \pmod{8\mathbb{Z}_2[x]},$$

where  $c_{g+1}$  denotes the coefficient of  $x^{g+1}$  in  $-\sqrt{1-4x}$ . Let  $w_0(x)$  denote the partial sum of the power series of  $(1-4ux)^{-1/2}$  up to and including the term with  $x^g$ , and set

$$\tilde{w}(x) = x^g w_0(1/x) = x^g + 2ux^{g-1} + 6u^2x^{g-2} - \dots$$

Then

$$\tilde{w}(x)s(x) \equiv 2x^{2g+1} + (\text{terms up to } x^g) \pmod{8\mathbb{Z}_2[x]},$$

so  $\tilde{w}(x) \equiv w(x) \pmod{8\mathbb{Z}_2[x]}$ , where  $w(x)$  is the monic polynomial of degree  $g$  such that  $w(\theta)s(\theta) \in \mathbb{Q}_2 + \mathbb{Q}_2\theta + \dots + \mathbb{Q}_2\theta^g$ . Let  $Q \in J_l(\mathbb{Q}_2)$  denote the point such that  $2Q = [(4u, *) - \infty]$ . By Corollary 5.4, the image of  $Q$  in  $L_2^\square$  is given by the class of

$$\begin{aligned} (-1)^g w(\theta) &\sim (-1)^g \tilde{w}(\theta) \\ &= (-\theta)^g (1 + 2u\lambda^2 + 6u^2\lambda^4 + 20u^3\lambda^6 + 70u^4\lambda^8 + \dots) \\ &\sim 1 + 2\lambda^2 + 6\lambda^4 + 20\lambda^6 + 70\lambda^8 + \dots \\ &\sim 1 + \sum_{k=1}^{\infty} \lambda^{l+2k} \\ &\sim (1 + \lambda^{l+2})(1 + \lambda^{l+4})(1 + \lambda^{l+8}) \dots (1 + \lambda^{l+2^k}) \dots, \end{aligned}$$

where the product can be truncated as soon as  $2^k > l$ . (We have used that the valuation of the coefficient of  $x^n$  in  $(1-4x)^{-1/2}$  is 1 precisely when  $n$  is a power of 2.) □

We can generalize this result to certain curves of the form  $y^2 = 4x^l + A$ . Let  $A \in \mathbb{Z}$  with  $A \equiv 1 \pmod{8}$  and consider

$$C_{l,A}: y^2 = 4x^l + A.$$

Then  $C_{l,A}$  is  $\mathbb{Q}_2$ -isomorphic to  $C_l = C_{l,1}$ , since  $A$  is a square and an  $l$ th power in  $\mathbb{Q}_2$ . In particular, we still have  $L_2 = \mathbb{Q}_2(\lambda)$ , where now  $L = \mathbb{Q}[x]/\langle 4x^l + A \rangle$ , and the image of  $q(i_\infty(C_{l,A}(\mathbb{Q}_2)))$  in  $L_2^\square$  is the same as for  $C_l$ , namely  $Z \cup \{1\}$ .

**Proposition 7.3** *Let  $A$  be an integer satisfying  $A \equiv 1 \pmod{8}$ ; consider the curve  $C_{l,A}: y^2 = 4x^l + A$  over  $\mathbb{Q}$  with  $l = 2g + 1 \geq 5$ , with Jacobian  $J_{l,A}$ . Let further  $L = \mathbb{Q}[x]/\langle 4x^l + A \rangle$  and  $L_2 = L \otimes_{\mathbb{Q}} \mathbb{Q}_2 = \mathbb{Q}_2(\lambda)$  with  $\lambda = 2^{1/l}$ . If*

- (1) *the canonical map  $\text{Sel}_2 J_{l,A} \hookrightarrow L^\square \rightarrow L_2^\square$  is injective and*
- (2) *its image does not meet  $Z$ ,*

*then  $C_{l,A}(\mathbb{Q}) = \{\infty\}$  if  $A$  is not a square, and  $C_{l,A}(\mathbb{Q}) = \{\infty, (0, a), (0, -a)\}$  if  $A = a^2$ .*

*Proof* We apply Theorem 2.1 with  $A = J_{l,A}$ ,  $i = i_\infty$ ,  $\Gamma = \{0\}$  and  $X = C_{l,A}(\mathbb{Q}_2)$ . By Lemma 7.2,  $Z$  is the set of nontrivial images in  $L_2^\square$  of elements in  $q(i_\infty(C_{l,A}(\mathbb{Q}_2)))$ . So the assumptions here match the assumptions of Theorem 2.1, and we conclude that  $i_\infty(C_{l,A}(\mathbb{Q})) \subseteq \overline{\{0\}} = J_{l,A}(\mathbb{Q})_{\text{tors}}$ . Since  $C_{l,A}$  has good reduction at 2 and  $J_{l,A}(\mathbb{Q})[2]$  is trivial, we find that  $J_{l,A}(\mathbb{Q})_{\text{tors}}$  injects into  $J_{l,A}(\mathbb{F}_2)$ ; in particular,  $C_{l,A}(\mathbb{Q})$  will inject into  $C_{l,A}(\mathbb{F}_2)$ , which has three elements. Since each residue class in  $C_{l,A}(\mathbb{Q}_2)$  contains exactly one torsion point (namely,  $\infty$ ,  $(0, a)$  and  $(0, -a)$ , respectively, where  $a$  is a square root of  $A$  in  $\mathbb{Q}_2$ ), the claim follows.  $\square$

It is known that Fermat’s Last Theorem holds for a prime  $p \geq 3$  if (and only if) the curve  $y^2 = 4x^p + 1$  has only the obvious three rational points. So Proposition 7.3 gives a criterion for FLT for exponent  $p$  to hold, in terms of the 2-Selmer group of the Jacobian of this curve. We can deduce the following criterion.

**Proposition 7.4** *Let  $p \geq 5$  be a prime and set  $L = \mathbb{Q}(2^{1/p})$  and  $L_2 = \mathbb{Q}_2(2^{1/p})$ . Let  $r: \mathcal{O}_L^\square \rightarrow \mathcal{O}_{L_2}^\square$  denote the canonical map. If*

- (1)  $p^2 \nmid 2^{p-1} - 1$ ,
- (2) *the class number of  $L$  is odd, and*
- (3)  $\text{im}(r) \cap Z = \emptyset$  (where  $Z$  is as above),

*then Fermat’s Last Theorem holds for the exponent  $p$ .*

*Proof* Let  $f(x) = x^p + 1/4$ . Then  $f(x - 1/4) \equiv x^p \pmod{p\mathbb{Z}_p[x]}$ , and the first assumption  $p^2 \nmid 2^{p-1} - 1$  implies that the constant term is not divisible by  $p^2$ . This in turn implies that  $C_p: y^2 = 4x^p + 1$  is regular over  $\mathbb{Z}_p$  and the component group of the Néron model of the Jacobian  $J$  of  $C$  over  $\mathbb{Z}_p$  is trivial. By [14, Lemma 4.5] or [13, Prop. 3.2] (which applies equally to abelian varieties), the only ‘bad prime’ for the computation of  $\text{Sel}_2 J_p$  is 2. By the second assumption, the class group of  $L$  has odd order and therefore trivial 2-torsion. Together, the previous two sentences imply that the isomorphic image of  $\text{Sel}_2 J_p$  in  $L^\square$  is contained in the subgroup generated by  $\mathcal{O}_L^\square$  and the image of  $2^{1/p}$ . The map to  $L_2^\square$  decomposes as a direct sum of the map  $r$  and an isomorphism of 1-dimensional  $\mathbb{F}_2$ -vector spaces (since the class of  $\lambda = 2^{1/p}$  is not contained in the image of the (global or 2-adic) units). We note that  $r$  is injective: assume that  $u \in \mathcal{O}_L^\times$  is a square in  $\mathcal{O}_{L_2}$ . Since  $u$  is a unit, the extension  $L(\sqrt{u})/L$  is unramified at all places not dividing 2 or  $\infty$ . The extension is unramified at  $\infty$ , since  $N_{L/\mathbb{Q}}(u)$  must be 1 (it is a 2-adic square by assumption), so the image of  $u$  under the unique real embedding of  $L$  is positive. Finally, it is unramified (and even split) at the prime above 2. Since the class number is odd, there are no nontrivial everywhere unramified quadratic extensions of  $L$ , hence  $u$  must be a square. This implies that  $\text{Sel}_2 J \rightarrow L_2^\square$  is injective. Since  $Z$  is contained in  $\mathcal{O}_{L_2}^\square$ , assumption (3) implies that  $r(\text{Sel}_2 J) \cap Z = \emptyset$  as well. We can now apply Proposition 7.3 and conclude that  $C_p(\mathbb{Q}) = \{\infty, (0, 1), (0, -1)\}$ .

Now let  $F_p: u^p + v^p + w^p = 0$  denote the projective Fermat curve of exponent  $p$ . Then there is a non-constant morphism

$$\psi: F_p \longrightarrow C_p, \quad (u : v : w) \longmapsto (x, y) = \left( -\frac{uv}{w^2}, 2\frac{u^p}{w^p} + 1 \right).$$

So if  $P = (u : v : w) \in F_p(\mathbb{Q})$ , then either  $w = 0$  (if  $\psi(P) = \infty$ ) or else  $uv = 0$  (if  $\psi(P) = (0, \pm 1)$ ), so  $P$  is a trivial point.  $\square$

Note that by Remark 4.2, the criterion formulated in the proposition above is equivalent to what we would obtain when using the 2-Selmer group  $\text{Sel}_2 J_p$  instead of  $\mathcal{O}_L^\square$ .

We can improve on Proposition 7.4 a bit. Note that if  $u, u' \in \mathcal{O}_L$  are units with  $u$  positive (in the unique real embedding of  $L$ ), then the Hilbert symbol  $(u, u')_v$  is 1 for all places  $v$  distinct from the place  $\lambda$  above 2. The product formula for the Hilbert symbol implies that  $(u, u')_\lambda = 1$  as well. There are the two positive global units  $\lambda - 1$  and  $(1 - \lambda + \lambda^2)/(1 + \lambda)$ . Multiplying the latter by the square  $(1 + \lambda)^2$ , we obtain  $1 + \lambda^3$ . So if  $u \in \mathcal{O}_{L_2}^\times$  and we can show that  $(\lambda - 1, u)_\lambda = -1$  or  $(1 + \lambda^3, u)_\lambda = -1$ , then  $u$  cannot be in the image of  $\mathcal{O}_L^\times$ .

**Lemma 7.5** *We work in  $L_2 = \mathbb{Q}_2(\lambda)$  with  $\lambda^l = 2$  as before. If  $1 \leq m < l$ , then we have*

$$(\lambda - 1, 1 + \lambda^{2l-m})_\lambda = -1 \quad \text{and} \quad (1 + \lambda^3, 1 + \lambda^{2l-m})_\lambda = \begin{cases} 1 & \text{if } 3 \nmid m, \\ -1 & \text{if } 3 \mid m. \end{cases}$$

*Proof* We first consider  $\lambda - 1$ . Note that  $(-1, 1 + \lambda^{2l-m})_\lambda = (-1, 1 + 2^{2l-m})_2 = 1$ , so we can as well work with  $(1 - \lambda, 1 + \lambda^{2l-m})_\lambda$ . We have for  $n \geq (l - 1)/2$  that

$$(1 + \lambda^n)^2 - (1 - \lambda)(\lambda^n)^2 = 1 + \lambda^{2n+1} + \lambda^{l+n} \sim (1 + \lambda^{2n+1})(1 + \lambda^{l+n})$$

is a norm from  $L_2(\sqrt{1 - \lambda})$ , which implies that

$$(1 - \lambda, 1 + \lambda^{2l-m})_\lambda = (1 - \lambda, 1 + \lambda^{2l-(m+1)/2})_\lambda$$

when  $1 \leq m < l$  is odd. For even  $m$ , we have

$$1 \sim (1 + \lambda^{l-m/2})^2 = 1 + \lambda^{2l-m} + \lambda^{2l-m/2} \sim (1 + \lambda^{2l-m})(1 + \lambda^{2l-m/2}),$$

which implies that

$$(1 - \lambda, 1 + \lambda^{2l-m})_\lambda = (1 - \lambda, 1 + \lambda^{2l-m/2})_\lambda$$

when  $1 \leq m < l$  is even. An easy induction then shows that

$$(1 - \lambda, 1 + \lambda^{2l-m})_\lambda = (1 - \lambda, 1 + \lambda^{2l-1})_\lambda$$

for all  $1 \leq m < l$ . Finally, this last symbol is  $-1$ : clearly, an element is a norm from  $L_2(\sqrt{1 + \lambda^{2l-1}})$  if and only if it has the form  $x^2 - (1 + \lambda^{2l-1})y^2$ . Substituting  $(x, y) \leftarrow (\lambda^{l-1}x + y, y)$  and dividing by  $\lambda^{2l-2}$ , we see that norms have the form  $x^2 + \lambda xy - \lambda y^2$ . If the norm is integral, then  $x$  and  $y$  must be in  $\mathcal{O}_{L_2}$  as well. Considering

the equation

$$1 - \lambda = x^2 + \lambda xy - \lambda y^2$$

modulo  $\lambda^2$ , we see that it has no solution.

Now we consider  $1 + \lambda^3$ . For even  $1 \leq m < l$  we have in the same way as above that

$$(1 + \lambda^3, 1 + \lambda^{2l-m})_\lambda = (1 + \lambda^3, 1 + \lambda^{2l-m/2})_\lambda .$$

For  $n \geq (l - 1)/2$ , we have the norms

$$(1 + \lambda^n)^2 - (1 + \lambda^3)(\lambda^n)^2 = 1 - \lambda^{2n+3} + \lambda^{l+n} \sim (1 + \lambda^{2n+3})(1 + \lambda^{l+n}) ,$$

leading to

$$(1 + \lambda^3, 1 + \lambda^{2l-m})_\lambda = (1 - \lambda, 1 + \lambda^{2l-(m+3)/2})_\lambda$$

when  $1 \leq m < l$  is odd. By induction again, we see that

$$(1 + \lambda^3, 1 + \lambda^{2l-m})_\lambda = \begin{cases} (1 + \lambda^3, 1 + \lambda^{2l-1})_\lambda & \text{if } 3 \nmid m, \\ (1 + \lambda^3, 1 + \lambda^{2l-3})_\lambda & \text{if } 3 \mid m. \end{cases}$$

Let  $a \in L_2$  satisfy  $a^2 - a + \lambda^2 = 0$  (such  $a$  exist by Hensel's Lemma). Then  $1 + \lambda^3 = 1^2 + \lambda \cdot 1 \cdot a - \lambda \cdot a^2$  is a norm from  $L_2(\sqrt{1 + \lambda^{2l-1}})$ , so the first symbol is 1. In a similar way as before, we see that norms from  $L_2(\sqrt{1 + \lambda^{2l-3}})$  are of the form  $x^2 + \lambda^2 xy - \lambda y^2$ . A consideration modulo  $\lambda^4$  shows that this can never equal  $1 + \lambda^3$ , so the second symbol is  $-1$ .  $\square$

**Corollary 7.6** *Let  $p \geq 5$  be a prime and set  $L = \mathbb{Q}(2^{1/p})$  and  $L_2 = \mathbb{Q}_2(2^{1/p})$ . As before,  $r: \mathcal{O}_L^\square \rightarrow \mathcal{O}_{L_2}^\square$  denotes the canonical map. If*

- (1)  $p^2 \nmid 2^{p-1} - 1$ ,
- (2) *the class number of  $L$  is odd, and*
- (3)  $4 \nmid \lfloor \log_2 p \rfloor$  *or*  $z \notin \text{im}(r)$ , *where  $z$  is the last element listed in Lemma 7.2,*

*then Fermat's Last Theorem holds for the exponent  $p$ .*

*Proof* We only have to show that the third condition here implies that  $\text{im}(r) \cap Z = \emptyset$ . By Lemma 7.5, we have (with  $l = p$ )

$$(\lambda - 1, 1 + \lambda^{p+2})_\lambda = (\lambda - 1, 1 + \lambda^{2p-1})_\lambda = -1 ,$$

which implies that the first two elements of  $Z$  can never be images of global units. We also have  $(\lambda - 1, z)_\lambda = (-1)^{\lfloor \log_2 p \rfloor}$ , so we can also rule out  $z$  when  $\lfloor \log_2 p \rfloor$  is odd. So we can now assume that  $\lfloor \log_2 p \rfloor \equiv 2 \pmod 4$ . Then by Lemma 7.5 again,



we find that  $(1 + \lambda^3, z)_\lambda = -1$  (note that every other term in the sequence  $(p - 2^k)_k$  is divisible by 3), and we can again rule out  $z$ .  $\square$

**Corollary 7.7** *FLT holds for exponents 5, 7, 11, 13, 17, 19 and, assuming the Generalized Riemann Hypothesis, also for exponents 23, 29, 31, 37, 41, 43, 47, 53 and 59.*

*Proof* We use Magma [3] to check the assumptions (assuming GRH where indicated to speed up the computation of the class group). It turns out that the class group of  $\mathbb{Q}(2^{1/p})$  is trivial for all primes considered. We note that  $p = 17, 19, 23, 29, 31$  are the only primes  $p$  up to 59 that satisfy  $4 \mid \lfloor \log_2 p \rfloor$ , so we need a basis of  $\mathcal{O}_L^\square$  only for these primes; for the remaining ones it suffices to know that the class number is odd.  $\square$

*Remark 7.8* Computations show that the class group of  $\mathbb{Q}(2^{1/n})$  is trivial for all  $n \leq 50$  (assuming GRH for  $n \geq 20$ ), regardless whether  $n$  is prime or not. According to class group heuristics [20, Sect. 4.1], the 2-torsion in the class group of a number field with unit rank  $u$  should behave like the cokernel of a random linear map  $\mathbb{F}_2^{n+u} \rightarrow \mathbb{F}_2^n$  for large  $n$  (at least in absence of special effects leading to systematically occurring elements of order 2). Such a map is surjective with probability  $> 1 - 2^{-u}$ , so noting that  $u = (p - 1)/2$  in the case of interest, the ‘probability’ that the class number of  $L$  is odd for all  $p$  is  $> 1 - 2^{-29}$  (assuming we know it for  $p \leq 59$ ). See also [9].

We also remark that when the first condition  $p^2 \nmid 2^{p-1} - 1$  is not satisfied, the criterion does still work when we replace  $\mathcal{O}_L^\square$  by the larger subgroup  $L(\{p\}, 2)$  of  $L^\square$  represented by elements generating ideals of the form  $I_1^2 I_2$  with  $I_2$  supported on the ideals above  $p$ . In this case, however, we also have to check that the map to  $L_2^\square$  is injective. A similar remark applies to the case when the class group does have even order.

## 8 An Application to Certain Generalized Fermat Equations

Recall the following statement, which is Lemma 3.1 and Prop. 3.3 in [7].

**Proposition 8.1 (Dahmen and Siksek)** *Let  $p$  be an odd prime. If the only rational points on the curve*

$$C'_p: 5y^2 = 4x^p + 1$$

*are the obvious three (namely  $\infty, (1, 1)$  and  $(1, -1)$ ), then the only primitive integral solutions of the generalized Fermat equation  $x^5 + y^5 = z^p$  are the trivial ones:*

$$(x, y, z) = \pm(0, 1, 1), \pm(1, 0, 1), \pm(1, -1, 0).$$

Dahmen and Siksek verify the assumption on  $C'_p(\mathbb{Q})$  when  $p \in \{7, 19\}$  and, assuming GRH, also when  $p \in \{11, 13\}$ . We will use our approach to extend the range of primes  $p$  for which it can be shown that  $C'_p(\mathbb{Q})$  has only the obvious three rational points.

So we now consider the curves  $C'_l$ , with  $l = 2g + 1$  odd, but not necessarily prime. The corresponding étale algebra is still  $L = \mathbb{Q}(\lambda)$  with  $\lambda = 2^{1/l}$  (since  $C'_l$  is the quadratic twist by 5 of  $y^2 = 4x^l + 1$ ), but the descent map is now given on a point on the Jacobian with Mumford representation  $[a, b]$  by the class of  $-5a(\theta)$  (instead of  $-a(\theta)$ ) if the degree of  $a$  is odd.

It is still the case that  $C'_l$  has good reduction mod 2 (replacing  $y$  by  $2y + 1$  and dividing by 4 gives  $5(y^2 + y) = x^l - 1$ ) and that there is no nontrivial 2-torsion in  $J'_l(\mathbb{Q}_2)$  nor in  $J'_l(\mathbb{F}_2)$ , where  $J'_l$  denotes the Jacobian of  $C'_l$ . We therefore have a statement similar to Corollary 7.1. Note that we have again three residue disks, centered at  $\infty$ ,  $(1, 1)$  and  $(1, -1)$ , respectively.

If  $P_0 \in C'_l(\mathbb{Q})$ , then we write  $D_{P_0}$  for the residue disk centered at  $P_0$ . We let  $\varphi_{P_0}: D_0 \rightarrow D_{P_0}$  be a parameterization of  $D_{P_0}$  such that  $\varphi(0) = P_0$  (and such that  $i_\infty \circ \varphi_\infty$  is odd).

**Corollary 8.2** *We have*

$$q(i_\infty(D_\infty)) = q(i_\infty(\varphi_\infty(2\mathbb{Z}_2^\times))) \cup \{0\}$$

$$q(i_{(1,1)}(D_{(1,1)})) = q(i_{(1,1)}(\varphi_{(1,1)}(2\mathbb{Z}_2^\times \cup 4\mathbb{Z}_2^\times))) \cup \{0\}$$

*Proof* This again follows from Corollary 3.9. □

The main difference with the case discussed in the previous section is that, if  $l \geq 7$ , the two points  $(1, \pm 1)$  do not map to points of finite order in  $J'_l$  under the embedding that sends  $\infty$  to zero. So from now on, we assume that  $l \geq 7$ . Note that the rank of  $J'_5(\mathbb{Q})$  is zero (the 2-Selmer group is trivial), so it is almost immediate that  $C'_5(\mathbb{Q}) = \{\infty, (1, \pm 1)\}$ .

We first consider the image of  $C'_l(\mathbb{Q}_2)$  in  $J'_l(\mathbb{Q}_2)/2J'_l(\mathbb{Q}_2)$  under  $q \circ i_\infty$ .

**Lemma 8.3** *In terms of representatives in  $L_2^\times$ , we have*

- (1)  $\mu_2(q(i_\infty(D_\infty))) = \{1, 1 + \lambda^{2l-1}\}$ .
- (2)  $\mu_2(q(i_\infty(D_{(1,1)}))) = \{5(1 + \lambda^2), 5(1 + \lambda^2 + \lambda^{l+2})\}$ .

*Proof* By Corollary 8.2, we know that  $q(i_\infty(D_\infty)) = q(i_\infty(\varphi_\infty(2\mathbb{Z}_2^\times))) \cup \{0\}$ , where we can choose  $\varphi_\infty$  such that  $x(\varphi_\infty(t)) = 5t^{-2}$ . So let  $u \in \mathbb{Z}_2^\times$ , then  $\mu_2(i_\infty(\varphi_\infty(2u)))$  is represented by

$$5 \left( \frac{5}{4u^2} + \lambda^{-2} \right) = \left( \frac{5}{2u} \right)^2 \left( 1 + \frac{4u^2}{5} \lambda^{-2} \right) \sim 1 + \lambda^{2l-2} \sim 1 + \lambda^{2l-1} .$$

This proves (1).

Now let  $P \in D_{(1,1)}$ . We can choose  $\varphi_{(1,1)}$  such that  $x(\varphi_{(1,1)}(t)) = 1 + t$ . If  $u \in \mathbb{Z}_2^\times$ , then  $\mu_2(i_\infty(\varphi_{(1,1)}(2u)))$  is represented by

$$5(1 + 2u + \lambda^{-2}) \sim 5(1 + \lambda^2 + u\lambda^{l+2}) \sim 5(1 + \lambda^2 + \lambda^{l+2}),$$

and for any  $u \in \mathbb{Z}_2$ ,  $\mu_2(i_\infty(\varphi_{(1,1)}(4u)))$  is represented by

$$5(1 + 4u + \lambda^{-2}) \sim 5(1 + \lambda^2 + u\lambda^{2l+2}) \sim 5(1 + \lambda^2).$$

This proves (2). □

Now we consider the embedding  $i_{(1,1)}$ .

**Lemma 8.4** *In terms of representatives in  $L_2^\times$ , we have*

$$\mu_2(q(i_{(1,1)}(D_{(1,1)}))) = \{1, 1 + \lambda^{l+2}/(1 + \lambda^2), \sigma, \sigma'\},$$

where  $\sigma = \mu_2(Q)$  for the point  $Q \in J_l(\mathbb{Q}_2)$  such that  $2Q = \varphi_{(1,1)}(4)$  and  $\sigma' = \mu_2(Q')$  where  $2Q' = \varphi_{(1,1)}(-4)$ .

*Proof* We make use of Corollary 8.2 again, which tells us that it suffices to consider points  $P$  with  $x$ -coordinates  $1 + 2u$  or  $1 + 4u$ , where  $u \in \mathbb{Z}_2^\times$ . If  $x = 1 + 2u$ , then we have  $\pi_2(i_{(1,1)}(P)) = \pi_2(i_\infty(P)) - \pi_2(i_\infty((1, 1)))$ , which by the computation in the proof of Lemma 8.3 is represented by

$$5(1 + \lambda^2) \cdot 5(1 + \lambda^2 + \lambda^{l+2}) \sim 1 + \frac{\lambda^{l+2}}{1 + \lambda^2}.$$

If  $x = 1 + 4u$ , then  $i_{(1,1)}(P)$  is divisible by 2 in  $J'_l(\mathbb{Q}_2)$ , so we have to look at  $\pi_2(Q)$  where  $2Q = P$ , for suitable values of  $u$ . Since  $i_{(1,1)}(P) \in K_2 \setminus K_3$ , we have  $\tau(i_{(1,1)}(P)) = 1$ , so by Corollary 3.2,  $\pi_2(Q)$  depends only on  $u \pmod 4$ , so the two values  $u = 1$  and  $u = -1$  are sufficient. □

In practice, it appears that  $\sigma = \sigma'$  in all cases, which would be implied by the difference of the images of any pair chosen from the relevant points being divisible by 4. We know this difference is in  $K_3$ , but we did not exclude the possibility that it is only divisible by 2 and not by 4.

We can now formulate a criterion.

**Proposition 8.5** *Consider  $C'_l: 5y^2 = 4x^l + 1$ , with Jacobian  $J'_l$ , where  $l = 2g + 1$  is odd and  $l \geq 7$ . Recall that  $L = \mathbb{Q}(2^{1/l})$ ; let  $S \subseteq L^\square$  be a finite subgroup that contains the image of  $\text{Sel}_2 J'_l$ . Assume that*

- (1) *the canonical map  $S \hookrightarrow L^\square \rightarrow L_2^\square$  is injective, and*
- (2) *its image does not meet the set  $Z'$  consisting of the classes of*

$$1 + \lambda^{2l-1}, \quad 1 + \frac{\lambda^{l+2}}{1 + \lambda^2}, \quad \sigma, \quad \sigma'$$

*in  $L_2^\square$ .*

Then  $C'_l(\mathbb{Q}) = \{\infty, (1, 1), (1, -1)\}$ .

In particular, if  $l = p$  is a prime, then the generalized Fermat equation

$$x^5 + y^5 = z^p$$

has no nontrivial coprime integral solutions.

*Proof* Note that Lemmas 8.3 and 8.4 imply that  $Z' \cup \{1\}$  is the union of the sets  $Y$  occurring in Algorithm 4 when it is applied to the curve  $C'_l$ , so the assumptions imply that the algorithm will not return FAIL. (There cannot be any elements in  $\text{Sel}_2 C$  other than the images of the known points, since this would lead to a non-trivial intersection of  $Z'$  with the image of  $S$ .) The set returned by the algorithm can contain at most one point in each 2-adic residue disk. Since there are only three such disks, the known points must account for all rational points on  $C'_l$ .  $\square$

Computing  $\sigma$  and  $\sigma'$  for many values of  $l$ , it appears that their images in  $L_2^\square$  are represented uniformly by an infinite product

$$(1 + \lambda^{l+2})(1 + \lambda^{l+6})(1 + \lambda^{l+8})(1 + \lambda^{l+10})(1 + \lambda^{l+14})(1 + \lambda^{l+18})(1 + \lambda^{l+22}) \dots,$$

but it is not obvious which rule is behind the sequence  $(2, 6, 8, 10, 14, 18, 22, \dots)$ . However, extending it further and consulting the OEIS [19] gives exactly one hit, namely A036554, the sequence of ‘numbers  $n$  whose binary representation ends in an odd number of zeros’, i.e., such that  $v_2(n)$  is odd. So we propose the following.

*Conjecture 1*  $\mu_2(\sigma)$  (and also  $\mu_2(\sigma')$ ) is represented by

$$\prod_{n \geq 1, 2 \nmid v_2(n)} (1 + \lambda^{l+n}) \sim 1 + \frac{\lambda^l}{\prod_{k \geq 1} (1 + \lambda^{2^k})}.$$

We give a more concrete version of the criterion, following the considerations of Remark 4.2.

**Corollary 8.6** *Assume that  $l$  is prime and that  $l^2 \nmid 2^{l-1} - 1$ . Then a possible choice of the subgroup  $S$  in Proposition 8.5 is the subgroup of  $L(\{5\}, 2)$  consisting of elements mapping into the image of  $J'_l(\mathbb{Q}_5)$  in  $L_5^\square$ . In fact, the resulting criterion is equivalent to what would be obtained by taking  $S$  to be the image of  $\text{Sel}_2 J'_l$ .*

*Proof* As in the case discussed in the preceding section, the assumption  $l^2 \nmid 2^{l-1} - 1$  implies that the Tamagawa number at  $l$  is 1, so that we can reduce to  $\Sigma = \{2, 5\}$ . Furthermore, since 2 is totally ramified in  $L$  and  $L$  has odd degree, the norm of any element  $\alpha \in L^\times$  whose valuation with respect to the prime above 2 is odd will have odd 2-adic valuation and cannot be a square. This lets us reduce to  $L(\{5\}, 2)$ . Remark 4.2 now shows that using  $S$  is equivalent to using  $\text{Sel}_2 J'_l$  in the algorithm.  $\square$

We note that it is fairly easy to find  $S$ , given  $L(\{5\}, 2)$ , since the image of  $J'_l(\mathbb{Q}_5)$  in  $L_5^\square$  equals the image of  $J'_l(\mathbb{Q}_5)[2]$ , unless there are elements of order 4 in  $J'_l(\mathbb{Q}_5)$ .

We can easily exclude this by checking that the images of an  $\mathbb{F}_2$ -basis of  $J'_1(\mathbb{Q}_5)[2]$  are independent.

We carried out the computations necessary to test the criterion of Proposition 8.5 in the version of Corollary 8.6. This results in the following.

**Theorem 8.7** *For  $7 \leq p \leq 53$  prime, we have (assuming GRH when  $p \geq 23$ )*

$$C'_p(\mathbb{Q}) = \{\infty, (1, 1), (1, -1)\} .$$

*In particular, the generalized Fermat equation  $x^5 + y^5 = z^p$  has only the trivial coprime integral solutions.*

## 9 An ‘Elliptic Chabauty’ Example

In this section, we apply our approach to ‘Elliptic Curve Chabauty’. The curve in the following result comes up in the course of trying to find all primitive integral solutions to the Generalized Fermat Equation  $x^2 + y^3 = z^{25}$ . It is a hyperelliptic curve over  $\mathbb{Q}$  of genus 4; it can be shown that the Mordell-Weil group of its Jacobian has rank 4 (generators of a finite-index subgroup can be found), so that Chabauty’s method does not apply directly to the curve.

**Theorem 9.1** *Let  $C$  be the smooth projective curve given by the affine equation*

$$y^2 = 81x^{10} + 420x^9 + 1380x^8 + 1860x^7 + 3060x^6 - 66x^5 + 3240x^4 - 1740x^3 + 1320x^2 - 480x + 69 .$$

*If GRH holds, then  $C(\mathbb{Q})$  consists of the two points at infinity only.*

*Proof* As a first step, we compute the fake 2-Selmer set as in [5]. We obtain a one-element set (this requires local information only at the primes 2, 3 and 29). Using [14, Lemma 6.3], we then show that the points in  $C(\mathbb{Q}_2)$  whose image in  $L_2^\square/\mathbb{Q}_2^\square$  is the image of the unique element of the fake 2-Selmer set are those whose  $x$ -coordinate has 2-adic valuation  $\leq -3$ . This set is the union of two half residue disks (the maximal residue disks contain the points  $P$  such that  $v_2(x(P)) \leq -2$ ) that are mapped to each other by the hyperelliptic involution, so it is sufficient to consider just one of them, say the disk that contains  $P_0 = \infty_9$ , the point at infinity such that  $(y/x^5)(P_0) = 9$ .

The splitting field of the polynomial  $f$  on the right hand side of the curve equation contains three pairwise non-conjugate subfields  $k$  of degree 10 over which  $f$  is divisible by a monic polynomial  $g \in k[x]$  of degree 4. If  $P$  is any rational point on  $C$ , then it follows that  $g(x(P))$  is a square in  $k$  (this is because the image of  $P$  in the fake 2-Selmer set is the same as that of  $P_0$ ), so we obtain a point  $(\xi, \eta) \in H(k)$  with  $\xi \in \mathbb{Q}$  (or  $\xi = \infty$ ) where  $H$  is the smooth projective curve given by  $y^2 = g(x)$ .

We can parameterize the image of the residue disk around  $P_0$  by a pair of Laurent series  $((2t)^{-1}, \sqrt{g((2t)^{-1}})$  (where we can take the square root to have leading term  $t^{-2}/4$ ). Since  $H(k)$  contains the two points at infinity,  $H$  is isomorphic to an elliptic curve  $E$  over  $k$ ; we take the isomorphism so that it sends  $\infty_1 \in H(k)$  to the origin of  $E$ . We then obtain a Laurent series  $\xi(t) \in k((t))$  that gives the  $x$ -coordinate of the image on  $E$  of the point whose parameter is  $t$ .

For the following, we take  $k$  to be the field generated by a root of

$$x^{10} + 75x^6 - 50x^5 + 100x^3 + 625x^2 + 1250x + 645 ;$$

the polynomial  $g$  and the curves  $H$  and  $E$  are taken with respect to this field. We next compute the 2-Selmer group of  $E$  over  $k$ . There is exactly one point of order 2 in  $E(k)$ , which means that we have to work with a quadratic extension of  $k$ . This is where we use GRH, which allows us to find the relevant arithmetic information for this field of degree 20 faster. The Selmer group has  $\mathbb{F}_2$ -dimension 6 (so the bound for the rank of  $E(k)$  is 5). We check that it injects into  $E(k_2)/2E(k_2)$ , where  $k_2 = k \otimes_{\mathbb{Q}} \mathbb{Q}_2$ ; note that this splits as a product of two extensions of  $\mathbb{Q}_2$ , both of ramification index 2 and one of residue class degree 1, the other of residue class degree 4.

In the context of our method, we consider the curve that is the (desingularization of) the curve over  $\mathbb{Q}$  in  $A = R_{k/\mathbb{Q}}E$  (the latter denotes the Weil restriction of scalars) that corresponds to the set of points on  $H$  whose  $x$ -coordinate is rational. We have  $E(k_2) \cong A(\mathbb{Q}_2)$ , so we can use arithmetic on  $E$  over  $k$  and its completions for the computations. We check that  $n_{\text{tors}} = 1$  (the map from  $E(k_2)[2]$  to  $E(k_2)/2E(k_2)$  is injective). By Lemma 3.7, a suitable value of  $m$  is  $m = 4$ , provided  $5 \geq n_4$  in the notation of the lemma. Note that in this situation halving points is easy, since doubling a point corresponds to an explicit map of degree 4 on the  $x$ -coordinate. If  $P$  is in our half residue disk, then  $i(P) + T$  (where  $T$  is the point of order 2 in  $E(k) = A(\mathbb{Q})$ ) is not divisible by 2, and its image in  $A(\mathbb{Q}_2)/2A(\mathbb{Q}_2)$  is the same as that of  $T$ . So we only have to determine  $q(i(P))$  for a suitable selection of points  $P$  in our disk. We write  $P_t$  for the point corresponding to the parameter  $t \in 2\mathbb{Z}_2$ .

We compute  $Y_t = q(i(P_t))$  for  $t \in \{\pm 4, \pm 8, \pm 16\}$ . Note that this can be done solely in terms of the  $x$ -coordinate  $\xi(t)$ . We find that  $\tau(i(P_t)) = v_2(t) - 1$  for these values and that  $Y_{-t} = Y_t$ . This implies that  $n_m = m - 1$  for  $2 \leq m \leq 4$  and that  $q$  of the disk in question is the union  $Y_4 \cup Y_8 \cup Y_{16}$ . In fact, it turns out that this union is equal to  $Y_8$ , and we verify that  $Y_8$  meets the image of the 2-Selmer group only in the image of the global torsion. By Theorem 2.6, this then implies that there can be no other point than  $P_0$  in our (half) residue disk. The claim follows. □

We note that the two other possible choices of  $k$  also lead to elliptic curves with a 2-Selmer rank of 5 (this is unconditional for one of the choices, where the curve  $E$  happens to have full 2-torsion over  $k$ ), but for these other fields, the condition that the image of the disk under  $q$  meets the image of the Selmer group only in the image of the global torsion is not satisfied. We also remark that we have been unable to find five independent points in  $E(k)$  (for any of the three possible choices of  $k$  and  $E$ ), so that we could not apply the standard Elliptic Curve Chabauty method.

Another application of our ‘Selmer group Chabauty’ approach in the setting of Elliptic Curve Chabauty was made in [8]. We use this to show that there are no unexpected points on the elliptic curve  $X_0(11)$  defined over certain number fields of degree 12 and such that the image under the  $j$ -map is in  $\mathbb{Q}$ . This is a vital step in the proof that the only nontrivial primitive integral solutions of the Generalized Fermat Equation  $x^2 + y^3 = z^{11}$  are  $(x, y, z) = (\pm 3, -2, 1)$ . The situation is similar to what happens for the example presented here: we can compute the 2-Selmer group of  $X_0(11)$  over the fields of interest, but we are unable to produce enough independent points to meet the upper bound on the rank, so we cannot apply the standard method.

**Acknowledgements** I would like to thank Bjorn Poonen for useful discussions and MIT for its hospitality during a visit of two weeks in May 2015, when these discussions took place. All computations were done using the computer algebra system Magma [3].

## References

1. J.S. Balakrishnan, R.W. Bradshaw, K.S. Kedlaya, Explicit Coleman integration for hyperelliptic curves, in *Algorithmic Number Theory*. Lecture Notes in Computer Science, vol. 6197 (Springer, Berlin, 2010), pp. 16–31
2. M. Bhargava, B.H. Gross, The average size of the 2-Selmer group of Jacobians of hyperelliptic curves having a rational Weierstrass point, in *Automorphic Representations and L-Functions*. Tata Institute of Fundamental Research Studies in Mathematics, vol. 22 (Tata Institute of Fundamental Research, Mumbai, 2013), pp. 23–91
3. W. Bosma, J. Cannon, C. Playoust, The Magma algebra system. I. The user language. *J. Symb. Comput.* **24**, 235–265 (1997)
4. N. Bruin, E.V. Flynn, Exhibiting SHA[2] on hyperelliptic Jacobians. *J. Number Theory* **118**, 266–291 (2006)
5. N. Bruin, M. Stoll, Two-cover descent on hyperelliptic curves. *Math. Comput.* **78**, 2347–2370 (2009)
6. N. Bruin, M. Stoll, The Mordell–Weil sieve: proving non-existence of rational points on curves. *LMS J. Comput. Math.* **13**, 272–306 (2010)
7. S.R. Dahmen, S. Siksek, Perfect powers expressible as sums of two fifth or seventh powers. *Acta Arith.* **164**, 65–100 (2014)
8. N. Freitas, B. Naskręcki, M. Stoll, The generalized Fermat equation with exponents 2, 3,  $n$ . Preprint (2017). arXiv:1703.05058 [math.NT]
9. W. Ho, A. Shankar, I. Varma, Odd degree number fields with odd class number. Preprint (2016). arXiv:1603.06269
10. W.G. McCallum, On the method of Coleman and Chabauty. *Math. Ann.* **299**, 565–596 (1994)
11. B. Poonen, M. Stoll, Most odd degree hyperelliptic curves have only one rational point. *Ann. Math. (2)* **180**, 1137–1166 (2014)
12. E.F. Schaefer, 2-Descent on the Jacobians of hyperelliptic curves. *J. Number Theory* **51**, 219–232 (1995)
13. E.F. Schaefer, M. Stoll, How to do a  $p$ -descent on an elliptic curve. *Trans. Am. Math. Soc.* **356**, 1209–1231 (2004)
14. M. Stoll, Implementing 2-descent for Jacobians of hyperelliptic curves. *Acta Arith.* **98**, 245–277 (2001)

15. M. Stoll, Independence of rational points on twists of a given curve. *Compos. Math.* **142**, 1201–1214 (2006)
16. M. Stoll, Finite descent obstructions and rational points on curves. *Algebra Number Theory* **1**, 349–391 (2007)
17. M. Stoll, Uniform bounds for the number of rational points on hyperelliptic curves of small Mordell-Weil rank. *J. Eur. Math. Soc. Preprint*. arXiv:1307.1773 (to appear)
18. R. Taylor, A. Wiles, Ring-theoretic properties of certain Hecke algebras. *Ann. Math. (2)* **141**, 553–572 (1995)
19. The On-Line Encyclopedia of Integer Sequences, <http://oeis.org>
20. A. Venkatesh, J.S. Ellenberg, Statistics of number fields and function fields, in *Proceedings of the International Congress of Mathematicians*, vol. II (Hindustan Book Agency, New Delhi, 2010)
21. A. Wiles, Modular elliptic curves and Fermat’s last theorem. *Ann. Math. (2)* **141**, 443–551 (1995)



# An Explicit Theory of Heights for Hyperelliptic Jacobians of Genus Three



Michael Stoll

**Abstract** We develop an explicit theory of Kummer varieties associated to Jacobians of hyperelliptic curves of genus 3, over any field  $k$  of characteristic  $\neq 2$ . In particular, we provide explicit equations defining the Kummer variety  $\mathcal{K}$  as a subvariety of  $\mathbb{P}^7$ , together with explicit polynomials giving the duplication map on  $\mathcal{K}$ . A careful study of the degenerations of this map then forms the basis for the development of an explicit theory of heights on such Jacobians when  $k$  is a number field. We use this input to obtain a good bound on the difference between naive and canonical height, which is a necessary ingredient for the explicit determination of the Mordell-Weil group. We illustrate our results with two examples.

**Keywords** Kummer variety • Hyperelliptic curve • Genus 3 • Canonical height

*Subject Classifications:* 14H40, 14H45, 11G10, 11G50, 14Q05, 14Q15

## 1 Introduction

The goal of this paper is to take up the approaches used to deal with Jacobians and Kummer surfaces of curves of genus 2 by Cassels and Flynn [4] and by the author [14, 16] and extend them to hyperelliptic curves of genus 3. We always assume that the base field  $k$  has characteristic  $\neq 2$ . A hyperelliptic curve  $\mathcal{C}$  over  $k$  of genus 3 is then given by an equation of the form  $y^2 = f(x)$ , where  $f$  is a squarefree polynomial of degree 7 or 8 with coefficients in  $k$ ; we take  $\mathcal{C}$  to be the smooth projective curve determined by this affine equation. We denote the Jacobian variety of  $\mathcal{C}$  by  $\mathcal{J}$ . Identifying points with their negatives on  $\mathcal{J}$ , we obtain the Kummer variety of  $\mathcal{J}$ . It is known that the morphism  $\mathcal{J} \rightarrow \mathbb{P}^7$  given

---

M. Stoll (✉)

Mathematisches Institut, Universität Bayreuth, 95440 Bayreuth, Germany  
e-mail: [Michael.Stoll@uni-bayreuth.de](mailto:Michael.Stoll@uni-bayreuth.de)

by the linear system  $|2\Theta|$  on  $\mathcal{J}$  (where  $\Theta$  denotes the theta divisor) induces an isomorphism of the Kummer variety with the image of  $\mathcal{J}$  in  $\mathbb{P}^7$ ; we denote the image by  $\mathcal{K} \subset \mathbb{P}^7$ . Our first task is to find a suitable basis of the Riemann-Roch space  $L(2\Theta)$  and to give explicit equations defining  $\mathcal{K}$ , thereby completing earlier work by Stubbs [18], Duquesne [5] and Müller [9, 11]. To this end, we make use of the canonical identification of  $\mathcal{J}$  with  $\mathcal{X} = \text{Pic}^4(\mathcal{C})$  and realize the complement of  $\Theta$  in  $\mathcal{X}$  as the quotient of an explicit 6-dimensional variety  $\mathcal{V}$  in  $\mathbb{A}^{15}$  by the action of a certain group  $\Gamma$ . This allows us to identify the ring of regular functions on  $\mathcal{X} \setminus \Theta$  with the ring of  $\Gamma$ -invariants in the coordinate ring of  $\mathcal{V}$ . In this way, we obtain a natural basis of  $L(2\Theta)$ , and we find the quadric and the 34 quartics that define  $\mathcal{K}$ ; see Sect. 2. We give the relation between the coordinates chosen here and those used in previous work and discuss how transformations of the curve equation induced by the action of  $\text{GL}(2)$  on  $(x, z)$  act on our coordinates; see Sect. 3. We then give a recipe that allows to decide whether a  $k$ -rational point on  $\mathcal{K}$  comes from a  $k$ -rational point on  $\mathcal{J}$  (Sect. 4).

The next task is to describe the maps  $\mathcal{K} \rightarrow \mathcal{K}$  and  $\text{Sym}^2 \mathcal{K} \rightarrow \text{Sym}^2 \mathcal{K}$  induced by multiplication by 2 and by  $\{P, Q\} \mapsto \{P + Q, P - Q\}$  on  $\mathcal{J}$ . We use the approach followed in [14]: we consider the action of a double cover of the 2-torsion subgroup  $\mathcal{J}[2]$  on the coordinate ring of  $\mathbb{P}^7$ . This induces an action of  $\mathcal{J}[2]$  itself on forms of even degree. We use the information obtained on the various eigenspaces and the invariant subspaces in particular to obtain an explicit description of the duplication map  $\underline{\delta}$  and of the sum-and-difference map on  $\mathcal{K}$ . The study of the action of  $\mathcal{J}[2]$  is done in Sects. 5 and 6; the results on the duplication map and on the sum-and-difference map are obtained in Sects. 7 and 8, respectively. In Sect. 9, we then study the degeneration of these maps that occur when we allow the curve to acquire singularities. This is relevant in the context of bad reduction and is needed as input for the results on the height difference bound.

We then turn to the topic motivating our study, which is the canonical height  $\hat{h}$  on the Jacobian, and, in particular, a bound on the difference  $h - \hat{h}$  between naive and canonical height. Such a bound is a necessary ingredient for the determination of generators of the Mordell-Weil group  $\mathcal{J}(k)$  (where  $k$  now is a number field; in practice, usually  $k = \mathbb{Q}$ ), given generators of a finite-index subgroup. The difference  $h - \hat{h}$  can be expressed in terms of the local ‘loss of precision’ under  $\underline{\delta}$  at the various primes of bad reduction and the archimedean places of  $k$ . In analogy with [14], we obtain an estimate for this local ‘loss of precision’ in terms of the valuation of the discriminant of  $f$ . This is one of the main results of Sect. 10, together with a statement on the structure of the local ‘height correction function’, which is analogous to that obtained in [16, Thm. 4.1]. These results allow us to obtain reasonable bounds for the height difference. We illustrate this by determining generators of the Mordell-Weil group of the Jacobian of the curve  $y^2 = 4x^7 - 4x + 1$ . We then use this result to determine the set of integral solutions of the equation  $y^2 - y = x^7 - x$ , using the method of [3]; see Sect. 11.

In addition, we show in Sect. 12 how one can obtain better bounds (for a modified naive height) when the polynomial defining the curve is not primitive. As an example, we determine explicit generators of the Mordell-Weil group of the

Jacobian of the curve given by the binomial coefficient equation

$$\binom{y}{2} = \binom{x}{7}.$$

We have made available at [17] files that can be read into Magma [1] and provide explicit representations of the quartics defining the Kummer variety, the matrices giving the action of 2-torsion points, the polynomials defining the duplication map and the matrix of bi-quadratic forms related to the ‘sum-and-difference map’. We have also made available the file `Kum3-verification.magma`, which, when loaded into Magma, will perform the computations necessary to verify a number of claims made throughout the paper. These claims are marked by a star, like this★.

## 2 The Kummer Variety

We consider a hyperelliptic curve of genus 3 over a field  $k$  of characteristic different from 2, given by the affine equation

$$\mathcal{C}: y^2 = f_8x^8 + f_7x^7 + \dots + f_1x + f_0 = f(x),$$

where  $f$  is a squarefree polynomial of degree 7 or 8. (We do not assume that  $\mathcal{C}$  has a Weierstrass point at infinity, which would correspond to  $f$  having degree 7.) Let  $F(x, z)$  denote the octic binary form that is the homogenization of  $f$ ;  $F$  is squarefree. Then the equation  $y^2 = F(x, z)$  defines a smooth model of  $\mathcal{C}$  in the weighted projective plane  $\mathbb{P}_{1,4,1}^2$ . Here  $x$  and  $z$  have weight 1 and  $y$  has weight 4. We denote the hyperelliptic involution on  $\mathcal{C}$  by  $\iota$ , so that  $\iota(x : y : z) = (x : -y : z)$ .

As in the introduction, we denote the Jacobian variety of  $\mathcal{C}$  by  $\mathcal{J}$ . We would like to find an explicit version of the map

$$\mathcal{J} \longrightarrow \mathbb{P}^7$$

given by the linear system of twice the theta divisor; it embeds the Kummer variety  $\mathcal{J}/\{\pm 1\}$  into  $\mathbb{P}^7$ . We denote the image by  $\mathcal{K}$ .

We note that the canonical class  $\mathfrak{W}$  on  $\mathcal{C}$  has degree 4. Therefore  $\mathcal{J} = \text{Pic}_{\mathcal{C}}^0$  is canonically isomorphic to  $\mathcal{X} = \text{Pic}_{\mathcal{C}}^4$ , with the isomorphism sending  $\mathfrak{D}$  to  $\mathfrak{D} + \mathfrak{W}$ . Then the map induced by  $\iota$  on  $\mathcal{X}$  corresponds to multiplication by  $-1$  on  $\mathcal{J}$ . There is a canonical theta divisor on  $\text{Pic}_{\mathcal{C}}^0$  whose support consists of the divisor classes of the form  $[(P_1) + (P_2)] - \mathfrak{m}$ , where  $\mathfrak{m}$  is the class of the polar divisor  $(x)_{\infty}$ ; we have  $\mathfrak{W} = 2\mathfrak{m}$ . The support of the theta divisor is the locus of points on  $\mathcal{X}$  that are not represented by divisors in general position, where an effective divisor  $\mathfrak{D}$  on  $\mathcal{C}$  is *general position* unless there is a point  $P \in \mathcal{C}$  such that  $\mathfrak{D} \geq (P) + (\iota P)$ . This can be seen as follows. The image on  $\mathcal{X}$  of a point  $[(P_1) + (P_2)] - \mathfrak{m}$  on the theta divisor is represented by all effective divisors of the form  $(P_1) + (P_2) + (P) + (\iota P)$  for an arbitrary point  $P \in \mathcal{C}$ . If  $P_2 \neq \iota P_1$ , then the Riemann-Roch Theorem implies that

the linear system containing these divisors is one-dimensional, and so *all* divisors representing our point on  $\mathcal{X}$  have this form; in particular, there is no representative divisor in general position. If  $P_2 = \iota P_1$ , then the linear system has dimension 2 and consists of all divisors of the form  $(P) + (\iota P) + (P') + (\iota P')$ , none of which is in general position.

We identify  $\mathcal{J}$  and  $\mathcal{X}$ , and we denote the theta divisor on  $\mathcal{J}$  and its image on  $\mathcal{X}$  by  $\Theta$ . We write  $L(n\Theta)$  for the Riemann-Roch space  $L(\mathcal{X}, n\Theta) \cong L(\mathcal{J}, n\Theta)$ , where  $n \geq 0$  is an integer. It is known that  $\dim L(n\Theta) = n^3$ . Since  $\Theta$  is symmetric, the negation map acts on  $L(n\Theta)$  (via  $\phi \mapsto (P \mapsto \phi(-P))$ ), and it makes sense to speak of even and odd functions in  $L(n\Theta)$  (with respect to this action). We write  $L(n\Theta)^+$  for the subspace of even functions. It is known that  $\dim L(n\Theta)^+ = n^3/2 + 4$  for  $n$  even and  $\dim L(n\Theta)^+ = (n^3 + 1)/2$  for  $n$  odd.

We can parameterize effective degree 4 divisors in general position as follows. Any such divisor  $\mathcal{D}$  is given by a binary quartic form  $A(x, z)$  specifying the image of  $\mathcal{D}$  on  $\mathbb{P}^1$  under the hyperelliptic quotient map  $\pi: \mathcal{C} \rightarrow \mathbb{P}^1, (x : y : z) \mapsto (x : z)$ , together with another quartic binary form  $B(x, z)$  such that  $y = B(x, z)$  on the points in  $\mathcal{D}$ , with the correct multiplicity. (Note that by the ‘general position’ condition,  $y$  is uniquely determined by  $x$  and  $z$  for each point in the support of  $\mathcal{D}$ .) More precisely, we must have that

$$B(x, z)^2 - A(x, z)C(x, z) = F(x, z) \tag{1}$$

for a suitable quartic binary form  $C(x, z)$ . We then have a statement analogous to that given in [4, Chap. 4] for  $\text{Pic}^3$  of a curve of genus 2; see Lemma 2.1 below. Before we can formulate it, we need some notation.

We let  $Q$  be the ternary quadratic form  $x_2^2 - x_1x_3$ . We write

$$D = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 0 \end{pmatrix} \tag{2}$$

for the associated symmetric matrix (times 2) and

$$\Gamma = \text{SO}(Q) = \{\gamma \in \text{SL}(3) : \gamma D \gamma^T = D\};$$

then  $-\Gamma = \text{O}(Q) \setminus \text{SO}(Q)$ , and  $\pm\Gamma = \text{O}(Q)$ . We have the following elements in  $\Gamma$  (for arbitrary  $\lambda$  and  $\mu$  in the base field):

$$t_\lambda = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda^{-1} \end{pmatrix}, \quad n_\mu = \begin{pmatrix} 1 & \mu & \mu^2 \\ 0 & 1 & 2\mu \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix};$$

these elements generate  $\Gamma$ .

**Lemma 2.1** *Two triples  $(A, B, C)$  and  $(A', B', C')$  satisfying (1) specify the same point on  $\mathcal{X}$  if and only if  $(A', B', C') = (A, B, C)\gamma$  for some  $\gamma \in \Gamma$ . They represent*

opposite points (with respect to the involution on  $\mathcal{X}$  induced by  $\iota$ ) if and only if the relation above holds for some  $\gamma \in -\Gamma$ .

*Proof* We first show that two triples specifying the same point are in the same  $\Gamma$ -orbit. Let  $\mathfrak{D}$  and  $\mathfrak{D}'$  be the effective divisors of degree 4 given by  $A(x, z) = 0, y = B(x, z)$  and by  $A'(x, z) = 0, y = B'(x, z)$ , respectively. By assumption,  $\mathfrak{D}$  and  $\mathfrak{D}'$  are linearly equivalent, and they are both in general position. If  $\mathfrak{D}$  and  $\mathfrak{D}'$  share a point  $P$  in their supports, then subtracting  $P$  from both  $\mathfrak{D}$  and  $\mathfrak{D}'$ , we obtain two effective divisors of degree 3 in general position that are linearly equivalent. Since such divisors are non-special, they must be equal, hence  $\mathfrak{D} = \mathfrak{D}'$ . So  $A$  and  $A'$  agree up to scaling, and  $B' - B$  is a multiple of  $A$ :

$$A' = \lambda A, \quad B' = B + \mu A, \quad C' = \lambda^{-1}(C + 2\mu B + \mu^2 A);$$

then  $(A', B', C') = (A, B, C)n_{\mu}t_{\lambda}$ . So we can now suppose that the supports of  $\mathfrak{D}$  and  $\mathfrak{D}'$  are disjoint. Then, denoting by  $\iota\mathfrak{D}'$  the image of  $\mathfrak{D}'$  under the hyperelliptic involution,  $\mathfrak{D} + \iota\mathfrak{D}'$  is a divisor of degree 8 in general position, which is in twice the canonical class, so it is linearly equivalent to  $4m$ . Since the Riemann-Roch space of that divisor on  $\mathcal{C}$  is generated (in terms of the affine coordinates obtained by setting  $z = 1$ ) by  $1, x, x^2, x^3, x^4, y$ , there is a function of the form  $y - \tilde{B}(x, 1)$  with  $\tilde{B}$  homogeneous of degree 4 that has divisor  $\mathfrak{D} + \iota\mathfrak{D}' - 4m$ . Equivalently,  $\mathfrak{D} + \iota\mathfrak{D}'$  is the intersection of  $\mathcal{C}$  with the curve given by  $y = \tilde{B}(x, z)$ . This implies that  $\tilde{B}^2 - F$  is a constant times  $AA'$ . Up to scaling  $A'$  and  $C'$  by  $\lambda$  and  $\lambda^{-1}$  for a suitable  $\lambda$  (this corresponds to acting on  $(A', B', C')$  by  $t_{\lambda} \in \Gamma$ ), we have

$$\tilde{B}^2 - AA' = F,$$

so that  $(A, \tilde{B}, A')$  corresponds to  $\mathfrak{D}$  and  $(A', -\tilde{B}, A)$  corresponds to  $\mathfrak{D}'$ . The argument above (for the case  $\mathfrak{D} = \mathfrak{D}'$ ) shows that  $(A, B, C)$  and  $(A, \tilde{B}, A')$  are in the same  $\Gamma$ -orbit, and the same is true of  $(A', B', C')$  and  $(A', -\tilde{B}, A)$ . Finally,

$$(A', -\tilde{B}, A) = (A, \tilde{B}, A')w.$$

Conversely, it is easy to see that the generators of  $\Gamma$  given above do not change the linear equivalence class of the associated divisor: the first two do not even change the divisor, and the third replaces  $\mathfrak{D}$  by the linearly equivalent divisor  $\iota\mathfrak{D}'$ , where  $\mathfrak{D} + \mathfrak{D}' \sim 2\mathfrak{W}$  is the divisor of  $y - B(x, z)$  on  $\mathcal{C}$ .

For the last statement, it suffices to observe that  $(A, -B, C)$  gives the point opposite to that given by  $(A, B, C)$ ; the associated matrix is  $-t_{-1} \in -\Gamma$ .  $\square$

We write  $A, B, C$  as follows.

$$\begin{aligned} A(x, z) &= a_4x^4 + a_3x^3z + a_2x^2z^2 + a_1xz^3 + a_0z^4 \\ B(x, z) &= b_4x^4 + b_3x^3z + b_2x^2z^2 + b_1xz^3 + b_0z^4 \\ C(x, z) &= c_4x^4 + c_3x^3z + c_2x^2z^2 + c_1xz^3 + c_0z^4 \end{aligned}$$

and use  $a_0, \dots, a_4, b_0, \dots, b_4, c_0, \dots, c_4$  as affine coordinates on  $\mathbb{A}^{15}$ . We arrange these coefficients into a matrix

$$L = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & a_4 \\ b_0 & b_1 & b_2 & b_3 & b_4 \\ c_0 & c_1 & c_2 & c_3 & c_4 \end{pmatrix}. \tag{3}$$

Then  $\gamma \in \pm\Gamma$  acts on  $\mathbb{A}^{15}$  via multiplication by  $\gamma^\top$  on the left on  $L$ . Since there is a multiplicative group sitting inside  $\Gamma$  acting by  $(A, B, C) \cdot \lambda = (\lambda A, B, \lambda^{-1}C)$ , any  $\Gamma$ -invariant polynomial must be a linear combination of monomials having the same number of  $a_i$  and  $c_j$ . Hence in any term of a homogeneous  $\Gamma$ -invariant polynomial of degree  $d$ , the number of factors  $b_i$  has the same parity as  $d$ . This shows that such a  $\Gamma$ -invariant polynomial is even with respect to  $\iota$  if  $d$  is even, and odd if  $d$  is odd.

It is not hard to see that there are no  $\Gamma$ -invariant polynomials of degree 1: by the above, they would have to be a linear combination of the  $b_i$ , but the involution  $(A, B, C) \mapsto (C, -B, A) = (A, B, C)w$  negates all the  $b_i$ . It is also not hard to check that the space of invariants of degree 2 is spanned by the coefficients of the quadratic form

$$B_l^2 - A_l C_l \in \text{Sym}^2\langle x_0, x_1, x_2, x_3, x_4 \rangle,$$

where

$$\begin{aligned} A_l &= a_0x_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 \\ B_l &= b_0x_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 \\ C_l &= c_0x_0 + c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 \end{aligned}$$

are linear forms in five variables. We write

$$B_l^2 - A_l C_l = \sum_{0 \leq i < j \leq 4} \eta_{ij} x_i x_j,$$

so that  $\eta_{ii} = b_i^2 - a_i c_i$  and for  $i < j$ ,  $\eta_{ij} = 2b_i b_j - a_i c_j - a_j c_i$ . Up to scaling, the quadratic form corresponds to the symmetric matrix

$$L^\top D L = \begin{pmatrix} 2\eta_{00} & \eta_{01} & \eta_{02} & \eta_{03} & \eta_{04} \\ \eta_{01} & 2\eta_{11} & \eta_{12} & \eta_{13} & \eta_{14} \\ \eta_{02} & \eta_{12} & 2\eta_{22} & \eta_{23} & \eta_{24} \\ \eta_{03} & \eta_{13} & \eta_{23} & 2\eta_{33} & \eta_{34} \\ \eta_{04} & \eta_{14} & \eta_{24} & \eta_{34} & 2\eta_{44} \end{pmatrix}, \tag{4}$$

and the image  $\mathcal{Q}$  of the map  $q: \mathbb{A}^{15} \rightarrow \text{Sym}^2 \mathbb{A}^5$  given by this matrix consists of the matrices of rank at most 3; it is therefore defined by the 15 different quartics obtained as  $4 \times 4$ -minors of this matrix.

Scaling  $x$  by  $\lambda$  corresponds to scaling  $a_j, b_j, c_j$  by  $\lambda^j$ . This introduces another grading on the coordinate ring of our  $\mathbb{A}^{15}$ ; we call the corresponding degree the *weight*. We then have  $\text{wt}(a_j) = \text{wt}(b_j) = \text{wt}(c_j) = j$  and therefore  $\text{wt}(\eta_{ij}) = i + j$ . The 15 quartics defining  $\mathcal{Q}$  have weights

$$12, 13, 14, 14, 15, 15, 16, 16, 16, 17, 17, 18, 18, 19, 20 .$$

We will reserve the word *degree* for the degree in terms of the  $\eta_{ij}$ ; then it makes sense to set  $\text{deg}(a_j) = \text{deg}(b_j) = \text{deg}(c_j) = \frac{1}{2}$ .

We let  $\mathcal{V} \subset \mathbb{A}^{15}$  be the affine variety given by (1). The defining equations of  $\mathcal{V}$  then read

$$\begin{aligned} b_0^2 - a_0c_0 &= f_0 \\ 2b_0b_1 - (a_0c_1 + a_1c_0) &= f_1 \\ 2b_0b_2 + b_1^2 - (a_0c_2 + a_1c_1 + a_2c_0) &= f_2 \\ 2b_0b_3 + 2b_1b_2 - (a_0c_3 + a_1c_2 + a_2c_1 + a_3c_0) &= f_3 \\ 2b_0b_4 + 2b_1b_3 + b_2^2 - (a_0c_4 + a_1c_3 + a_2c_2 + a_3c_1 + a_4c_0) &= f_4 \\ 2b_1b_4 + 2b_2b_3 - (a_1c_4 + a_2c_3 + a_3c_2 + a_4c_1) &= f_5 \\ 2b_2b_4 + b_3^2 - (a_2c_4 + a_3c_3 + a_4c_2) &= f_6 \\ 2b_3b_4 - (a_3c_4 + a_4c_3) &= f_7 \\ b_4^2 - a_4c_4 &= f_8 . \end{aligned}$$

In terms of the  $\eta_{ij}$ , we have

$$\begin{aligned} \eta_{00} = f_0, \quad \eta_{01} = f_1, \quad \eta_{02} + \eta_{11} = f_2, \quad \eta_{03} + \eta_{12} = f_3, \quad \eta_{04} + \eta_{13} + \eta_{22} = f_4, \\ \eta_{14} + \eta_{23} = f_5, \quad \eta_{24} + \eta_{33} = f_6, \quad \eta_{34} = f_7, \quad \eta_{44} = f_8 ; \end{aligned}$$

in particular, the image of  $\mathcal{V}$  under  $q$  is a linear ‘slice’  $\mathcal{W}$  of  $\mathcal{Q}$ , cut out by the nine linear equations above (recall that the  $\eta_{ij}$  are coordinates on the ambient space  $\text{Sym}^2 \mathbb{A}^5$  of  $\mathcal{Q}$ ). It is then natural to define  $\text{deg}(f_j) = 1$  and  $\text{wt}(f_j) = j$ .

By Lemma 2.1, the quotient  $\mathcal{V}/\Gamma$  of  $\mathcal{V}$  by the action of  $\Gamma$  can be identified with  $\mathcal{U} := \mathcal{X} \setminus \Theta$ , the complement of the theta divisor in  $\mathcal{X}$ . Since the map  $q$  is given by  $\pm\Gamma$ -invariants, we obtain a surjective morphism  $\mathcal{X} \setminus \kappa(\Theta) \rightarrow \mathcal{W}$ . We will see that it is actually an isomorphism.

Functions in the Riemann-Roch space  $L(n\Theta)$  will be represented by  $\Gamma$ -invariant polynomials in the  $a_i, b_i, c_i$ . Similarly, functions in the even part  $L(n\Theta)^+$  of this space are represented by  $\pm\Gamma$ -invariant polynomials. A  $\Gamma$ -invariant polynomial that is homogeneous of degree  $n$  in the  $a_i, b_i, c_i$  will conversely give rise to a function in  $L(n\Theta)$ . Modulo the relations defining  $\mathcal{V}$ , there are six independent such invariants

of degree 2. We choose

$$\eta_{02}, \eta_{03}, \eta_{04}, \eta_{13}, \eta_{14}, \eta_{24}$$

as representatives. As mentioned above, invariants of even degree are  $\pm\Gamma$ -invariant and so give rise to even functions on  $\mathcal{X}$  with respect to  $\iota$ , whereas invariants of odd degree give rise to odd functions on  $\mathcal{X}$ . Together with the constant function 1, we have found seven functions in  $L(2\Theta) = L(2\Theta)^+$ . Since  $\dim L(2\Theta) = 2^3 = 8$ , we are missing one function. We will see that is given by some quadratic form in the  $\eta_{ij}$  above, with the property that it does not grow faster than them when we approach  $\Theta$ .

To find this quadratic form, we have to find out what  $(\eta_{02} : \eta_{03} : \dots : \eta_{24})$  tends to as we approach the point represented by  $(x_1, y_1) + (x_2, y_2) + \mathfrak{m}$  on  $\mathcal{X}$ . A suitable approximation, taking  $y = \ell(x)$  to be the line interpolating between the two points,

$$B(x, 1) = \lambda(x - x_0)(x - x_1)(x - x_2) + \ell(x) ,$$

$$A_0(x) = (x - x_1)(x - x_2), \varphi_{\pm}(x) = (f(x) \pm \ell(x)^2)/A_0(x)^2, \psi(x) = \ell(x)/A_0(x), \text{ and}$$

$$A(x, 1) = A_0(x)(\lambda^2(x - x_0)^2 + (2\lambda\psi(x_0) - \varphi'_+(x_0))(x - x_0) - \varphi_-(x_0) + O(\lambda^{-1})) ,$$

shows that★

$$\begin{aligned} \eta_{02} &= -\lambda^2(x_1x_2)^2 + O(\lambda) \\ \eta_{03} &= \lambda^2(x_1 + x_2)x_1x_2 + O(\lambda) \\ \eta_{04} &= -\lambda^2x_1x_2 + O(\lambda) \\ \eta_{13} &= -\lambda^2(x_1^2 + x_2^2) + O(\lambda) \\ \eta_{14} &= \lambda^2(x_1 + x_2) + O(\lambda) \\ \eta_{24} &= -\lambda^2 + O(1) \end{aligned}$$

as  $\lambda \rightarrow \infty$ . There are various quadratic expressions in these that grow at most like  $\lambda^3$ , namely

$$\begin{aligned} 2\eta_{04}\eta_{24} + \eta_{13}\eta_{24} - \eta_{14}^2, \quad \eta_{03}\eta_{24} - \eta_{04}\eta_{14}, \quad \eta_{02}\eta_{24} - \eta_{04}^2, \\ \eta_{02}\eta_{14} - \eta_{03}\eta_{04}, \quad 2\eta_{02}\eta_{04} + \eta_{02}\eta_{13} - \eta_{03}^2 \end{aligned}$$

(they provide five independent even functions in  $L(3\Theta)$  modulo  $L(2\Theta)$ ) and

$$\eta = \eta_{02}\eta_{24} - \eta_{03}\eta_{14} + \eta_{04}^2 + \eta_{04}\eta_{13} , \tag{5}$$



which in fact only grows like  $\lambda^2$  and therefore gives us the missing basis element of  $L(2\Theta)$ . We find that★

$$\eta = \lambda^2 \frac{G(x_1, x_2) - 2y_1y_2}{(x_1 - x_2)^2} + O(\lambda),$$

where

$$G(x_1, x_2) = 2 \sum_{j=0}^4 f_{2j}(x_1x_2)^j + (x_1 + x_2) \sum_{j=0}^3 f_{2j+1}(x_1x_2)^j.$$

(Note the similarity with the fourth Kummer surface coordinate in the genus 2 case; see [4].)

The map  $\mathcal{X} \rightarrow \mathbb{P}^7$  we are looking for is then given by

$$(1 : \eta_{24} : \eta_{14} : \eta_{04} : \eta_{04} + \eta_{13} : \eta_{03} : \eta_{02} : \eta).$$

We use  $(\xi_1, \dots, \xi_8)$  to denote these coordinates (in the given order). The reason for setting  $\xi_5 = \eta_{04} + \eta_{13}$  rather than  $\eta_{13}$  is that this leads to nicer formulas later on. For example, we then have the simple quadratic relation

$$\xi_1\xi_8 - \xi_2\xi_7 + \xi_3\xi_6 - \xi_4\xi_5 = 0. \tag{6}$$

Regarding degree and weight, we have, writing  $\underline{\xi} = (\xi_1, \xi_2, \dots, \xi_8)$ , that

$$\deg(\underline{\xi}) = (0, 1, 1, 1, 1, 1, 1, 2) \quad \text{and} \quad \text{wt}(\underline{\xi}) = (0, 6, 5, 4, 4, 3, 2, 8).$$

It is known that the image  $\mathcal{K}$  of the Kummer variety in  $\mathbb{P}^7$  of a generic hyperelliptic Jacobian of genus 3 is given by a quadric and 34 independent quartic relations that are not multiples of the quadric; see [11, Thm. 3.3]. (For this, we can work over an algebraically closed field, so that we can change coordinates to move one of the Weierstrass points to infinity so that we are in the setting of [11].) The quadric is just (6). It is also known [11, Prop. 3.1] that  $\mathcal{K}$  is defined by quartic equations. Since there are 36 quartic multiples of the quadric (6), the space of quartics in eight variables has dimension 330 and the space  $L(8\Theta)^+$  has dimension 260, there must be at least 34 further independent quartics vanishing on  $\mathcal{K}$ : the space of quartics vanishing on  $\mathcal{K}$  is the kernel of  $\text{Sym}^4 L(2\Theta) \rightarrow L(8\Theta)^+$ , which has dimension  $\geq 70$ . We can find these quartics as follows.

There are 15 quartic relations in  $(\xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \xi_6, \xi_7)$  coming from the quartics defining  $\mathcal{Q}$ . They are given by the  $4 \times 4$  minors of the matrix (4), which restricted to  $\mathcal{V}$  is

$$M = \begin{pmatrix} 2f_0\xi_1 & f_1\xi_1 & \xi_7 & \xi_6 & \xi_4 \\ f_1\xi_1 & 2(f_2\xi_1 - \xi_7) & f_3\xi_1 - \xi_6 & \xi_5 - \xi_4 & \xi_3 \\ \xi_7 & f_3\xi_1 - \xi_6 & 2(f_4\xi_1 - \xi_5) & f_5\xi_1 - \xi_3 & \xi_2 \\ \xi_6 & \xi_5 - \xi_4 & f_5\xi_1 - \xi_3 & 2(f_6\xi_1 - \xi_2) & f_7\xi_1 \\ \xi_4 & \xi_3 & \xi_2 & f_7\xi_1 & 2f_8\xi_1 \end{pmatrix}. \tag{7}$$

Since these relations do not involve  $\xi_8$ , they cannot be multiples of the quadratic relation. We find 55 further independent quartics vanishing on  $\mathcal{X}$  (and thence a basis of the ‘new’ space of quartics that are not multiples of the quadratic relation) by searching for polynomials of given degree and weight that vanish on  $\mathcal{V}$  when pulled back to  $\mathbb{A}^{15}$ . Removing those that are multiples of the invariant quadric, we obtain quartics with the following 34 pairs of degree and weight:

- deg = 4: wt = 12, 13, 14, 14, 15, 15, 16, 16, 16, 17, 17, 18, 18, 19, 20;
- deg = 5: wt = 17, 18, 18, 19, 19, 20, 20, 20, 21, 21, 22, 22, 23;
- deg = 6: wt = 22, 23, 24, 24, 25, 26.

(Recall that ‘degree’ refers to the degree in terms of the original  $\eta_{ij}$ .) These quartics are given in the file `Kum3-quartics.magma` at [17]. The quartics are scaled so that their coefficients are in  $\mathbb{Z}[f_0, \dots, f_8]$ . The 15 quartics of degree 4 are exactly those obtained as  $4 \times 4$ -minors of the matrix  $M$  above.

**Lemma 2.2** *Let  $f_0, \dots, f_8 \in k$  be arbitrary. Then the 70 quartics constructed as described above are linearly independent over  $k$ .*

*Proof* We can find<sup>★</sup> 70 monomials such that the  $70 \times 70$ -matrix formed by the coefficients of the quartics with respect to these monomials has determinant  $\pm 1$ .  $\square$

Note that regarding  $k$ , this is a slight improvement over [11, Lemma 3.2], where  $k$  was assumed to have characteristic  $\neq 2, 3, 5$ .

We now show that these quartics indeed give all the relations.

**Lemma 2.3** *The natural map  $\text{Sym}^2 L(4\Theta)^+ \rightarrow L(8\Theta)^+$  is surjective.*

*Proof* Mumford shows [13, §4, Thm. 1] that  $\text{Sym}^2 L(4\Theta) \rightarrow L(8\Theta)$  is surjective. The proof can be modified to give the corresponding result for the even subspaces, as follows (notations as in [13]). We work with the even functions  $\delta_{a+b} + \delta_{-a-b}$  and

$\delta_{a-b} + \delta_{-a+b}$ . This gives

$$\begin{aligned} & \sum_{\eta \in \mathbb{Z}_2} l(\eta)(\delta_{a+b+\eta} + \delta_{-a-b-\eta}) * (\delta_{a-b+\eta} + \delta_{-a+b-\eta}) \\ &= \left( \sum_{\eta \in \mathbb{Z}_2} l(\eta)q_1(b + \eta) \right) \left( \sum_{\eta \in \mathbb{Z}_2} l(\eta)(\delta_{a+\eta} + \delta_{-a-\eta}) \right) \\ & \quad + \left( \sum_{\eta \in \mathbb{Z}_2} l(\eta)q_1(a + \eta) \right) \left( \sum_{\eta \in \mathbb{Z}_2} l(\eta)(\delta_{b+\eta} + \delta_{-b-\eta}) \right). \end{aligned}$$

We fix the homomorphism  $l: \mathbb{Z}_2 \rightarrow \{\pm 1\}$  and the class of  $a \pmod{K(\delta)}$ . By (\*) in [13, p. 339] there is some  $b$  in this class such that  $\sum_{\eta} l(\eta)q(b + \eta) \neq 0$ . Taking  $a = b$ , we see that

$$\Delta(b) := \sum_{\eta} l(\eta)(\delta_{b+\eta} + \delta_{-b-\eta})$$

is in the image. Using this, we see that for all other  $a$  in the class,  $\Delta(a)$  is also in the image. Inverting the Fourier transform, we find that all  $\delta_a + \delta_{-a}$  are in the image, which therefore consists of all even functions. □

**Corollary 2.4** *The natural map  $\text{Sym}^4 L(2\Theta) \rightarrow L(8\Theta)^+$  is surjective.*

*Proof* Note that  $L(2\Theta) = L(2\Theta)^+$ , so the image of  $\text{Sym}^4 L(2\Theta) \rightarrow L(8\Theta)$  is contained in the even subspace. Since there is exactly one quadratic relation, the map  $\text{Sym}^2 L(2\Theta) \rightarrow L(4\Theta)^+$  is not surjective, but has a one-dimensional cokernel. We will see below in Sect. 7 that this cokernel is generated by the image of a function  $\mathcal{E}$  such that  $\xi_i \xi_j \mathcal{E}$  (for all  $i, j$ ) and  $\mathcal{E}^2$  can be expressed as quartics in the  $\xi_i$ . This implies that the image of the map in the statement contains the image of  $\text{Sym}^2 L(4\Theta)^+$ , and surjectivity follows from Lemma 2.3. Note that once we have found  $\mathcal{E}$  explicitly, the assertions relating to it made above can be checked directly and without relying on the considerations leading to the determination of  $\mathcal{E}$ . □

**Theorem 2.5** *Let  $k$  be a field of characteristic different from 2 and let  $F \in k[x, z]$  be homogeneous of degree 8 and squarefree. Then the image  $\mathcal{K}$  in  $\mathbb{P}^7$  of the Kummer variety associated to the Jacobian variety of the hyperelliptic curve  $y^2 = F(x, 1)$  is defined by the quadric (6) and the 34 quartics constructed above.*

*Proof* By Corollary 2.4 the dimension of the space of quartics vanishing on  $\mathcal{K}$  is 70. By Lemma 2.2 the quadric and the 34 quartics give rise to 70 independent quartics vanishing on  $\mathcal{K}$ . By [11, Prop. 3.1]  $\mathcal{K}$  can be defined by quartics, so the claim follows. □

This improves on [11, Thm. 3.3] by removing the genericity assumption (and allowing characteristic 3 or 5).

To conclude this section, we determine the images of some special points on  $\mathcal{J}$  under the map to  $\mathcal{K}$ .

The discussion on p. 672 shows that on a point  $[(x_1, y_1) + (x_2, y_2) + \mathfrak{m}] \in \Theta$ , the map restricts to

$$\left( 0 : 1 : -(x_1 + x_2) : x_1 x_2 : x_1^2 + x_1 x_2 + x_2^2 \right. \\ \left. : -(x_1 + x_2)x_1 x_2 : (x_1 x_2)^2 : \frac{2y_1 y_2 - G(x_1, x_2)}{(x_1 - x_2)^2} \right).$$

If we write  $(X - x_1)(X - x_2) = \sigma_0 X^2 + \sigma_1 X + \sigma_2$ , then this can be written as

$$(0 : \sigma_0^2 : \sigma_0 \sigma_1 : \sigma_0 \sigma_2 : \sigma_1^2 - \sigma_0 \sigma_2 : \sigma_1 \sigma_2 : \sigma_2^2 : \xi_8),$$

where, rewriting  $((x_1 - x_2)^2 \xi_8 - G(x_1, x_2))^2 = 4F(x_1, 1)F(x_2, 1)$ , we have that

$$\begin{aligned} & (\sigma_1^2 - 4\sigma_0 \sigma_2) \xi_8^2 \\ & + (4f_0 \sigma_0^4 - 2f_1 \sigma_0^3 \sigma_1 + 4f_2 \sigma_0^3 \sigma_2 - 2f_3 \sigma_0^2 \sigma_1 \sigma_2 + 4f_4 \sigma_0^2 \sigma_2^2 \\ & \quad - 2f_5 \sigma_0 \sigma_1 \sigma_2^2 + 4f_6 \sigma_0 \sigma_2^3 - 2f_7 \sigma_1 \sigma_2^3 + 4f_8 \sigma_2^4) \xi_8 \\ & + (-4f_0 f_2 + f_1^2) \sigma_0^6 + 4f_0 f_3 \sigma_0^5 \sigma_1 - 2f_1 f_3 \sigma_0^5 \sigma_2 - 4f_0 f_4 \sigma_0^4 \sigma_1^2 \\ & \quad + (-4f_0 f_5 + 4f_1 f_4) \sigma_0^4 \sigma_1 \sigma_2 + (-4f_0 f_6 + 2f_1 f_5 - 4f_2 f_4 + f_3^2) \sigma_0^4 \sigma_2^2 \\ & \quad + 4f_0 f_5 \sigma_0^3 \sigma_1^3 + (8f_0 f_6 - 4f_1 f_5) \sigma_0^3 \sigma_1^2 \sigma_2 + (8f_0 f_7 - 4f_1 f_6 + 4f_2 f_5) \sigma_0^3 \sigma_1 \sigma_2^2 \\ & \quad + (-2f_1 f_7 - 2f_3 f_5) \sigma_0^3 \sigma_2^3 - 4f_0 f_6 \sigma_0^2 \sigma_1^4 + (-12f_0 f_7 + 4f_1 f_6) \sigma_0^2 \sigma_1^3 \sigma_2 \\ & \quad + (-16f_0 f_8 + 8f_1 f_7 - 4f_2 f_6) \sigma_0^2 \sigma_1^2 \sigma_2^2 + (8f_1 f_8 - 4f_2 f_7 + 4f_3 f_6) \sigma_0^2 \sigma_1 \sigma_2^3 \\ & \quad + (-4f_2 f_8 + 2f_3 f_7 - 4f_4 f_6 + f_5^2) \sigma_0^2 \sigma_2^4 + 4f_0 f_7 \sigma_0 \sigma_1^5 \\ & \quad + (16f_0 f_8 - 4f_1 f_7) \sigma_0 \sigma_1^4 \sigma_2 + (-12f_1 f_8 + 4f_2 f_7) \sigma_0 \sigma_1^3 \sigma_2^2 \\ & \quad + (8f_2 f_8 - 4f_3 f_7) \sigma_0 \sigma_1^2 \sigma_2^3 + (-4f_3 f_8 + 4f_4 f_7) \sigma_0 \sigma_1 \sigma_2^4 - 2f_3 f_7 \sigma_0 \sigma_2^5 \\ & \quad - 4f_0 f_8 \sigma_1^6 + 4f_1 f_8 \sigma_1^5 \sigma_2 - 4f_2 f_8 \sigma_1^4 \sigma_2^2 + 4f_3 f_8 \sigma_1^3 \sigma_2^3 - 4f_4 f_8 \sigma_1^2 \sigma_2^4 \\ & \quad + 4f_5 f_8 \sigma_1 \sigma_2^5 + (-4f_6 f_8 + f_7^2) \sigma_2^6 \\ & = 0. \end{aligned}$$

(This is similar to the quartic defining the Kummer surface in the genus 2 case.) The image on  $\mathcal{K}$  of the theta divisor is a surface of degree 12 in  $\mathbb{P}^6 = \mathbb{P}^7 \cap \{\xi_1 = 0\}$ ; the intersection of  $\mathcal{K}$  with the hyperplane  $\xi_1 = 0$  is twice the image of  $\Theta$ . (The

equation above is cubic in the middle six coordinates and  $\xi_8$ , so we get three times the degree of the Veronese surface. It is known that  $\mathcal{K}$  has degree 24.)

When  $(x_2, y_2)$  approaches  $(x_1, -y_1)$ , then the last coordinate tends to infinity, whereas the remaining ones stay bounded, so the origin on  $\mathcal{J}$  is mapped to

$$o := (0 : 0 : 0 : 0 : 0 : 0 : 0 : 1) .$$

Points in  $\mathcal{J}[2]$  are represented by factorizations  $F = GH$  with  $d = \deg G$  even, compare Sect. 5 below. Writing

$$G = g_d x^d + g_{d-1} x^{d-1} z + \dots + g_0 z^d \text{ and } H = h_{8-d} x^{8-d} + h_{7-d} x^{7-d} z + \dots + h_0 z^{8-d} ,$$

we see that a 2-torsion point represented by  $(G, H)$  with  $\deg G = 2$  maps to

$$(0 : g_2^2 : g_1 g_2 : g_0 g_2 : g_1^2 - g_0 g_2 : g_0 g_1 : g_0^2 : g_0^3 h_6 + g_0^2 g_2 h_4 + g_0 g_2^2 h_2 + g_2^3 h_0) . \tag{8}$$

A 2-torsion point represented by  $(G, H)$  with  $\deg G = 4$  maps to

$$\begin{aligned} (1 : g_2 h_4 + g_4 h_2 : g_1 h_4 + g_4 h_1 : g_0 h_4 + g_4 h_0 & \tag{9} \\ : g_0 h_4 + g_4 h_0 + g_1 h_3 + g_3 h_1 : g_0 h_3 + g_3 h_0 : g_0 h_2 + g_2 h_0 & \\ : (g_0 h_4 + g_4 h_0)^2 + (g_0 h_2 + g_2 h_0)(g_2 h_4 + g_4 h_2) + (g_1 h_0 - g_0 h_1)(g_4 h_3 - g_3 h_4) ; & \end{aligned}$$

this is obtained by taking  $(A, B, C) = (G, 0, H)$  in our original parameterization.

### 3 Transformations

We compare our coordinates for the Kummer variety with those of Stubbs [18], Duquesne [5] and Müller [9] in the special case  $f_8 = 0$ . In this case there is a rational Weierstrass point at infinity, and we can fix the representation of a point outside of  $\Theta$  by requiring that  $A$  vanishes at infinity and that  $\deg B(x, 1) < \deg A(x, 1)$ . For a generic point  $P$  on  $\mathcal{J}$ ,  $\deg A(x, 1) = 3$ ; let  $(x_j, y_j)$  for  $j = 1, 2, 3$  be the three points in the effective divisor  $D$  such that  $P = [D - 3 \cdot \infty]$ . Generically, the three points are distinct. Then

$$A(x, 1) = (x - x_1)(x - x_2)(x - x_3)$$

and  $B(x, 1)$  is the interpolation polynomial such that  $B(x_j, 1) = y_j$  for  $j = 1, 2, 3$ . We obtain the  $c_j$  from  $C = (B^2 - F)/A$  by polynomial division. This leads to★

$$\begin{aligned}
 \xi_1 &= \kappa_1 \\
 \xi_2 &= -f_7\kappa_2 \\
 \xi_3 &= f_7\kappa_3 \\
 \xi_4 &= -f_7\kappa_4 \\
 \xi_5 &= f_4\kappa_1 + f_5\kappa_2 + 2f_6\kappa_3 + 3f_7\kappa_4 - \kappa_5 \\
 \xi_6 &= f_3\kappa_1 + f_4\kappa_2 + f_5\kappa_3 - \kappa_6 \\
 \xi_7 &= f_2\kappa_1 - f_4\kappa_3 - 3f_5\kappa_4 - \kappa_7 \\
 \xi_8 &= -f_2f_7\kappa_2 - f_3f_7\kappa_3 - f_4f_7\kappa_4 + f_7\kappa_8
 \end{aligned}$$

where  $\kappa_1, \kappa_2, \dots, \kappa_8$  are the coordinates used by the other authors.

We consider the effect of a transformation of the curve equation. First suppose that  $\tilde{F}(x, z) = F(x + \lambda z, z)$  (corresponding to a shift of the  $x$ -coordinate in the affine equation). A point represented by a triple  $(A(x, z), B(x, z), C(x, z))$  of polynomials will correspond to the point  $(\tilde{A}(x, z), \tilde{B}(x, z), \tilde{C}(x, z))$  with  $\tilde{A}(x, z) = A(x + \lambda z, z)$  and analogously for  $\tilde{B}$  and  $\tilde{C}$ . We obtain★

$$\begin{aligned}
 \tilde{\xi}_1 &= \xi_1 \\
 \tilde{\xi}_2 &= \xi_2 + 3\lambda f_7 \xi_1 + 12\lambda^2 f_8 \xi_1 \\
 \tilde{\xi}_3 &= \xi_3 + 2\lambda \xi_2 + 3\lambda^2 f_7 \xi_1 + 8\lambda^3 f_8 \xi_1 \\
 \tilde{\xi}_4 &= \xi_4 + \lambda \xi_3 + \lambda^2 \xi_2 + \lambda^3 f_7 \xi_1 + 2\lambda^4 f_8 \xi_1 \\
 \tilde{\xi}_5 &= \xi_5 + \lambda(2f_5 \xi_1 + 3\xi_3) + \lambda^2(6f_6 \xi_1 + 3\xi_2) + 17\lambda^3 f_7 \xi_1 + 34\lambda^4 f_8 \xi_1 \\
 \tilde{\xi}_6 &= \xi_6 + \lambda(3\xi_4 + \xi_5) + \lambda^2(f_5 \xi_1 + 3\xi_3) + \lambda^3(2f_6 \xi_1 + 2\xi_2) + 5\lambda^4 f_7 \xi_1 + 8\lambda^5 f_8 \xi_1 \\
 \tilde{\xi}_7 &= \xi_7 + \lambda(f_3 \xi_1 + 2\xi_6) + \lambda^2(2f_4 \xi_1 + 3\xi_4 + \xi_5) + \lambda^3(4f_5 \xi_1 + 2\xi_3) \\
 &\quad + \lambda^4(6f_6 \xi_1 + \xi_2) + 9\lambda^5 f_7 \xi_1 + 12\lambda^6 f_8 \xi_1 \\
 \tilde{\xi}_8 &= \xi_8 + \lambda(f_3 \xi_2 + 2f_5 \xi_4 + 3f_7 \xi_7) \\
 &\quad + \lambda^2(3f_3 f_7 \xi_1 + 2f_4 \xi_2 + f_5 \xi_3 + 6f_6 \xi_4 + 3f_7 \xi_6 + 12f_8 \xi_7) \\
 &\quad + \lambda^3((12f_3 f_8 + 6f_4 f_7) \xi_1 + 4f_5 \xi_2 + 4f_6 \xi_3 + 17f_7 \xi_4 + f_7 \xi_5 + 16f_8 \xi_6) \\
 &\quad + \lambda^4((24f_4 f_8 + 11f_5 f_7) \xi_1 + 8f_6 \xi_2 + 12f_7 \xi_3 + 46f_8 \xi_4 + 6f_8 \xi_5) \\
 &\quad + \lambda^5((44f_5 f_8 + 18f_6 f_7) \xi_1 + 16f_7 \xi_2 + 32f_8 \xi_3) \\
 &\quad + \lambda^6((68f_6 f_8 + 29f_7^2) \xi_1 + 32f_8 \xi_2) + 148\lambda^7 f_7 f_8 \xi_1 + 148\lambda^8 f_8^2 \xi_1 .
 \end{aligned}$$

For the transformation given by  $\tilde{F}(x, z) = F(z, x)$ , we have

$$\tilde{a}_j = a_{4-j}, \quad \tilde{b}_j = b_{4-j}, \quad \tilde{c}_j = c_{4-j}$$

and therefore

$$(\tilde{\xi}_1, \tilde{\xi}_2, \tilde{\xi}_3, \tilde{\xi}_4, \tilde{\xi}_5, \tilde{\xi}_6, \tilde{\xi}_7, \tilde{\xi}_8) = (\xi_1, \xi_7, \xi_6, \xi_4, \xi_5, \xi_3, \xi_2, \xi_8) .$$

More generally, consider an element

$$\sigma = \begin{pmatrix} r & s \\ t & u \end{pmatrix} \in \text{GL}(2)$$

acting by  $(x, z) \mapsto (rx + sz, tx + uz)$ . Let  $\Sigma \in \text{GL}(5)$  be the matrix whose columns are the coefficients of  $(rx + sz)^j(tx + uz)^{4-j}$ , for  $j = 0, 1, 2, 3, 4$  (this is the matrix giving the action of  $\sigma$  on the fourth symmetric power of the standard representation of  $\text{GL}(2)$ ). Recall the matrix  $L$  from (3) whose rows contain the coefficients of  $A$ ,  $B$  and  $C$ . Then the effect on our variables  $a_i, b_i, c_i$  is given by  $L \mapsto L\Sigma^\top$ . With  $D$  as in (2), we have  $L^\top DL = M$  with  $M$  as in (7). So the effect of  $\sigma$  on  $M$  is given by  $M \mapsto \Sigma M \Sigma^\top$ . Note that  $\tilde{\xi}_1 = \xi_1$  and that we can extract  $\tilde{\xi}_2, \dots, \tilde{\xi}_7$  from  $M$ ; to get  $\tilde{\xi}_8$  when  $\xi_1$  is not invertible, we can perform a generic computation and then specialize.

This allows us to reduce our more general setting to the situation when there is a Weierstrass point at infinity: we adjoin a root of  $F(x, 1)$ , then we shift this root to zero and invert. This leads to an equation with  $f_8 = 0$ . This was used to obtain the matrix representing the action of an even 2-torsion point in Sect. 5 below.

### 4 Lifting Points to the Jacobian

Let  $P \in \mathcal{H}(k)$  be a  $k$ -rational point on the Kummer variety. We want to decide if  $P = \kappa(P')$  for a  $k$ -rational point  $P'$  on the Jacobian  $\mathcal{J}$ . Consider an odd function  $h$  on  $\mathcal{J}$  (i.e., such that  $h(-Q) = -h(Q)$  for  $Q \in \mathcal{J}$ ) such that  $h$  is defined over  $k$ ; then  $h(P') \in k$  (or  $h$  as a pole at  $P'$ ). Since  $h^2$  is an even function, it descends to a function on  $\mathcal{H}$ , and we must have that  $h^2(P) = h^2(P') = h(P')^2$  is a square in  $k$ . Conversely, any non-zero odd function  $h$  on  $\mathcal{J}$  will generically separate the two points in the fiber of the double cover  $\mathcal{J} \rightarrow \mathcal{H}$ , so if  $h^2(P)$  is a non-zero square in  $k$ , then this implies that  $P$  lifts to a  $k$ -rational point on  $\mathcal{J}$ .

So we will now exhibit some odd functions that we can use to decide if a point lifts. Since  $L(2\Theta)$  consists of even functions only, we look at  $L(3\Theta)$ , which has dimension  $3^3 = 27$ . Its subspace of even functions has dimension 14 and is spanned

by  $\xi_1, \dots, \xi_8$ , the five quadratics

$$\xi_2(\xi_4 + \xi_5) - \xi_3^2, \quad \xi_2\xi_6 - \xi_3\xi_4, \quad \xi_2\xi_6 - \xi_4^2, \quad \xi_3\xi_6 - \xi_4\xi_7, \quad (\xi_4 + \xi_5)\xi_7 - \xi_6^2$$

and a further function, which can be taken to be  $\star$

$$2(2f_0\xi_2^2 - f_1\xi_2\xi_3 + 2f_2\xi_2\xi_4 - f_3\xi_2\xi_6 + 2f_4\xi_2\xi_7 - f_5\xi_3\xi_7 + 2f_6\xi_4\xi_7 - f_7\xi_6\xi_7 + 2f_8\xi_7^2) - 7\xi_2\xi_4\xi_7 + \xi_2\xi_5\xi_7 + \xi_2\xi_6^2 + \xi_3^2\xi_7 + 4\xi_3\xi_4\xi_6 - 2\xi_3\xi_5\xi_6 + \xi_4^3 - 5\xi_4^2\xi_5 + 2\xi_4\xi_5^2.$$

The subspace of odd functions has dimension 13. We obtain a ten-dimensional subspace of this space by considering the coefficients of  $A_l \wedge B_l \wedge C_l$ , which is an expression of degree 3, of odd degree in  $B$  and invariant even under  $SL(3)$  acting on  $(A, B, C)$ . (One can check that there are no further  $\Gamma$ -invariants of degree 3.) These coefficients are given by the  $3 \times 3$ -minors of the matrix  $L$  of (3). If we denote the minor corresponding to  $0 \leq i < j < k \leq 4$  by  $\mu_{ijk}$ , then we find that

$$\mu_{ijk}^2 = \eta_{ii}\eta_{jk}^2 + \eta_{jj}\eta_{ik}^2 + \eta_{kk}\eta_{ij}^2 - 4\eta_{ii}\eta_{jj}\eta_{kk} - \eta_{ij}\eta_{ik}\eta_{jk}. \tag{10}$$

If  $L_{ijk}$  is the corresponding  $3 \times 3$  submatrix of  $L$ , then we have that

$$\mu_{ijk}^2 = \det(L_{ijk})^2 = -\frac{1}{2} \det(L_{ijk}^\top DL_{ijk})$$

with  $D$  as in (2). We also have that  $L^\top DL = M$ , where  $M$  is the matrix corresponding to the quadratic form  $B_l^2 - A_l C_l$  given in (7). We can express this by saying that  $\mu_{ijk}^2$  is  $-\frac{1}{2}$  times the corresponding principal minor of  $M$ . In the same way, one sees that  $\mu_{ijk}\mu_{i'j'k'}$  is  $-\frac{1}{2}$  times the minor of  $M$  given by selecting rows  $i, j, k$  and columns  $i', j', k'$ . This shows that if one  $\mu_{ijk}^2(P)$  is a non-zero square in  $k$ , then all  $\mu_{i'j'k'}^2(P)$  are squares in  $k$ . All ten of them vanish simultaneously if and only if  $A, B$  and  $C$  are linearly dependent (this is equivalent to the rank of  $B_l^2 - A_l C_l$  being at most 2). The dimension of the space spanned by  $A, B$  and  $C$  cannot be strictly less than 2, since this would imply that  $F$  is a constant times a square, which contradicts the assumption that  $F$  is squarefree. So we can write  $A, B$  and  $C$  as linear combinations of two polynomials  $A'$  and  $C'$ , and after a suitable change of basis, we find that  $F = B^2 - AC = A'C'$ . This means that the point is the image of a 2-torsion point on  $\mathcal{J}$ , and it will always lift.

So for a point  $P$  in  $\mathcal{X}(k)$  with  $\xi_1 = 1$  (hence outside the theta divisor) to lift to a point in  $\mathcal{J}(k)$ , it is necessary that all these expressions, when evaluated at  $P$ , are squares in  $k$ , and sufficient that one of them gives a non-zero square. For points with  $\xi_1 = 0$ , we can use the explicit description of the image of  $\Theta$  given in Sect. 2.

Let  $\mathcal{V}'$  be the quotient of  $\mathcal{V}$  by the action of the subgroup of  $\Gamma$  generated by the elements of the form  $t_\lambda$  and  $n_\mu$ ; then the points of  $\mathcal{V}'$  correspond to effective divisors of degree 4 on  $\mathcal{C}$  in general position. Geometrically, the induced map



$\mathcal{V}' \rightarrow \mathcal{X} \setminus \Theta$  is a conic bundle: for a point on  $\mathcal{X}$  outside the theta divisor, all effective divisors representing it are in general position, and the corresponding linear system has dimension 1 by the Riemann-Roch Theorem, so the fibers are Severi-Brauer varieties of dimension 1. If  $\mathcal{C}$  has a  $k$ -rational point  $P$ , then the bundle has a section (and so is in fact a  $\mathbb{P}^1$ -bundle), since we can select the unique representative containing  $P$  in its support. If  $k$  is a number field and  $\mathcal{C}$  has points over every completion of  $k$ , then all the conics in fibers above  $k$ -rational points on  $\mathcal{X} \setminus \Theta$  have points over all completions of  $k$  and therefore are isomorphic to  $\mathbb{P}^1$  over  $k$ . We can check whether a  $k$ -defined divisor representing a lift of  $P$  to a  $k$ -rational point on  $\mathcal{J}$  exists and find one in this case in the following way. We assume that  $P$  is not in the image of  $\Theta$  and is not the image of a 2-torsion point. We are looking for a matrix  $\tilde{L} \in \mathbb{A}^{15}(k)$  representing a lift  $P' \in \mathcal{J}(k)$  of  $P$ . Since we exclude 2-torsion, the matrix  $\tilde{L}$  must have rank 3, and there is a minor  $\mu_{ijk}$  such that  $\mu_{ijk}^2(P) = \mu_{ijk}(P')^2$  is a non-zero square in  $k$ . The rank of  $M(P) = \tilde{L}^\top D \tilde{L}$  is also 3, so both  $L(\tilde{P})$  and  $M(P)$  have the same 2-dimensional kernel. We can compute the kernel from  $M(P)$  and then we find the space generated by the rows of  $\tilde{L}$  as its annihilator, which is simply given by rows  $i, j, k$  of  $M(P)$ . If we find an invertible  $3 \times 3$  matrix  $U$  with entries in  $k$  such that  $M_{ijk}(P) = U^\top D U$  (where  $M_{ijk}$  is the principal  $3 \times 3$  submatrix of  $M$  given by rows and columns  $i, j, k$ ), then we can find a suitable matrix  $\tilde{L}$  whose rows are in the space generated by rows  $i, j, k$  of  $M(P)$  and such that  $\tilde{L}_{ijk} = U$ . Then  $\tilde{L}^\top D \tilde{L} = M(P)$ , so  $\tilde{L}$  gives us the desired representative. Finding  $U$  is equivalent to finding an isomorphism between the quadratic forms given by

$$(x_1, x_2, x_3)M_{ijk}(P)(x_1, x_2, x_3)^\top \quad \text{and} \quad 2x_1x_3 - 2x_2^2,$$

for whose existence a necessary condition is that  $\det M_{ijk}(P) = -2\mu_{ijk}^2(P)$  is a square times  $\det D = -2$ . Given this, the problem comes down to finding a point on the conic given by the first form (which is the conic making up the fiber above  $P'$  or  $-P'$ ) and then parameterizing the conic using lines through the point.

*Remark 4.1* One can check<sup>★</sup> that the following three expressions are a possible choice for the missing three basis elements of the odd subspace of  $L(3\Theta)$ :

$$\begin{aligned} &\xi_2\mu_{012} - \xi_3\mu_{013} + \xi_5\mu_{014} \\ &\xi_3\mu_{014} - (\xi_4 + \xi_5)\mu_{024} + \xi_4\mu_{123} + \xi_6\mu_{034} \\ &\xi_5\mu_{034} - \xi_6\mu_{134} + \xi_7\mu_{234} \end{aligned}$$

## 5 The Action of the 2-Torsion Subgroup on $\mathcal{H}$

We follow the approach taken in [14] and consider the action of the 2-torsion subgroup of  $\mathcal{J}$  on  $\mathcal{H}$  and the ambient projective space. Note that translation by a 2-torsion point commutes with negation on  $\mathcal{J}$ , so the translation descends to

an automorphism of  $\mathcal{K}$ , and since  $2\Theta$  is linearly equivalent to its translate, this automorphism actually is induced by an automorphism of the ambient  $\mathbb{P}^7$ .

We will see that this projective representation of  $\mathcal{J}[2] \simeq (\mathbb{Z}/2\mathbb{Z})^6$  can be lifted to a representation of a central extension of  $\mathcal{J}[2]$  by  $\mu_2$  on the space of linear forms in the coordinates  $\xi_1, \dots, \xi_8$ . This representation is irreducible. In the next section, we consider this representation and the induced representations on the spaces of quadratic and quartic forms in  $\xi_1, \dots, \xi_8$ , whereas in this section, we obtain an explicit description of the action of  $\mathcal{J}[2]$  on  $\mathbb{P}^7$ .

There is a natural bijection between the 2-torsion subgroup  $\mathcal{J}[2]$  of the Jacobian and the set of unordered partitions of the set  $\Omega \subset \mathbb{P}^1$  of zeros of  $F$  into two subsets of even cardinality. The torsion point  $T$  corresponding to a partition  $\{\Omega_1, \Omega_2\}$  is

$$\left[ \sum_{\omega \in \Omega_1} (\omega, 0) \right] - \frac{\#\Omega_1}{2} \mathfrak{m} = \left[ \sum_{\omega \in \Omega_2} (\omega, 0) \right] - \frac{\#\Omega_2}{2} \mathfrak{m} .$$

Since  $\#\Omega = 8$  is divisible by 4, the quantity  $\varepsilon(T) = (-1)^{\#\Omega_1/2} = (-1)^{\#\Omega_2/2}$  is well-defined. We say that  $T$  is *even* if  $\varepsilon(T) = 1$  and *odd* if  $\varepsilon(T) = -1$ . By definition, the even 2-torsion points are the 35 points corresponding to a partition into two sets of four roots, together with the origin, and the odd 2-torsion points are the 28 points corresponding to a partition into subsets of sizes 2 and 6. The Weil pairing of two torsion points  $T$  and  $T'$  represented by  $\{\Omega_1, \Omega_2\}$  and  $\{\Omega'_1, \Omega'_2\}$ , respectively, is given by

$$e_2(T, T') = (-1)^{\#(\Omega_1 \cap \Omega'_1)} .$$

It is then easy to check that

$$e_2(T, T') = \varepsilon(T)\varepsilon(T')\varepsilon(T + T') . \tag{11}$$

Note that  $\text{Pic}_\mathcal{C}^0$  is canonically isomorphic to  $\text{Pic}_\mathcal{C}^2$  (by adding the class of  $\mathfrak{m}$ ), which contains the theta characteristics. (A divisor class  $\mathcal{D} \in \text{Pic}_\mathcal{C}^2$  is a *theta characteristic* if  $2\mathcal{D} = \mathcal{W}$ .) In this way, the theta characteristics are identified with the 2-torsion points, and the odd (resp., even) theta characteristics correspond to the odd (resp., even) 2-torsion points.

Using the transformations described in Sect. 3 and the matrices obtained by Duquesne [5] representing the translation by a 2-torsion point, we find the corresponding matrices in our setting for an even nontrivial 2-torsion point. The matrices corresponding to odd 2-torsion points can then also be derived. For each factorization  $F = GH$  into two forms of even degree, there is a matrix  $M_{(G,H)}$  whose entries are polynomials with integral coefficients in the coefficients of  $G$  and  $H$  and whose image in  $\text{PGL}(8)$  gives the action of the corresponding 2-torsion point. These entries are too large to be reproduced here, but are given in the file `Kum3-torsionmats.magma` at [17].

The matrices satisfy the relations★

$$M_{(G,H)}^2 = \text{Res}(G, H)I_8 \quad \text{and} \quad \det M_{(G,H)} = \text{Res}(G, H)^4, \tag{12}$$

where Res denotes the resultant of two binary forms. Let

$$S = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

be the matrix corresponding to the quadratic relation (6) satisfied by points on the Kummer variety.

**Definition 5.1** We will write  $\langle \cdot, \cdot \rangle_S$  for the pairing given by  $S$ . Concretely, for vectors  $\underline{\xi} = (\xi_1, \dots, \xi_8)$  and  $\underline{\zeta} = (\zeta_1, \dots, \zeta_8)$ , we have

$$\langle \underline{\xi}, \underline{\zeta} \rangle_S = \xi_1\zeta_8 - \xi_2\zeta_7 + \xi_3\zeta_6 - \xi_4\zeta_5 - \xi_5\zeta_4 + \xi_6\zeta_3 - \xi_7\zeta_2 + \xi_8\zeta_1.$$

One checks★ that for all  $G, H$  as above,

$$(SM_{(G,H)})^\top = (-1)^{(\deg G)/2} SM_{(G,H)}.$$

If  $T \neq 0$  is even, then all corresponding matrices  $M_{(G,H)}$  are equal; we denote this matrix by  $M_T$ . In this case, also the resultant  $\text{Res}(G, H)$  depends only on  $T$ ; we write it  $r(T)$ , so that we have  $M_T^2 = r(T)I_8$ . For  $T$  odd and represented by  $(G, H)$  with  $\deg G = 2$ , we have  $M_{(\lambda G, \lambda^{-1}H)} = \lambda^2 M_{(G,H)}$ . As a special case, we have  $M_{(1,F)} = I_8$ . For  $T \neq 0$  even, the entry in the upper right corner of  $M_T$  is 1, for all other 2-torsion points, this entry is zero.

For a 2-torsion point  $T \in \mathcal{J}[2]$ , if we denote by  $M_T$  the matrix corresponding to one of the factorizations defining  $T$ , we therefore have (using that  $S = S^\top = S^{-1}$ )

$$(SM_T)^\top = \varepsilon(T)SM_T, \quad \text{or equivalently,} \quad M_T = \varepsilon(T)SM_T^\top S.$$

This implies (using that  $M_{T'}M_T$  is, up to scaling, a matrix corresponding to  $T + T'$ )

$$\begin{aligned} M_T M_{T'} &= \varepsilon(T)SM_T^\top S \cdot \varepsilon(T')SM_{T'}^\top S \\ &= \varepsilon(T)\varepsilon(T')S(M_{T'}M_T)^\top S = \varepsilon(T)\varepsilon(T')\varepsilon(T + T')M_{T'}M_T. \end{aligned}$$

Using (11), we recover the well-known fact that

$$M_T M_{T'} = e_2(T, T') M_{T'} M_T . \tag{13}$$

Since  $M_T^2$  is a scalar matrix, the relation given above implies that the quadratic relation is invariant (up to scaling) under the action of  $\mathcal{J}[2]$  on  $\mathbb{P}^7$ :

$$M_T^T S M_T = \text{Res}(G, H) S .$$

## 6 The Action on Linear, Quadratic and Quartic Forms

We work over an algebraically closed field  $k$  of characteristic different from 2. The first result describes a representation of a central extension  $G$  of  $\mathcal{J}[2]$  on the space of linear forms that lifts the action on  $\mathbb{P}^7$ .

**Lemma 6.1** *There is a subgroup  $G$  of  $\text{SL}(8)$  and an exact sequence*

$$0 \longrightarrow \mu_2 \longrightarrow G \longrightarrow \mathcal{J}[2] \longrightarrow 0$$

*induced by the standard sequence*

$$0 \longrightarrow \mu_8 \longrightarrow \text{SL}(8) \longrightarrow \text{PSL}(8) \longrightarrow 0$$

*and the embedding  $\mathcal{J}[2] \rightarrow \text{PSL}(8)$  given by associating to  $T$  the class of any matrix  $M_T$ .*

*Proof* Let  $T \in \mathcal{J}[2]$  and let  $M_T \in \text{GL}(8)$  be any matrix associated to  $T$ . Then  $M_T^2 = cI_8$  with some  $c$  (compare (12)), and we let  $\tilde{M}_T$  denote one of the two matrices  $\gamma M_T$  where  $\gamma^2 c = \varepsilon(T)$ . Then  $\tilde{M}_T \in \text{SL}(8)$ , since (again by (12))

$$\det \tilde{M}_T = \gamma^8 \det M_T = (\varepsilon(T)c^{-1})^4 c^4 = 1 .$$

Since any two choices of  $M_T$  differ only by scaling,  $\tilde{M}_T$  is well-defined up to sign. Among the lifts of the class of  $M_T$  in  $\text{PSL}(8)$  to  $\text{SL}(8)$ ,  $\pm \tilde{M}_T$  are characterized by the relation  $\tilde{M}_T^2 = \varepsilon(T)I_8$ . We now set

$$G = \{ \pm \tilde{M}_T : T \in \mathcal{J}[2] \} .$$

It is clear that  $G$  surjects onto the image of  $\mathcal{J}[2]$  in  $\text{PSL}(8)$  and that the map is two-to-one. It remains to show that  $G$  is a group. So let  $T, T' \in \mathcal{J}[2]$ . Then  $\tilde{M}_T \tilde{M}_{T'}$

is a matrix corresponding to  $T + T'$ . Since (using (13) and (11))

$$\begin{aligned} (\tilde{M}_T \tilde{M}_{T'})^2 &= \tilde{M}_T \tilde{M}_{T'} \tilde{M}_T \tilde{M}_{T'} = e_2(T, T') \tilde{M}_T^2 \tilde{M}_{T'}^2 \\ &= e_2(T, T') \varepsilon(T) \varepsilon(T') I_8 = \varepsilon(T + T') I_8, \end{aligned}$$

we find that  $\tilde{M}_T \tilde{M}_{T'} \in G$ . □

*Remark 6.2* Note that the situation here is somewhat different from the situation in genus 2, as discussed in [14]. In the even genus hyperelliptic case, the theta characteristics live in  $\text{Pic}^{\text{odd}}$  rather than in  $\text{Pic}^{\text{even}}$  and can therefore not be identified with the 2-torsion points. The effect is that there is no map  $\varepsilon: \mathcal{J}[2] \rightarrow \mu_2$  that induces the Weil pairing as in (11), so that we have to use a fourfold covering of  $\mathcal{J}[2]$  in  $\text{SL}(4)$  rather than a double cover.

We now proceed to a study of the representations of  $G$  on linear, quadratic and quartic forms on  $\mathbb{P}^7$  that are induced by  $G \subset \text{SL}(8)$ . The representation  $\rho_1$  on the space  $V_1$  of linear forms is the standard representation. For its character  $\chi_1$ , we find that

$$\chi_1(\pm I_8) = \pm 8 \quad \text{and} \quad \chi_1(\pm \tilde{M}_T) = 0 \quad \text{for all } T \neq 0.$$

This follows from the observation that  $T$  can be written as  $T = T' + T''$  with  $e_2(T', T'') = -1$ . Since  $\pm \tilde{M}_T = \tilde{M}_{T'} \tilde{M}_{T''} = -\tilde{M}_{T''} \tilde{M}_{T'}$ , the trace of  $\tilde{M}_T$  must be zero. We deduce that  $\rho_1$  is irreducible. ( $\rho_1$  is essentially the representation  $V(\delta)$  in [13], where  $\delta = (2, 2, 2)$  in our case.)

The representation  $\rho_2$  on the space  $V_2$  of quadratic forms is the symmetric square of  $\rho_1$ . Since  $\pm I_8$  act trivially on even degree forms,  $\rho_2$  descends to a representation of  $\mathcal{J}[2]$ . Its character  $\chi_2$  is given by

$$\begin{aligned} \chi_2(0) &= 36 \quad \text{and} \\ \chi_2(T) &= \frac{1}{2}(\chi_1(\tilde{M}_T)^2 + \chi_1(\tilde{M}_T^2)) = \frac{1}{2}(0 + 8\varepsilon(T)) = 4\varepsilon(T) \quad \text{for } T \neq 0. \end{aligned}$$

Since  $\mathcal{J}[2]$  is abelian, this representation has to split into a direct sum of one-dimensional representations. Define the character  $\chi_T$  of  $\mathcal{J}[2]$  by  $\chi_T(T') = e_2(T, T')$ . Then the above implies that

$$\rho_2 = \bigoplus_{T:\varepsilon(T)=1} \chi_T. \tag{14}$$

So for each even  $T \in \mathcal{J}[2]$ , there is a one-dimensional eigenspace of quadratic forms such that the action of  $T'$  is given by multiplication with  $e_2(T, T')$ . For  $T = 0$ , this eigenspace is spanned by the invariant quadratic (6).

**Definition 6.3** We set

$$y_0 = 2(\xi_1\xi_8 - \xi_2\xi_7 + \xi_3\xi_6 - \xi_4\xi_5) ;$$

this is the quadratic form corresponding to  $S$ , since  $y_0(\underline{\xi}) = \underline{\xi}S\underline{\xi}^\top = \langle \underline{\xi}, \underline{\xi} \rangle_S$ . For nontrivial even  $T$ , we denote by  $y_T$  the form in the eigenspace corresponding to  $T$  that has coefficient 1 on  $\xi_8^2$ . We will see that this makes sense, i.e., that this coefficient is always nonzero.

**Lemma 6.4** *For every nontrivial even 2-torsion point  $T$ , the matrix corresponding to the quadratic form  $y_T$  is the symmetric matrix  $SM_T$ . In particular, if  $T$  corresponds to a factorization  $F = GH$  into two polynomials of degree 4, then the coefficients of  $y_T$  are polynomials in the coefficients of  $G$  and  $H$  with integral coefficients, and the coefficients of the monomials  $\xi_i\xi_j$  with  $i \neq j$  are divisible by 2.*

*Proof* We show that  $\tilde{M}_{T'}^\top(SM_T)\tilde{M}_{T'} = e_2(T, T')SM_T$ . We use that  $\tilde{M}_{T'}^2 = \varepsilon(T')I_8$ ,  $S\tilde{M}_{T'} = \varepsilon(T')\tilde{M}_{T'}^\top S$  and the fact that the Weil pairing is given by commutators. This gives that

$$\tilde{M}_{T'}^\top SM_T \tilde{M}_{T'} = \varepsilon(T')S\tilde{M}_{T'}M_T\tilde{M}_{T'} = \varepsilon(T')e_2(T, T')SM_T\tilde{M}_{T'}^2 = e_2(T, T')SM_T$$

as desired, so  $SM_T$  gives a quadratic form in the correct eigenspace. Since the upper right entry of  $M_T$  is 1, the lower right entry, which corresponds to the coefficient of  $\xi_8^2$ , of  $SM_T$  is 1, so that we indeed obtain  $y_T$ . □

We can express  $y_T$  as  $y_T(\underline{\xi}) = \langle \underline{\xi}, \underline{\xi}M_T^\top \rangle_S$ .

*Remark 6.5* Note that if  $T$  is an odd 2-torsion point, represented by the factorization  $(G, H)$ , then the same argument shows that the alternating bilinear form corresponding to the matrix  $SM_{(G,H)}$  is multiplied by  $e_2(T, T')$  under the action of  $T' \in \mathcal{J}[2]$ .

We set

$$(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6, \varepsilon_7, \varepsilon_8) = (1, -1, 1, -1, -1, 1, -1, 1) ;$$

these are the entries occurring in  $S$  along the diagonal from upper right to lower left.

**Corollary 6.6** *Let  $T$  be a nontrivial even 2-torsion point with image on  $\mathcal{K}$  given by*

$$(1 : \tau_2 : \tau_3 : \tau_4 : \tau_5 : \tau_6 : \tau_7 : \tau_8) .$$

Then

$$y_T = \xi_8^2 + 2 \sum_{j=2}^8 \varepsilon_j \tau_j \xi_{9-j} \xi_8 + (\text{terms not involving } \xi_8) .$$

A similar statement is true for  $T = 0$  if we take coordinates  $(0 : \dots : 0 : 1)$ : we have that  $y_0 = 2\xi_1\xi_8 + (\text{terms not involving } \xi_8)$ .

*Proof* The last column of  $M_T$  has entries  $1, \tau_2, \dots, \tau_8$  (since  $M_T$  maps the origin to the image of  $T$  and has upper right entry 1). Multiplication by  $S$  from the left reverses the order and introduces the signs  $\varepsilon_j$ . Since the coefficients of  $y_T$  of monomials involving  $\xi_8$  are given by the entries of the last column of  $SM_T$  by Lemma 6.4, the claim follows.  $\square$

We define a pairing on the space  $V_1 \otimes V_1$  of bilinear forms as follows. If the bilinear forms  $\phi$  and  $\phi'$  are represented by matrices  $A$  and  $A'$  with respect to our standard basis  $\xi_1, \dots, \xi_8$  of  $V_1$ , then  $\langle \phi, \phi' \rangle = \frac{1}{8} \text{Tr}(A^\top A')$  (the scaling has the effect of giving the standard quadratic form norm 1).

For an even 2-torsion point  $T$ , we write  $\tilde{y}_T$  for the symmetric bilinear form corresponding to the matrix  $\tilde{S}M_T$  (this is well-defined up to sign) and  $\tilde{z}_T$  for the symmetric bilinear form corresponding to  $\tilde{S}M_T^\top = \tilde{M}_T S$ . Also,  $z_T$  will denote the form corresponding to  $SM_T^\top = M_T S$ . Then, since  $S(M_T S)S = SM_T$ , we have the relation  $z_T(\underline{\xi}) = y_T(\underline{\xi}S)$ ; explicitly,

$$z_T(\xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \xi_6, \xi_7, \xi_8) = y_T(\xi_8, -\xi_7, \xi_6, -\xi_5, -\xi_4, \xi_3, -\xi_2, \xi_1) .$$

**Lemma 6.7** *For all even 2-torsion points  $T$  and  $T'$ , we have that*

$$\langle \tilde{z}_T, \tilde{y}_{T'} \rangle = \begin{cases} 1 & \text{if } T = T', \\ 0 & \text{if } T \neq T'. \end{cases}$$

*Equivalently,*

$$\langle z_T, y_{T'} \rangle = \begin{cases} r(T) & \text{if } T = T', \\ 0 & \text{if } T \neq T'. \end{cases}$$

*Here we restrict the scalar product defined above to  $V_2 \subset V_1 \otimes V_1$ .*

*Proof* The claim is that  $\text{Tr}((\tilde{S}M_T^\top)^\top (\tilde{S}M_{T'}))$  is zero if  $T \neq T'$  and that it equals 8 if  $T = T'$ . We have that

$$\text{Tr}((\tilde{S}M_T^\top)^\top (\tilde{S}M_{T'})) = \text{Tr}(\tilde{M}_T S^2 \tilde{M}_{T'}) = \text{Tr}(\tilde{M}_T \tilde{M}_{T'}) = \pm \text{Tr}(\tilde{M}_{T+T'}) .$$

If  $T \neq T'$ , then this trace is zero, as we had already seen. If  $T = T'$ , then the matrix  $\pm \tilde{M}_{T+T'}$  is  $I_8$ , so the result is 8 as desired.  $\square$

This allows us to express the  $\xi_j^2$  in terms of the  $y_T$ . We set  $r(0) = 1$  and  $M_0 = I_8$ . We denote the coefficient of  $\xi_i \xi_j$  in a quadratic form  $q \in V_2$  by  $[\xi_i \xi_j]q$ .

**Lemma 6.8** *For every  $j \in \{1, 2, \dots, 8\}$ , we have that*

$$\xi_j^2 = \sum_{T:\varepsilon(T)=1} \frac{[\xi_{9-j}^2]y_T}{8r(T)} y_T .$$

Similarly, for  $1 \leq i < j \leq 8$ , we have that

$$2\xi_i\xi_j = \varepsilon_i\varepsilon_j \sum_{T:\varepsilon(T)=1} \frac{[\xi_{9-i}\xi_{9-j}]y_T}{8r(T)} y_T.$$

*Proof* We have by Lemma 6.7 that

$$\begin{aligned} \xi_j^2 &= \sum_{T:\varepsilon(T)=1} \langle \tilde{z}_T, \xi_j^2 \rangle \tilde{y}_T = \sum_{T:\varepsilon(T)=1} \frac{\langle z_T, \xi_j^2 \rangle}{r(T)} y_T \\ &= \sum_{T:\varepsilon(T)=1} \frac{[\xi_j^2]z_T}{8r(T)} y_T = \sum_{T:\varepsilon(T)=1} \frac{[\xi_{9-j}^2]y_T}{8r(T)} y_T. \end{aligned}$$

In the same way, we have for  $i \neq j$  that

$$\begin{aligned} 2\xi_i\xi_j &= \sum_{T:\varepsilon(T)=1} 2\langle \tilde{z}_T, \xi_i\xi_j \rangle \tilde{y}_T = \sum_{T:\varepsilon(T)=1} 2\frac{\langle z_T, \xi_i\xi_j \rangle}{r(T)} y_T \\ &= \sum_{T:\varepsilon(T)=1} \frac{[\xi_i\xi_j]z_T}{8r(T)} y_T = \varepsilon_i\varepsilon_j \sum_{T:\varepsilon(T)=1} \frac{[\xi_{9-i}\xi_{9-j}]y_T}{8r(T)} y_T. \end{aligned}$$

(Note that  $8\langle z_T, \xi_i\xi_j \rangle$  is half the coefficient of  $\xi_i\xi_j$  in  $z_T$ .) □

**Corollary 6.9** *We have that*

$$\sum_{T:\varepsilon(T)=1} \frac{1}{8r(T)} y_T(\underline{\xi}) y_T(\underline{\zeta}) = \left( \sum_{j=1}^8 \varepsilon_j \xi_j \zeta_{9-j} \right)^2 = \langle \underline{\xi}, \underline{\zeta} \rangle_S^2.$$

*In particular, setting  $\underline{\zeta} = \underline{\xi}$ , we obtain that*

$$\sum_{T:\varepsilon(T)=1} \frac{1}{8r(T)} y_T^2 = y_0^2 = 4(\xi_1\xi_8 - \xi_2\xi_7 + \xi_3\xi_6 - \xi_4\xi_5)^2.$$

*Proof* We compute using Lemma 6.8:

$$\begin{aligned} &\sum_{T:\varepsilon(T)=1} \frac{1}{8r(T)} y_T(\underline{\xi}) y_T(\underline{\zeta}) \\ &= \sum_{i=1}^8 \xi_i^2 \sum_{T:\varepsilon(T)=1} \frac{[\xi_i^2]y_T(\underline{\xi})}{8r(T)} y_T(\underline{\zeta}) + \sum_{1 \leq i < j \leq 8} \xi_i \xi_j \sum_{T:\varepsilon(T)=1} \frac{[\xi_i \xi_j]y_T(\underline{\xi})}{8r(T)} y_T(\underline{\zeta}) \end{aligned}$$



$$\begin{aligned}
 &= \sum_{i=1}^8 \xi_i^2 \zeta_{9-i}^2 + 2 \sum_{1 \leq i < j \leq 8} \varepsilon_i \varepsilon_j \xi_i \xi_j \zeta_{9-i} \zeta_{9-j} \\
 &= \left( \sum_{j=1}^8 \varepsilon_j \xi_j \zeta_{9-j} \right)^2 .
 \end{aligned}$$

□

Now we consider the representation  $\rho_4$  of  $\mathcal{J}[2]$  on the space  $V_4$  of quartic forms. For its character  $\chi_4$ , we have the general formula

$$\begin{aligned}
 \chi_4(T) = \frac{1}{24}(\chi_1(\tilde{M}_T)^4 + 8\chi_1(\tilde{M}_T)\chi_1(\tilde{M}_T^3) + 3\chi_1(\tilde{M}_T^2)^2 \\
 + 6\chi_1(\tilde{M}_T)^2\chi_1(\tilde{M}_T^2) + 6\chi_1(\tilde{M}_T^4)) .
 \end{aligned}$$

This gives us that

$$\chi_4(0) = 330 \quad \text{and} \quad \chi_4(T) = 10 \quad \text{for } T \neq 0 .$$

We deduce that

$$\rho_4 = \chi_0^{\oplus 15} \oplus \bigoplus_{T \neq 0} \chi_T^{\oplus 5} . \tag{15}$$

### 7 The Duplication Map and the Missing Generator of $L(4\Theta)^+$

We continue to work over a field  $k$  of characteristic  $\neq 2$ . We also continue to assume that  $F \in k[x, z]$  is squarefree, so that  $\mathcal{C}$  is a smooth hyperelliptic curve of genus 3 over  $k$ .

Consider the commutative diagram

$$\begin{array}{ccc}
 \mathcal{J} & \xrightarrow{\cdot 2} & \mathcal{J} \\
 \downarrow \kappa & & \downarrow \kappa \\
 \mathcal{H} & \xrightarrow{\delta} & \mathcal{H} \hookrightarrow \mathbb{P}^7 ,
 \end{array}$$

where the map in the top row is multiplication by 2 and  $\delta$  is the endomorphism of  $\mathcal{H}$  induced by it. Pulling back a hyperplane section to the copy of  $\mathcal{J}$  on the right, we obtain a divisor in the class of  $2\Theta$ . Pulling it further back to the copy on the left, we obtain a divisor in the class of the pull-back of  $2\Theta$  under duplication, which is the class of  $8\Theta$  ( $\Theta$  is symmetric, so pulling back under multiplication

by  $n$  multiplies its class by  $n^2$ ). The combined map from the left  $\mathcal{J}$  to  $\mathbb{P}^7$  then is given by an 8-dimensional subspace of  $L(8\Theta)^+$ ; by Corollary 2.4 this means that  $\delta$  is given by eight quartic forms in  $\underline{\xi}$ . Since  $\delta$  maps  $o$ , the image of the origin on  $\mathcal{H}$ , to itself, we can normalize these quartics so that they evaluate to  $(0, \dots, 0, 1)$  on  $(0, \dots, 0, 1)$ . We use  $\underline{\delta} = (\delta_1, \dots, \delta_8)$  to denote these quartic forms; they are determined up to adding a quartic form vanishing on  $\mathcal{H}$ . We write  $E_4 \subset V_4$  for the subspace of quartics vanishing on  $\mathcal{H}$ . Note that we can test whether a given homogeneous polynomial in  $\underline{\xi}$  vanishes on  $\mathcal{H}$  by pulling it back to  $\mathcal{W}$  or to  $\mathbb{A}^{15}$  and checking whether it vanishes on  $\mathcal{V}$ .

We now determine the structure of  $E_4$  as a representation of  $\mathcal{J}$  [2] and we identify the space generated by  $\underline{\delta}$  in  $V_4/E_4$ .

**Lemma 7.1**

- (1) *The restriction of  $\rho_4$  to  $E_4$  splits as  $\rho_4|_{E_4} = \chi_0^{\oplus 7} \oplus \bigoplus_{T \neq 0} \chi_T$ .*
- (2) *The images of  $\delta_1, \dots, \delta_8$  form a basis of the quotient  $V_4^{\mathcal{J}[2]}/E_4^{\mathcal{J}[2]}$  of invariant subspaces.*

*Proof*

- (1) The dimension of  $E_4$  is 70 by Theorem 2.5, and a subspace of dimension 36 is given by  $y_0V_2$ . The latter splits in the same way as  $\rho_2$  does. Since for the generic curve, the Galois action is transitive on the odd 2-torsion points and on the nontrivial even 2-torsion points, the multiplicities of all odd characters and those of all nontrivial even characters in  $\rho_4|_{E_4}$  have to agree. The only way to make the numbers come out correctly is as indicated.
- (2) Since the result of duplicating a point is unchanged when a 2-torsion point is added to it, the images of all  $\delta_j$  in  $V_4/E_4$  must lie in the same eigenspace of the  $\mathcal{J}$  [2]-action. Since  $K$  spans  $\mathbb{P}^7$  and the duplication map  $\delta: K \rightarrow K$  is surjective, the images of the  $\delta_j$  in  $V_4/E_4$  must be linearly independent. So they must live in an eigenspace of dimension at least eight. The only such eigenspace is that of the trivial character, which has dimension exactly  $8 = 15 - 7$  by the first part.

□

We see that the 36 quartic forms  $y_T^2$  for  $T$  an even 2-torsion point are in the invariant subspace of  $V_4$  of dimension 15. Let  $\mathcal{T}_{\text{even}}$  denote the finite  $k$ -scheme whose geometric points are the 36 even 2-torsion points (we can consider  $\mathcal{T}_{\text{even}}$  as a subscheme of  $\mathcal{J}$  or of  $\mathcal{H}$ ), and denote by  $k_{\text{even}}$  its coordinate ring; this is an étale  $k$ -algebra of dimension 36. Then  $y: T \mapsto y_T$  can be considered as a quadratic form with coefficients in  $k_{\text{even}}$  and  $r: T \mapsto r(T)$  is an element of  $k_{\text{even}}^\times$ .

**Lemma 7.2** *The 36 coefficients  $c_{ii} = [\xi_i^2]y$ , for  $1 \leq i \leq 8$ , and  $c_{ij} = \frac{1}{2}[\xi_i \xi_j]y$ , for  $1 \leq i < j \leq 8$ , constitute a  $k$ -basis of  $k_{\text{even}}$ .*

*Proof* We define further elements of  $k_{\text{even}}$  by

$$\tilde{c}_{ii} = \frac{1}{8r}[\xi_{9-i}^2]y \quad \text{and} \quad \tilde{c}_{ij} = \frac{\varepsilon_i \varepsilon_j}{8r}[\xi_{9-i} \xi_{9-j}]y .$$

$$\begin{aligned}
 q_1 &= \xi_1 \xi_8^3 + 2(-f_2 \xi_2 + f_3 \xi_3 - f_4 \xi_4 - f_4 \xi_5 + f_5 \xi_6 - f_6 \xi_7) \xi_1 \xi_8^2 + \dots \\
 q_2 &= \xi_2 \xi_8^3 + (4f_8(-f_0 \xi_2 + f_2 \xi_4 + f_4 \xi_7) - 2f_3 f_8 \xi_6 - f_3 f_7 \xi_7) \xi_1 \xi_8^2 + \dots \\
 q_3 &= \xi_3 \xi_8^3 + (f_7(-2f_0 \xi_2 + 2f_2 \xi_4 + f_3 \xi_6) + 2f_8(-2f_0 \xi_3 + 4f_1 \xi_4 - 2f_2 \xi_6 - f_3 \xi_7)) \xi_1 \xi_8^2 + \dots \\
 q_4 &= \xi_4 \xi_8^3 + (-2f_0 f_7 \xi_3 + (12f_0 f_8 + f_1 f_7) \xi_4 - 2f_1 f_8 \xi_6) \xi_1 \xi_8^2 + \dots \\
 q_5 &= \xi_5 \xi_8^3 + ((4f_0 f_6 - 2f_1 f_5) \xi_2 + (-2f_0 f_7 - 2f_1 f_6 + 2f_2 f_5) \xi_3 \\
 &\quad + (4f_0 f_8 + 4f_1 f_7 + 4f_2 f_6 - 5f_3 f_5) \xi_4 \\
 &\quad + (-2f_1 f_8 - 2f_2 f_7 + 2f_3 f_6) \xi_6 + (4f_2 f_8 - 2f_3 f_7) \xi_7) \xi_1 \xi_8^2 + \dots \\
 q_6 &= \xi_6 \xi_8^3 + (f_0(-2f_5 \xi_2 - 4f_6 \xi_3 + 8f_7 \xi_4 - 4f_8 \xi_6) + f_1(f_5 \xi_3 + 2f_6 \xi_4 - 2f_8 \xi_7)) \xi_1 \xi_8^2 + \dots \\
 q_7 &= \xi_7 \xi_8^3 + (4f_0(f_4 \xi_2 + f_6 \xi_4 - f_8 \xi_7) - f_1 f_3 \xi_2 - 2f_0 f_5 \xi_3) \xi_1 \xi_8^2 + \dots \\
 q_8 &= \xi_8^4 + 16(f_1 f_8(f_1 \xi_2 - f_2 \xi_3 + f_3 \xi_4) + f_0 f_7(f_5 \xi_4 - f_6 \xi_6 + f_7 \xi_7)) \xi_1 \xi_8^2 + \dots \\
 q_9 &= 2(f_7 \xi_6 - 4f_8 \xi_7) \xi_1 \xi_8^2 + \dots \\
 q_{10} &= 2(f_5 \xi_4 - f_6 \xi_6 + f_7 \xi_7) \xi_1 \xi_8^2 + \dots \\
 q_{11} &= 2(f_3 \xi_3 + 2f_4 \xi_4 - 2f_4 \xi_5 + f_5 \xi_6) \xi_1 \xi_8^2 + \dots \\
 q_{12} &= 2(f_1 \xi_2 - f_2 \xi_3 + f_3 \xi_4) \xi_1 \xi_8^2 + \dots \\
 q_{13} &= 2(-4f_0 \xi_2 + f_1 \xi_3) \xi_1 \xi_8^2 + \dots \\
 q_{14} &= (3\xi_4 - \xi_5) \xi_1 \xi_8^2 + \dots \\
 q_{15} &= \xi_1^2 \xi_8^2 + \dots = (\xi_1 \xi_8 - \xi_2 \xi_7 + \xi_3 \xi_6 - \xi_4 \xi_5)^2
 \end{aligned}$$

Fig. 1 A basis of the  $\mathcal{J}[2]$ -invariant subspace of  $V_4$

Lemma 6.8 can be interpreted as saying that

$$\text{Tr}_{k_{\text{even}}/k}(\tilde{C}_{ij} C_{i'j'}) = \begin{cases} 1 & \text{if } (i, j) = (i', j'), \\ 0 & \text{otherwise.} \end{cases}$$

This shows that the given elements are linearly independent over  $k$ . □

We can compute the structure constants of  $k_{\text{even}}$  with respect to this basis and use this to express  $y^2$  in terms of the basis again. Extracting coefficients, we obtain 36 quartic forms with coefficients in  $k$  that all lie in the 15-dimensional space of invariants under  $\mathcal{J}[2]$ . We check<sup>★</sup> that they indeed span a space of this dimension and that we get a subspace of dimension 7 of quartics vanishing on the Kummer variety.

It turns out<sup>★</sup> that the quartics in  $V_4^{\mathcal{J}[2]}$  that vanish on  $\mathcal{X}$  are exactly those that do not contain terms cubic or quartic in  $\xi_8$ . Forms spanning the complementary space are uniquely determined modulo  $E_4^{\mathcal{J}[2]}$  by fixing the terms of higher degree

in  $\xi_8$ . We take  $q_j = \xi_j \xi_8^3 + (\deg_{\xi_8} \leq 2)$  for  $j = 1, \dots, 8$ . Then the  $q_j$  can be chosen so that they have coefficients in  $\mathbb{Z}[f_0, \dots, f_8]$ . To fix  $q_j$  completely, it suffices to specify in addition the coefficients of  $\xi_1 \xi_i \xi_8^2$  for  $1 \leq i \leq 7$ . One possibility is to choose them as given in Fig. 1, which includes  $q_9, \dots, q_{15}$  in the ideal of  $\mathcal{X}$ , where  $E_4^{\mathcal{J}[2]} = \langle q_9, q_{10}, \dots, q_{15} \rangle$ . These quartics can be obtained from `Kum3-invariants.magma` at [17].

We can now identify the duplication map on  $\mathcal{X}$ .

**Theorem 7.3** *The polynomials*

$$(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7, \delta_8) = (4q_1, 4q_2, 4q_3, 4q_4, 4q_5, 4q_6, 4q_7, q_8)$$

in  $V_4^{\mathcal{J}[2]}$  (with  $q_j$  as above) have the following properties.

- (1)  $\delta_j \in \mathbb{Z}[f_0, f_1, \dots, f_8][\xi_1, \xi_2, \dots, \xi_8]$  for all  $1 \leq j \leq 8$ .
- (2)  $(\delta_1, \delta_2, \dots, \delta_8)(0, 0, \dots, 0, 1) = (0, 0, \dots, 0, 1)$ .
- (3) With  $y_T$  as defined earlier for an even 2-torsion point with image

$$(1 : \tau_2 : \tau_3 : \tau_4 : \tau_5 : \tau_6 : \tau_7 : \tau_8)$$

on  $\mathcal{X}$ , we have that

$$y_T^2 \equiv \delta_8 - \tau_2 \delta_7 + \tau_3 \delta_6 - \tau_4 \delta_5 - \tau_5 \delta_4 + \tau_6 \delta_3 - \tau_7 \delta_2 + \tau_8 \delta_1 = \langle \underline{\tau}, \underline{\delta} \rangle_S \pmod{E_4^{\mathcal{J}[2]}}$$

where  $\underline{\tau} = (1, \tau_2, \dots, \tau_8)$  and  $\underline{\delta} = (\delta_1, \dots, \delta_8)$ .

- (4) The  $\delta_j$  do not vanish simultaneously on  $\mathcal{X}$ .
- (5) The map  $\delta: \mathcal{X} \rightarrow \mathcal{X}$  given by  $(\delta_1 : \dots : \delta_8)$  is the duplication map on  $\mathcal{X}$ .

*Proof*

- (1) This can be verified using the explicit polynomials.
- (2) This is obvious.
- (3) We compare the coefficients of  $\xi_j \xi_8^3$  on both sides. Since by Corollary 6.6,

$$y_T = \xi_8^2 + 2\varepsilon_2 \tau_2 \xi_7 \xi_8 + 2\varepsilon_3 \tau_3 \xi_6 \xi_8 + \dots + 2\varepsilon_8 \tau_8 \xi_1 \xi_8 + (\text{terms not involving } \xi_8),$$

we find that

$$y_T^2 = \xi_8^4 + 4\varepsilon_2 \tau_2 \xi_7 \xi_8^3 + \dots + 4\varepsilon_8 \tau_8 \xi_1 \xi_8^3 + (\text{terms of degree } \leq 2 \text{ in } \xi_8)$$

and the right hand side has the same form. So the difference is a form in  $V_4^{\mathcal{J}[2]}$  of degree at most 2 in  $\xi_8$ , which implies that it is in  $E_4^{\mathcal{J}[2]}$ .

- (4) Let  $\underline{\xi} \in k^8 \setminus \{0\}$  be coordinates of a point in  $\mathcal{X}$ . Then  $\underline{\delta}(\underline{\xi}) = 0$  implies by (3) that  $y_T(\underline{\xi}) = 0$  for all even 2-torsion points  $T$  (note that  $y_0$  vanishes on all of  $\mathcal{X}$ ). Lemma 6.8 then shows that  $\underline{\xi} = 0$  as well, since  $8r(T) \neq 0$  in  $k$ . This contradicts our choice of  $\underline{\xi}$ .

(5) By (4),  $\delta$  is a morphism  $\mathcal{H} \rightarrow \mathbb{P}^7$ , and by Lemma 7.1 (2)  $\delta$  differs from the duplication map by post-composing with an automorphism  $\alpha$  of  $\mathbb{P}^7$ . We show<sup>★</sup> that on a generic point,  $\delta$  coincides with the duplication map; this proves that  $\alpha$  is the identity. We use the action of  $GL(2)$  on  $(x, z)$  (and scaling on  $y$ ) to reduce to the case that  $F(x, 1)$  is monic of degree 7. A generic point  $P$  on  $\mathcal{J}$  can then be represented by  $(A, B, C)$  such that  $A(x, 1)$  is monic of degree 3 and squarefree and  $B(x, 1)$  is of degree  $\leq 2$ . After making a further affine transformation, we can assume that  $A(x, 1) = x(x - 1)(x - a)$  for some  $a \in k$ . The corresponding point on  $\mathcal{H}$  is then

$$\begin{aligned} \kappa(P) = & (1 : -a - 1 : a : 0 : -ac_3 - c_1 : -c_0 \\ & : (a + 1)c_0 + 2b_0b_2 : -(a^2 + a + 1)c_0 - 2(a + 1)b_0b_2), \end{aligned}$$

where  $B(x, 1) = b_0 + b_1x + b_2x^2$ ,  $C(x, 1) = c_0 + c_1x + c_2x^2 + c_3x^3 - x^4$ . We compute  $2P$  in terms of its Mumford representation using Cantor’s algorithm as implemented in Magma and find  $\kappa(2P)$ . On the other hand, we compute  $\delta(\kappa(P))$ . Both points are equal, which proves the claim.  $\square$

The quartics  $\underline{\delta} = (\delta_1, \dots, \delta_8)$  are given in the file `Kum3-deltas.magma` at [17].

The canonical map from  $V_2 = \text{Sym}^2 L(2\Theta)$  to  $L(4\Theta)$  has non-trivial one-dimensional kernel, spanned by the quadric  $y_0$  vanishing on  $\mathcal{H}$ . Since the dimension of the even part  $L(4\Theta)^+$  of  $L(4\Theta)$  is  $36 = \dim V_2$ , the map  $V_2 \rightarrow L(4\Theta)^+$  has a one-dimensional cokernel. Looking at the action of  $\mathcal{J}[2]$  on  $L(4\Theta)^+$ , it is clear that this space splits as a direct sum of the image of  $V_2$  and a one-dimensional invariant subspace. We will identify a generator of the latter.

**Lemma 7.4** *The image of  $q_1$  in  $L(8\Theta)$  is the square of an element  $\mathcal{E} \in L(4\Theta)^+$  that is invariant under the action of  $\mathcal{J}[2]$ .*

*Proof* We pull back  $q_1$  to a polynomial function on the affine space  $\mathbb{A}^{15}$  that parameterizes the triples of polynomials  $(A, B, C)$ . We find<sup>★</sup> that this polynomial is the square of some other polynomial  $p$  that can be written as a quadratic in the components of  $A_l \wedge B_l \wedge C_l$ . So  $p$  is invariant under  $\pm\Gamma$ , which means that it gives an element  $\mathcal{E}$  of  $L(4\Theta)^+$ .  $\square$

To make  $\mathcal{E}$  more explicit, we note that  $p$  can be expressed as a cubic in the  $\xi_j$ . Taking into account that  $\xi_1 = 1$  on the affine space, we find that (up to the choice of a sign)

$$\begin{aligned} \xi_1 \mathcal{E} = & (-8f_0f_4f_8 + 2f_0f_5f_7 + 2f_1f_3f_8)\xi_1^3 - 4f_0f_6\xi_1^2\xi_2 + (-4f_0f_7 + 2f_1f_6)\xi_1^2\xi_3 \\ & + (-4f_0f_8 + 2f_1f_7 - 4f_2f_6 + f_3f_5)\xi_1^2\xi_4 + (12f_0f_8 - f_1f_7)\xi_1^2\xi_5 \\ & + (-4f_1f_8 + 2f_2f_7)\xi_1^2\xi_6 - 4f_2f_8\xi_1^2\xi_7 + 6f_0\xi_1\xi_2^2 - 3f_1\xi_1\xi_2\xi_3 \\ & + 6f_2\xi_1\xi_2\xi_4 - f_3\xi_1\xi_2\xi_6 - 2f_3\xi_1\xi_3\xi_4 + 2f_4\xi_1\xi_3\xi_6 - f_5\xi_1\xi_3\xi_7 \end{aligned}$$

$$\begin{aligned}
 &+ 4f_4\xi_1\xi_4^2 - 2f_4\xi_1\xi_4\xi_5 - 2f_5\xi_1\xi_4\xi_6 + 6f_6\xi_1\xi_4\xi_7 - 3f_7\xi_1\xi_6\xi_7 \\
 &+ 6f_8\xi_1\xi_7^2 - 11\xi_2\xi_4\xi_7 + \xi_2\xi_5\xi_7 + 2\xi_2\xi_6^2 + 2\xi_3^2\xi_7 + 5\xi_3\xi_4\xi_6 \\
 &- 3\xi_3\xi_5\xi_6 + 2\xi_4^3 - 7\xi_4^2\xi_5 + 3\xi_4\xi_5^2 .
 \end{aligned}$$

We obtain similar cubic expressions for  $\xi_j\mathcal{E}$  with  $j \in \{2, 3, \dots, 8\}$  by multiplying the polynomial above by  $\xi_j$ , then adding a suitable linear combination of the quartics vanishing on  $\mathcal{K}$  so that we obtain something that is divisible by  $\xi_1$ . These cubics are given in the file `Kum3-Xip01s.magma` at [17]. With this information, we can evaluate  $\mathcal{E}$  on any given set  $\underline{\xi}$  of coordinates of a point on  $\mathcal{K}$ : we find an index  $j$  with  $\xi_j \neq 0$  and evaluate  $\mathcal{E}$  as  $(\xi_j\mathcal{E})/\xi_j$ .

This gives us a basis of  $L(4\Theta)^+$  consisting of  $\mathcal{E}$  and the quadratic monomials in the  $\xi_j$  minus one of the monomials  $\xi_j\xi_{9-j}$ . Alternatively, we can use the basis consisting of  $\mathcal{E}$  and the  $y_T$  for the 35 nonzero even 2-torsion points  $T$ .

### 8 Sum and Difference on the Kummer Variety

In this section,  $k$  continues to be a field of characteristic  $\neq 2$  and  $F$  to be squarefree.

We consider the composition

$$\mathcal{J} \times \mathcal{J} \xrightarrow{(+,-)} \mathcal{J} \times \mathcal{J} \xrightarrow{(\kappa,\kappa)} \mathcal{K} \times \mathcal{K} \longrightarrow \mathbb{P}^7 \times \mathbb{P}^7 \xrightarrow{\text{Segre}} \mathbb{P}^{63} \xrightarrow{\text{symm.}} \mathbb{P}^{35}$$

where ‘symm.’ is the symmetrization map that sends a matrix  $A$  to  $A + A^T$  and we identify the Segre map with the multiplication map

$$(\text{column vectors}) \times (\text{row vectors}) \longrightarrow \text{matrices} .$$

Pulling back hyperplanes to  $\mathcal{J} \times \mathcal{J}$ , we see that the map is given by sections of  $4\text{pr}_1^*\Theta + 4\text{pr}_2^*\Theta$ , hence symmetric bilinear forms on  $L(4\Theta)$ . The map is invariant under negation of either one of the arguments, therefore the bilinear forms only involve even sections. The map can be described by a symmetric matrix  $B$  of such bilinear forms such that in terms of coordinates  $(w_j)$  and  $(z_j)$  of the images  $\kappa(P + Q)$  and  $\kappa(P - Q)$  of  $P \pm Q$  on  $\mathcal{K}$ , we have that (up to scaling)  $w_iz_j + w_jz_i = 2B_{ij}(\kappa(P), \kappa(Q))$ . We normalize by requiring that  $B_{88}(o, o) = 1$ , where  $o = (0, \dots, 0, 1)$ .

We write  $\tilde{V}_2$  for  $L(4\Theta)^+$ ; then  $B$  can be interpreted as an element  $\beta$  of the tensor product  $\tilde{V}_2 \otimes \tilde{V}_2 \otimes V_2^*$ . The last factor  $V_2^*$  can be identified with the space of symmetric  $8 \times 8$  matrices (whose entries are thought of representing  $\frac{1}{2}(w_iz_j + w_jz_i)$  for coordinates  $\underline{w}$  and  $\underline{z}$  of points in  $\mathbb{P}^7$ ) by specifying that a quadratic form  $q \in V_2$  evaluates on such a matrix to  $b(\underline{w}, \underline{z})$  where  $b$  is the bilinear form such that  $q(\underline{x}) = b(\underline{x}, \underline{x})$ . If  $M$  is the matrix of  $b$  and  $B$  is the matrix corresponding to the unordered pair  $\{\underline{w}, \underline{z}\}$ , then the pairing is  $\text{Tr}(M^T B) = 8\langle M, B \rangle$ . Put differently, we obtain the  $(i, j)$ -entry of the matrix by evaluating at the quadratic form  $\xi_i\xi_j$ .

The 2-torsion group  $\mathcal{J}[2]$  acts on each factor, and  $\beta$  must be invariant under the action of  $\mathcal{J}[2] \times \mathcal{J}[2]$  such that  $(T, T')$  acts via  $(T, T', T + T')$  on the three factors (shifting  $P$  by  $T$  and  $Q$  by  $T'$  shifts  $P \pm Q$  by  $T + T'$ ).

We use the basis of  $\tilde{V}_2$  given by  $\mathcal{E}$  and  $y_T$  for the nonzero even 2-torsion points  $T$  (suitably extending  $k$  if necessary); for  $V_2^*$  we use the basis dual to  $(y_T)_{T \text{ even}}$ , which is given by the linear forms

$$y_T^*: v \mapsto \frac{1}{r(T)} \langle z_T, v \rangle .$$

If  $T_1, T_2, T_3$  are even 2-torsion points, then the effect of  $(T, T')$  acting on the corresponding basis element of the triple tensor product is to multiply it by

$$e_2(T, T_1)e_2(T', T_2)e_2(T + T', T_3) = e_2(T, T_1 + T_3)e_2(T', T_2 + T_3) .$$

If this basis element occurs in  $\beta$  with a nonzero coefficient, then this factor must be 1 for all  $T, T'$ , which means that  $T_1 = T_2 = T_3$ . This shows that we must have that

$$\beta = \sum_{T \neq 0} a_T (y_T \otimes y_T \otimes y_T^*) + a_0 (\mathcal{E} \otimes \mathcal{E} \otimes y_0^*) .$$

If we evaluate at the origin in the first component, we obtain (using that  $\mathcal{E}$  vanishes there and that  $y_T(o) = 1$  for  $T \neq 0$  even) that

$$\beta_o = \sum_{T \neq 0} a_T (y_T \otimes y_T^*) .$$

This corresponds to taking  $P = O$ , resulting in the pair  $\pm Q$  leading to  $\{\kappa(Q), \kappa(Q)\}$ . So, taking  $\underline{\xi}$  as coordinates of  $Q$  and using that  $B_{88}(o, o) = 1$ , the  $(i, j)$ -component of this expression, evaluated at  $\underline{\xi}$  in the (now) first component of  $\beta_o$ , must be  $\xi_i \xi_j$ , up to a multiple of  $y_0$ :

$$\xi_i \xi_j \equiv \sum_{T \neq 0} a_T y_T^*(\xi_i \xi_j) \cdot y_T \text{ mod } y_0 .$$

In other words,  $\beta_o$ , interpreted as a linear map  $V_2 \rightarrow \tilde{V}_2$ , is the canonical map; in particular, it sends  $y_T$  to  $y_T$  for all even  $T \neq 0$ , and so  $a_T = 1$  for all  $T \neq 0$ . It only remains to find  $a_0$ ; then  $\beta$  is completely determined. We consider the image of  $\beta$  in  $\text{Sym}^2 \tilde{V}_2 \otimes V_2^*$ , which corresponds to taking  $P = Q$ . This results in the unordered pair  $\{2P, O\}$ , represented (according to our normalization) by the symmetric matrix that is zero everywhere except in the last row and column, where it has entries

$\frac{1}{2}\delta_1, \dots, \frac{1}{2}\delta_7, \delta_8$ . We obtain (recall that  $\mathcal{E}^2 = q_1$  and  $\delta_1 = 4q_1$ ) that

$$\sum_{T \neq 0} y_T^2 \otimes y_T^*(\xi_i \xi_j) + a_0 q_1 \otimes y_0^*(\xi_i \xi_j) = \begin{cases} 0 & \text{if } i, j < 8; \\ \frac{1}{2}\delta_i & \text{if } i < j = 8; \\ \delta_8 & \text{if } i = j = 8. \end{cases}$$

Evaluating at  $y_0 = 2(\xi_1 \xi_8 - \xi_2 \xi_7 + \xi_3 \xi_6 - \xi_4 \xi_5)$ , we find that

$$a_0 q_1 = \delta_1 = 4q_1 .$$

This shows that  $a_0 = 4$ . (Note that if we evaluate at  $y_T$ , we recover the relation

$$y_T^2 = \sum_{j=1}^7 \frac{1}{2}\delta_j \cdot [\xi_j \xi_8] y_T + \delta_8 \cdot [\xi_8^2] y_T = \sum_{j=1}^7 \varepsilon_{9-j} \tau_{9-j} \delta_j + \delta_8 .)$$

We have shown:

**Lemma 8.1** *The element  $\beta \in \tilde{V}_2 \otimes \tilde{V}_2 \otimes V_2^*$  is given by*

$$\beta = \sum_{T \neq 0} y_T \otimes y_T \otimes y_T^* + 4 \mathcal{E} \otimes \mathcal{E} \otimes y_0^* .$$

In terms of matrices, we have that

$$2B(\underline{\xi}, \underline{\zeta}) = \sum_{T \neq 0} \frac{y_T(\underline{\xi}) y_T(\underline{\zeta})}{4r(T)} M_T S + \mathcal{E}(\underline{\xi}) \mathcal{E}(\underline{\zeta}) S . \tag{16}$$

To get the expression for  $B$ , note that  $y_T^*$  corresponds to the matrix

$$(y_T^*(\xi_i \xi_j))_{i,j} = \frac{1}{r(T)} ((z_T, \xi_i \xi_j))_{i,j} = \frac{1}{8r(T)} M_T S .$$

The resulting matrix of bi-quadratic forms corresponding to the first summand in (16) has entries that can be written as elements of  $\mathbb{Z}[f_0, \dots, f_8][\underline{\xi}, \underline{\zeta}]$ . The entries are given in the file `Kum3-biquforms.magma` at [17]. More precisely, let

$$q = \xi_1(f_3 f_5 \xi_4 + f_1 f_7 \xi_5) + f_1 \xi_2 \xi_3 + f_3 \xi_2 \xi_6 + f_5 \xi_3 \xi_7 + f_7 \xi_6 \xi_7 + (\xi_4 + \xi_5) \xi_8 ,$$

then the entries of

$$B(\underline{\xi}, \underline{\zeta}) - \frac{1}{2}(q(\underline{\xi})q(\underline{\zeta}) + \mathcal{E}(\underline{\xi})\mathcal{E}(\underline{\zeta}))S$$

are (up to addition of multiples of  $y_0(\underline{\xi})$  and  $y_0(\underline{\zeta})$ ) in  $\mathbb{Z}[f_0, \dots, f_8][\underline{\xi}, \underline{\zeta}]$ . (Note that  $q \equiv \mathcal{E} \pmod{(2, y_0)}$  so that the term in parentheses is divisible by 2.)



We can now use the matrix  $B$  to perform ‘pseudo-addition’ on  $\mathcal{K}$  in complete analogy to the case of genus 2 described in [6]. This means that given  $\kappa(P), \kappa(Q)$  and  $\kappa(P - Q)$ , we can find  $\kappa(P + Q)$ . This in turn can be used to compute multiples of points on  $\mathcal{K}$  by a variant of the usual divide-and-conquer scheme (‘repeated squaring’).

We can make the upper left entry of  $B$  completely explicit.

**Lemma 8.2** *Recall that  $\langle \cdot, \cdot \rangle_S$  denotes the bilinear form corresponding to the matrix  $S$ . We have that*

$$B_{11}(\underline{\xi}, \underline{\zeta}) \equiv \langle \underline{\xi}, \underline{\zeta} \rangle_S^2 \pmod{(y_0(\underline{\xi}), y_0(\underline{\zeta}))} .$$

*Proof* This follows from  $\langle z_T, \xi_1^2 \rangle = [\xi_8^2]y_T = 1$  (for  $T \neq 0$ ) and Corollary 6.9:

$$B_{11}(\underline{\xi}, \underline{\zeta}) \equiv \sum_{T \neq 0} \frac{y_T(\underline{\xi})y_T(\underline{\zeta})}{8r(T)} = \langle \underline{\xi}, \underline{\zeta} \rangle_S^2 .$$

□

**Corollary 8.3** *For two points  $P, Q \in \mathcal{J}$  with images  $\kappa(P), \kappa(Q) \in \mathcal{K}$ , we have that*

$$P \pm Q \in \Theta \iff \langle \kappa(P), \kappa(Q) \rangle_S = 0 .$$

*Proof* The bilinear form associated to  $S$  vanishes if and only if  $B_{11}(\kappa(P), \kappa(Q))$  vanishes, which means that  $\xi_1(P + Q)\xi_1(P - Q) = 0$ , which in turn is equivalent to  $P + Q \in \Theta$  or  $P - Q \in \Theta$ . □

This is analogous to the duality between the Kummer Surface and the Dual Kummer Surface in the case of a curve of genus 2, see [4, Thm. 4.3.1]. The difference is that here the Kummer variety is self-dual.

We can now also describe the locus of vanishing of  $y_T$  on  $\mathcal{K}$ .

**Corollary 8.4** *Let  $T \neq 0$  be an even 2-torsion point. Then for  $P \in \mathcal{J}$ , we have that  $y_T(\kappa(P)) = 0$  if and only if  $2P + T \in \Theta$ .*

*Proof* This is because  $y_T^2 = \langle \kappa(T), \underline{\delta} \rangle_S$  (up to scaling). □

For  $T = 0$ , we get that  $\mathcal{E}(\kappa(P)) = 0$  if and only if  $2P \in \Theta$ . This is because  $4\mathcal{E}^2 = \delta_1$ .

## 9 Further Properties of the Duplication and the Sum-and-Difference Maps

With a view of considering bad reduction later, we now allow  $k$  to be any field and  $F \in k[x, z]$  to be any binary form of degree 8; in particular,  $F = 0$  is allowed. Note that the relations deduced so far are valid over  $\mathbb{Z}[f_0, \dots, f_8]$  and so can be specialized

to any  $k$  and  $F$ . In this context,  $\mathcal{K}$  denotes the variety in  $\mathbb{P}_k^7$  defined by the specializations of the quadric and the 34 quartics that define the Kummer variety in the generic case, and  $\delta$  denotes the rational map (which now may have base points) from  $\mathcal{K}$  to itself given by the quartics  $\underline{\delta}$ . We can also still consider factorizations  $F = GH$  into two factors of degree 4 (if  $F = 0$ , we take both of the factors to be the zero form of degree 4) and obtain points on  $\mathcal{K}$  that are specializations of the images of 2-torsion points. We will call equivalence classes of such factorizations (up to scaling) ‘nontrivial even 2-torsion points’ for simplicity, even though they do not in general arise from points of order 2 on some algebraic group. If  $T$  is such a nontrivial even 2-torsion point, then we denote the corresponding point on  $\mathcal{K}$  by  $\kappa(T)$ . We normalize the coordinates of  $\kappa(T)$  such that the first coordinate is 1. We also have the associated quadratic form  $y_T$ . If  $F = 0$ , we obtain for example  $\kappa(T) = (1 : 0 : \dots : 0)$  for the unique nontrivial even 2-torsion point, with associated quadratic form  $y_T = \xi_8^2$ .

We now state explicit criteria for the vanishing of  $\underline{\delta}$  at a point on  $\mathcal{K}$ . We first exhibit a necessary condition. For the following, we assume  $k$  to be algebraically closed and of characteristic  $\neq 2$ .

*Remark 9.1* Note that in characteristic 2 we have that  $\delta_1 = \dots = \delta_7 = 0$  and  $\delta_8 = y_T^2$  on  $\mathcal{K}$  for all  $T$ , where

$$y_T = \xi_8^2 + f_6 f_8 \xi_7^2 + f_4 f_8 \xi_6^2 + f_2 f_8 \xi_5^2 + f_4 f_6 \xi_4^2 + f_2 f_6 \xi_3^2 + f_2 f_4 \xi_2^2 + f_2 f_4 f_6 f_8 \xi_1^2,$$

which is the square of a linear form over  $k$  when  $k$  is perfect. Let  $\mathcal{L}$  denote the hyperplane defined by this linear form. Then  $\delta$  restricts to a morphism on  $\mathcal{K} \setminus \mathcal{L}$ , which is constant with image the origin  $(0 : \dots : 0 : 1)$ .

Assume for now that  $F \neq 0$  and write

$$F = F_0^2 F_1 \quad \text{with } F_1 \text{ squarefree.}$$

We define  $\mathcal{T}(F)$  to be the set of nontrivial even 2-torsion points  $T$  associated to factorizations  $(G, H)$  with  $G$  and  $H$  both divisible by  $F_0$ . So  $\mathcal{T}(F)$  is in bijection with the unordered partitions of the roots of  $F_1$  into two sets of equal size. We also define  $\mathcal{T}(0)$  to be the one-element set  $\{T\}$ , where  $T$  corresponds to the factorization  $0 = 0 \cdot 0$ .

**Lemma 9.2** *With the notation introduced above, the following statements are equivalent for a point on  $\mathcal{K}$  with coordinate vector  $\underline{\xi}$ :*

- (i) *For all  $T \in \mathcal{T}(F)$ , we have that  $\langle \kappa(T), \underline{\delta}(\underline{\xi}) \rangle_S = 0$ .*
- (ii) *For all  $T \in \mathcal{T}(F)$ , we have that  $\langle \kappa(T), \underline{\xi} \rangle_S = 0$ .*

*In particular,  $\underline{\delta}(\underline{\xi}) = 0$  implies that  $\langle \kappa(T), \underline{\xi} \rangle_S = 0$  for all  $T \in \mathcal{T}(F)$ .*

*Proof* By Theorem 7.3 (3), we have for all  $T \in \mathcal{T}(F)$  that  $y_T(\underline{\xi})^2 = \langle \kappa(T), \underline{\delta}(\underline{\xi}) \rangle_S$ , so (i) is equivalent to  $y_T(\underline{\xi}) = 0$  for all  $T \in \mathcal{T}(F)$ . When  $F = 0$ , we have  $y_T = \xi_8^2$  and  $\kappa(T) = (1 : 0 : \dots : 0)$  for the unique  $T \in \mathcal{T}(F)$ , so  $y_T(\underline{\xi}) = 0$  is equivalent

to  $\xi_8 = 0$ , which is equivalent to  $\langle \kappa(T), \underline{\xi} \rangle_S = 0$ . If, at the other extreme,  $F$  is squarefree, then one checks<sup>★</sup> that the coordinate vectors of the points in  $\mathcal{T}(F)$  are linearly independent, which implies that (i) is equivalent to  $\underline{\delta}(\underline{\xi}) = 0$  and (ii) is equivalent to  $\underline{\xi} = 0$ . The claim then follows from Theorem 7.3 (4).

We now assume that  $F \neq 0$  and write  $F = F_0^2 F_1$  as above with  $F_1$  squarefree and  $F_0$  non-constant. We check by an explicit computation<sup>★</sup> that

(\*) the  $y_T$  for  $T \in \mathcal{T}(F)$  form a basis of the symmetric square of the space spanned by the linear forms  $\langle \kappa(T), \cdot \rangle_S$  for  $T \in \mathcal{T}(F)$ .

This implies that the vanishing of the  $y_T$  is equivalent to (ii). To verify (\*), we can apply a transformation moving the roots of  $F_0$  to an initial segment of  $(0, \infty, 1, a)$  (where  $a \in k \setminus \{0, 1\}$ ). The most involved case is when  $\deg F_0 = 1$ . We can then take  $F_0 = x$  and find that the linear forms given by the  $T \in \mathcal{T}(F)$  span  $\langle \xi_4, \xi_6, \xi_7, \xi_8 \rangle$  and that the  $10 \times 10$  matrix whose rows are the coefficient vectors of the  $y_T$  with respect to the monomials of degree 2 in these four variables has determinant a power of two times a power of  $\text{disc}(F_1)$ , hence is invertible. The other cases are similar, but simpler. □

This prompts the following definition.

**Definition 9.3** We write  $\mathcal{H}_{\text{good}}$  for the open subscheme

$$\mathcal{H} \setminus \{P : \langle \kappa(T), P \rangle_S = 0 \text{ for all } T \in \mathcal{T}(F)\}$$

of  $\mathcal{H}$ .

Lemma 9.2 now immediately implies the following.

**Corollary 9.4** The rational map  $\delta$  on  $\mathcal{H}$  restricts to a morphism  $\mathcal{H}_{\text{good}} \rightarrow \mathcal{H}_{\text{good}}$ .

We will now consider the ‘bad’ subset  $\mathcal{H} \setminus \mathcal{H}_{\text{good}}$  of  $\mathcal{H}$  in more detail, in particular in relation to the base locus of  $\delta$ , which it contains according to Corollary 9.4. We begin with a simple sufficient condition for a point to be in the base locus.

**Lemma 9.5** Assume that  $F(x, z)$  is divisible by  $z^2$ . Let  $\underline{\xi}$  be the coordinate vector of a point on  $\mathcal{H}$  such that  $\xi_2 = \xi_3 = \xi_4 = \xi_8 = 0$ . Then  $\underline{\delta}(\underline{\xi}) = 0$ .

*Proof* Plugging  $f_7 = f_8 = \xi_2 = \xi_3 = \xi_4 = \xi_8 = 0$  into the expressions for the  $\delta_j$  gives zero<sup>★</sup>. □

We set

$$\mathcal{L}_\infty = \{(\xi_1 : \dots : \xi_8) \in \mathbb{P}^7 : \xi_2 = \xi_3 = \xi_4 = \xi_8 = 0\}.$$

Using the formulas given in Sect. 3 for the action on  $\underline{\xi}$ , one sees easily that  $\mathcal{L}_\infty$  is invariant under scaling of  $x$  and also under shifting  $(x, z) \mapsto (x + \lambda z, z)$  (always assuming that  $f_7 = f_8 = 0$ ), which together generate the stabilizer of  $\infty$  in  $\text{PGL}(2)$ .

For  $F$  with a multiple root at some point  $a \in \mathbb{P}^1$ , let  $\tilde{F}$  be the result of acting on  $F$  by a linear substitution  $\phi$  that moves  $a$  to  $\infty$ ; then  $\tilde{F}$  is divisible by  $z^2$ . We write

$\mathcal{L}_a \subset \mathbb{P}^7$  for the image of  $\mathcal{L}_\infty$  under the automorphism of  $\mathbb{P}^7$  induced by  $\phi^{-1}$ . Since the stabilizer of  $\infty$  in  $\text{PGL}(2)$  leaves  $\mathcal{L}_\infty$  invariant, this definition of  $\mathcal{L}_a$  does not depend on the choice of  $\phi$ . For example,

$$\mathcal{L}_0 = \{(\xi_1 : \dots : \xi_8) \in \mathbb{P}^7 : \xi_4 = \xi_6 = \xi_7 = \xi_8 = 0\}.$$

We write  $A(F) \subset \mathbb{P}^1$  for the set of multiple roots of  $F$ . This is all of  $\mathbb{P}^1$  when  $F = 0$ . Otherwise,  $A(F)$  consists of the roots of  $F_0$  when  $F = F_0^2 F_1$  with  $F_1$  squarefree.

**Corollary 9.6** *If  $P \in \mathcal{H} \cap \mathcal{L}_a$  for some  $a \in A(F)$ , then  $\underline{\delta}(P) = 0$ .*

*Proof* This follows from Lemma 9.5 by applying a suitable automorphism of  $\mathbb{P}^1$ . □

So the base locus of  $\delta$  contains  $\mathcal{H} \cap \bigcup_{a \in A(F)} \mathcal{L}_a$ . When  $F$  is not a nonzero square, we can show that this is exactly the ‘bad set’  $\mathcal{H} \setminus \mathcal{H}_{\text{good}}$ .

**Lemma 9.7** *Assume that  $F$  is not of the form  $F = H^2$  with  $H \neq 0$ . Let  $P$  be in the ‘bad set’  $\mathcal{H} \setminus \mathcal{H}_{\text{good}}$ . Then  $P \in \mathcal{L}_a$  for some  $a \in A(F)$ . In particular,*

$$\mathcal{H}_{\text{good}} = \mathcal{H} \setminus \bigcup_{a \in A(F)} \mathcal{L}_a,$$

and  $\mathcal{H} \setminus \mathcal{H}_{\text{good}} = \mathcal{H} \cap \bigcup_{a \in A(F)} \mathcal{L}_a$  is the base locus of  $\delta$ .

*Proof* Let  $\underline{\xi}$  be a coordinate vector for  $P$ . We write  $F = F_0^2 F_1$  with  $F_1$  squarefree. We split the proof into various cases according to the factorization type of  $F_0$ . If  $F_0$  is constant, there is nothing to prove. Otherwise we move the roots of  $F_0$  to an initial segment of  $(0, \infty, 1)$ .

1.  $F_0 = x$ . In this case the assumption is equivalent to  $\xi_4 = \xi_6 = \xi_7 = \xi_8 = 0$  (compare the proof of Lemma 9.2), so that  $P \in \mathcal{L}_0$ .
2.  $F_0 = x^2$ . The assumption is that  $\xi_7 = \xi_8 = 0$ ; using the equations defining  $\mathcal{H}$  this implies<sup>★</sup> that  $\xi_4 = \xi_6 = 0$ , so  $P \in \mathcal{L}_0$ .
3.  $F_0 = x^3$ . The assumption is that  $\xi_8 = 0$ , which implies<sup>★</sup> that  $\xi_7 = \xi_6 = \xi_4 = 0$ , so  $P \in \mathcal{L}_0$ .
4.  $F_0 = xz$ . In this case the assumption is that  $\xi_4 = \xi_8 = 0$ , which then implies<sup>★</sup> that  $\xi_6 = \xi_7 = 0$  or  $\xi_2 = \xi_3 = 0$ , and so  $P \in \mathcal{L}_0$  or  $P \in \mathcal{L}_\infty$ .
5.  $F_0 = x^2 z$ . The assumption is that  $\xi_8 = 0$ , which leads to<sup>★</sup>  $P \in \mathcal{L}_0$  or  $P \in \mathcal{L}_\infty$ .
6.  $F_0 = xz(x - z)$ . A similar computation shows<sup>★</sup> that  $P \in \mathcal{L}_0 \cup \mathcal{L}_1 \cup \mathcal{L}_\infty$ .
7.  $F = 0$ . Here the assumption is that  $\xi_8 = 0$ . The intersection  $\mathcal{H} \cap \{\xi_8 = 0\}$  is defined<sup>★</sup> by the  $2 \times 2$ -minors of the matrix

$$\begin{pmatrix} \xi_2 & \xi_3 & \xi_4 \\ \xi_3 & \xi_4 + \xi_5 & \xi_6 \\ \xi_4 & \xi_6 & \xi_7 \end{pmatrix},$$

which therefore has rank 1 when evaluated on any point in  $\mathcal{H} \cap \{\xi_8 = 0\}$ . If  $\xi_2 = 0$ , then this implies that  $\xi_3 = \xi_4 = 0$  as well, so that  $P \in \mathcal{L}_\infty$ . Otherwise,

we can make a transformation shifting  $x/z$  by  $\lambda$  as in Sect. 3 that makes  $\tilde{\xi}_7 = 0$  ( $\tilde{\xi}_7$  is a polynomial of degree 4 in  $\lambda$  with leading coefficient  $\xi_2$ , so we can find a suitable  $\lambda$ , since  $k$  is assumed to be algebraically closed). Then we get that  $\tilde{\xi}_8 = \tilde{\xi}_7 = \tilde{\xi}_6 = \tilde{\xi}_4 = 0$ , so the image point is in  $\mathcal{L}_0$ , hence  $P \in \mathcal{L}_\lambda$ .

The last statement follows, since Corollary 9.4 shows that the base scheme of  $\delta$  is contained in  $\mathcal{H} \setminus \mathcal{H}_{\text{good}}$  and Corollary 9.6 shows that it contains the intersection of  $\mathcal{H}$  with the union of the  $\mathcal{L}_a$ . □

We now consider the case  $F = F_0^2 \neq 0$ . Then the curve  $y^2 = F(x, z) = F_0(x, z)^2$  splits into the two components  $y = \pm F_0(x, z)$ . The points on  $\mathcal{H}$  correspond to linear equivalence classes of effective divisors of degree 4, modulo the action of the hyperelliptic involution. So there are three distinct possibilities how the points can be distributed among the two components: two on each, one and three, or all four on the same component. In the last case, we have  $B \equiv \pm F_0 \pmod A$ , and we can change the representative so that  $B = \pm F_0$ , which makes  $C = 0$ . So the two components of  $\text{Pic}^4(\mathcal{C})$  consisting of classes of divisors whose support is contained in one of the two components of  $\mathcal{C}$  map to a single point  $\omega \in \mathcal{H}$ , which one can check<sup>★</sup> coincides with  $\kappa(T)$  for the single  $T \in \mathcal{T}(F)$ ; it satisfies  $\underline{\delta}(\omega) = 0$ .

Now a point  $P$  on the component of  $\mathcal{H}$  corresponding to the distribution of one and three points on the two components, if it is not in the base scheme of  $\delta$ , must satisfy  $\delta(P) = \omega$ . So for such points we have  $\underline{\delta}(\delta(P)) = 0$ , but  $\underline{\delta}(P) \neq 0$ . Let  $\xi$  be coordinates for a point  $P$  with  $\delta(P) = \omega = \kappa(T)$ . Then  $\langle \kappa(T), \underline{\delta}(\xi) \rangle_S = \langle \kappa(T), \kappa(T) \rangle_S = 0$  (all points on  $\mathcal{H}$  satisfy  $\langle \underline{\xi}, \xi \rangle_S = y_0(\xi) = 0$ ). By Lemma 9.2, this is equivalent to  $\langle \kappa(T), \xi \rangle_S = 0$ . We write  $\mathcal{E}$  for the hyperplane given by  $\langle \kappa(T), \xi \rangle_S = 0$ . So in this case  $\mathcal{H}_{\text{good}} = \mathcal{H} \setminus \mathcal{E}$ , and  $P \in \mathcal{H} \cap \mathcal{E} = \mathcal{H} \setminus \mathcal{H}_{\text{good}}$  does not necessarily imply that  $\underline{\delta}(P) = 0$ . But we still have the following.

**Lemma 9.8** *Assume that  $F = F_0^2$  with  $F_0 \neq 0$ . If  $P \in \mathcal{H}$  with  $\underline{\delta}(P) = 0$ , then  $P \in \mathcal{L}_a$  for some  $a \in A(F)$  (which here is simply the set of roots of  $F_0$ ).*

*Proof* We can again assume that the roots of  $F_0$  are given by an initial segment of  $(0, \infty, 1, a)$  (with  $a \neq \infty, 0, 1$ ). We consider the various factorization types of  $F_0$  in turn; they are represented by

$$F_0 = x^4, \quad x^3z, \quad x^2z^2, \quad x^2z(x-z) \quad \text{and} \quad xz(x-z)(x-az).$$

The computations<sup>★</sup> are similar to those done in the proof of Lemma 9.7. The most involved case is when  $F_0$  has four distinct roots. To deal with it successfully, we make use of the Klein Four Group of automorphisms of the set of roots of  $F_0$ . □

We now have a precise description of the base scheme of the duplication map  $\delta$  on  $\mathcal{H}$ , which is given by the quartic forms  $\underline{\delta}$ .

**Proposition 9.9** *Let  $k$  be an algebraically closed field of characteristic  $\neq 2$  and let  $F \in k[x, z]$  be homogeneous of degree 8. We denote by  $\mathcal{H}$  and  $\underline{\delta}$  the objects associated to  $F$ .*

- (1) The base locus of  $\delta$  is  $\mathcal{K} \cap \bigcup_{a \in A(F)} \mathcal{L}_a$ .
- (2) The base locus of  $\delta \circ \delta$  is  $\mathcal{K} \setminus \mathcal{K}_{\text{good}}$ ;  $\delta$  can be iterated indefinitely on  $\mathcal{K}_{\text{good}}$ .
- (3) If  $F$  is not of the form  $F = F_0^2$  with  $F_0 \neq 0$ , then the base locus of  $\delta$  is  $\mathcal{K} \setminus \mathcal{K}_{\text{good}}$ .

*Proof*

- (1) Corollary 9.6 shows that the condition is sufficient. Conversely, if  $\underline{\delta}(P) = 0$ , then Lemmas 9.2, 9.7 and 9.8 show that  $P \in \mathcal{L}_a$  for some multiple root  $a$  of  $F$ .
- (2) The second statement is Corollary 9.4. In view of (3), it is sufficient to consider the case  $F = F_0^2 \neq 0$  for the first statement. If  $P \in \mathcal{K} \setminus \mathcal{K}_{\text{good}}$  is not in the base locus of  $\delta$ , then  $\delta(P) = \omega$ , which is in the base locus of  $\delta$ , so  $P$  is in the base locus of  $\delta \circ \delta$ . Conversely, if  $P$  is in the base locus of  $\delta \circ \delta$ , then  $P$  cannot be in  $\mathcal{K}_{\text{good}}$  by the second statement.
- (3) This follows from Corollary 9.4 and Lemma 9.7. □

We can state a property of the ‘add-and-subtract’ morphism that is similar to that of  $\delta$  given in Corollary 9.4. We write  $\alpha: \text{Sym}^2 \mathcal{K} \rightarrow \text{Sym}^2 \mathcal{K}$  for the map given by the matrix  $B$  as defined in Sect. 8; this is defined for arbitrary  $F \in k[x, z]$ , homogeneous of degree 8. In general  $\alpha$  is only a rational map.

**Lemma 9.10** *Let  $k$  be an algebraically closed field of characteristic different from 2 and let  $F \in k[x, z]$  be homogeneous of degree 8. We denote by  $\mathcal{K}$  and  $\underline{\delta}$  the objects associated to  $F$ . Then  $\alpha$  restricts to a morphism  $\text{Sym}^2 \mathcal{K}_{\text{good}} \rightarrow \text{Sym}^2 \mathcal{K}_{\text{good}}$ .*

*Proof* Note that generically,  $\alpha \circ \alpha = \text{Sym}^2 \delta$ ; this comes from the fact that

$$\{(P + Q) + (P - Q), (P + Q) - (P - Q)\} = \{2P, 2Q\} .$$

If we write  $\underline{\xi} * \underline{\xi}'$  for the symmetric matrix  $\underline{\xi}^\top \cdot \underline{\xi}' + \underline{\xi}'^\top \cdot \underline{\xi}$ , then this relation shows that

$$\underline{\xi} * \underline{\xi}' = 2B(\underline{\xi}, \underline{\xi}') \implies \underline{\delta}(\underline{\xi}) * \underline{\delta}(\underline{\xi}') = 2B(\underline{\zeta}, \underline{\zeta}') , \tag{17}$$

up to a scalar factor, which we find to be 1 by taking  $\underline{\xi} = \underline{\xi}' = (0, \dots, 0, 1)$ . This is then a relation that is valid over  $\mathbb{Z}[f_0, \dots, f_8]$ .

Now let  $\underline{\xi}$  and  $\underline{\xi}'$  be projective coordinate vectors of points in  $\mathcal{K}_{\text{good}}$  and write  $2B(\underline{\xi}, \underline{\xi}') = \underline{\zeta} * \underline{\zeta}'$  for suitable vectors  $\underline{\zeta}, \underline{\zeta}'$ . Then by Corollary 9.4,  $\underline{\delta}(\underline{\xi})$  and  $\underline{\delta}(\underline{\xi}')$  both do not vanish, so  $\underline{\delta}(\underline{\xi}) * \underline{\delta}(\underline{\xi}') \neq 0$ . This implies that  $\underline{\zeta}, \underline{\zeta}' \neq 0$ , which shows that  $\alpha$  is defined on  $\mathcal{K}_{\text{good}}$ . If the point given by  $\underline{\zeta} * \underline{\zeta}'$  were not in  $\text{Sym}^2 \mathcal{K}_{\text{good}}$ , then iterating  $\alpha$  at most four more times would produce zero by Proposition 9.9 (2), contradicting the fact that  $\delta$  can be iterated indefinitely on the points represented by  $\underline{\xi}$  and  $\underline{\xi}'$ . □

### 10 Heights

We now take  $k$  to be a number field (or some other field of characteristic  $\neq 2$  with a collection of absolute values satisfying the product formula, for example a function field in one variable). We also assume again that  $F \in k[x, z]$  is a squarefree binary octic form. Then  $\mathcal{C}$  is a curve of genus 3 over  $k$ , and we have the Jacobian  $\mathcal{J}$  and the Kummer variety  $\mathcal{K}$  associated to  $\mathcal{C}$ . We define the *naive height* on  $\mathcal{J}$  and on  $\mathcal{K}$  to be the standard height on  $\mathbb{P}^7$  with respect to the coordinates  $(\xi_1 : \dots : \xi_8)$ . We denote it by

$$h(P) = \sum_v n_v \log \max\{|\xi_1(P)|_v, \dots, |\xi_8(P)|_v\} \quad \text{for } P \in \mathcal{J}(k) \text{ or } \mathcal{K}(k),$$

where  $v$  runs through the places of  $k$ , the absolute values  $|\cdot|_v$  extend the standard absolute values on  $\mathbb{Q}$  and  $n_v = [K_v : \mathbb{Q}_w]$ , where  $w$  is the place of  $\mathbb{Q}$  lying below  $v$ , so that we have the product formula

$$\prod_v |\alpha|_v^{n_v} = 1 \quad \text{for all } \alpha \in k^\times.$$

Then by general theory (see for example [7, Part B]) the limit

$$\hat{h}(P) = \lim_{n \rightarrow \infty} \frac{h(nP)}{n^2}$$

exists and differs from  $h(P)$  by a bounded amount. This is the *canonical height* of  $P$ . One of our goals in this section will be to find an explicit bound for

$$\beta = \sup_{P \in \mathcal{J}(k)} (h(P) - \hat{h}(P)).$$

We refer to [12] for a detailed study of heights in the case of Jacobians of curves of genus 2, with input from [14] and [16]. We will now proceed to obtain some comparable results in our case of hyperelliptic genus 3 Jacobians. Most of this is based on the following telescoping series trick going back to Tate: we write

$$\hat{h}(P) = \lim_{n \rightarrow \infty} 4^{-n} h(2^n P) = h(P) + \sum_{n=0}^{\infty} 4^{-(n+1)} (h(2^{n+1} P) - 4h(2^n P))$$

and split the term  $h(2P) - 4h(P)$  into local components as follows:

$$h(2P) - 4h(P) = \sum_v n_v \left( \max_j \log |\delta_j(\underline{\xi}(P))|_v - 4 \max_j \log |\xi_j(P)|_v \right) = \sum_v n_v \varepsilon_v(P)$$

with  $\varepsilon_v(P) = \max_j \log |\delta_j(\underline{\xi}(P))|_v - 4 \max_j \log |\xi_j(P)|_v$ , which is independent of the scaling of the coordinates  $\underline{\xi}(P)$  and so can be defined for all  $P \in \mathcal{J}(k_v)$  or  $\mathcal{K}(k_v)$ . Then  $\varepsilon_v: \mathcal{K}(k_v) \rightarrow \mathbb{R}$  is continuous, so (since  $\mathcal{K}(k_v)$  is compact) it is bounded. If  $-\gamma_v \leq \inf_{P \in \mathcal{K}(k_v)} \varepsilon_v(P)$ , then we have that

$$\beta \leq \sum_v n_v \sum_{n=0}^{\infty} 4^{-(n+1)} \gamma_v = \frac{1}{3} \sum_v n_v \gamma_v .$$

So we will now obtain estimates for  $\gamma_v$ . We follow closely the strategy of [14]. Note that writing

$$\mu_v(P) = \sum_{n=0}^{\infty} 4^{-(n+1)} \varepsilon_v(2^n P) = \lim_{n \rightarrow \infty} 4^{-n} \max_j \log |\underline{\delta}^{2^n}(\underline{\xi}(P))|_v - \max_j \log |\xi_j(P)| ,$$

we also have that

$$\hat{h}(P) = h(P) + \sum_v n_v \mu_v(P) .$$

We assume that the polynomial defining the curve  $\mathcal{C}$  has coefficients in the ring of integers of  $k$ . Then the matrices  $M_T$  defined in Sect. 5 for even 2-torsion points have entries that are algebraic integers. We use  $\mathcal{O}$  to denote the ring of all algebraic integers. Let  $\underline{\xi}$  be coordinates of a point on  $\mathcal{K}$ . Then Theorem 7.3 (3) tells us that for all even 2-torsion points  $T \neq 0$ , we have that

$$y_T(\underline{\xi})^2 \in \mathcal{O} \delta_1(\underline{\xi}) + \mathcal{O} \delta_2(\underline{\xi}) + \dots + \mathcal{O} \delta_8(\underline{\xi})$$

and Lemma 6.8 tells us that (note that the coefficient of  $\xi_{9-j}^2$  in  $y_0$  is zero)

$$\xi_j^2 \in \sum_{T \neq 0, \text{even}} \frac{1}{8r(T)} \mathcal{O} y_T(\underline{\xi}) .$$

**Lemma 10.1** *Let  $v$  be a non-archimedean place of  $k$ . Then for  $P \in \mathcal{K}(k_v)$ , we have that*

$$\log |2^6 \text{disc}(F)|_v \leq \log \min_T |2^6 r(T)|_v \leq \varepsilon_v(P) \leq 0 ,$$

where  $T$  runs through the non-trivial even 2-torsion points.

*Proof* Let  $\underline{\xi}$  be coordinates for  $P$  and write  $d_j = \delta_j(\underline{\xi})$  for  $j = 1, \dots, 8$ . Then for all even  $T \neq 0$ ,

$$|y_T(\underline{\xi})|_v^2 \leq \max_j |d_j|_v$$



and

$$|\xi_j|_v^4 \leq \max_T |8r(T)|_v^{-2} |y_T(\underline{\xi})|_v^2 \leq \max_T |8r(T)|_v^{-2} \max_j |d_j|_v .$$

So

$$\varepsilon_v(P) = \log \max_j |d_j|_v - 4 \log \max_j |\xi_j|_v \geq \log \min_T |2^6 r(T)^2|_v .$$

Since  $r(T)^2$  divides the discriminant  $\text{disc}(F)$ , the first inequality on the left also follows. The upper bound follows from the fact that the polynomials  $\delta_j$  have integral coefficients. □

Since  $\varepsilon_v(P)$  is an integral multiple of the logarithm of the absolute value of a uniformizer  $\pi_v$ , we can sometimes gain a little bit by using

$$\varepsilon_v(P) \geq - \left\lfloor \max_T v(|2^6 r(T)^2|) \right\rfloor \log |\pi_v|_v ,$$

where  $v$  denotes the  $v$ -adic additive valuation, normalized so that  $v(\pi_v) = 1$ .

*Example 10.2* For the curve

$$y^2 = 4x^7 - 4x + 1$$

over  $\mathbb{Q}$  and  $v = 2$ , the discriminant bound gives  $\star \varepsilon_2(P) \geq -22 \log 2$ , since the discriminant of the polynomial on the right hand side (considered as a dehomogenized binary octic form) has 2-adic valuation 16. To get a better bound, we consider the resultants  $r(T)$ . If we write

$$f(x) = 4x^7 - 4x + 1 = 4g(x)h(x)$$

with  $g$  and  $h$  monic of degree 3 and 4, respectively, then  $r(T) = 2^8 \text{Res}(g, h)$ . From the Newton Polygon of  $f$  we see that all roots  $\theta$  of  $f$  satisfy  $v_2(\theta) = -2/7$ . This gives  $v_2(r(T)) \geq 32/7$ . Since the product of all 35 resultants  $r(T)$  is the tenth power of the discriminant, we must have equality. We get that  $\varepsilon_2(P) \geq -(15 + \frac{1}{7}) \log 2$ , which can be improved to  $-15 \log 2$ , so that we have the bound  $-\mu_2 \leq 5 \log 2$ .

**Corollary 10.3** *Assume that  $k = \mathbb{Q}$ . Then we have that*

$$\beta \leq \frac{1}{3} \log |2^6 \text{disc}(F)| + \frac{1}{3} \gamma_\infty .$$

To get a bound on  $\gamma_\infty$ , we use the archimedean triangle inequality. We write  $\tau_j(T)$  for the coordinates of a non-trivial even 2-torsion point  $T$  (with  $\tau_1(T) = 1$ ) and  $v_j(T)$  for the coefficients in the formula for  $\xi_j^2$ , so that we have

$$\xi_j^2 = \sum_T v_j(T) y_T .$$

**Lemma 10.4** *Let  $v$  be an archimedean place of  $k$ . Then we have that*

$$\gamma_v \leq \log \max_j \left( \sum_T |v_j(T)|_v \sqrt{\sum_{i=1}^8 |\tau_i(T)|_v} \right)^2 .$$

*Proof* Similarly as in the non-archimedean case, we have that

$$|y_T(\underline{\xi})|_v^2 \leq \sum_{j=1}^8 |\tau_j(T)|_v \max_j |d_j|_v$$

and

$$\max_j |\xi_j|_v^2 \leq \max_j \sum_T |v_j(T)|_v |y_T(\underline{\xi})|_v .$$

Combining these gives the result. □

As in [12, Sect. 16B], we can refine this result somewhat. Define a function

$$f: \mathbb{R}_{\geq 0}^8 \longrightarrow \mathbb{R}_{\geq 0}^8, \quad (d_1, \dots, d_8) \longmapsto \left( \sqrt{\sum_T |v_j(T)|_v \sqrt{\sum_{i=1}^8 |\tau_i(T) d_{9-i}|_v}} \right)_{1 \leq j \leq 8} .$$

We write  $\|(x_1, \dots, x_8)\|_\infty = \max\{|x_1|, \dots, |x_8|\}$  for the maximum norm.

**Lemma 10.5** *Define a sequence  $(b_n)$  in  $\mathbb{R}_{\geq 0}^8$  by*

$$b_0 = (1, \dots, 1) \quad \text{and} \quad b_{n+1} = f(b_n) .$$

*The  $(b_n)$  converges to a limit  $b$ , and we have that*

$$-\mu_v(P) \leq \frac{4^N}{4^N - 1} \log \|b_N\|_\infty$$

*for all  $N \geq 1$  and all  $P \in \mathcal{J}(\mathbb{C})$ . In particular,  $\sup -\mu_v(\mathcal{J}(\mathbb{C})) \leq \log \|b\|_\infty$ .*

*Proof* See the proof of [12, Lemma 16.1]. □

*Example 10.6* For the curve

$$y^2 = 4x^7 - 4x + 1 ,$$

the bound  $\gamma_\infty/3$  is 1.15134, whereas with  $N = 8$ , we obtain the considerably better bound  $-\mu_\infty \leq 0.51852$ .

We can improve this a little bit more if  $k_v = \mathbb{R}$ , by making use of the fact that the coordinates of the points involved are real, but the  $\tau_i(T)$  may be non-real. This can give a better bound on

$$|y_T^2|_v \leq \max_{|\delta_i| \leq d_i} \left| \sum_{i=1}^8 \varepsilon_i \tau_i(T) \delta_{9-i} \right|_v .$$

For the curve above, this improves★ the upper bound for  $-\mu_\infty$  to 0.43829.

Now we show that in the most common cases of bad reduction, there is in fact no contribution to the height difference bound. This result is similar to [16, Prop. 5.2].

**Lemma 10.7** *Let  $v$  be a non-archimedean place of  $k$  of odd residue characteristic. Assume that the reduction of  $F$  at  $v$  has a simple root and that the model of  $\mathcal{C}$  given by  $y^2 = F(x, z)$  is regular at  $v$ . Then  $\mu_v(P) = \varepsilon_v(P) = 0$  for all  $P \in \mathcal{J}(k_v)$ .*

Note that the assumptions on the model are satisfied when  $v(\text{disc}(F)) = 1$ .

*Proof* We work with a suitable unramified extension  $K$  of  $k_v$ , so that the reduction  $\bar{F}$  of  $F$  splits into linear factors over the residue field. We denote the ring of integers of  $K$  by  $\mathcal{O}$ . By assumption,  $\bar{F}$  has a simple root, which by Hensel’s Lemma lifts to a root of  $F$  in  $\mathbb{P}^1(K)$ . We can use a transformation defined over  $\mathcal{O}$  to move this root of  $F$  to  $\infty$ . Then we have  $f_8 = 0$  and  $v(f_7) = 0$ . We can further scale  $F$  (at the cost of at most a further quadratic unramified extension) so that  $f_7 = 1$ .

Assume that  $P \in \mathcal{J}(K)$  has  $\varepsilon_v(P) \neq 0$  and let  $\xi$  be normalized coordinates for  $\kappa(P) \in \mathcal{X}(K)$  (i.e., such that the coordinates are in  $\mathcal{O}$  and at least one of them is in  $\mathcal{O}^\times$ ). By Proposition 9.9, the reduction of  $P$  must lie in some  $\mathcal{L}_a$  where  $a \neq \infty$  is a multiple root of  $\bar{F}$ . We can shift  $a$  to 0; then the coordinates  $\xi_4, \xi_6, \xi_7$  and  $\xi_8$  have positive valuation. We also have  $v(f_0) = 1$  (this is because the model is regular at the point  $(0 : 0 : 1)$  in the reduction) and  $v(f_1) \geq 1$  (since  $a = 0$  is a multiple root of  $\bar{F}$ ).

Now assume first that  $v(\xi_1) = 0$ ; then we can scale  $\xi$  such that  $\xi_1 = 1$ . We consider the quantity  $\mu_{034}$  introduced in Sect. 4; its value on  $P$  is in  $K$ . By (10), we have that

$$\begin{aligned} \mu_{034}^2 &= \eta_{00}\eta_{34}^2 + \eta_{33}\eta_{04}^2 + \eta_{44}\eta_{03}^2 - 4\eta_{00}\eta_{33}\eta_{44} - \eta_{03}\eta_{04}\eta_{34} \\ &= f_0 + (f_6 - \xi_2)\xi_4^2 - \xi_6\xi_4 \end{aligned}$$

(note that  $\eta_{44} = f_8 = 0, \eta_{34} = f_7 = 1, \eta_{33} = f_6 - \eta_{24}, \eta_{24} = \xi_2, \eta_{04} = \xi_4$  and  $\eta_{03} = \xi_6$ ). Now since  $v(f_0) = 1, v(\xi_4) \geq 1$  and  $v(\xi_6) \geq 1$ , we find that  $2v(\mu_{034}) = 1$ , a contradiction.

So we must have that  $v(\xi_1) > 0$ . One can check★ that

$$\begin{aligned} v_1 &= (\xi_4 - \xi_5)\mu_{013} + \xi_7\mu_{123}, \\ v_2 &= \xi_3\mu_{014} - \xi_4\mu_{024} \quad \text{and} \\ v_3 &= \xi_2\mu_{024} - \xi_4\mu_{134} \end{aligned}$$

are functions in  $L(4\Theta)$ , which are clearly odd, so their squares can be written as quartics in the  $\xi_j$  by Lemma 2.3. Let  $I$  be the square of the ideal generated by  $f_0, f_1, \xi_1, \xi_4, \xi_6, \xi_7, \xi_8$ ; then anything in  $I$  has valuation at least 2. We find<sup>★</sup> that modulo  $I$ ,

$$v_1^2 \equiv f_0 \xi_5^4, \quad v_2^2 \equiv f_0 \xi_3^4, \quad v_3^2 \equiv f_0 \xi_2^4.$$

Since (at least) one of  $\xi_2, \xi_3, \xi_5$  is a unit and  $v(f_0) = 1$ , we obtain a contradiction again.

Therefore  $\varepsilon_v(P) = 0$  for all  $P \in \mathcal{J}(K)$ , which implies that  $\mu_v(P) = 0$  as well. □

*Example 10.8* The discriminant of the curve

$$\mathcal{C}: y^2 = 4x^7 - 4x + 1$$

is<sup>★</sup>  $2^{28} \cdot 19 \cdot 223 \cdot 44909$ . Lemma 10.7 now implies that  $\varepsilon_v(P) = 0$  for all  $P \in \mathcal{J}(\mathbb{Q}_v)$  for all places  $v$  except 2 and  $\infty$ , including the bad primes 19, 223 and 44909. So, using Examples 10.2 and 10.6, we obtain the bound

$$h(P) \leq \hat{h}(P) + 5 \log 2 + 0.43829 \leq \hat{h}(P) + 3.90403$$

for all  $P \in \mathcal{J}(\mathbb{Q})$ .

To compute the canonical height  $\hat{h}(P)$  for some point  $P \in \mathcal{J}(\mathbb{Q})$  (say, for a hyperelliptic curve  $\mathcal{C}$  of genus 3 defined over  $\mathbb{Q}$ ), we can use any of the approaches described in [12], except the most efficient one (building on Proposition 14.3 in loc. cit.), since we have so far no general bound on the denominator of  $\mu_p / \log p$  in terms of the discriminant. A little bit of care is needed, since contrary to the genus 2 situation,  $\varepsilon_v = 0$  and  $\mu_v = 0$  are not necessarily equivalent—there can be a difference when the reduction of  $F$  is a constant times a square—so the criterion for a point to be in the subgroup on which  $\mu_v = 0$  has to be taken as  $\overline{\kappa(P)} \in \mathcal{K}_{\text{good}}(\mathbb{F})$ , where  $\overline{\kappa(P)}$  is the reduction of  $\kappa(P)$  at  $v$  and  $\mathbb{F}$  is the residue class field.

We can describe the subset on which  $\mu_v = 0$  and show that it is a subgroup and that  $\mu_v$  factors through the quotient.

**Theorem 10.9** *Let  $v$  be a non-archimedean place of  $k$  of odd residue characteristic. Write  $\mathcal{J}(k_v)_{\text{good}}$  for the subset of  $\mathcal{J}(k_v)$  consisting of the points  $P$  such that  $\kappa(P)$  reduces to a point in  $\mathcal{K}_{\text{good}}(\mathbb{F})$ . Then  $\mathcal{J}(k_v)_{\text{good}} = \{P \in \mathcal{J}(k_v) : \mu_v(P) = 0\}$  is a subgroup of finite index of  $\mathcal{J}(k_v)$ , and  $\varepsilon_v$  and  $\mu_v$  factor through the quotient  $\mathcal{J}(k_v) / \mathcal{J}(k_v)_{\text{good}}$ .*

*Proof* That  $\mathcal{J}(k_v)_{\text{good}}$  is a group follows from Lemma 9.10: If  $P_1$  and  $P_2$  are in  $\mathcal{J}(k_v)_{\text{good}}$ , then  $P_1 \pm P_2$  reduce to a point in  $\mathcal{K}_{\text{good}}$  as well. This subgroup contains the kernel of reduction, which is of finite index, so it is itself of finite index. That  $\mathcal{J}(k_v)_{\text{good}} = \{P \in \mathcal{J}(k_v) : \mu_v(P) = 0\}$  follows from the results of Sect. 9.

It remains to show that  $\mu_v$  (and therefore also  $\varepsilon_v$ , since  $\varepsilon_v(P) = 4\mu_v(P) - \mu_v(2P)$ ) factors through the quotient group. Let  $P, P' \in \mathcal{J}(k_v)$  and let  $\underline{\xi}$  and  $\underline{\xi}'$  be coordinate vectors for  $\kappa(P)$  and  $\kappa(P')$ , respectively. We can then choose coordinate vectors  $\underline{\zeta}$  and  $\underline{\zeta}'$  for  $\kappa(P'+P)$  and  $\kappa(P'-P)$ , respectively, such that  $\underline{\zeta} * \underline{\zeta}' = 2B(\underline{\xi}, \underline{\xi}')$ . Iterating the implication in (17) then gives

$$\underline{\delta}(\underline{\zeta}) * \underline{\delta}(\underline{\zeta}') = 2B(\underline{\delta}(\underline{\xi}), \underline{\delta}(\underline{\xi}')) ,$$

and we can iterate this relation further. If  $\underline{\alpha}$  is a vector or matrix, then we write  $|\underline{\alpha}|_v$  for the maximum of the  $v$ -adic absolute values of the entries of  $\underline{\alpha}$ . Define

$$\varepsilon_v(P, P') = \log |2B(\underline{\xi}, \underline{\xi}')|_v - 2 \log |\underline{\xi}|_v - 2 \log |\underline{\xi}'|_v$$

(this does not depend on the scaling of the coordinate vectors) and note that  $|\underline{\zeta} * \underline{\zeta}'|_v = |\underline{\zeta}|_v \cdot |\underline{\zeta}'|_v$  (here we use that the residue characteristic is odd). We then see that  $\mu_v(\bar{P}) = \bar{0}$  implies  $\mu_v(P + Q) = \mu_v(Q)$  for all  $Q \in \mathcal{J}(k_v)$  in the same way as in the proof of [12, Lemma 3.7]. □

## 11 An Application

We consider the curve

$$\mathcal{C}' : y^2 - y = x^7 - x ,$$

which is isomorphic to the curve

$$\mathcal{C} : y^2 = 4x^7 - 4x + 1 ,$$

which we have been using as our running example. Our results can now be used to determine a set of generators for the Mordell-Weil group  $\mathcal{J}(\mathbb{Q})$ . This is the key ingredient for the method that determines the set of integral points on a hyperelliptic curve as in [3]. We carry out the necessary computations and thence find all the integral solutions of the equation  $y^2 - y = x^7 - x$ .

A 2-descent on the Jacobian  $\mathcal{J}$  of  $\mathcal{C}$  as described in [15] and implemented in Magma [1] shows that the rank of  $\mathcal{J}(\mathbb{Q})$  is at most 4. We have  $\# \mathcal{J}(\mathbb{F}_3) = 94$  and  $\# \mathcal{J}(\mathbb{F}_7) = 911$ , which implies that  $\mathcal{J}(\mathbb{Q})$  is torsion free (the torsion subgroup injects into  $\mathcal{J}(\mathbb{F}_p)$  for  $p$  an odd prime of good reduction). We have the obvious points  $(0, \pm 1), (\pm 1, \pm 1), (\pm\omega, \pm 1), (\pm\omega^2, \pm 1)$  on  $\mathcal{C}$ , where  $\omega$  denotes a primitive cube root of unity, together with the point at infinity. We can check that the rational divisors of degree zero on  $\mathcal{C}$  supported in these points generate a subgroup  $G$  of  $\mathcal{J}(\mathbb{Q})$  of rank 4, which already shows that  $\mathcal{J}(\mathbb{Q}) \cong \mathbb{Z}^4$ . Computing canonical heights, either with an approach as in [12] or with the more general algorithms due

independently to Holmes [8] and Müller [10], we find that an LLL-reduced basis of the lattice  $(G, \hat{h})$  is given by

$$P_1 = [(0, 1) - \infty], \quad P_2 = [(1, 1) - \infty], \quad P_3 = [(-1, 1) - \infty],$$

$$P_4 = [(1, -1) + (\omega, -1) + (\omega^2, -1) - 3 \cdot \infty]$$

with height pairing matrix

$$M \approx \begin{pmatrix} 0.17820 & 0.01340 & -0.05683 & 0.08269 \\ 0.01340 & 0.81995 & -0.34461 & -0.26775 \\ -0.05683 & -0.34461 & 0.98526 & 0.37358 \\ 0.08269 & -0.26775 & 0.37358 & 1.07765 \end{pmatrix}.$$

We can bound the covering radius  $\rho$  of this lattice by  $\rho^2 \leq 0.50752$ . Using Example 10.8, it follows that if  $G \neq \mathcal{J}(\mathbb{Q})$ , then there must be a point  $P \in \mathcal{J}(\mathbb{Q}) \setminus G$  satisfying

$$h(P) \leq \rho^2 + \beta \leq 0.50752 + 3.90403 = 4.41155,$$

so that we can write  $\kappa(P) = (\xi_1 : \xi_2 : \dots : \xi_8) \in \mathcal{X}(\mathbb{Q})$  with coprime integers  $\xi_j$  such that  $|\xi_j| \leq \lfloor e^{4.41155} \rfloor = 82$ . We can enumerate all points in  $\mathcal{X}(\mathbb{Q})$  up to this height bound and check that no such point lifts to a point in  $\mathcal{J}(\mathbb{Q})$  that is not in  $G$ . (Compare [16, §7] for this approach to determining the Mordell-Weil group.) We have therefore proved the following.

**Proposition 11.1** *The group  $\mathcal{J}(\mathbb{Q})$  is free abelian of rank 4, generated by the points  $P_1, P_2, P_3$  and  $P_4$ .*

A Mordell-Weil sieve computation as described in [2] shows that any unknown rational point on  $\mathcal{C}$  must differ from one of the eleven known points

$$\infty, (-1, \pm 1), (0, \pm 1), \left(\frac{1}{4}, \pm \frac{1}{64}\right), (1, \pm 1), (5, \pm 559)$$

by an element of  $B \cdot \mathcal{J}(\mathbb{Q})$ , where

$$B = 2^6 \cdot 3^3 \cdot 5^3 \cdot 7^2 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 37 \cdot 43 \cdot 47 \cdot 53 \cdot 61 \cdot 71 \cdot 79 \cdot 83 \cdot 97$$

$$\approx 1.1 \cdot 10^{32}.$$

In particular, we know that every rational point is in the same coset modulo  $2 \mathcal{J}(\mathbb{Q})$  as one of the known points. For each of these cosets (there are five such cosets: the points with  $x$ -coordinate  $1/4$  are in the same coset as those with  $x$ -coordinate 0), we compute a bound for the size of the  $x$ -coordinate of an integral point on  $\mathcal{C}$  with the method given in [3]. This shows that

$$\log |x| \leq 2 \cdot 10^{1229}$$

for any such point  $(x, y)$ . On the other hand, using the second stage of the Mordell-Weil sieve as explained in [3], we obtain a lattice  $L \subset \mathbb{Z}^4$  of index  $\approx 2.3 \cdot 10^{2505}$  such that the minimal squared euclidean length of a nonzero element of  $L$  is roughly  $2.55 \cdot 10^{1252}$  and such that every rational point on  $\mathcal{C}$  differs from one of the known points by an element in the image of  $L$  in  $\mathcal{J}(\mathbb{Q})$  under the isomorphism  $\mathbb{Z}^4 \xrightarrow{\cong} \mathcal{J}(\mathbb{Q})$  given by the basis above. This is more than sufficient to produce a contradiction to the assumption that there is an integral point we do not already know. We have therefore proved:

**Theorem 11.2** *The only points in  $\mathcal{C}(\mathbb{Q})$  with integral  $x$ -coordinate are*

$$(-1, \pm 1), (0, \pm 1), (1, \pm 1), (5, \pm 559).$$

*In particular, the only integral solutions of the equation*

$$y^2 - y = x^7 - x$$

*are  $(x, y) = (-1, 0), (-1, 1), (0, 0), (0, 1), (1, 0), (1, 1), (5, 280)$  and  $(5, -279)$ .*

## 12 Quadratic Twists

Let  $F$  be a squarefree octic binary form over a field  $k$  not of characteristic 2 and let  $c \in k^\times$ . Then the Kummer varieties  $\mathcal{K}$  and  $\mathcal{K}^{(c)}$  associated to  $F$  and to  $cF$ , respectively, are isomorphic, with an isomorphism from the former to the latter being given by

$$(\xi_1 : \xi_2 : \xi_3 : \dots : \xi_7 : \xi_8) \longmapsto (\xi_1 : c\xi_2 : c\xi_3 : \dots : c\xi_7 : c^2\xi_8).$$

We can therefore use  $\mathcal{K}$  as a model for the Kummer variety associated to the curve  $\mathcal{C}^{(c)}: y^2 = cF(x, z)$ . This will in general change the naive height of a point  $P \in \mathcal{J}^{(c)}(\mathbb{Q})$ , but will not affect the canonical height, which is insensitive to automorphisms of the ambient  $\mathbb{P}^7$ . The duplication map is preserved by the isomorphism. This implies that the height difference bounds of Lemmas 10.1 and 10.5 for  $F$  apply to  $\mathcal{K}$ , even when  $\mathcal{K}$  is used as the Kummer variety of  $\mathcal{C}^{(c)}$ . This is because these bounds are valid for all  $k_v$ -points on  $\mathcal{K}$ , regardless of whether they lift to points in  $\mathcal{J}(k_v)$  or not. Note, however, that the result of Lemma 10.7 does *not* carry over: in the interesting case,  $c$  has odd valuation at  $v$ , and so we are in effect looking at (certain) points on  $\mathcal{J}$  defined over a ramified quadratic extension of  $k_v$ . Since in terms of the original valuation, the possible values of the valuation on this larger field are now in  $\frac{1}{2}\mathbb{Z}$ , the argument in the proof of Lemma 10.7 breaks down.

When working with this model, one has to modify the criterion for a point to lift to  $\mathcal{J}(k)$  by multiplying the  $\mu_{ijk}$  by  $c$ .

As an example, consider the curve given by

$$\begin{pmatrix} y \\ 2 \end{pmatrix} = \begin{pmatrix} x \\ 7 \end{pmatrix}.$$

It is isomorphic to the curve

$$\mathcal{C}: y^2 = 70(x^7 - 14x^5 + 49x^3 - 36x + 630) = 70F(x, 1)$$

where  $F$  is the obvious octic binary form. The 2-Selmer rank of its Jacobian  $\mathcal{J}$  is 9,  $\mathcal{J}(\mathbb{Q})$  is torsion free, and the subgroup  $G$  of  $\mathcal{J}(\mathbb{Q})$  generated by differences of the 27 small rational points on  $\mathcal{C}$  has rank 9 with LLL-reduced basis

$$\begin{aligned} & [(-2, 210) - \infty], \quad [(1, 210) - \infty], \quad [(3, 210) - \infty], \quad [(2, 210) - \infty], \\ & [(-3, 210) - \infty], \quad [(4, 630) - \infty], \quad [(-\frac{5}{2}, -\frac{1785}{8}) + (3, 210) + (4, 630) - 3\infty], \\ & [(0, 210) - \infty], \quad [(6, 3570) - \infty]. \end{aligned}$$

We would like to show that these points are actually generators of  $\mathcal{J}(\mathbb{Q})$ .

Using the Kummer variety associated to  $70F$ , we obtain the following bound for  $\mu_v$  at the bad primes and infinity (using the valuations of the resultants  $r(T)$ , Lemma 10.7 and Lemma 10.5):

$$\begin{aligned} \mu_2 &\geq -6 \log 2, \quad \mu_3 \geq -\frac{10}{3} \log 3, \quad \mu_5 \geq -\frac{10}{3} \log 5, \quad \mu_7 \geq -\frac{8}{3} \log 7, \\ \mu_{13} &= 0, \quad \mu_{17} \geq -\frac{2}{3} \log 17, \quad \mu_{15717742643} = 0, \quad \mu_\infty \geq -0.6152. \end{aligned}$$

The resulting bound  $\approx 20.88$  for  $h - \hat{h}$  is *much* too large to be useful.

However, using the Kummer variety associated to  $F$ , we find that

$$\begin{aligned} \mu_2 &\geq -\frac{10}{3} \log 2, \quad \mu_3 \geq -\frac{10}{3} \log 3, \quad \mu_5 \geq -\frac{2}{3} \log 5, \quad \mu_7 = 0, \\ \mu_{13} &= 0, \quad \mu_{17} \geq -\frac{2}{3} \log 17, \quad \mu_{15717742643} = 0, \quad \mu_\infty \geq -0.6152. \end{aligned}$$

This gives a bound of  $\approx 9.55$  (now for a different naive height), which is already a lot better, but still a bit too large for practical purposes. Now one can check that for a point  $P \in \mathcal{J}(\mathbb{Q}_p)$  with  $p \in \{5, 17\}$ , we always have  $\kappa(2P) \in \mathcal{K}_{\text{good}}$ . This implies that we get a better estimate

$$h(2P) \leq \hat{h}(2P) + \frac{10}{3} \log 6 + 0.6152 \leq \hat{h}(2P) + 6.588$$

for  $P \in \mathcal{J}(\mathbb{Q})$ . A further study of the situation at  $p = 3$  reveals that  $\mu_3$  factors through the component group  $\Phi$  of the Néron model of  $\mathcal{J}$  over  $\mathbb{Z}_3$ , which has the



structure  $\mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ , and that the minimum of  $\mu_3$  on  $2\Phi$  is  $-\frac{5}{3} \log 3$ . This leads to

$$h(2P) \leq \hat{h}(2P) + 4.757. \tag{18}$$

We enumerate all points  $P$  in  $\mathcal{J}(\mathbb{Q})$  such that  $h(P) \leq \log 2000$  using a  $p$ -adic lattice-based approach with  $p = 277$ , as follows. For each of the 10,965,233 points  $\kappa(0) \neq Q \in \mathcal{H}(\mathbb{F}_p)$  that are in the image of  $\mathcal{J}(\mathbb{F}_p)$ , we construct a sublattice  $L_Q$  of  $\mathbb{Z}^8$  such that for every point  $P \in \mathcal{J}(\mathbb{Q})$  such that  $\kappa(P)$  reduces mod  $p$  to  $Q$ , every integral coordinate vector for  $\kappa(P)$  is in  $L_Q$  and such that  $(\mathbb{Z}^8 : L_Q) \geq p^{11}$ . We then search for short vectors in  $L_Q$ , thus obtaining all points of multiplicative naive height  $\leq 2000$ . Note that all these points are smooth on  $\mathcal{H}$  over  $\mathbb{F}_p$ , since  $\#\mathcal{J}(\mathbb{F}_p)$  is odd. This computation took about two CPU weeks. For points reducing to the origin, we see that the quadratic equation satisfied by points on  $\mathcal{H}$  forces  $\xi_1$  to be divisible by  $p^2 > 2000$ , so  $\xi_1 = 0$ , and every such point must be on the theta divisor. A point  $P = [P_1 + P_2 - 2 \cdot \infty] \in \mathcal{J}(\mathbb{Q})$  reduces to the origin if and only if the points  $P_1$  and  $P_2$  reduce to opposite points; in particular, the polynomial whose roots are the  $x$ -coordinates of  $P_1$  and  $P_2$  reduces to a square mod  $p$ . Since the coefficients are bounded by  $7 = \lfloor 2000/p \rfloor$ , divisibility of the discriminant by  $p$  implies that the discriminant vanishes, so that  $P_1 = P_2$ , and the point  $P$  does not reduce to the origin, after all.

We find no point  $P$  such that  $0 < \hat{h}(P) < \hat{h}(P_1) \approx 1.619$ , where  $P_1$  is a known point of minimal positive canonical height, and no points  $P$  outside  $G$  such that  $\hat{h}(P) < 2.844 \approx \log 2000 - 4.757$ . Since the bound (18) is only valid on  $2\mathcal{J}(\mathbb{Q})$ , this implies that there are no points  $P \in \mathcal{J}(\mathbb{Q})$  with  $0 < \hat{h}(P) < 0.711 =: m$ . Using the bound (see [6])

$$I \leq \left\lfloor \sqrt{\frac{\gamma_9^9 \det(M)}{m^9}} \right\rfloor \leq 1787$$

for the index of the known subgroup in  $\mathcal{J}(\mathbb{Q})$ , where  $\gamma_9$  denotes the Hermite constant for 9-dimensional lattices and  $M$  is the height pairing matrix of the basis of the known subgroup of  $\mathcal{J}(\mathbb{Q})$ , we see that it suffices to rule out all primes up to 1787 as possible index divisors. We therefore check that the known subgroup  $G$  is in fact saturated at all those primes with the method already introduced in [6]: to verify saturation at  $p$ , we find sufficiently many primes  $q$  of good reduction such that  $\#\mathcal{J}(\mathbb{F}_q)$  is divisible by  $p$  (usually nine such primes will suffice) and check that the kernel of the natural map

$$G/pG \longrightarrow \prod_q \mathcal{J}(\mathbb{F}_q)/p\mathcal{J}(\mathbb{F}_q)$$

is trivial. This computation takes a few CPU days; the most time-consuming task is to find  $\#\mathcal{J}(\mathbb{F}_q)$  for all primes  $q$  up to  $q = 322\,781$  (which is needed for  $p = 1471$ ). This gives the following result.

**Theorem 12.1** *The points  $[P_j - \infty]$  freely generate  $\mathcal{J}(\mathbb{Q})$ , where the  $P_j \in \mathcal{C}(\mathbb{Q})$  are the points with the following  $x$ -coordinates and positive  $y$ -coordinate:*

$$-3, -2, -\frac{5}{2}, 0, 1, 2, 3, 4, 6.$$

In principle, one could now try to determine the set of integral points on  $\mathcal{C}$  with the method we had already used for  $y^2 - y = x^7 - x$ . However, a Mordell-Weil sieve computation with a group of rank 9 is a rather daunting task, which we prefer to leave to the truly dedicated reader.

**Acknowledgements** I would like to thank Steffen Müller for helpful comments on a draft version of this paper and for pointers to the literature. The necessary computations were performed using the Magma computer algebra system [1].

## References

1. W. Bosma, J. Cannon, C. Playoust, The Magma algebra system. I. The user language. *J. Symb. Comput.* **24**, 235–265 (1997)
2. N. Bruin, M. Stoll, The Mordell-Weil sieve: proving non-existence of rational points on curves. *LMS J. Comput. Math.* **13**, 272–306 (2010)
3. Y. Bugeaud, M. Mignotte, S. Siksek, M. Stoll, Sz. Tengely, Integral points on hyperelliptic curves. *Algebra & Number Theory* **2**(8), 859–885 (2008)
4. J.W.S. Cassels, E.V. Flynn, *Prolegomena to a Middlebrow Arithmetic of Curves of Genus 2* (Cambridge University Press, Cambridge, 1996)
5. S. Duquesne, Calculs effectifs des points entiers et rationnels sur les courbes, Thèse de doctorat, Université Bordeaux, 2001
6. E.V. Flynn, N.P. Smart, Canonical heights on the Jacobians of curves of genus 2 and the infinite descent. *Acta Arith.* **79**(4), 333–352 (1997)
7. M. Hindry, J.H. Silverman, *Diophantine Geometry. An Introduction*. Springer GTM, vol. 201 (Springer, New York, 2000)
8. D. Holmes, Computing Néron–Tate heights of points on hyperelliptic Jacobians. *J. Number Theory* **132**(6), 1295–1305 (2012)
9. J.S. Müller, Computing canonical heights on Jacobians, PhD thesis, Universität Bayreuth, 2010
10. J.S. Müller, Computing canonical heights using arithmetic intersection theory. *Math. Comput.* **83**, 311–336 (2014)
11. J.S. Müller, Explicit Kummer varieties of hyperelliptic Jacobian threefolds. *LMS J. Comput. Math.* **17**, 496–508 (2014)
12. J.S. Müller, M. Stoll, Canonical heights on genus two Jacobians. *Algebra & Number Theory* **10**(10), 2153–2234 (2016)
13. D. Mumford, On the equations defining abelian varieties. I. *Invent. Math.* **1**, 287–354 (1966)
14. M. Stoll, On the height constant for curves of genus two. *Acta Arith.* **90**, 183–201 (1999)

15. M. Stoll, Implementing 2-descent for Jacobians of hyperelliptic curves. *Acta Arith.* **98**, 245–277 (2001)
16. M. Stoll, On the height constant for curves of genus two, II. *Acta Arith.* **104**, 165–182 (2002)
17. M. Stoll, Magma files with relevant data, <http://www.mathe2.uni-bayreuth.de/stoll/magma/index.html>
18. A.G.J. Stubbs, Hyperelliptic curves, PhD thesis, University of Liverpool, 2000

# Some Recent Developments in Spectrahedral Computation



**Thorsten Theobald**

**Abstract** Spectrahedra are the feasible sets of semidefinite programming and provide a central link between real algebraic geometry and convex optimization. In this expository paper, we review some recent developments on effective methods for handling spectrahedra. In particular, we consider the algorithmic problems of deciding emptiness of spectrahedra, boundedness of spectrahedra as well as the question of containment of a spectrahedron in another one. These problems can profitably be approached by combinations of methods from real algebra and optimization.

**Keywords** Spectrahedron • Spectrahedral computation • Real algebraic geometry • Convex algebraic geometry • Containment

**Subject Classifications** 14Q20, 52A20, 68W30, 90C22

## 1 Introduction

In the last decade tremendous developments around the connections between algebraic geometry, convexity and optimization have brought the geometric concept of a *spectrahedron* into the focus of research activities. A spectrahedron, whose terminology is due to Ramana and Goldman [33], is the feasible region of a semidefinite program. Hence, spectrahedra are a natural generalization of polyhedra (which are the feasible sets of linear programs). Spectrahedra are basic semialgebraic sets and provide a major concept in modern computational real algebraic geometry [4, 14, 30].

---

T. Theobald (✉)

Goethe-Universität, FB 12 – Institut für Mathematik, Postfach 11 19 32,  
60054 Frankfurt am Main, Germany  
e-mail: [theobald@math.uni-frankfurt.de](mailto:theobald@math.uni-frankfurt.de)

© Springer International Publishing AG, part of Springer Nature 2017  
G. Böckle et al. (eds.), *Algorithmic and Experimental Methods  
in Algebra, Geometry, and Number Theory*,  
[https://doi.org/10.1007/978-3-319-70566-8\\_30](https://doi.org/10.1007/978-3-319-70566-8_30)

717

Formally, let  $\mathcal{S}_k$  be the set of real symmetric  $k \times k$ -matrices,  $\mathcal{S}_k^+ \subseteq \mathcal{S}_k$  be the subset of positive semidefinite matrices, and  $\mathcal{S}_k[x]$  be the set of symmetric  $k \times k$ -matrices with polynomial entries in  $x = (x_1, \dots, x_n)$ . For  $A_0, \dots, A_n \in \mathcal{S}_k$ , denote by  $A(x)$  the *linear (matrix) pencil*  $A(x) = A_0 + x_1 A_1 + \dots + x_n A_n \in \mathcal{S}_k[x]$ . The set

$$S_A = \{x \in \mathbb{R}^n : A(x) \succeq 0\} \quad (1)$$

is called a *spectrahedron*, where  $A(x) \succeq 0$  denotes positive semidefiniteness of the matrix  $A(x)$ .

Recent work by a number of authors have advanced a theory of spectrahedral computation. In this expository paper, we review some of these developments, equipped with a view towards real and convex algebraic geometry. A particular focus will then be given on the question whether one given spectrahedron is contained in another one.

Precisely, given linear matrix pencils  $A(x)$  and  $B(x)$  we consider the following problems:

**Emptiness:** Is  $S_A$  empty?

**Boundedness:** Is  $S_A$  bounded?

**Containment:** Does  $S_A \subseteq S_B$  hold?

Most of the results discussed here come from the work of Helton, Kellner, Klep, McCullough, Schweighofer, Trabant as well as the author. Rather than to focus on complete coverage, our goal is to provide an insightful window into these research developments. Most proofs are omitted and can be found in the original papers.

The paper is structured as follows. In Sect. 2, we introduce polyhedra and spectrahedra and highlight some occurrences of spectrahedra in real and convex algebraic geometry. In Sect. 3, we discuss some fundamental algorithmic problems, in particular the emptiness and boundedness problem. Then, in Sect. 4, we deal with fundamental aspects of the containment problem. Section 5 is devoted to hierarchical semidefinite approaches to the containment problem.

## 2 From Polyhedra to Spectrahedra

Starting from polyhedra as a classical cornerstone of mathematics (see the monographs of Grünbaum [11] or Ziegler [38]), we then introduce some basic notions of spectrahedra.

### 2.1 Polyhedra and Polytopes

For a matrix  $A \in \mathbb{R}^{m \times n}$  and a vector  $b \in \mathbb{R}^m$ , the set  $P = \{x \in \mathbb{R}^n : b + Ax \geq 0\}$  is called a *polyhedron*. Geometrically,  $P$  is the intersection of a finite number of halfspaces ( $\mathcal{H}$ -presentation of a polyhedron, or, for short,  $\mathcal{H}$ -polyhedron). If the

polyhedron  $P$  is a bounded set, then  $P$  is called a *polytope*. Polytopes can also be represented as the convex hull of finitely many points,  $P = \text{conv}\{p^{(1)}, \dots, p^{(l)}\}$  with  $p^{(1)}, \dots, p^{(l)} \in \mathbb{R}^n$  ( $\mathcal{V}$ -presentation of a polytope or, for short,  $\mathcal{V}$ -polytope).

As an occurrence of polyhedra in real algebraic geometry, let us state Handelman’s Theorem [12], which provides a characterization of the positive polynomials on a given polytope. And under a degree restriction it gives a polyhedron of solutions, since all conditions are linear.

**Theorem 2.1 (Handelman)** *Let  $g_1, \dots, g_m \in \mathbb{R}[x]$  be affine-linear polynomials such that  $K = \{x \in \mathbb{R}^n : g_1(x) \geq 0, \dots, g_m(x) \geq 0\}$  is non-empty and bounded, that is, a polytope. Any polynomial  $p \in \mathbb{R}[x]$  which is strictly positive on  $K$  can be written as a finite sum*

$$p = \sum_{\beta} c_{\beta} \prod_{j=1}^m g_j^{\beta_j} \tag{2}$$

with coefficients  $c_{\beta} \geq 0$  ( $\beta \in \mathbb{N}_0^m$ ). For a fixed upper bound  $t$  on the degree, where  $t \geq \deg p$ , the set of solutions  $(c_{\beta})_{|\beta| \leq t}$  of

$$p = \sum_{|\beta| \leq t} c_{\beta} \prod_{j=1}^m g_j^{\beta_j}$$

is a polyhedron.

The latter condition can be transformed into an optimization version to find lower bounds for  $p$  on  $K$ .

Though polytopes and polyhedra are defined by linear inequalities, they have a rich geometric and combinatorial structure. Denote by  $V(P)$  the set of vertices (i.e., 0-dimensional faces) of a polytope  $P$ , and by  $F(P)$  the set of facets (i.e., faces of codimension 1). By McMullen’s Upper bound Theorem [28], any  $n$ -dimensional polytope with  $k$  vertices has at most

$$\binom{k - \lceil \frac{n}{2} \rceil}{\lfloor \frac{n}{2} \rfloor} + \binom{k - 1 - \lceil \frac{n-1}{2} \rceil}{\lfloor \frac{n-1}{2} \rfloor} \tag{3}$$

facets. This bound, which is of inherent importance for polyhedral computation software such as `polymake` [9], is sharp for neighborly polytopes, that is, for polytopes with the property that every set of at most  $\lfloor n/2 \rfloor$  vertices is the vertex set of a face of  $P$ . For example, cyclic polytopes are neighborly. And, dual to the statement, the maximum number of vertices of any  $n$ -dimensional polytope with  $k$  facets is given by (3) as well, with equality for dually neighborly polytopes.

## 2.2 Spectrahedra

We build upon the terminology from the Introduction. Specifically, for  $A_0, \dots, A_n \in \mathcal{S}_k$ , let  $S_A = \{x \in \mathbb{R}^n : A(x) = A_0 + \sum_{i=1}^n x_i A_i \succeq 0\}$  denote the *spectrahedron* as defined in (1). The inequality  $A_0 + \sum_{i=1}^n x_i A_i \succeq 0$  is called a *linear matrix inequality (LMI)*. Since the operator  $A(\cdot)$  is linear, any spectrahedron is a convex set.

*Example 2.2* Figure 1 shows the example of the ellipsope

$$S_A = \left\{ x \in \mathbb{R}^3 : \begin{pmatrix} 1 & x_1 & x_2 \\ x_1 & 1 & x_3 \\ x_2 & x_3 & 1 \end{pmatrix} \succeq 0 \right\}$$

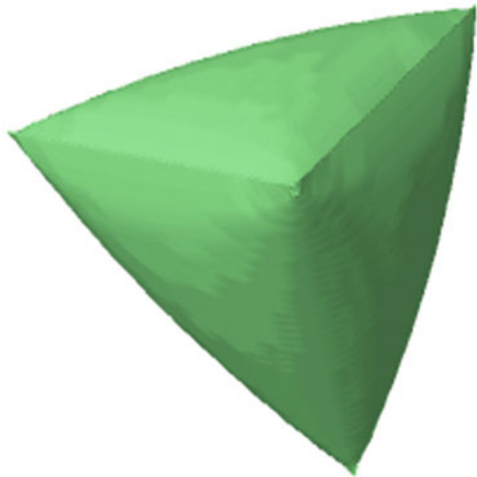
(see, e.g., [26]).

Note that every polyhedron  $P = \{x \in \mathbb{R}^n : b + Ax \geq 0\}$  can be regarded as a spectrahedron,

$$P = P_A = \left\{ x \in \mathbb{R}^n : A(x) = \begin{pmatrix} a_1(x) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & a_k(x) \end{pmatrix} \succeq 0 \right\}, \tag{4}$$

where  $a_i(x)$  denotes the  $i$ -th entry of the vector  $b + Ax$ .  $P_A$  contains the origin in its interior if and only if the inequalities can be scaled so that  $b$  is the all-ones vector  $\mathbb{1}_k$  in  $\mathbb{R}^k$ . In this case,  $A(x)$  is called the *normal form* of the polyhedron  $P_A$ .

**Fig. 1** Visualization of an ellipsope



*Example 2.3* The unit disc  $\{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\}$  is a spectrahedron. This follows from setting

$$A_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and observing that

$$A(x) = \begin{pmatrix} 1 + x_1 & x_2 \\ x_2 & 1 - x_1 \end{pmatrix}$$

is positive semidefinite if and only if  $1 - x_1^2 - x_2^2 \geq 0$ .

Every spectrahedron  $S$  is a basic closed semialgebraic set. This can be seen by writing  $S = \{x \in \mathbb{R}^n : p_i(x) \geq 0, i \in I\}$  where the  $p_i(x)$  are the principal minors of  $A(x)$ , indexed by the set  $I \in 2^{\{1, \dots, k\}} \setminus \{0\}$ . A slightly more concise representation is given by the following well-known statement, where  $I_k$  denotes the  $k \times k$  identity matrix.

**Proposition 2.4** *Any spectrahedron  $S = S_A$  is a basic closed semialgebraic set. In particular, given the modified characteristic polynomial*

$$t \mapsto \det(A(x) + tI_k) =: t^k + \sum_{i=0}^{k-1} p_i(x)t^i, \tag{5}$$

$S$  has the representation  $S = \{x \in \mathbb{R}^n : p_i(x) \geq 0, 0 \leq i \leq k - 1\}$ .

*Proof* Denoting by  $\lambda_1(x), \dots, \lambda_k(x)$  the eigenvalues of the linear pencil  $A(x)$ , we observe

$$\det(A(x) + tI_k) = (t + \lambda_1(x)) \cdots (t + \lambda_k(x)).$$

Since  $A(x)$  is symmetric, all  $\lambda_i(x)$  are real, for any  $x \in \mathbb{R}^n$ . Comparing the coefficients then shows

$$p_{k-i}(x) = \sum_{t_1 < \dots < t_i} \lambda_{t_1}(x) \cdots \lambda_{t_i}(x), \quad 1 \leq i \leq k.$$

Now “ $\subseteq$ ” of the desired representation follows from the fact that positive semidefiniteness of  $A(x)$  at a given  $x \in \mathbb{R}^n$  implies non-negativity of all eigenvalues  $\lambda_1(x), \dots, \lambda_k(x)$  and thus non-negativity of all  $p_i(x)$ . Conversely, if for a given



$x \in \mathbb{R}^n$  we have  $p_i(x) \geq 0$  for all  $i$ , then the modified characteristic polynomial has no sign changes. Thus, by Descartes’ rule of signs, it has no positive roots, and therefore  $A(x)$  is positive semidefinite.  $\square$

It is an open question to provide good effective criteria to test whether a given convex semialgebraic set is a spectrahedron or the linear projection of a spectrahedron. Recently, the conjecture that every convex semialgebraic set would be the linear projection of a spectrahedron (“Helton-Nie conjecture”) has been disproven by Scheiderer [34].

### 2.3 Spectrahedra in Real and Convex Algebraic Geometry

Spectrahedra occur in many places of real and convex algebraic geometry. We point out three connections to the algorithmic problems mentioned in Sect. 3.

**Non-negative Polynomials and Sums of Squares** A polynomial  $p = \sum_{\alpha} c_{\alpha} x^{\alpha} \in \mathbb{R}[x] = \mathbb{R}[x_1, \dots, x_n]$  is called a *sum of squares (sos)* if it can be written as a finite sum  $\sum_i u_i(x)^2$  with polynomials  $u_i \in \mathbb{R}[x]$ . The total degree  $\deg p$  of an sos-polynomial  $p$  is even. Sum of squares polynomials are ubiquitous in real and convex algebraic geometry and provide a fundamental sufficient condition for the property that a polynomial  $p$  is non-negative. In order to phrase the sos-property in terms of a spectrahedral property, let  $y$  denote the  $\binom{n+\deg p/2}{n}$ -dimensional vector of all monomials in  $x$  up to half of the total degree of  $p$ . And for some  $m \geq 0$  and  $k = \binom{n+\deg p/2}{n}$ , let  $A(w) = A_0 + \sum_{i=1}^m w_i A_i$  be a matrix pencil spanning the subspace in  $\mathcal{S}_k$  defined by the equations

$$c_{\alpha} = \sum_{\beta+\gamma=\alpha} z_{\beta,\gamma} \quad \text{for all } \alpha \text{ of total degree at most } \deg p \tag{6}$$

in the symmetric matrix of variables  $Z = (z_{\beta,\gamma})_{|\beta|,|\gamma| \leq \deg p/2}$ .

**Proposition 2.5** *A polynomial  $p \in \mathbb{R}[x]$  can be written as a sum of squares if and only if the spectrahedron  $S_A$  is non-empty.*

*Proof* The comparison of the coefficients in (6) is satisfied if and only if there exists a matrix  $Z$  with  $y^T Z y = c$ , where  $c$  is the coefficient vector of  $p$ . Since  $Z$  has a Choleski decomposition  $LL^T$  if and only if it is positive semidefinite, the claim follows.  $\square$

**Computation of Amoebas** For an ideal  $I = \langle f_1, \dots, f_r \rangle \subseteq \mathbb{C}[z] = \mathbb{C}[z_1, \dots, z_n]$ , the *algebraic amoeba* (or *unlog amoeba*)  $\mathcal{A}_I$  is the image of its zero set  $\mathcal{V}(I)$  under the absolute value map, that is,  $\mathcal{A}_I = \{|z| : z \in \mathcal{V}(I)\}$ . Given  $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}_{>0}^n$ , the amoeba membership problem asks whether  $\lambda \in \mathcal{A}_I$ .

For  $f \in \mathbb{C}[z]$ , let  $\Re(f)$  and  $\Im(f) \in \mathbb{R}[x, y]$  be given through

$$f(x + iy) = \Re(f)(x, y) + i\Im(f)(x, y).$$

Now consider the ideal  $J$  generated by the set of polynomials

$$\Re(f_j), \Im(f_j) \quad 1 \leq j \leq r, \quad x_k^2 + y_k^2 - \lambda_k^2, \quad 1 \leq k \leq n.$$

By the real Nullstellensatz, we have  $\lambda \in \mathcal{A}_I$  unless there exists a polynomial  $G \in J$  and an sos-polynomial  $H$  such that  $G + H + 1 = 0$ . Given a fixed degree bound, the set of all the certificates satisfying that bound defines a spectrahedron, and thus the amoeba membership problem can be approached through a hierarchy of spectrahedral feasibility problems, see [37].

**Non-negative Biquadratic Forms** Given a biquadratic form

$$F(x, y) = \sum_{(i,j,s,t) \in \Lambda} b_{ijkl} x_i y_j x_s y_t$$

with  $\Lambda = \{(i, j, s, t) : 1 \leq i, s \leq k, 1 \leq j, t \leq l\}$  and real coefficients  $b_{ijkl}$ , we ask whether  $F$  is non-negative. We can assume that the coefficients satisfy the symmetry condition  $b_{ijkl} = b_{kjil}$  and  $b_{ijkl} = b_{ilkj}$ .

In order to phrase this question as a containment problem of spectrahedra, set  $n = \binom{k+l}{2}$ . For notational convenience, we can then identify  $x = (x_1, \dots, x_n)$  with a matrix  $X \in \mathcal{S}_k$ . Let  $A(X) = X$  and  $B(X) \in \mathcal{S}_l[X]$  be given by  $b_{j,t}(X) = \sum_{1 \leq i,s \leq k} b_{ijst} x_{is}$ ,  $1 \leq j, t \leq l$ .

**Proposition 2.6** *The biquadratic form  $F$  is non-negative if and only if the spectrahedron  $S_A$  is contained in the spectrahedron  $S_B$ .*

*Proof* If  $S_A \subseteq S_B$  then any positive semidefinite matrix  $X$  satisfies  $B(X) \succeq 0$ , and thus for every  $(x, y) \in \mathbb{R}^k \times \mathbb{R}^l$  we have  $F(x, y) = y^T B(x x^T) y \geq 0$ . Hence,  $F$  is positive semidefinite.

Conversely, let  $F(x, y)$  be a positive semidefinite biquadratic form. Since any positive semidefinite matrix  $X$  can be written as a finite sum  $X = \sum_i x^{(i)} (x^{(i)})^T$  with vectors  $x^{(i)} \in \mathbb{R}^k$ , linearity implies  $y^T B(X) y = \sum_i y^T B(x^{(i)} (x^{(i)})^T) y = \sum_i F(x^{(i)}, y) \geq 0$  for any  $y \in \mathbb{R}^l$ . Hence,  $B(X) \succeq 0$ .  $\square$

### 3 Fundamental Algorithmic Concepts

In the early years, spectrahedra were mainly considered within optimization frameworks. The stronger focus on the geometry of these sets has established new connections to real algebraic geometry and effective computation.

### 3.1 Infeasibility Certificates

Given a linear matrix pencil  $A(x) \in \mathcal{S}_k[x]$ , we study the question whether  $S_A = \emptyset$ .

*Remark 3.1* For polytopes  $P_A = \{x \in \mathbb{R}^n : b + Ax \geq 0\}$ , the question whether  $P_A$  is non-empty can be phrased as a linear program and thus can be decided in polynomial time for a rational input polytope. Also note that even deciding whether a polytope has an interior point can be decided by a linear program as well (see, e.g., [18, Example 4.3]).

Testing whether  $S_A = \emptyset$  can be regarded as the complement of a *semidefinite feasibility problem* (SDFP), which asks whether for a given linear pencil  $A(x)$  the spectrahedron  $S_A$  is nonempty. While semidefinite programs (with rational input data) can be approximated in polynomial time (see [6]), the complexity of SDFP is open, see [32]. In practice, however, SDFPs can numerically be solved efficiently by semidefinite programming.

In view of the classical Nullstellensätze and Positivstellensätze from real algebraic geometry, it is a natural question how to certify the emptiness of a spectrahedron. For polytopes, the classical Farkas' Lemma (see, e.g., [35, Cor. 7.1e]) characterizes the emptiness of a polytope in terms of an identity of affine functions coming from a geometric cone condition.

**Theorem 3.2** *A polyhedron  $P = \{x \in \mathbb{R}^n : Ax + b \geq 0\}$  is empty if and only if the constant polynomial  $-1$  can be written as  $-1 = \sum_i s_i(Ax + b)_i$  with  $s_i \geq 0$ ; or, equivalently, if  $-1$  can be written as  $-1 = c + \sum_i s_i(Ax + b)_i$  with  $c \geq 0$ ,  $s_i \geq 0$ .*

Let  $A(x) \in \mathcal{S}_k[x]$ .  $A(x)$  is called *feasible* if the spectrahedron  $S_A$  is non-empty. Further,  $A(x)$  is called *strongly feasible* if  $A(x)$  is feasible and there exists an  $x \in \mathbb{R}^n$  with  $A(x) \succ 0$ . In relation to this, the spectrahedron  $S_A$  is called *strongly empty* if  $A(x)$  it is not strongly feasible.

In order to extend Farkas' Lemma to spectrahedra, denote by  $C_A$  the convex cone in  $\mathcal{S}_k[x]$  defined by

$$\begin{aligned} C_A &= \{c + \langle A, S \rangle : c \geq 0, S \in \mathcal{S}_k^+\} \\ &= \{c + \sum_i u_i^T A u_i : c \geq 0, u_i \in \mathbb{R}^n\}, \end{aligned}$$

where  $\langle A, S \rangle = \text{Tr}(AS)$  is the dot product underlying the Frobenius norm and  $\text{Tr}$  denotes the trace of a matrix. Since  $A = A(x)$  is a linear pencil in  $\mathcal{S}_k[x]$ , every element in  $C_A$  is a linear polynomial which is non-negative on the spectrahedron  $S_A$ .

**Theorem 3.3 (Sturm [36])** *Given  $A(x) \in \mathcal{S}_k[x]$ , the spectrahedron  $S_A$  is strongly empty if and only if  $-1 \in C_A$ .*

An exact characterization for the emptiness of  $S_A$  can be established in terms of a quadratic module associated to  $A(x)$ . Recall that a subset  $M$  of a commutative ring  $R$  with 1 is called a *quadratic module* if it satisfies the conditions

$$1 \in M, M + M \subseteq M \text{ and } a^2M \subseteq M \text{ for any } a \in R.$$

Given a linear pencil matrix  $A = A(x)$ , denote by  $M_A$  the quadratic module in  $\mathbb{R}[x]$

$$M_A = \{s + \langle A, S \rangle : s \in \Sigma[x], S \in \mathbb{R}[x]^{k \times k} \text{ an sos-matrix}\} \tag{7}$$

$$= \{s + \sum_i u_i^T A u_i : s \in \Sigma[x], u_i \in \mathbb{R}[x]^k\}, \tag{8}$$

where  $\Sigma[x]$  denotes the subset of sums of squares of polynomials within  $\mathbb{R}[x]$  and an sos-matrix is a matrix polynomial of the form  $P^T P$  for some matrix polynomial  $P$ . Note that if a polynomial  $f \in \mathbb{R}[x]$  is contained in  $M_A$  then it is non-negative on  $S_A$ . Further, denote by  $M_A^{(t)}$  the truncated quadratic module

$$\begin{aligned} M_A^{(t)} &= \{s + \langle A, S \rangle : s \in \Sigma[x] \cap \mathbb{R}[x]_{2t}, S \in \mathbb{R}[x]_{2t}^{k \times k} \text{ sos-matrix}\} \\ &= \{s + \sum_i u_i^T A u_i : s \in \Sigma[x]_{2t}, u_i \in \mathbb{R}[x]_t^k\} \subseteq \mathbb{R}[x]_{2t+1}, \end{aligned}$$

where  $\mathbb{R}[x]_t$  denotes the set of polynomials of total degree at most  $t$ .

**Theorem 3.4 (Klep, Schweighofer [23])** *For  $A(x) \in \mathcal{S}_k[x]$ , the following are equivalent:*

1. *The spectrahedron  $S_A$  is empty.*
2.  $-1 \in M_A$ .
3.  $-1 \in M_A^{(2^{\min\{n, k-1\}})}$ .

The third of these statements provides the ground for a computational treatment in terms of algebraic certificates for infeasibility. Namely, the question whether such a representation of bounded degree exists can be formulated as a semidefinite feasibility problem.

In order to carry out this formulation as a semidefinite program, set  $t = 2^{\min\{n, k-1\}}$ . Then the value

$$\max \{ \gamma \in \mathbb{R} : -1 - \gamma = s + \langle A, S \rangle, s \in \Sigma[x] \cap \mathbb{R}[x]_{2t}, S \in \mathbb{R}[x]_{2t}^{k \times k} \text{ sos-matrix} \}$$

coincides with the value of the semidefinite program

$$\begin{aligned} &\max \gamma \\ &s.t. \quad -1 - \gamma = \text{Tr}(P_1 X) + \text{Tr}(Q_1 Y) \\ &\quad \quad 0 = \text{Tr}(P_i X) + \text{Tr}(Q_i Y) \quad \text{for } 2 \leq i \leq m_w := \binom{n+2t+1}{2t+1}, \\ &\quad \quad X \geq 0, Y \geq 0. \end{aligned} \tag{9}$$

Here, denoting by  $w = w(x)$  and  $y = y(x)$  the vectors of monomials in  $x_1, \dots, x_n$  of degrees up to  $2t + 1$  and  $t$  in lexicographic order,  $Q_i$  is defined through  $y(x)y(x)^T = \sum_{i=1}^{m_w} Q_i w_i(x)$ . And, setting  $m_y = \binom{n+t}{t}$ , the permutation matrix  $P \in \mathbb{R}^{k m_y \times k m_y}$  is given via  $P(I_k \otimes y(x)) = y(x) \otimes I_k$ , and the matrices  $P_i$  are defined through

$$P(I_k \otimes y(x)) \cdot A(x) \cdot (P(I_k \otimes y(x)))^T = \sum_{i=1}^{m_w} P_i w_i(x) \in \mathbb{R}[x]^{k m_y \times k m_y}.$$

Hence,  $-1 \in M_A^{(2 \min\{n, k-1\})}$  if and only if the objective value of (9) is non-negative. This decision problem is a semidefinite feasibility problem, since the property of a non-negative linear objective function can also be viewed as an additional linear constraint.

*Example 3.5* Let

$$A(x) = \begin{pmatrix} 1+x & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & x \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix} + x \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Since  $\min\{n, k-1\} = \{1, 3-1\} = 1$ , we can assume  $y = y(x) = (1, x)^T$ . We obtain

$$Q_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad Q_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Q_3 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad Q_4 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

and the matrices  $P_1, \dots, P_4$  are

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Since the positive semidefinite matrices

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & \frac{3}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{2} & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

provide a feasible solution of the semidefinite program (9) with objective value 0, we see that the spectrahedron  $S_A$  is empty. By a Choleski factorization

$$X = LL^T \quad \text{with } L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \\ 0 & \sqrt{2}/2 & \sqrt{6}/2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{6}/2 & \sqrt{2}/2 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

we can deduce from the semidefinite program (9) that  $u_1 = (1, 0, 0)^T$ ,  $u_2 = (0, \sqrt{2}, \sqrt{2}/2)^T$ ,  $u_3 = (0, \sqrt{6}/2x, \sqrt{6}/2)^T$ ,  $u_4 = (0, \sqrt{2}/2x, 0)^T$  provides the desired algebraic certificate  $-1 \in M_A$ , where the  $u_i$  are as in (8). We remark that  $u_4$  can be omitted due to  $u_4^T A(x) u_4 = 0$ .

**Origin in the Interior** We shortly point out a fine point which explains a technical assumption in later statements. Clearly, if the constant matrix  $A_0$  of a pencil  $A(x)$  is positive semidefinite then the origin is contained in the spectrahedron  $S_A$ . However, in general it is *not* true that  $A_0$  is positive definite if and only if the origin is contained in the interior of  $S_A$ . Fortunately, by [33, Corollary 5], if a spectrahedron  $S_A$  is full-dimensional, then there exists a so-called reduced linear pencil that is positive definite exactly on the interior of  $S_A$ . Hence, in the case of a reduced pencil we have  $0 \in \text{int } S_A$  if and only if  $A_0 \succ 0$ . Moreover, for arbitrary dimension of  $S_A$ , we have  $0 \in \text{int } S_A$  if and only if there is a linear pencil  $A'(x)$  with the same positivity domain such that  $A'_0 = I_k$  (see [14]). Such a pencil is called *monic*.

### 3.2 Boundedness

In order to certify that a given spectrahedron is bounded, the quadratic module (7) is applied as well. Recall that a quadratic module  $M \subseteq \mathbb{R}[x]$  is called *archimedean* if it contains a polynomial of the form  $N - \sum_{i=1}^n x_i^2$  for some  $N > 0$ .

**Theorem 3.6 (Klep, Schweighofer [23])** *Given  $A(x) \in \mathcal{S}_k[x]$ , the spectrahedron  $S_A$  is bounded if and only if  $M_A$  is archimedean.*

*Example 3.7* In order to show that the spectrahedron  $S_A$  of

$$A(x) = \begin{pmatrix} x & 1 & 0 \\ 1 & x & 0 \\ 0 & 0 & -x + 2 \end{pmatrix}$$

is bounded, we ask for  $u \in \mathbb{R}[x]^3$  and sos-polynomials  $s_0, s_1$  with

$$N - \sum_{i=1}^3 x_i^2 = u^T A u + s_1^2(-x + 2) + s_0$$

for some  $N > 0$ . The choice  $u = (x - \frac{1}{2}, -x + 1, 0)^T$ ,  $s_1 = 2x^2 + \frac{17}{4}$ ,  $s_0 = 0$  and  $N = \frac{17}{2}$  gives an algebraic certificate for the boundedness of  $S_A$ .

There exist spectrahedra whose elements have coordinates of double-exponential size in the number of variables and whose distance to the origin grows double-exponentially in the number of variables (see [1, 33]). Hence, in general one cannot expect to have a certificate of polynomial size for the boundedness of the spectrahedron.

## 4 Containment Problems

As a next step in the class of algorithmic problems on spectrahedra, we consider containment problems: Given two linear pencils  $A(x) \in \mathcal{S}_k[x]$  and  $B(x) \in \mathcal{S}_l[x]$ , is  $S_A \subseteq S_B$ ?

Containment problems of convex sets are a classical topic in convex geometry (see, e.g., Gritzmann and Klee for the containment of polytopes and a number of computational aspects [10]). In the context of spectrahedra, the study of algorithmic approaches and relaxations has been initiated by Ben-Tal and Nemirovski [2] who investigated the case where  $S_A$  is a cube and  $S_B$  is an arbitrary spectrahedron (“matrix cube problem”). Figure 2 visualizes an ellipsope in a ball.

### 4.1 Complexity of Containment Problems for Spectrahedra

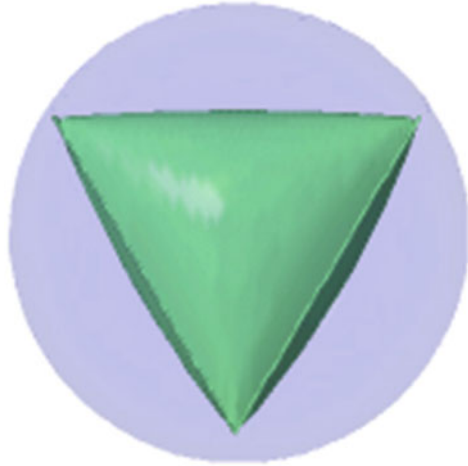
It is useful to start from the case of polytopes. Here, it is well-known that the computational complexity of deciding containment of a given polytope  $P$  in a given polytope  $Q$  strongly depends on the type of input representations. We assume that all input data is given in terms of rational numbers, and the dimension is part of the input.

**Proposition 4.1** [7] *The following problems can be decided in polynomial time.*

1. Given  $\mathcal{H}$ -polytopes  $P$  and  $Q$ , is  $P \subseteq Q$ ?
2. Given  $\mathcal{V}$ -polytopes  $P$  and  $Q$ , is  $P \subseteq Q$ ?
3. Given a  $\mathcal{V}$ -polytope  $P$  and an  $\mathcal{H}$ -polytope  $Q$ , is  $P \subseteq Q$ ?

*In contrast to this, deciding whether an  $\mathcal{H}$ -polytope is contained in a  $\mathcal{V}$ -polytope is co-NP-complete.*

**Fig. 2** Visualization of an elliptope in a ball



In [21], this classification has been extended to containment problems involving polytopes and spectrahedra, where the spectrahedra are given by a linear pencil with rational entries. The main hardness results are given by the subsequent Theorems 4.2 and Proposition 4.3.

**Theorem 4.2** [21] *The following problems are co-NP-hard:*

1. Given a spectrahedron  $S_A$  and a  $\mathcal{V}$ -polytope  $Q$ , is  $S_A \subseteq Q$ ?
2. Given an  $\mathcal{H}$ -polytope  $P$  and a spectrahedron  $S_B$ , is  $P \subseteq S_B$ ?

*The latter hardness statement persists if the  $\mathcal{H}$ -polytope is a standard cube or if the outer spectrahedron is a ball.*

Since deciding whether a given rational matrix is positive semidefinite can be done in polynomial time, it can be decided in polynomial time whether a  $\mathcal{V}$ -polytope is contained in a spectrahedron. As mentioned earlier, the question “Can semidefinite feasibility problems SDFP be solved in polynomial time?” is an open complexity question. Consequently, the following statement on containment of a spectrahedron in an  $\mathcal{H}$ -polytope does not give a complete answer concerning polynomial solvability of these containment questions in the Turing machine model. If the additional inequalities were non-strict, then we had to decide a finite set of problems from the complement of the class SDFP.



**Table 1** Computational complexity of containment problems, where the rows refer to the inner set and the columns to the outer set

	$\mathcal{H}$	$\mathcal{V}$	$\mathcal{S}$
$\mathcal{H}$	P	co-NP-complete	co-NP-hard
$\mathcal{V}$	P	P	P
$\mathcal{S}$	“SDFP”	co-NP-hard	co-NP-hard

“SDFP” refers to the formulations through semidefinite feasibility problems as described in Proposition 4.3

**Proposition 4.3** [21] *The problem of deciding whether a spectrahedron is contained in an  $\mathcal{H}$ -polytope can be formulated by the complement of semidefinite feasibility problems (involving also strict inequalities), whose sizes are polynomial in the input data.*

Since Theorem 4.2 also implies that deciding containment of a spectrahedron in a spectrahedron is co-NP-hard, all the relevant cases are covered. See Table 1 for a condensed presentation, where  $\mathcal{H}$ ,  $\mathcal{V}$  and  $\mathcal{S}$  stand for  $\mathcal{H}$ -polytope,  $\mathcal{V}$ -polytope and spectrahedron, respectively.

For the computational question of deciding whether a spectrahedron is a polyhedron see Bhardwaj et al. [3], and for sos-based approaches to the NP-hard containment problem of deciding whether an  $\mathcal{H}$ -polytope is contained in a  $\mathcal{V}$ -polytope see Kellner and Theobald [20].

## 4.2 From Farkas-Type Characterizations for Polytopes to Relaxations for Spectrahedra

In this section, we present some recent results on semidefinite relaxations which provide a sufficient criterion for the containment problem of spectrahedra. Here, relaxation means that some conditions are omitted from the original problem in order to obtain a more tractable, semidefinite formulation.

It is helpful to start from the containment problem for pairs of  $\mathcal{H}$ -polytopes, which by Proposition 4.1 can be decided in polynomial time. Indeed, as a consequence of the affine form of Farkas’ Lemma, this can be achieved by solving a linear program, as stated by the following necessary and sufficient criterion (see, e.g., [21]). Recall that a real matrix with non-negative entries is called right stochastic if each row sums to one.

**Proposition 4.4** *Let  $P_A = \{x \in \mathbb{R}^n : \mathbb{1}_k + Ax \geq 0\}$  and  $P_B = \{x \in \mathbb{R}^n : \mathbb{1}_l + Bx \geq 0\}$  be polytopes. Then  $P_A \subseteq P_B$  if and only if there exists a right stochastic matrix  $C$  with  $B = CA$ .*

For the treatment of containment of spectrahedra, a good starting point is the sufficient criterion given by Helton et al. [15]. As earlier, let  $A(x) \in \mathcal{S}_k[x]$  and  $B(x) \in \mathcal{S}_l[x]$  be linear pencils. In the subsequent statement, the indeterminate matrix  $C = (C_{ij})_{i,j=1}^k$  is a symmetric  $kl \times kl$ -matrix where the  $C_{ij}$  are  $l \times l$ -blocks.

**Theorem 4.5** ([15, Theorem 4.3], See Also [21, Theorems 4.3 and 4.4]) *Let  $A(x) \in \mathcal{S}_k[x]$  and  $B(x) \in \mathcal{S}_l[x]$  be linear pencils. If one of the systems*

$$C = (C_{ij})_{i,j=1}^k \succeq 0, \quad \forall p = 0, \dots, n : B_p = \sum_{i,j=1}^k a_{ij}^p C_{ij} \tag{10}$$

or

$$C = (C_{ij})_{i,j=1}^k \succeq 0, \quad B_0 - \sum_{i,j=1}^k a_{ij}^0 C_{ij} \succeq 0, \quad \forall p = 1, \dots, n : B_p = \sum_{i,j=1}^k a_{ij}^p C_{ij} \tag{11}$$

is feasible, then  $S_A \subseteq S_B$ . Here,  $a_{ij}^p$  denotes the  $(i, j)$ -entry of  $A_p$ .

Note that whenever (10) is satisfied, condition (11) is satisfied as well. However, (11) contains an additional sos-condition. An elementary proof of Theorem 4.5 was given in [21]—here, we provide a slight variant of that proof.

*Proof of Theorem 4.5.* For  $x \in S_A$ , the last two conditions in (11) imply

$$B(x) = B_0 + \sum_{p=1}^n x_p B_p \succeq \sum_{i,j=1}^k a_{ij}^0 C_{ij} + \sum_{p=1}^n \sum_{i,j=1}^k x_p a_{ij}^p C_{ij} = \sum_{i,j=1}^k (A(x))_{ij} C_{ij}. \tag{12}$$

For any block matrices  $S = (S_{ij})_{ij}$  and  $T = (T_{ij})_{ij}$ , consisting of  $k \times k$  blocks of size  $p \times p$  and  $q \times q$ , the *Khatri-Rao product* of  $S$  and  $T$  is defined as the block-wise Kronecker product of  $S$  and  $T$ , i.e.,

$$S * T = (S_{ij} \otimes T_{ij})_{ij} \in \mathcal{S}_{kpq}.$$

If both  $S$  and  $T$  are positive semidefinite, then the Khatri-Rao product  $S * T$  is positive semidefinite as well, see [27, Theorem 5].

In our situation, we have  $p = 1$  and  $q = l$ , and the Khatri-Rao product

$$A(x) * C = ((A(x))_{ij} \otimes C_{ij})_{i,j=1}^k = ((A(x))_{ij} C_{ij})_{i,j=1}^k$$

is positive semidefinite. And since  $B(x)$  is given in (12) as a sum of submatrices of  $A(x) * C$ , we obtain that  $B(x)$  is positive semidefinite, i.e.,  $x \in S_B$ .

When starting from system (10), the inequality chain in (12) becomes an equality, and the remaining part of the proof remains valid.  $\square$

For both systems (10) and (11) the feasibility depends on the linear pencil representation of the sets involved. If  $S_B$  is contained in the positive orthant, a stronger version can be given.

**Corollary 4.6** *Let  $A(x) \in \mathcal{S}_k[x]$  and  $B(x) \in \mathcal{S}_l[x]$  be linear pencils and let  $S_A$  be contained in the non-negative orthant. If the system*

$$C = (C_{ij})_{i,j=1}^k \geq 0, \quad B_0 - \sum_{i,j=1}^k a_{ij}^0 C_{ij} \geq 0, \quad \forall p = 1, \dots, n : B_p - \sum_{i,j=1}^k a_{ij}^p C_{ij} \geq 0 \tag{13}$$

is feasible, then  $S_A \subseteq S_B$ .

*Proof* Since  $S_A$  is contained in the non-negative orthant, any  $x \in S_A$  has non-negative coordinates, and hence,

$$B(x) = B_0 + \sum_{p=1}^n x_p B_p \succeq \sum_{i,j=1}^k a_{ij}^0 C_{ij} + \sum_{p=1}^n \sum_{i,j=1}^k x_p a_{ij}^p C_{ij} = \sum_{i,j=1}^k (A(x))_{ij} C_{ij} .$$

□

The version (13) is strictly stronger than system (10). There are cases, where a solution to the condition (13) exists, even though the original system (10) is infeasible.

### 4.3 Exact Cases of the Relaxation

It turns out that the sufficient semidefinite criteria (10) and (11) even provide exact containment characterizations in several important cases.

Recall the normal form for polyhedral spectrahedra introduced in Sect. 2, and let us also introduce a normal form for the class of centered and aligned ellipsoids. Here, an ellipsoid is called *centered* if it is centrally symmetric, and it is called *aligned* if its axes are aligned to the directions of the coordinate axes. A centered and aligned ellipsoid with semi-axes of lengths  $a_1, \dots, a_n$  can be written as the spectrahedron  $S_A$  of the monic linear pencil

$$A(x) = I_{n+1} + \sum_{p=1}^n \frac{x_p}{a_p} (E_{p,n+1} + E_{n+1,p}), \tag{14}$$

where  $E_{ij}$  denotes the matrix with a one in position  $(i, j)$  and zeros elsewhere. This representation is called the *normal form of the ellipsoid*. If  $a_1 = \dots = a_n$ , this gives the *normal form of a ball*. The exact characterizations also use the following

extended form  $S_{\widehat{A}}$  of a spectrahedron  $S_A$ . Given a linear pencil  $A(x) \in \mathcal{S}_k[x]$ , we call the linear pencil with an additional 1 on the diagonal

$$\widehat{A}(x) = \begin{pmatrix} 1 & 0 \\ 0 & A(x) \end{pmatrix} \in \mathcal{S}_{k+1}[x] \tag{15}$$

the *extended linear pencil* of  $S_A = S_{\widehat{A}}$ . Note that the spectrahedra  $S_A = S_{\widehat{A}}$  coincide. The entries of  $\widehat{A}_p$  in the pencil  $\widehat{A}(x) = \widehat{A}_0 + \sum_{p=1}^n x_p \widehat{A}_p$  are denoted by  $\widehat{a}_{ij}^p$  for  $i, j = 0, \dots, k$ .

**Theorem 4.7** [21] *Let  $A(x) \in \mathcal{S}_k[x]$  and  $B(x) \in \mathcal{S}_l[x]$  be monic linear pencils. In the following cases the criteria (10) as well as (11) are necessary and sufficient for the inclusion  $S_A \subseteq S_B$ :*

1. if  $A(x)$  and  $B(x)$  are normal forms of centered and aligned ellipsoids,
2. if  $A(x)$  and  $B(x)$  are normal forms of a ball and an  $\mathcal{H}$ -polyhedron, respectively,
3. if  $B(x)$  is the normal form of a polytope,
4. if  $\widehat{A}(x)$  is the extended form of a spectrahedron and  $B(x)$  is the normal form of a polyhedron.

Recently, Fritz, Netzer and Thom have shown the following exactness result which distinguishes the simplex situation within the situation that  $S_A$  is a polytope.

**Theorem 4.8** [8, Cor. 5.3] *For a fixed polytope  $S_A$ , the criterion (10) is exact for any spectrahedron  $S_B$  if and only if  $S_A$  is a simplex, and this statement is independent of the representing pencil of the polytope  $S_A$ .*

Note that all the exactness statements presented in this section refer to exact characterizations of the containment problem in terms of a formulation as semidefinite program. Similar to the case of the infeasibility certificates in Sect. 3, when it comes to actually solving the semidefinite programs, in case of employing numerical solvers this involves additional numerical aspects.

## 5 Sufficient Semidefinite Hierarchies for Containment of Spectrahedra

In this section, we present two hierarchical approaches for the containment problem in terms of polynomial matrix inequalities (PMI). The underlying PMI hierarchy was developed by Kojima [24], Hol and Scherer [17], as well as Henrion and Lasserre [16], and it generalizes the Lasserre hierarchy for polynomial optimization [25]. We then discuss the relation of the two approaches for containment to each other as well as the connection to positive maps.

### 5.1 From the Sufficient Criterion to a Moment Hierarchy of Sufficient Criteria

As before, let  $A(x) \in \mathcal{S}_k[x]$  and  $B(x) \in \mathcal{S}_l[x]$ , and assume that  $S_A \neq \emptyset$ . By definition of a positive semidefinite matrix, we have  $S_A \subseteq S_B$  if and only if the infimum  $\mu$  of the polynomial optimization problem

$$\begin{aligned} \mu &= \inf z^T B(x)z \\ \text{s.t. } A(x) &\geq 0 \\ g(z) &:= z^T z - 1 = 0 \end{aligned} \tag{16}$$

in the variables  $(x, z) = (x_1, \dots, x_n, z_1, \dots, z_l)$  is non-negative (cf. [22] for improved numerical stability). Setting  $G_A(x, z)$  to be the matrix with blocks  $A(x)$  as well as the two  $1 \times 1$ -blocks  $g(z)$  and  $-g(z)$ , the constraints can be written as  $G_A(x, z) \geq 0$ .

The general framework of moment relaxations for PMIs translates the optimization problem into a semidefinite hierarchy as a relaxation to problem (16). Assuming, for ease of notation, that we are working over the variables  $x = (x_1, \dots, x_n)$ , let  $y = (y_\alpha)$  be a real sequence indexed by the monomials in  $x$ . Let  $M(y)$  be the infinite moment matrix defined by  $(M(y))_{\alpha, \beta} = L_y([x][x]^T)_{\alpha, \beta} = y_{\alpha+\beta}$ , where  $[x]$  is the infinite vector of monomials in  $x_1, \dots, x_n$  and  $L_y$  is the linearization operator that maps a monomial  $x^\alpha$  to the associated moment variable  $y_\alpha$ .  $M_t(y)$  denotes the truncated moment matrix that contains only entries  $(M(y))_{\alpha, \beta}$  with  $|\alpha|, |\beta| \leq t$ .

The positive semidefiniteness constraint on a matrix polynomial  $G(x) \in \mathcal{S}_k[x]$  is captured by the *localizing matrices*. The truncated localizing matrix  $M_t(Gy)$  is defined as  $M_t(Gy) = L_y([x]_t[x]_t^T \otimes G(x))$ , where application of the linearization operator  $L_y$  is component-wise. If  $d_G$  denotes the highest degree of a polynomial appearing in  $G(x)$ , then only linearization variables coming from monomials of degree at most  $2t + d_G$  appear in  $M_t(Gy)$ .

For  $t \geq 2$ , the  $t$ -th relaxation of the polynomial optimization problem (16) becomes

$$\begin{aligned} \mu_{\text{mom}}(t) &= \inf L_y(z^T B(x)z) \\ \text{s.t. } M_t(y) &\geq 0 \\ M_{t-1}(G_A y) &\geq 0. \end{aligned} \tag{17}$$

Note that  $t = 2$  is the initial relaxation order. The sequence  $\mu_{\text{mom}}(t)$  for  $t \geq 2$  is monotone non-decreasing. If for some  $t^*$  the condition  $\mu_{\text{mom}}(t^*) \geq 0$  is satisfied, then  $S_A \subseteq S_B$ .

The following connection will be further refined and extended in Theorems 5.5 and 5.7.

**Theorem 5.1** [22] *Let  $S_A \neq \emptyset$ . Then  $\mu_{\text{mom}}(2) \geq 0$  (and thus  $\mu_{\text{mom}}(t) \geq 0$  for all  $t \geq 2$ ) if and only if the SDFP (10) has a solution  $C \geq 0$ , that is, if and only if the sufficient containment criterion in Theorem 4.5 is satisfied.*

### 5.2 The Hol-Scherer Hierarchy

The background of the second hierarchical approach is provided by Hol-Scherer’s Positivstellensatz. In order to characterize matrix polynomials which are positive semidefinite on a spectrahedron, we consider a generalization of the quadratic module (7) for a matrix polynomial  $G \in \mathcal{S}_k[x]$ . For any  $l \geq 0$ , let

$$M_{G,l} = \{S_0 + \langle S, G \rangle_l : S_0 \in \mathbb{R}[x]^{l \times l} \text{ sos-matrix, } S \in \mathbb{R}[x]^{kl \times kl} \text{ sos-matrix}\},$$

where for matrices  $U = (U_{ij})_{i,j=1}^l \in \mathcal{S}_{kl}$  and  $V \in \mathcal{S}_k$  the  $l^{\text{th}}$  scalar product is defined by

$$\langle U, V \rangle_l = (\langle U_{ij}, V \rangle)_{i,j=1}^l \in \mathcal{S}_l.$$

**Proposition 5.2 (Hol, Scherer [17])** *Let  $G(x)$  be a matrix polynomial in  $\mathcal{S}_k[x]$ . Further assume that there exists a polynomial  $p(x) = s(x) + \langle S(x), G(x) \rangle$  for some sos-polynomial  $s(x) \in \mathbb{R}[x]$  and some sos-matrix  $S(x) \in \mathcal{S}_k[x]$ , such that the level set  $\{x \in \mathbb{R}^n : p(x) \geq 0\}$  is compact. Then every matrix polynomial  $F \in \mathcal{S}_l[x]$  which is positive semidefinite on  $\{x \in \mathbb{R}^n : G(x) \geq 0\}$  is contained in the quadratic module  $M_{G,l}$ .*

As before, let  $A(x) \in \mathcal{S}_k[x]$  and  $B(x) \in \mathcal{S}_l[x]$  be linear pencils, and consider for  $t \geq 0$  the truncated quadratic module

$$M_{A,l}^{(t)} = \{S_0 + \langle S, A \rangle_l : S \in \mathbb{R}[x]_{2t}^{l \times l} \text{ sos-matrix, } S \in \mathbb{R}[x]_{2t}^{kl \times kl} \text{ sos-matrix}\}. \quad (18)$$

**Proposition 5.3 ([19, 22])** *Let  $A(x) \in \mathcal{S}_k[x]$ ,  $B[x] \in \mathcal{S}_l[x]$  be linear pencils.*

1. *If  $B(x) \in M_{A,l}^{(t)}$  for some  $t \geq 0$ , then  $S_A \subseteq S_B$ .*
2. *Let  $S_A$  be bounded and  $B(x)$  be a reduced pencil. If  $S_A$  is contained in the interior of  $S_B$  then there exists some  $t \geq 0$  such that  $B(x) \in M_{A,l}^{(t)}$ .*

For computational purposes and to relate the hierarchy to the moment approach in Sect. 5.1, it is useful to pass over to a robust optimization version. First note that  $S_A \subseteq S_B$  if and only there exists some  $\lambda \geq 0$  with

$$B(x) - \lambda I_l \geq 0 \quad \text{for all } x \in S_A.$$

Now we consider the hierarchy of optimization problems

$$\begin{aligned} \lambda_{\text{sos}}(t) = \sup \lambda \\ \text{s.t. } B(x) - \lambda I_l - (\{S_{i,j}(x), A(x)\})_{i,j=1}^l \text{ sos-matrix} \\ S(x) = (S_{i,j}(x))_{i,j=1}^l \in \mathcal{S}_{kl}[x] \text{ sos-matrix,} \end{aligned} \tag{19}$$

where  $S(x)$  has  $l \times l$  blocks of size  $k \times k$  with entries of degree at most  $2t \geq 0$ . Given some  $t \geq 0$ , we observe that  $\lambda_{\text{sos}}(t) \geq 0$  implies that  $S_A \subseteq S_B$ .

**Theorem 5.4 ([22])** *Let  $A(x) \in \mathcal{S}_k[x]$  be a linear pencil such that the spectrahedron  $S_A$  is bounded. Then the optimal values of the moment relaxation (17) and of the sos-relaxation (19) converge from below to the optimal value of the polynomial optimization problem (16), i.e.,  $\mu_{\text{mom}}(t) \uparrow \mu$  and  $\lambda_{\text{sos}}(t) \uparrow \mu$  as  $t \rightarrow \infty$ .*

The following theorem shows that that the sufficient criteria coming from the hierarchies of relaxations are at least as strong as the criterion (10) by showing that feasibility of the criterion (10) implies  $\mu(t) \geq 0$  and  $\lambda_{\text{sos}}(t) \geq 0$  in the initial relaxation steps of the semidefinite hierarchies (17) and (19). From this relation, we get that in some cases already the initial relaxation step of the hierarchies gives an exact answer to the containment problem; see Sect. 5.3.

**Theorem 5.5 ([22])** *Let  $S_A \neq \emptyset$ . Then for the properties*

1. *the SDFP (10) has a solution  $C \geq 0$ ,*
2.  *$\lambda_{\text{sos}}(0) \geq 0$ ,*
3.  *$\mu_{\text{mom}}(2) \geq 0$ ,*
4.  *$S_A \subseteq S_B$ ,*

*we have the implications  $1 \iff 2 \implies 3 \implies 4$ .*

For further aspects on the Hol-Scherer hierarchy for containment see also Kellner’s dissertation [19].

### 5.3 (Completely) Positive Maps

We briefly discuss the connection of the hierarchies to the theory of positive maps and completely positive maps. For background on positive and completely positive maps see, e.g., [31].

**Definition 5.6** Given two linear subspaces  $\mathcal{A} \subseteq \mathbb{R}^{k \times k}$  and  $\mathcal{B} \subseteq \mathbb{R}^{l \times l}$ , a linear map  $\Phi : \mathcal{A} \rightarrow \mathcal{B}$  is called *positive* if  $\Phi(A) \geq 0$  for any  $A \in \mathcal{A}$  with  $A \geq 0$ .

The map  $\Phi$  is called *d-positive* if the map  $\Phi_d : \mathbb{R}^{d \times d} \otimes \mathcal{A} \rightarrow \mathbb{R}^{d \times d} \otimes \mathcal{B}$ ,  $M \otimes A \mapsto M \otimes \Phi(A)$  is positive, i.e., if  $M \otimes \Phi(A) \geq 0$  whenever  $M \otimes A \geq 0$ . And  $\Phi$  is called *completely positive* if  $\Phi_d$  is positive for all  $d \geq 1$ .

As explained in the following, checking positivity of a map on a subspace is equivalent to checking containment for spectrahedra. This does not only provide a structural connection, but also allows to apply the hierarchy for the containment question to positivity questions of maps on subspaces, such as the ones in [13]. Note that for the special case of detecting positivity of a map on the whole space, Nie has recently shown that this can be done by solving a finite number of semidefinite relaxations [29].

For simplicity, we restrict to the situation that  $A_0, \dots, A_n$  are linearly independent and that  $S_A$  is bounded. Let the linear map  $\Phi_{AB} : \mathcal{A} \rightarrow \mathcal{B}$  be defined through

$$\Phi_{AB}(A_p) = B_p \quad \text{for } 0 \leq p \leq n.$$

Then the following extension of Theorem 5.5 states the connection of the semidefinite hierarchies with positive and completely positive maps.

**Theorem 5.7** [22] *Let  $A_0, \dots, A_n$  be linearly independent and  $S_A$  be non-empty and bounded. Then for the properties*

1.  $\Phi_{AB}$  is completely positive,
2. the SDFP (10) has a solution  $C \succeq 0$ ,
3.  $\lambda_{\text{sos}}(0) \geq 0$ ,
4.  $\mu_{\text{mom}}(2) \geq 0$ ,
5.  $S_A \subseteq S_B$ ,
6.  $\Phi_{AB}$  is positive,

*we have the implications  $1 \iff 2 \iff 3 \implies 4 \implies 5 \iff 6$ . If  $\mathcal{A}$  contains a positive definite matrix, then the implication  $1 \iff 2$  is an equivalence.*

Note that Theorem 5.4 implies a partial converse of the implication  $3 \implies 4$ . Namely, if  $\emptyset \neq S_A \subseteq S_B$  and  $S_A$  is bounded, then  $\mu_{\text{mom}}(t) \uparrow \mu \geq 0$  for  $t \rightarrow \infty$ .

Theorem 5.7 allows to extend the exactness results from Theorem 4.7 to the initial step of the hierarchy (17).

*Remark 5.8* It is well-known that the map  $\Phi_{AB}$  connects to the characterization of biquadratic forms in Proposition 2.6 (see [5]). A positive linear map  $\Phi : \mathcal{S}_k \rightarrow \mathcal{S}_l$  is completely positive if and only if  $\Phi$  can be written as  $\Phi(A) = \sum_s V_s^T A V_s$  for some matrices  $V_s \in \mathbb{R}^{k \times l}$  if and only if the corresponding biquadratic form  $F(x, y)$  is a sum of squares of bilinear forms,  $F(x, y) = \sum_s (x^T V_s y)^2$ .

## 6 Final Remarks

We have reviewed some recent developments on fundamental algorithmic problems in spectrahedral computation. While containment question for spectrahedra are co-NP-hard in general, the hierarchical relaxation techniques give a practical way of certifying containment. For detailed experiments of the two approaches (17)



and (19), see [21] and [22]. In practice, the sufficient criteria perform well already for small relaxation orders.

While in many situations the running times of the two hierarchical approaches for containment are comparable, the number of linearization variables in the moment approach (17) does not depend on the size  $k$  of the pencil  $A(x)$ . Therefore, for problems with relatively large  $k$ , this approach to the containment problem seems to be superior to the approach based on Hol-Scherer's hierarchy.

**Acknowledgements** The author was partially supported through DFG grant 1333/3-1 within the Priority Program 1489 "Algorithmic and Experimental Methods in Algebra, Geometry, and Number Theory."

He would also like to thank an anonymous referee for careful reading and helpful suggestions.

## References

1. F. Alizadeh, Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM J. Optim.* **5**(1), 13–51 (1993)
2. A. Ben-Tal, A. Nemirovski, On tractable approximations of uncertain linear matrix inequalities affected by interval uncertainty. *SIAM J. Optim.* **12**(3), 811–833 (2002)
3. A. Bhardwaj, P. Rostalski, R. Sanyal, Deciding polyhedrality of spectrahedra. *SIAM J. Optim.* **25**(3), 1873–1884 (2015)
4. G. Blekherman, P.A. Parrilo, R.R. Thomas, *Semidefinite Optimization and Convex Algebraic Geometry* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 2013)
5. M.-D. Choi, Positive semidefinite biquadratic forms. *Linear Algebra Appl.* **12**(2), 95–100 (1975)
6. E. de Klerk, *Aspects of Semidefinite Programming*, Applied Optimization, vol. 65 (Kluwer Academic Publishers, Dordrecht, 2002)
7. R.M. Freund, J.B. Orlin, On the complexity of four polyhedral set containment problems. *Math. Program.* **33**(2), 139–145 (1985)
8. T. Fritz, T. Netzer, A. Thom, Spectrahedral containment and operator systems with finite-dimensional realization (2016). Preprint, arXiv:1609.07908
9. E. Gawrilow, M. Joswig, polymake: A framework for analyzing convex polytopes, in *Polytopes—Combinatorics and Computation*, ed. by G. Kalai, G.M. Ziegler (Birkhäuser, Basel, 2000), pp. 43–74
10. P. Gritzmann, V. Klee, On the complexity of some basic problems in computational convexity. I. Containment problems. *Discrete Math.* **136**(1–3), 129–174 (1994)
11. B. Grünbaum, *Convex Polytopes*, 2nd edn. Graduate Texts in Mathematics, vol. 221 (Springer, New York, 2003)
12. D. Handelman, Representing polynomials by positive linear functions on compact convex polyhedra. *Pac. J. Math.* **132**(1), 35–62 (1988)
13. T. Heinosaari, M.A. Jivulescu, D. Reeb, M.M. Wolf, Extending quantum operations. *J. Math. Phys.* **53**(10), 102208 (2012)
14. J.W. Helton, V. Vinnikov, Linear matrix inequality representation of sets. *Commun. Pure Appl. Math.* **60**(5), 654–674 (2007)
15. J.W. Helton, I. Klep, S. McCullough, The matricial relaxation of a linear matrix inequality. *Math. Program.* **138**(1–2, Ser. A), 401–445 (2013)
16. D. Henrion, J.B. Lasserre, Convergent relaxations of polynomial matrix inequalities and static output feedback. *IEEE Trans. Autom. Control* **51**(2), 192–202 (2006)

17. C.W.J. Hol, C.W. Scherer, Sum of squares relaxations for polynomial semidefinite programming, in *Proceedings of Symposium in Mathematical Theory of Networks and Systems, Leuven, 2004*
18. M. Joswig, T. Theobald, *Polyhedral and Algebraic Methods in Computational Geometry*. Universitext (Springer, London, 2013)
19. K. Kellner, Positivstellensatz certificates for containment of polyhedra and spectrahedra. PhD thesis, Goethe University, Frankfurt am Main, 2015
20. K. Kellner, T. Theobald, Sum of squares certificates for containment of  $\mathcal{H}$ -polytopes in  $\mathcal{V}$ -polytopes. *SIAM J. Discrete Math.* **30**(2), 763–776 (2016)
21. K. Kellner, T. Theobald, C. Trabant, Containment problems for polytopes and spectrahedra. *SIAM J. Optim.* **23**(2), 1000–1020 (2013)
22. K. Kellner, T. Theobald, C. Trabant, A semidefinite hierarchy for containment of spectrahedra. *SIAM J. Optim.* **25**(2), 1013–1033 (2015)
23. I. Klep, M. Schweighofer, An exact duality theory for semidefinite programming based on sums of squares. *Math. Oper. Res.* **38**(3), 569–590 (2013)
24. M. Kojima, Sums of squares relaxations of polynomial semidefinite programs. Technical report, Research Report B-397, Mathematical and Computing Sciences Tokyo Institute of Technology, Tokyo, 2003
25. J.B. Lasserre, Global optimization with polynomials and the problem of moments. *SIAM J. Optim.* **11**(3), 796 (2001)
26. M. Laurent, S. Poljak, On a positive semidefinite relaxation of the cut polytope. *Linear Algebra Appl.* **223**, 439–461 (1995)
27. S. Liu, Matrix results on the Khatri-Rao and Tracy-Singh products. *Linear Algebra Appl.* **289**(1–3), 267–277 (1999)
28. P. McMullen, The maximum numbers of faces of a convex polytope. *Mathematika* **17**, 179–184 (1970)
29. J. Nie, X. Zhang, Positive maps and separable matrices. *SIAM J. Optim.* **26**(2), 1236–1256 (2016)
30. G. Pataki, The geometry of semidefinite programming, in *Handbook of Semidefinite Programming* (Kluwer Academic Publishers, Boston, MA, 2000), pp. 29–65
31. V. Paulsen, *Completely Bounded Maps and Operator Algebras* (Cambridge University Press, Cambridge, 2003)
32. M. Ramana, An exact duality theory for semidefinite programming and its complexity implications. *Math. Program.* **77**(1, Ser. A), 129–162 (1997)
33. M. Ramana, A.J. Goldman, Some geometric results in semidefinite programming. *J. Global Optim.* **7**(1), 33–50 (1995)
34. C. Scheiderer, Semidefinitely representable convex sets (2016). Preprint. arXiv:1612.07048
35. A. Schrijver, *Theory of Linear and Integer Programming*. Wiley-Interscience Series in Discrete Mathematics (Wiley, Chichester, 1986)
36. J.F. Sturm, Theory and algorithms of semidefinite programming, in *High Performance Optimization*, ed. by H. Frenk, K. Roos, K. Terlaky, S. Zhang. Applied Optimization, vol. 33 (Kluwer Academic Publishers, Dordrecht, 2000), pp. 1–194
37. T. Theobald, T. de Wolff, Approximating amoebas and coamoebas by sums of squares. *Math. Comp.* **84**(291), 455–473 (2015)
38. G.M. Ziegler, *Lectures on Polytopes*. Graduate Texts in Mathematics (Springer, New York, 1995)

# Topics on Modular Galois Representations Modulo Prime Powers



Panagiotis Tsaknias and Gabor Wiese

**Abstract** This article surveys modularity, level raising and level lowering questions for two-dimensional representations modulo prime powers of the absolute Galois group of the rational numbers. It contributes some new results and describes algorithms and a database of modular forms orbits and higher congruences.

**Keywords** Modular forms • Galois representations • Modularity higher congruences • Level raising • Level lowering • Database

**Subject Classifications** 11F33, 11F80

## 1 Introduction

The Fontaine-Mazur conjecture relates  $\ell$ -adic ‘geometric’ Galois representations with objects from geometry. In the 2-dimensional case over  $\mathbb{Q}$  much progress has been achieved ([12, Thm.1.2.4(2)] and [16]):

**Theorem 1.1 (Emerton, Kisin)** *Let  $\ell > 2$ , let  $E/\mathbb{Q}_\ell$  be a finite extension and let  $\rho : \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow \text{GL}_2(E)$  be an irreducible, finitely ramified, odd Galois representation which is de Rham at  $\ell$  with distinct Hodge-Tate weights. Assume that the residual representation  $\overline{\rho}$  satisfies certain local conditions.*

*Then a twist of  $\rho$  is attached to some newform.*

In fact, the level and the weight of the newform can be read off from  $\rho$ .

The picture for mod  $\ell$  representations is even more complete: Serre type modularity conjectures relate 2-dimensional Galois representations with  $\overline{\mathbb{F}}_\ell$ -coefficients with modular forms over  $\overline{\mathbb{F}}_\ell$ . Serre’s original modularity conjecture has been established by Khare and Wintenberger [14]:

---

P. Tsaknias • G. Wiese (✉)

Université du Luxembourg, Unité de Recherche en Mathématiques, Maison du nombre, 6, avenue de la Fonte, L-4364 Esch-sur-Alzette, Luxembourg  
e-mail: [p.tsaknias@gmail.com](mailto:p.tsaknias@gmail.com); [gabor.wiese@uni.lu](mailto:gabor.wiese@uni.lu)

© Springer International Publishing AG, part of Springer Nature 2017  
G. Böckle et al. (eds.), *Algorithmic and Experimental Methods in Algebra, Geometry, and Number Theory*,  
[https://doi.org/10.1007/978-3-319-70566-8\\_31](https://doi.org/10.1007/978-3-319-70566-8_31)

741

**Theorem 1.2 (Khare, Wintenberger, Kisin)** *Let  $\bar{\rho} : \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow \text{GL}_2(\overline{\mathbb{F}}_\ell)$  be an odd irreducible Galois representation. Then  $\bar{\rho}$  is attached to (the reduction of) some newform.*

Also in this case, the level and the weight of one of the (infinitely many) newforms attached to  $\bar{\rho}$  can be read off from  $\bar{\rho}$ : the level is the prime-to- $\ell$  Artin conductor, and there is a formula for the weight given by Serre. In fact, it had been known for a long time that if  $\bar{\rho}$  is attached to *some* newform then it also is to one with a certain predicted weight and level. Since the predicted weight and level are minimal (except for two cases in the weight, see [11]), the process of finding a newform with predicted invariants is called *level lowering* or *weight lowering*. The quest for attached newforms with non-minimal levels is accordingly called *level raising*. These three questions have been completely solved (with a tiny exception when  $\ell = 2$  and the minimal weight 1 is concerned).

With the  $\ell$ -adic and the mod  $\ell$  cases of irreducible odd 2-dimensional representations of  $G_{\mathbb{Q}} := \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$  essentially settled, it is natural to wonder what happens in between, i.e. modulo prime powers. Quite some research has been done, but the picture is far from clear. In fact, very basic questions are still open.

## 1.1 Modularity Modulo Prime Powers

Let us consider a representation (continuous like all representations in this paper)

$$\rho : G_{\mathbb{Q}} \rightarrow \text{GL}_2(\mathcal{O}/\lambda^m)$$

with  $\mathcal{O}$  the valuation ring of a finite extension of  $\mathbb{Q}_\ell$  and  $\lambda$  its valuation ideal. It turns out that the *modularity* of  $\rho$  follows from known results if one supposes that the residual representation  $\bar{\rho}$  is absolutely irreducible and odd and satisfies certain technical local conditions. This was surely known to many experts and had been, for instance, discussed on mathoverflow. We make this precise in Sect. 2.

An important point here is that one needs to use the right notion of *modularity*. This difficulty is not visible when only working  $\ell$ -adically or modulo  $\ell$ . The second author together with Chen and Kiming introduced in [7] three notions of modularity modulo prime powers: *strong modularity*, *weak modularity*, *dc-weak modularity*. These three notions stem from three notions of Hecke eigenforms modulo prime powers, also called *strong*, *weak*, and *dc-weak*, which we briefly explain now.

Throughout this article, we understand by a *Hecke eigenform*  $f$  with coefficients in a ring  $R$  (they are all normalised and almost all cuspidal without this being said explicitly) a ring homomorphism  $f : \mathbb{T} \rightarrow R$ , where  $\mathbb{T}$  is a Hecke algebra (to be specified very soon). We often think of  $f$  as the  $q$ -expansion  $\sum_{n \geq 1} f(T_n)q^n \in R[[q]]$ . Let  $\mathbb{T}_k(\Gamma)$  be the full Hecke algebra, generated as a ring by all Hecke operators  $T_n$ , acting faithfully on the space of holomorphic cusp forms  $S_k(\Gamma)$  of weight  $k$  and level  $\Gamma$ .

A *weak* Hecke eigenform of weight  $k$  and level  $\Gamma$  with coefficients in  $R$  is a ring homomorphism  $f : \mathbb{T}_k(\Gamma) \rightarrow R$ . It is called *strong* if there exists an order  $\mathcal{O}$  in a number field together with a ring homomorphism  $\pi : \mathcal{O} \rightarrow R$  such that  $f$  factors as  $\mathbb{T}_k(\Gamma) \rightarrow \mathcal{O} \xrightarrow{\pi} R$ . By embedding the order  $\mathcal{O}$  into  $\mathbb{C}$ , the first arrow leads to  $\mathbb{T}_k(\Gamma) \rightarrow \mathbb{C}$ , a holomorphic Hecke eigenform. In simple terms, strong Hecke eigenforms with coefficients in  $R$  are those that are obtained by applying  $\pi$  to the coefficients of a holomorphic eigenform.

Put  $S_{\leq b}(\Gamma) = \bigoplus_{k=1}^b S_k(\Gamma)$  and let  $\mathbb{T}_{\leq b}(\Gamma)$  be the full Hecke algebra acting faithfully on it. A ring homomorphism  $f : \mathbb{T}_{\leq b}(\Gamma) \rightarrow R$  is called a *dc-weak eigenform* of level  $\Gamma$  (and weights  $\leq b$ ; in fact,  $b$  will not play any role as long as it is large enough). A dc-weak eigenform can hence have contributions from many different weights, as is the case for divided congruences, which is what the abbreviation ‘dc’ stands for. If  $R$  is a finite field or  $\overline{\mathbb{F}}_\ell$ , all three notions coincide by the Deligne–Serre lifting lemma (for a presentation in the setup used here, see [7, Lemma 16]), but they are different in general. An example that strong is stronger than weak modulo  $\ell^m$  for  $m > 1$ , even if one allows the weight to change (but not the level), is given in [15, §2.5].

As rings of coefficients  $R$ , we take in this article rings of the form  $\mathcal{O}/\lambda^m$  where  $\mathcal{O}$  is the valuation ring of a finite field extension of  $\mathbb{Q}_\ell$ ,  $\lambda$  is its valuation ideal and  $m$  a positive integer. Many results can be and are phrased in this way. However, a general difficulty exists: we often need to compare two eigenforms, one with coefficients in  $\mathcal{O}_1/\lambda_1^{m_1}$ , the other one with coefficients in  $\mathcal{O}_2/\lambda_2^{m_2}$ . One then needs to find a ring containing both. In order for such a ring to exist, it is necessary that  $\lambda_i^{m_i} \cap \mathbb{Z}_\ell$  for  $i = 1, 2$  both yield the same power of  $\ell$ , say  $\ell^m$ . This led the second author together with Taixés i Ventosa [21] to introduce the ring

$$\overline{\mathbb{Z}/\ell^m\mathbb{Z}} = \overline{\mathbb{Z}_\ell}/\{x \in \overline{\mathbb{Z}_\ell} \mid v(x) > m - 1\},$$

where  $v$  denotes the normalised valuation, i.e.  $v(\ell) = 1$ . We always consider  $\overline{\mathbb{Z}/\ell^m\mathbb{Z}}$  with the discrete topology. We have  $\overline{\mathbb{Z}/\ell\mathbb{Z}} = \overline{\mathbb{F}}_\ell$  and for the valuation ring  $\mathcal{O}$  of any finite extension of  $\mathbb{Q}_\ell$  with absolute ramification index  $e$  and valuation ideal  $\lambda$ , the quotient  $\mathcal{O}/\lambda^{e(m-1)+1}$  injects into  $\overline{\mathbb{Z}/\ell^m\mathbb{Z}}$ . This quotient is the smallest one that extends  $\mathbb{Z}/\ell^m\mathbb{Z}$ . The ring  $\overline{\mathbb{Z}/\ell^m\mathbb{Z}}$  is a local  $\mathbb{Z}/\ell^m\mathbb{Z}$ -algebra of Krull dimension 0 with residue field  $\overline{\mathbb{F}}_\ell$  and the ring extension  $\mathbb{Z}/\ell^m\mathbb{Z} \subseteq \overline{\mathbb{Z}/\ell^m\mathbb{Z}}$  is integral. Any finitely generated subring  $R$  of  $\overline{\mathbb{Z}/\ell^m\mathbb{Z}}$  is contained in some ring  $\mathcal{O}/\lambda^{e(m-1)+1}$  as above. These are free as  $\mathbb{Z}/\ell^m\mathbb{Z}$ -modules, but this is not true for all finite subrings of  $\overline{\mathbb{Z}/\ell^m\mathbb{Z}}$ . A Hecke eigenform with coefficients in  $\overline{\mathbb{Z}/\ell^m\mathbb{Z}}$  shall simply be called a *modulo  $\ell^m$  Hecke eigenform*. Dc-weak (and hence also weak) Hecke eigenforms modulo  $\ell^m$  have attached Galois representations, under the condition that the residual representation is absolutely irreducible (see [7, Theorem 3]).

### 1.2 Weight Lowering and Finiteness

Let us recall now that in a fixed prime-to- $\ell$  level, there are only finitely many modular Galois representations with coefficients in  $\overline{\mathbb{F}}_\ell$ , which can all be realised—up to twist—in weights up to  $\ell + 1$  and that there is an explicit recipe for the minimal weight.

A natural question is whether there is a recipe for a minimal weight for strong eigenforms modulo  $\ell^m$ , i.e. whether (almost) all the (prime-indexed) Hecke eigenvalues of a given strong eigenform  $f$  modulo  $\ell^m$  in weight  $k$  and level  $N$  (prime to  $\ell$ ) also occur for a strong eigenform  $g$  modulo  $\ell^m$  in the same level  $N$  and a ‘low’ or ‘minimal’ weight that can be calculated from the restriction to a decomposition group at  $\ell$  of the Galois representation attached to  $f$  (under the assumption of residual absolute irreducibility).

This question seems to be very difficult. One is then led to consider the question, for fixed prime-to  $\ell$  level  $N$ , whether the set

$$\left\{ \sum_{n \geq 1} f(T_n)q^n \in \overline{\mathbb{Z}/\ell^m\mathbb{Z}}[[q]] \mid f \text{ strong eigenform modulo } \ell^m \text{ of level } N, \text{ any weight } \right\}$$

is finite. It can also be seen as the set of reductions modulo  $\ell^m$  of all holomorphic Hecke eigenforms in level  $N$  of any weight. The second author together with Kiming and Rustom conjectures that this is the case ([15, Conjecture 1]). As is shown in Theorem 2 of loc. cit., a positive answer to a question of Buzzard [5, Question 4.4] would indeed imply this. As an indication towards finiteness or the potential existence of a weight recipe as alluded to above, [15, Theorem 3], proved with the help of Frank Calegari, shows that for  $\ell \geq 5$ , there exists a bound  $B = B(N, \ell^m)$  such that the  $q$ -expansion of any *strong* Hecke eigenform modulo  $\ell^m$  of level  $N$ , but any weight, already occurs in weight  $k \leq B$  for some *weak* Hecke eigenform modulo  $\ell^m$  of level  $N$ . One should compare this with the level raising and level lowering results below, which also ‘only’ lead to weak forms.

Some first experimentation has led Kiming, Rustom and the second author to state the formula

$$B(N, \ell^m) = 2\ell^m + \ell^2 + 1$$

for  $m \geq 2$ . It is consistent with the available computational data, but should not be understood as a conjecture at this point.

### 1.3 Level Raising

Led by classical level raising results, one can hope that similar statements are true modulo  $\ell^m$ . It seems that part of the theory indeed carries over from modulo  $\ell$

to modulo  $\ell^m$  eigenforms. In Sect. 3 we prove a level raising result for weight 2 eigenforms on  $\Gamma_0(N)$ . For  $\ell > 2$ , this result is as general as possible. Only for  $\ell = 2$  some rare cases could not be proved.

Let  $f : \mathbb{T} \rightarrow \overline{\mathbb{Z}/\ell^m\mathbb{Z}}$  be a weak modulo  $\ell^m$  Hecke eigenform with Galois representation  $\rho$ . The main idea is to extend Ribet’s ‘classical’ geometric approach of level raising to our more general situation. For that we need to realise  $\rho$  on the Jacobian  $J$  of the appropriate modular curve. It is well known that the Hecke algebra acts faithfully on  $J$ . However, we need that  $\mathbb{T}/\ker(f)$ , i.e. the image of the weak eigenform, also acts faithfully on a subgroup of the Jacobian. The natural place is  $J(\mathbb{Q})[\ker(f)]$ . It turns out that this faithfulness does not seem to be that clear. In fact, we currently make use of the ‘multiplicity one’ property for the residual Galois representation on the Jacobian and, equivalently, the Gorenstein property of the residual Hecke algebra. Once this faithfulness is established, the proof proceeds by comparing the new and the old subvarieties in level  $Np$  as in Ribet’s original work [18]. One should expect similar limitations when extending level raising modulo  $\ell^m$  to higher weights, e.g. the weight will likely have to be less than  $\ell$  if one wants complete results in order to remain in the multiplicity one situation, where faithfulness is known and easily obtained from existing results.

### 1.4 Level Lowering and Other Results

Another natural domain is that of level lowering modulo  $\ell^m$ . The principal idea is that one should always be able to find an eigenform giving rise to a given Galois representation when the level is equal to the Artin conductor of the representation. An immediate difficulty is then, of course, to define an Artin conductor for Galois representations modulo  $\ell^m$ . It does not seem to be immediately clear how to do this because not every module over  $\mathbb{Z}/\ell^m\mathbb{Z}$  is free, so that there is no natural analog for the dimension (of, say, inertia invariants) used in the classical Artin conductor. Nevertheless, one can at least ask whether one can always find a modulo  $\ell^m$  Hecke eigenform of a level which is only divisible by primes ramifying in the representation. There are, indeed, two such results, one is due to Dummigan, and the other one due to Camporino and Pacetti. We quote both in Sect. 4. Dummigan’s result, similar to our level raising theorem, works geometrically on cohomology, whereas Camporino and Pacetti use the deformation theory of Galois representations. Both approaches currently seem to lead to some restrictions (a congruence condition for Dummigan, and unramified coefficients for Camporino–Pacetti).

Concerning generalisations along the lines of level lowering results modulo  $\ell$ , which are based on the use of Shimura curves, the mod  $\ell^m$  Galois representation must first be realised in the cohomology (or the Jacobian) of the appropriate Shimura curve. This is likely going to lead into faithfulness problems analogous to the one we solved in the level raising result by appealing to the Gorenstein or multiplicity one condition.

As a further instance of level lowering (though of a slightly different nature), we mention the following result from [7, Theorem 5]: Any dc-weak eigenform modulo  $\ell^m$  in level  $N\ell^r$  already arises from a dc-weak eigenform modulo  $\ell^m$  in level  $N$ , under the hypotheses  $\ell \geq 5$  and that the mod  $\ell$  reduction has an absolutely irreducible Galois representation. It is also shown that even if one starts with a strong eigenform modulo  $\ell^m$ , the one in level  $N$  will only be dc-weak, in general.

Another natural direction is to extend companionship results from eigenforms modulo  $\ell$  to  $\ell^m$ . This has been successfully performed by Adibhatla and Manoharmayum in [1] for odd  $\ell$  and ordinary modular forms with coefficients unramified at  $\ell$ , under certain conditions. In fact, that work is set in the more general world of Hilbert modular forms. Another companionship result modulo prime powers has been achieved by the first author together with Adibhatla [2].

## 1.5 Computations, Algorithm and Database

Next to the theoretical and structural motivation for studying modular forms and modularity questions modulo prime powers, there is also a strong computational driving force: realising  $\ell$ -adic modular forms on a computer is only possible up to a certain precision, i.e. one necessarily realises modular forms modulo  $\ell^m$ .

This also naturally leads to the questions studied in this article. For instance, if one wants to compute modulo which power of  $\ell$  a modular  $\ell$ -adic Galois representation  $\rho$  of conductor  $Np$  (with  $p$  a prime not dividing  $N$ ) becomes unramified at  $p$ , one can test whether the system of Hecke eigenvalues modulo  $\ell^m$  also occurs in level  $N/p$  for  $m = 1, 2, \dots$  until this fails. If it first fails at  $m + 1$ , then  $\rho$  modulo  $\ell^m$  is known to be unramified at  $p$ . In cases where level lowering modulo  $\ell^m$  is entirely proved, one also gets that  $\rho$  modulo  $\ell^{m+1}$  does ramify at  $p$ . The authors know of no other way of obtaining such information of an  $\ell$ -adic modular Galois representation.

The authors have developed several algorithmic tools for handling modular forms modulo  $\ell^m$  and they have set up a database. Section 5 contains a brief exposition of how to compute decompositions of commutative algebras into local factors in situations arising from Hecke algebras, and how to perform weak modularity tests explicitly. Finally, in Sect. 6 we describe features of the database of modular form orbits and higher congruences that we have developed.

## 2 Modularity

In this section we prove the following modularity theorem. This theorem has been known to the experts and is a pretty straight forward application of ‘bigR=bigT’ theorems.



**Theorem 2.1** *Let  $\ell \geq 5$  be a prime number, let  $\Sigma$  be a finite set of primes not containing  $\ell$  and let  $G_{\mathbb{Q}, \Sigma \cup \{\infty, \ell\}}$  be the Galois group of the maximal extension of  $\mathbb{Q}$  unramified outside  $\Sigma \cup \{\ell, \infty\}$ . Consider a continuous Galois representation*

$$\rho : G_{\mathbb{Q}, \Sigma \cup \{\infty, \ell\}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{Z}/\ell^m\mathbb{Z}})$$

such that the residual representation  $\overline{\rho}$  satisfies:

- $\overline{\rho}$  is odd,
- $\overline{\rho}|_{G_{\mathbb{Q}(\zeta_p)}}$  is absolutely irreducible,
- $\overline{\rho}|_{G_{\mathbb{Q}_\ell}} \not\sim \chi \otimes \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}$  and  $\overline{\rho}|_{G_{\mathbb{Q}_\ell}} \not\sim \chi \otimes \begin{pmatrix} 1 & * \\ 0 & \overline{\epsilon} \end{pmatrix}$ , for any  $\overline{\mathbb{F}}_\ell$ -valued character  $\chi$  of  $G_{\mathbb{Q}_\ell}$  and the mod  $\ell$  cyclotomic character  $\overline{\epsilon}$  (where  $*$  may or may not be zero).

Let  $N$  be the maximal positive integer divisible only by primes in  $\Sigma$  such that there is a newform of level  $N$  (and some weight) giving rise to  $\overline{\rho}$ .

Then  $\rho$  is dc-weakly modular of level  $N$ , i.e.  $\rho \cong \rho_f$  with  $f$  a dc-weak Hecke eigenform modulo  $\ell^m$  of level  $N$ .

In the exposition of the theory, we essentially follow Deo’s paper [8]. Let us assume the notation and the set-up from Theorem 2.1. Let  $\mathcal{O}$  be the valuation ring of a finite extension of  $\mathbb{Q}_\ell$  with ramification index  $e$ , valuation ideal  $\lambda$  and residue field  $\mathbb{F}$  such that (possibly after conjugation)  $\rho$  takes values in  $\mathrm{GL}_2(\mathcal{O}/\lambda^w) \subset \mathrm{GL}_2(\overline{\mathbb{Z}/\ell^m\mathbb{Z}})$  with  $w = e(m - 1) + 1$ . Let  $\mathbb{T}'_{\mathcal{O}}(\Gamma_1(N))$  be defined as the projective limit over  $b$  of  $\mathcal{O} \otimes \mathbb{T}'_{\leq b}(\Gamma_1(N))$  which are defined precisely like  $\mathbb{T}_{\leq b}(\Gamma_1(N))$ , but only take Hecke operators  $T_n$  with  $n$  coprime to  $N\ell$  into account. Similarly, like Deo we define the partially full Hecke algebra  $\mathbb{T}'_{\mathcal{O}}{}^{\mathrm{pf}}(\Gamma_1(N))$  as the projective limit of  $\mathcal{O} \otimes \mathbb{T}_{\leq b}{}^{\mathrm{pf}}(\Gamma_1(N))$  by using in addition the operators  $U_q$  for primes  $q \mid N$ . If we localise at the system of eigenvalues afforded by  $\overline{\rho}$ , we denote this by  $\overline{\rho}$  in the index. Accordingly, denote by  $R_{\overline{\rho}}$  the universal deformation ring of  $\overline{\rho}$  for the group  $G_{\mathbb{Q}, \Sigma \cup \{\infty, \ell\}}$  in the category of local profinite  $\mathcal{O}$ -algebras with residue field  $\mathbb{F}$ .

**Theorem 2.2 (Böckle, Diamond–Flach–Guo, Gouvêa–Mazur, Kisin)** *Assume the set-up of Theorem 2.1. Then  $R_{\overline{\rho}} \cong \mathbb{T}'_{\mathcal{O}}(\Gamma_1(N))_{\overline{\rho}}$ .*

This is Theorem 5 from [8]. Note that Deo works with pseudo-representations, but this comes down to the same thing here because we assume  $\overline{\rho}$  to be irreducible. In the proof, Deo essentially explains why the results of [9] allow to strengthen the conclusions of [3]. A similar discussion can also be found in [12, §7.3], where the theorem is, however, not stated in the form we need here. Alternatively, one can also invoke [12, Theorem 1.2.3] to an  $\ell$ -adic lift of  $\rho$ , provided such a lift exists. Recent work by Khare and Ramakrishna [13] provides a construction in the ordinary case.

We now apply Theorem 2.2. By assumption,  $\rho$  is a deformation of  $\overline{\rho}$  with the right ramification set, whence the universality leads to an  $\mathcal{O}$ -algebra homomorphism  $R_{\overline{\rho}} \rightarrow \mathcal{O}/\lambda^w$ , which we consider as an  $\mathcal{O}$ -algebra homomorphism

$$f : \mathbb{T}'_{\mathcal{O}}(\Gamma_1(N))_{\overline{\rho}} \rightarrow \mathcal{O}/\lambda^w.$$

By construction, the Galois representation associated with  $f$  is isomorphic to  $\rho$ .

In order to finish the proof of Theorem 2.1,  $f$  has to be extended to the full Hecke algebra in order to make it a genuine Hecke eigenform modulo  $\ell^m$ . Next we use that  $\mathbb{T}_{\mathcal{O}}^{\text{pf}}(\Gamma_1(N))_{\overline{\rho}}$  is finite over  $\mathbb{T}'_{\mathcal{O}}(\Gamma_1(N))_{\overline{\rho}}$ . This is proved in [8, Proposition 6]; one should note that the  $\Gamma_1(N)$ -new assumption is not necessary for this statement (see the proof of [8, Theorem 3]). This integrality allows us to extend  $f$  to an  $\mathcal{O}$ -algebra homomorphism

$$f : \mathbb{T}_{\mathcal{O}}^{\text{pf}}(\Gamma_1(N))_{\overline{\rho}} \rightarrow \tilde{\mathcal{O}}/\tilde{\lambda}^{\tilde{w}}$$

where  $\mathcal{O} \subseteq \tilde{\mathcal{O}}$  is the valuation ring of some finite extension of  $\mathbb{Q}_{\ell}$  with valuation ideal  $\tilde{\lambda}$  and ramification index  $\tilde{e}$  and  $\tilde{w} = \tilde{e}(m - 1) + 1$ . One is able to make this extension because one only needs to find one zero in some ring of the form  $\tilde{\mathcal{O}}/\tilde{\lambda}^{\tilde{w}}$  for any monic polynomial with coefficients in  $\mathcal{O}/\lambda^w$ ; that this is possible follows, for instance, by choosing any monic lift to  $\mathcal{O}$ . From the natural degeneracy map, we next get an  $\mathcal{O}$ -algebra homomorphism  $f : \mathbb{T}_{\mathcal{O}}^{\text{pf}}(\Gamma_1(N\ell))_{\overline{\rho}} \rightarrow \tilde{\mathcal{O}}/\tilde{\lambda}^{\tilde{w}}$ , which after choice of  $f(U_{\ell})$  leads to the  $\mathcal{O}$ -algebra homomorphism

$$f : \mathbb{T}_{\mathcal{O}}^{\text{pf}}(\Gamma_1(N\ell))_{\overline{\rho}}[[U_{\ell}]] \rightarrow \tilde{\mathcal{O}}/\tilde{\lambda}^{\tilde{w}}.$$

According to [8, Proposition 5], one can identify  $\mathbb{T}_{\mathcal{O}}^{\text{pf}}(\Gamma_1(N\ell))_{\overline{\rho}}[[U_{\ell}]]$  with a quotient of the full Hecke algebra  $\mathbb{T}_{\mathcal{O}}(\Gamma_1(N\ell))$ . We obtain thus an  $\mathcal{O}$ -algebra homomorphism

$$f : \mathbb{T}_{\mathcal{O}}(\Gamma_1(N\ell)) \rightarrow \tilde{\mathcal{O}}/\tilde{\lambda}^{\tilde{w}}.$$

As its image is finite, it will factor through  $\mathcal{O} \otimes \mathbb{T}_{\leq b}(\Gamma_1(N\ell))$  for a suitable weight bound  $b$ , so that we finally get a ring homomorphism

$$f : \mathbb{T}_{\leq b}(\Gamma_1(N\ell)) \rightarrow \tilde{\mathcal{O}}/\tilde{\lambda}^{\tilde{w}}.$$

This is the dc-weak eigenform that is needed to finish the proof of Theorem 2.1. Note that one can still remove  $\ell$  from the level of the final form because of [7, Theorem 5].

### 3 Level Raising via Modular Curves

Let  $p \nmid N$  be a rational prime. Then one has a natural inclusion map

$$S_k(\Gamma_0(N)) \oplus S_k(\Gamma_0(N)) \rightarrow S_k(\Gamma_0(Np)),$$

the image of which is called the  $p$ -old subspace. This subspace is stable under the action of  $\mathbb{T}_k(Np) := \mathbb{T}_k(\Gamma_0(Np))$  and so is its orthogonal complement under the Petersson inner product. This complementary subspace is called the  $p$ -new subspace

and we denote by  $\mathbb{T}_k^{p\text{-new}}(Np)$  the quotient of  $\mathbb{T}_k(Np)$  that acts faithfully on it. We will call this quotient the *p-new quotient* of  $\mathbb{T}_k(Np)$ . There is also the *p-old quotient* that is defined in the obvious way.

We can now state the main level raising result of this article.

**Theorem 3.1** *Let  $R$  be a local topological ring with maximal ideal  $\mathfrak{m}_R$ . Let  $\rho : G_{\mathbb{Q}} \rightarrow \text{GL}_2(R)$  be a continuous Galois representation that is modular, associated with a weak eigenform  $\theta : \mathbb{T}_2(N) \rightarrow R$ , and such that the residual representation  $\bar{\rho} : G_{\mathbb{Q}} \rightarrow \text{GL}_2(R/\mathfrak{m})$  is absolutely irreducible. If the characteristic of  $R/\mathfrak{m}$  is 2, assume the multiplicity one/Gorenstein condition that  $\bar{\rho}$  is not unramified at 2 with scalar Frobenius.*

*Let  $p \nmid N$  be a prime which satisfies the level raising condition for  $\rho$  by which we mean that  $\rho$  is unramified at  $p$  and*

$$\text{Tr}(\rho(\text{Frob}_p)) = \pm(p + 1).$$

*Then the image of  $\theta$  is a finite ring,  $R/\mathfrak{m}$  is a finite field and  $\rho$  is also associated with a weak eigenform  $\theta' : \mathbb{T}_2(Np) \rightarrow R$  which is new at  $p$ , i.e.  $\theta'$  factors through  $\mathbb{T}_2^{p\text{-new}}(Np)$ .*

In view of Lemma 3.4, the following corollary is essentially just an equivalent reformulation. Let  $\mathcal{O}$  be the ring of integers of a number field and  $\lambda$  a prime in  $\mathcal{O}$  above  $\ell$ .

**Corollary 3.2** *Let  $m \geq 1$  be an integer and  $\rho : G_{\mathbb{Q}} \rightarrow \text{GL}_2(\mathcal{O}/\lambda^m)$  be a continuous (for the discrete topology on  $\mathcal{O}/\lambda^m$ ) Galois representation that is modular, associated with a weak eigenform  $\theta : \mathbb{T}_2(N) \rightarrow \mathcal{O}/\lambda^m$ , and such that the residual representation  $\bar{\rho} : G_{\mathbb{Q}} \rightarrow \text{GL}_2(\mathcal{O}/\lambda)$  is absolutely irreducible. If  $\mathcal{O}/\lambda$  is of characteristic 2, assume the multiplicity one/Gorenstein condition that  $\bar{\rho}$  is not unramified at 2 with scalar Frobenius.*

*Let  $p$  be a prime which satisfies the level raising condition for  $\rho$ , which means here that*

$$(\ell N, p) = 1 \text{ and } \text{Tr}(\rho(\text{Frob}_p)) \equiv \pm(p + 1) \pmod{\lambda^m}.$$

*Then  $\rho$  is also associated with a weak eigenform  $\theta' : \mathbb{T}_2(Np) \rightarrow \mathcal{O}/\lambda^m$  which is new at  $p$ , i.e.  $\theta'$  factors through the  $\mathbb{T}_2^{p\text{-new}}(Np)$ .*

We remark that for  $m = 1$  this is Theorem 1 of [18]. Even if  $\theta$  is a strong eigenform, there is no guarantee that the weak eigenform of level new at  $p$  that one obtains in the end is strong.

**Corollary 3.3** *Let  $R$  be a local topological ring with maximal ideal  $\mathfrak{m}_R$  and let  $\rho : G_{\mathbb{Q}} \rightarrow \text{GL}_2(R)$  be a continuous Galois representation that is modular, has finite image and such that the residual representation  $\bar{\rho} : G_{\mathbb{Q}} \rightarrow \text{GL}_2(R/\mathfrak{m})$  is absolutely irreducible. If the characteristic of  $R/\mathfrak{m}$  is 2, assume the multiplicity one/Gorenstein condition that  $\bar{\rho}$  is not unramified at 2 with scalar Frobenius.*

Then there exists a positive set of primes  $p$  (coprime to  $N$ ) such that  $\rho$  is modular of level  $Np$  and new at  $p$ .

*Proof* This is proved as in [18]. The argument is that complex conjugation, as an involution, has trace 0 and determinant  $-1$ . By Chebotarev’s density theorem, there is a positive density set of primes  $p$  such that  $-1 = \det(\rho(\text{Frob}_p)) = p$  and  $p + 1 = 0 = \text{Tr}(\rho(\text{Frob}_p))$  in  $R$ .  $\square$

### 3.1 Jacobians of Modular Curves

In what follows we set  $\mathbb{T}_N := \mathbb{T}_2(\Gamma_0(N))$  and  $\mathbb{T}_{Np} := \mathbb{T}_2(\Gamma_0(Np))$ . The approach taken here is adapted from Ribet’s original one, i.e. it is based on the geometry of modular curves and their Jacobians. In this section we gather the necessary results from [18] that we need for the proof of the main result. Let  $N$  be a positive integer. Let  $X_0(N)$  be the modular curve of level  $N$  and  $J_0(N) := \text{Pic}^0(X_0(N))$  its Jacobian. There is a well defined action of the Hecke operators  $T_n$  on  $X_0(N)$  and hence, by functoriality, on  $J_0(N)$ , too. The dual of  $J_0(N)$  carries an action of the Hecke algebra as well and can be identified with  $S_2(\Gamma_0(N))$ . This implies that one has a faithful action of  $\mathbb{T}_N$  on  $J_0(N)$ .

Let now  $p$  be a prime not dividing  $N$ . In the same way one has an action of Hecke operators on  $X_0(Np)$  and its Jacobian  $J_0(Np)$  and the latter admits a faithful action of  $\mathbb{T}_{Np}$ . The moduli interpretation of  $X_0(N)$  and  $X_0(Np)$  allows us to define the two natural degeneracy maps  $\delta_1, \delta_p : X_0(Np) \rightarrow X_0(N)$  and their pullbacks  $\delta_1^*, \delta_p^* : J_0(N) \rightarrow J_0(Np)$ . The image of the map

$$\alpha : J_0(N) \times J_0(N) \rightarrow J_0(Np), \quad (x, y) \mapsto \delta_1^*(x) + \delta_p^*(y).$$

is by definition the  $p$ -old subvariety of  $J_0(Np)$ . We will denote it by  $A$ . The map  $\alpha$  is almost Hecke-equivariant:

$$\alpha \circ T_q = T_q \circ \alpha \text{ for every prime } q \neq p, \tag{1}$$

$$\alpha \circ \begin{pmatrix} T_p & p \\ -1 & 0 \end{pmatrix} = U_p \circ \alpha. \tag{2}$$

For the first equation to make sense one interprets the operator  $T_q$  on the left hand side of Eq. (1) as acting diagonally on  $J_0(N) \times J_0(N)$ . We also work under the notational convention  $T_q = U_q$  for primes  $q \mid N$ , but we write  $U_p$  in level  $Np$ . Consider also the kernel  $\text{Sh}$  of the map  $J_0(N) \rightarrow J_1(N)$  induced by  $X_1(N) \rightarrow X_0(N)$ . If we inject it into  $J_0(N) \times J_0(N)$  via  $x \mapsto (x, -x)$  then its image, which we will denote by  $\Sigma$ , is the kernel of  $\alpha$  (see Proposition 1 in [18]).

Let  $\Delta$  be the kernel of  $\begin{pmatrix} 1 + p & T_p \\ T_p & 1 + p \end{pmatrix} \in M^{2 \times 2}(\mathbb{T}_N)$  acting on  $J_0(N) \times J_0(N)$ . The group  $\Delta$  is finite and comes equipped with a perfect  $\mathbb{G}_m$ -valued skew-symmetric pairing. Furthermore  $\Sigma$  is a subgroup of  $\Delta$ , self orthogonal, and  $\Sigma \subseteq \Sigma^\perp \subseteq \Delta$ . One can also see  $\Delta/\Sigma$ , and therefore its subgroup  $\Sigma^\perp/\Sigma$ , as a subgroup of  $A$ .

Let  $B$  be the  $p$ -new subvariety of  $J_0(Np)$ . It is a complement of  $A$ , i.e.  $A + B = J_0(Np)$  and  $A \cap B$  is finite. The Hecke algebra acts faithfully on  $B$  through its  $p$ -new quotient and it turns out (see Theorem 2 in [18]) that

$$A \cap B \cong \Sigma^\perp/\Sigma. \tag{3}$$

as groups.

Furthermore  $\text{Sh}$ , and therefore  $\Sigma$  and its Cartier dual  $\Delta/\Sigma^\perp$ , are annihilated by the operators  $\eta_r = T_r - (r + 1) \in \mathbb{T}_N$  for all primes  $r \nmid Np$  (see Proposition 2 in [18]). In this context, we recall that a maximal ideal  $\mathfrak{m}$  of the Hecke algebra  $\mathbb{T}_N$  is called *Eisenstein* if  $T_r \pmod{\mathfrak{m}}$  equals the Frobenius traces of a two-dimensional reducible Galois representation at almost all primes  $r$ . This is in particular the case if  $\mathfrak{m}$  contains the operator  $T_r - (r + 1)$  for almost all primes  $r$ . Consequently, any maximal ideal in the support of the Hecke modules  $\Sigma$  and  $\Delta/\Sigma^\perp$  is Eisenstein.

### 3.2 Proof of Theorem 3.1

We assume the setting of Theorem 3.1. In particular, we assume that  $\rho$  satisfies the level raising condition at a prime  $p \nmid N$ , i.e. there is  $\epsilon \in \{\pm 1\}$  such that  $\theta(T_p) = \text{Tr}(\rho(\text{Frob}_p)) = \epsilon(p + 1)$ . Let  $\bar{\theta} : \mathbb{T}_N \rightarrow R/\mathfrak{m}_R$  be its reduction modulo  $\mathfrak{m}_R$  (which is associated with  $\bar{\rho}$ , the modulo  $\mathfrak{m}_R$  reduction of  $\rho$ ), and let  $I$  and  $\mathfrak{m}$  be the kernels of  $\theta$  and  $\bar{\theta}$ , respectively. It will be enough to find a weak eigenform  $\theta' : \mathbb{T}_{Np} \rightarrow R$  (i.e. a ring homomorphism) that agrees with  $\theta$  on  $T_q$  for all primes  $q \neq p$  and factors through  $\mathbb{T}_{Np}^{p\text{-new}}$  (hence, new at  $p$ ).

**Lemma 3.4** *The ideal  $\mathfrak{m}$  is the only maximal ideal of  $\mathbb{T}_N$  containing  $I$ . Moreover,  $\mathbb{T}_N/I$  is a finite subring of  $R$  of positive characteristic a prime power  $\ell^r$ .*

*Proof* Since  $\mathbb{T}_N$  is a  $\mathbb{Z}$ -Hecke algebra acting faithfully on  $S_2(N)$  we have that  $\mathbb{T}_N$  injects into  $M^{d \times d}(\mathbb{Z})$ , where  $d$  is the dimension of  $S_2(N)$ . We can therefore see every operator in  $\mathbb{T}_N$  as an integral matrix of dimension  $d$ . We recall that the eigenvalues of the operator  $T_n$  will correspond to the coefficients  $a_n(f)$  when  $f$  runs through the normalised eigenforms in  $S_2(N)$ .

Let  $g(X) \in \mathbb{Z}[X]$  be the characteristic polynomial of  $T_p$ . The hypothesis  $\theta(T_p) = \epsilon(p + 1)$  implies that  $T_p - \epsilon(p + 1) \in I$  and therefore  $m := g(\epsilon(p + 1)) \in I$ . Since  $p \nmid N$ , the Ramanujan-Petersson bounds guarantee that none of the eigenvalues of  $T_p$  is equal to  $\epsilon(p + 1)$  and therefore  $m$  is non-zero. We thus have that  $(m) \subseteq I$ . This makes the quotient  $\mathbb{T}/I$  finite.

Since  $\mathbb{T}/I$  is Artinian, it can be written as a direct product of Artinian local rings indexed by its finitely many maximal ideals. Assume it decomposes as a direct product of  $s$  local rings, with  $s \geq 1$ . The set containing the identity  $e_i$  of each component then forms a complete set (i.e.  $\sum_{i=1}^s e_i = 1$ ) of pairwise orthogonal (i.e.  $e_i e_j = 0$  for  $1 \leq i \neq j \leq s$ ) non-trivial (i.e.  $e_i \neq 0, 1$ ) idempotents for  $\mathbb{T}_N/I$ . The set  $\{\bar{e}_1, \dots, \bar{e}_s\}$  of their image through the injection of  $\mathbb{T}_N/I$  into  $R$  is clearly a complete set of pairwise orthogonal non-trivial idempotents, too. This implies that  $R$  is isomorphic to  $\prod_{i=1}^s \bar{e}_i R$ . But this cannot happen unless  $s = 1$  since  $R$  is local. Since  $s = 1$  we get that  $\mathbb{T}_N/I$  is local as well. The claims are then immediate.  $\square$

By the previous lemma, we have inclusions  $(\ell^r) \subseteq I \subseteq \mathfrak{m}$  with some prime power  $\ell^r > 1$ , giving rise to inclusions

$$V[\ell^r] := J_0(N)(\overline{\mathbb{Q}})[\ell^r] \supseteq V[I] := J_0(N)(\overline{\mathbb{Q}})[I] \supseteq V[\mathfrak{m}] := J_0(N)(\overline{\mathbb{Q}})[\mathfrak{m}].$$

**Lemma 3.5** *The support of  $V[I]$  is the singleton  $\mathfrak{m}$  and is hence non-Eisenstein.*

*Proof* As  $V[I] \supseteq V[\mathfrak{m}]$ , the maximal ideal  $\mathfrak{m}$  is in the support of  $V[I]$ . Since the representation  $\bar{\rho}$  is irreducible we get that  $\mathfrak{m}$  is non-Eisenstein (see for example Theorem 5.2c in [17]). Finally, Lemma 3.4 implies that  $\text{Supp}(V[I])$  is the singleton  $\{\mathfrak{m}\}$ .  $\square$

**Lemma 3.6** *The restriction of  $\alpha$  to  $V[I]$  is injective and its image  $\alpha(V[I])$  is stable under the action of  $\mathbb{T}_{Np}$ . In particular,  $U_p$  acts on  $\alpha(V[I])$  by multiplication by  $\epsilon$ .*

*Proof* Consider the image of  $V[I]$  (still denoted  $V[I]$ ) under the  $\mathbb{T}_N$ -equivariant embedding

$$J_0(N) \xrightarrow{x \mapsto (x, -\epsilon x)} J_0(N) \times J_0(N).$$

Next recall that the kernel  $\Sigma$  of  $J_0(N) \times J_0(N) \xrightarrow{\alpha} A \subseteq J_0(Np)$  is annihilated by almost all operators  $T_r - (r + 1)$  with  $r$  prime. The fact that the support of  $V[I]$  is non-Eisenstein from Lemma 3.5 shows that the intersection of  $\Sigma$  and  $V[I]$  is trivial, proving the injectivity of  $\alpha|_{V[I]}$ .

As  $\alpha$  commutes with the action of the Hecke operators  $T_n$  with  $n$  coprime to  $p$  (see Eq. (1)), it follows that  $\alpha(V[I])$  is stable under those operators. Here the level raising condition enters for proving the stability under  $U_p$ , as follows by using Eq. (2) for  $y \in V[I]$ :

$$\begin{aligned} U_p(y) &= U_p(\alpha(x, -\epsilon x)) = \alpha\left(\begin{pmatrix} T_p & p \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x \\ -\epsilon x \end{pmatrix}\right) = \alpha(T_p(x) - \epsilon px, -x) \\ &= \alpha(\epsilon(p + 1)x - \epsilon px, -x) = \alpha(\epsilon x, -x) = \epsilon \alpha(x, -\epsilon x) = \epsilon y. \end{aligned}$$

The final claim follows as well.  $\square$

The following proposition is a non-trivial input.

**Proposition 3.7** *The  $\mathbb{T}_N/I$ -module  $V[I]$  is faithful.*

*Proof* Due to the assumptions, Theorem 9.2 of [11] implies that  $V[\mathfrak{m}]$  is of dimension 2 as  $\mathbb{T}_N/\mathfrak{m}$ -module. By Nakayama’s Lemma, it follows that the localisation at  $\mathfrak{m}$  of the  $\ell$ -adic Tate module is free of rank 2 as  $(\mathbb{T}_N \otimes_{\mathbb{Z}} \mathbb{Z}_{\ell})_{\mathfrak{m}}$ -module and that  $\text{Hom}_{\mathbb{Z}_{\ell}}((\mathbb{T}_N \otimes_{\mathbb{Z}} \mathbb{Z}_{\ell})_{\mathfrak{m}}, \mathbb{Z}_{\ell})$  is free of rank 1 as  $(\mathbb{T}_N \otimes_{\mathbb{Z}} \mathbb{Z}_{\ell})_{\mathfrak{m}}$ -module, precisely as on p. 333 of [19]. Consequently,

$$V[\ell^r]_{\mathfrak{m}} \cong (\mathbb{T}_N/\ell^r\mathbb{T}_N)_{\mathfrak{m}}^2 \cong \text{Hom}_{\mathbb{Z}_{\ell}}((\mathbb{T}_N/\ell^r\mathbb{T}_N)_{\mathfrak{m}}, \mathbb{Z}/\ell^r\mathbb{Z})^2,$$

which implies by taking the  $\bar{I}$ -kernel with  $\bar{I}$  the ideal such that  $\mathbb{T}_N/I \cong (\mathbb{T}_N/\ell^r\mathbb{T}_N)/\bar{I}$  that

$$V[I] \cong \text{Hom}_{\mathbb{Z}}((\mathbb{T}_N/\ell^r\mathbb{T}_N)/\bar{I}, \mathbb{Z}/\ell^r\mathbb{Z})^2 = \text{Hom}_{\mathbb{Z}}(\mathbb{T}_N/I, \mathbb{Z}/\ell^r\mathbb{Z})^2,$$

showing that  $V[I]$  is faithful as  $\mathbb{T}_N/I$ -module. □

The authors do not know if the ‘multiplicity one’ or ‘Gorenstein’ condition is necessary. In the remaining case, the  $2'$ -torsion group scheme is ordinary, and hence by arguments as in Corollary 2.3 of [22] admits a nice decomposition as

$$0 \rightarrow (\mathbb{T}_N/\ell^r\mathbb{T}_N)_{\mathfrak{m}} \rightarrow V[\ell^r]_{\mathfrak{m}} \rightarrow \text{Hom}_{\mathbb{Z}}((\mathbb{T}_N/\ell^r\mathbb{T}_N)_{\mathfrak{m}}, \mathbb{Z}/\ell^r\mathbb{Z}) \rightarrow 0.$$

However, we do not know if this sequence remains exact after taking the  $\bar{I}$ -kernel. If this were the case, the additional assumption would be unnecessary.

**Lemma 3.8** *The action of  $\mathbb{T}_{Np}$  on  $\alpha(V[I])$  is given by a ring homomorphism  $\theta' : \mathbb{T}_{Np} \rightarrow R$  satisfying  $\theta'(T_q) = \theta(T_q)$  for all primes  $q \neq p$  and  $\theta'(U_p) = \epsilon$ . In particular,  $\theta$  and  $\theta'$  give rise to isomorphic Galois representations.*

*Proof* The faithfulness of  $V[I]$  as  $\mathbb{T}/I$ -module from Proposition 3.7 implies that  $\theta$  factors through a subring  $S$  of  $\text{End}(V[I])$ , which is also a subring of  $R$ . By Lemma 3.6 and Eq. (1), the action of  $\mathbb{T}_{Np}$  on  $\alpha(V[I])$  is also given by elements of  $S$ , leading to a ring homomorphism  $\theta' : \mathbb{T}_{Np} \rightarrow S \subseteq R$ . □

To finish the proof of Theorem 3.1, it remains to show that  $\theta'$  factors through the  $p$ -new quotient of  $\mathbb{T}_{Np}$ . To this end, it is enough to show that  $\alpha(V[I])$  is a subgroup of  $A \cap B$ . We again proceed according to Ribet. By the level raising condition,  $V[I]$ , when considered as a subgroup of  $J_0(N) \times J_0(N)$ , is a subgroup of  $\Delta$ , whence  $\alpha(V[I]) \subseteq \Delta/\Sigma$ . As  $\Delta/\Sigma^{\perp}$  is Eisenstein but  $\alpha(V[I])$  is not,  $\alpha(V[I])/\Sigma^{\perp} = 0$ . This implies  $\alpha(V[I]) \subseteq \Sigma^{\perp}/\Sigma = A \cap B$ , completing the proof of Theorem 3.1.

## 4 Level Lowering

In this section we give an overview of results about level lowering modulo prime powers. We start by the following simple observation: twisting an eigenform  $f$  by a

Dirichlet character  $\chi$  such that  $\chi \equiv 1 \pmod{\lambda^m}$  leads to an eigenform  $g = f \otimes \chi$ , which is congruent to  $f$  modulo  $\lambda^m$ . This idea leads to the following two level lowering results from the first author’s unpublished PhD thesis [20].

**Proposition 4.1 (Split Ramified Case)** *Let  $f \in S_k(\Gamma_1(M))$  be a newform such that the restriction to a decomposition group at  $p \neq \ell$  of the  $\ell$ -adic Galois representation attached to  $f$  is isomorphic to  $\chi_1 \oplus \chi_2$ , where both characters ramify. Let  $\lambda$  be a prime ideal of a number field containing the coefficients of  $f$ .*

*If  $\chi_1$  is unramified modulo  $\lambda^m$ , then there exists a normalised eigenform  $g \in S_k(\Gamma_1(M/p))$  such that  $f \equiv g \pmod{\lambda^m}$ .*

*Proof* We can decompose  $\chi_1 = \chi_{1,\text{unr}}\chi_{1,\text{ram}}$  into an unramified and a ramified character of  $G_{\mathbb{Q}_p}$ . As  $p \neq \ell$ , the order of  $\chi_{1,\text{ram}}$  is finite. By assumption,  $\chi_{1,\text{ram}} \equiv 1 \pmod{\lambda^m}$ , whence in particular the order of  $\chi_{1,\text{ram}}$  is a power of  $\ell$  because only roots of unity of  $\ell$ -power order vanish under reduction modulo  $\lambda$ . Thus  $\chi_{1,\text{ram}}$  is tamely ramified. By the local and the global Kronecker-Weber theorems,  $\chi_{1,\text{ram}}$  can be seen as a global Dirichlet character  $\tilde{\chi}_{1,\text{ram}}$  of conductor  $p$  the restriction of which to  $G_{\mathbb{Q}_p}$  equals  $\chi_{1,\text{ram}}$ .

Let now  $g$  be the newform corresponding to the twist  $f \otimes \tilde{\chi}_{1,\text{ram}}^{-1}$ . Then the restriction to a decomposition group at  $p$  of the  $\ell$ -adic Galois representation attached to  $g$  is isomorphic to  $\chi_{1,\text{unr}} \oplus \chi_2\chi_{1,\text{ram}}^{-1}$ . If  $\chi_2$  is tame (i.e. of conductor  $p$ ), then  $\chi_2\chi_{1,\text{ram}}^{-1}$  is either tame or unramified, and in any case its conductor divides  $p$ . If  $\chi_2$  is wild, i.e. it factors through  $\text{Gal}(\mathbb{Q}_p(\zeta_{p^r N})/\mathbb{Q}_p)$  with  $r \geq 2$  and  $p \nmid N$ , but not through  $\text{Gal}(\mathbb{Q}_p(\zeta_{p^{r-1} N})/\mathbb{Q}_p)$ , then also  $\chi_2\chi_{1,\text{ram}}^{-1}$  factors through  $\text{Gal}(\mathbb{Q}_p(\zeta_{p^r N})/\mathbb{Q}_p)$  but not through  $\text{Gal}(\mathbb{Q}_p(\zeta_{p^{r-1} N})/\mathbb{Q}_p)$ , whence the conductor of  $\chi_2\chi_{1,\text{ram}}^{-1}$  equals that of  $\chi_2$ . In both cases we hence find that the conductor of  $\chi_2\chi_{1,\text{ram}}^{-1}$  divides the conductor of  $\chi_2$ . Since the  $p$ -valuation of  $M$  equals the  $p$ -valuation of the conductor of  $\chi_2$  plus 1 (since the conductor of  $\chi_1$  is  $p$ ) and the  $p$ -valuation of the newform level of  $g$  is the  $p$ -valuation of the conductor of  $\chi_2\chi_{1,\text{ram}}^{-1}$ , it is clear that the newform level of  $g$  divides  $M/p$ . □

**Proposition 4.2 (Special Ramified Case)** *Let  $f \in S_k(\Gamma_1(M))$  be a newform such that the restriction to a decomposition group at  $p \neq \ell$  of the  $\ell$ -adic Galois representation attached to  $f$  is isomorphic to  $\chi \otimes \begin{pmatrix} \omega & * \\ 0 & 1 \end{pmatrix}$ , where  $\chi$  and  $*$  ramify and  $\omega$  is the  $\ell$ -adic cyclotomic character. Let  $\lambda$  be a prime ideal of a number field containing the coefficients of  $f$ .*

*If  $\chi$  is unramified modulo  $\lambda^m$ , then there exists a newform  $g \in S_k(\Gamma_1(M/p))$  such that  $f \equiv g \pmod{\lambda^m}$ .*

*Proof* The proof is essentially the same as in the split ramified case. Note, however, that the tameness of  $\chi$  implies that  $p^2$  exactly divides  $M$ , whence the newform level of  $g$  will be exactly  $M/p$ . □

These propositions may be useful in some situations. We also remark that the only Dirichlet character that is trivial modulo  $\ell^2$  in the sense of being equal to  $1 \in \overline{\mathbb{Z}}/\ell^2\overline{\mathbb{Z}}$  is the trivial one. That is just due to the fact that  $\lambda := 1 - \zeta_\ell$  is a uniformiser of  $\mathbb{Q}_\ell(\zeta_\ell)$ , whence  $\zeta_\ell \not\equiv 1 \pmod{(\lambda)^2}$ . This implies that the level does



not lower modulo  $\ell^m$  for any  $m \geq 2$  at primes  $p$  satisfying the hypothesis of one of the preceding propositions. We now quote the main result from [10], including the discussion in the last paragraph of that article.

**Theorem 4.3 (Dummigan)** *Let  $\ell$  be a prime. Let  $\ell + 2 > k \geq 2$  and let  $p$  be a prime not dividing  $N \in \mathbb{N}$  such that  $p \not\equiv 1 \pmod{\ell}$ . Let  $f \in S_k(\Gamma_1(Np))$  be an eigenform and let  $\lambda$  be a prime of the coefficient field of  $f$  above  $\ell$ . Suppose that the residual Galois representation of  $f$  modulo  $\lambda$  is irreducible.*

*If for some  $m \geq 1$  the Galois representation of  $f$  modulo  $\lambda^m$  is unramified at  $p$ , then there is a weak eigenform  $g$  of weight  $k$  and level  $\Gamma_1(N)$  such that  $f \pmod{\lambda^m}$  equals  $g$  at all coefficients the index of which is coprime to  $p$ .*

Dummigan also gives an explicit example where the resulting form  $g$  cannot be strong. We include another still unpublished result from [6] on level lowering, which is proved using the deformation theory of Galois representations.

**Theorem 4.4 (Pacetti-Camporino)** *Let  $\ell \geq 7$  be a prime. Let  $2 \leq k \leq \ell - 1$ . Let  $M$  be a positive integer. Let  $f \in S_k(\Gamma_1(M))$  be an eigenform with coefficients in  $K_f$ . Let  $\mathcal{O}_f$  be the ring of integers of  $K_f$ . Assume that*

- $\ell$  is unramified in  $\mathcal{O}_f$ , and
- $\mathrm{SL}_2(\mathcal{O}_f/\lambda)$  is a subgroup of the image of the mod  $\lambda$  representation attached to  $f$ .

*If  $p \mid M$  is a prime and  $m \geq 1$  is an integer such that the modulo  $\lambda^m$  Galois representation associated with  $f$  is unramified at  $p$ , then there is a weak eigenform  $g$  of weight  $k$  and level  $\Gamma_1(M/p)$  such that  $f \pmod{\lambda^m}$  equals  $g$  at all coefficients the index of which is coprime to  $p$ .*

This result is proved by first applying techniques of Ramakrishna: by introducing auxiliary primes in order to kill local obstructions, the authors construct an  $\ell$ -adic lift in which  $p$  remains unramified. They then prove and use a modularity lifting theorem to obtain that their lift is associated with some newform. Finally, they apply Theorem 4.3 to remove the auxiliary primes, which had been chosen in such a way that Dummigan’s theorem applies.

## 5 Computational Aspects

In this section, we describe various algorithms we have implemented and used in our computational study of higher congruences.

### 5.1 Some Commutative Algebra

We start by summarising some well known facts from commutative algebra. Let  $R$  be an Artinian ring, i.e. a ring in which every descending chain of ideals becomes stationary. In particular, for any ideal  $\mathfrak{a}$  of  $R$ , the sequence  $\mathfrak{a}^n$  becomes stationary,

i.e.  $\mathfrak{a}^n = \mathfrak{a}^{n+1}$  for all  $n$  “big enough”. We will then use the notation  $\mathfrak{a}^\infty$  for  $\mathfrak{a}^n$ . The following proposition is well known and easy to prove:

**Proposition 5.1** *Let  $R$  be an Artinian ring. Then every prime ideal of  $R$  is maximal and there are only finitely many maximal ideals in  $R$ . Moreover, the maximal ideal  $\mathfrak{m}$  is the only one containing  $\mathfrak{m}^\infty$ . Furthermore, if  $\mathfrak{m} \neq \mathfrak{n}$  are two maximal ideals, then for any  $k \in \mathbb{N} \cup \{\infty\}$ , the ideals  $\mathfrak{m}^k$  and  $\mathfrak{n}^k$  are coprime. The Jacobson radical  $\bigcap_{\mathfrak{m} \in \text{Spec}(R)} \mathfrak{m}$  is equal to the nilradical and consists of the nilpotent elements, and we have  $\bigcap_{\mathfrak{m} \in \text{Spec}(R)} \mathfrak{m}^\infty = (0)$ . Moreover, for every maximal ideal  $\mathfrak{m}$ , the ring  $R/\mathfrak{m}^\infty$  is local with maximal ideal  $\mathfrak{m}$  and is hence isomorphic to  $R_{\mathfrak{m}}$ , the localisation of  $R$  at  $\mathfrak{m}$ . Finally, by virtue of the Chinese Remainder Theorem we have the following isomorphism, referred to as local decomposition:*

$$R \xrightarrow{a \mapsto (\dots, a + \mathfrak{m}^\infty, \dots)} \prod_{\mathfrak{m} \in \text{Spec}(R)} R/\mathfrak{m}^\infty \cong \prod_{\mathfrak{m} \in \text{Spec}(R)} R_{\mathfrak{m}}.$$

**Definition 5.2** An idempotent of a ring  $R$  is an element  $e$  that satisfies  $e^2 = e$ . Two idempotents  $e, f$  are orthogonal if  $ef = 0$ . An idempotent  $e$  is primitive if it cannot be written as a sum of two idempotents both different from 0. A set of idempotents  $\{e_1, \dots, e_n\}$  is said to be complete if  $1 = \sum_{i=1}^n e_i$ .

In concrete terms, a complete set of primitive pairwise orthogonal idempotents is given by  $(1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$ .

**Proposition 5.3 (Newton Method/Hensel Lifting—Special Case)** *Let  $R$  be a ring and  $I$  be an ideal. Let  $f \in R[X]$  be a polynomial. We assume that there exist  $a \in R$  and a polynomial  $b \in R[X]$  such that  $1 = af(X) + b(X)f'(X)$ . Let further  $a_0 \in R$  be such that  $f(a_0) \in I^r$  for some  $r \geq 1$ . For  $n \geq 1$ , we make the following recursive definition:*

$$a_n := a_{n-1} - f(a_{n-1})b(a_{n-1}).$$

Then for all  $n \in \mathbb{N}$ , we have  $f(a_n) \in (I^r)^{2^n}$ . In particular, if  $\bigcap_{n \geq 1} I^n = 0$  then the sequence  $f(a_n)$  converges to 0 exponentially.

*Proof* This is a straight forward calculation with Taylor expansions of the polynomial. □

**Corollary 5.4 (Algorithmic Idempotent Lifting)** *Let  $R$  be a commutative  $\mathbb{Z}_\ell$ -algebra which is finitely generated as  $\mathbb{Z}_\ell$ -module. Let  $e_0 \in R/\ell R$  be an idempotent. For  $n \geq 1$ , make the following recursive definition:*

$$e_n := e_{n-1} - (e_{n-1}^2 - e_{n-1})(2e_{n-1} - 1) = 3e_{n-1}^2 - 2e_{n-1}^3. \tag{4}$$

Then  $e_n^2 \equiv e_n \pmod{\ell^{2^n} R}$  for all  $n \geq 0$ . Moreover, the  $e_n$  form a Cauchy sequence in  $R$  and thus converge to an idempotent  $e \in R$  ‘lifting’  $e_0$ , i.e. the image of  $e$  in  $R/\ell R$  is  $e_0$ .

*Proof* This is a simple application of the Newton method to the polynomial  $f(X) = X^2 - X$ . Note that we have  $f'(X) = 2X - 1$  and  $1 = -4(X^2 - X) + (2X - 1)(2X - 1)$ . □

The corollary thus tells us that any idempotent of  $R/\ell R$  lifts to an idempotent of  $R$ , and it tells us that the lift can be approximated by a simple recursion formula that is easy to implement and converges very rapidly. We shall now apply the preceding considerations to a commutative  $\mathbb{Z}_\ell$ -algebra  $\mathbb{T}$  which is free and finitely generated as a  $\mathbb{Z}_\ell$ -module. Let  $\overline{\mathbb{T}} = \mathbb{T} \otimes \mathbb{F}_\ell$  and  $\mathbb{T}_{\mathbb{Q}_\ell} = \mathbb{T} \otimes \mathbb{Q}_\ell$ . Note that  $\overline{\mathbb{T}}$  and  $\mathbb{T}_{\mathbb{Q}_\ell}$  are Artinian rings because they are finite dimensional vector spaces. The following well-known result follows from the above considerations together with some standard commutative algebra.

**Proposition 5.5** *The algebra  $\mathbb{T}$  is equidimensional (in the sense of Krull dimension) of dimension 1, i.e. any maximal ideal  $\mathfrak{m}$  strictly contains at least one minimal prime ideal  $\lambda$  and there is no prime ideal strictly in between the two. The maximal ideals of  $\mathbb{T}$  correspond bijectively under taking pre-images to the maximal ideals of  $\overline{\mathbb{T}}$ ; the same letter will be used to denote them. The minimal primes  $\lambda$  of  $\mathbb{T}$  are in bijection with the prime ideals of  $\mathbb{T}_{\mathbb{Q}_\ell}$  (all of which are maximal) under extension, for which the notation  $\lambda^{(e)}$  will be used. Under the correspondences, one has  $\overline{\mathbb{T}}_{\mathfrak{m}} \cong \mathbb{T}_{\mathfrak{m}} \otimes \mathbb{F}_\ell$  and  $\mathbb{T}_\lambda \cong \mathbb{T}_{\mathbb{Q}_\ell, \lambda^{(e)}}$ . By virtue of lifts of idempotents and Proposition 5.1, we have the local decompositions*

$$\mathbb{T} \cong \prod_{\mathfrak{m}} \mathbb{T}_{\mathfrak{m}}, \overline{\mathbb{T}} \cong \prod_{\mathfrak{m}} \overline{\mathbb{T}}_{\mathfrak{m}} \text{ and } \mathbb{T}_{\mathbb{Q}_\ell} \cong \prod_{\lambda} \mathbb{T}_{\mathbb{Q}_\ell, \lambda^{(e)}} \cong \prod_{\lambda} \mathbb{T}_\lambda,$$

where  $\mathfrak{m}$  runs through the maximal ideals of  $\mathbb{T}$  (and  $\overline{\mathbb{T}}$ ) and  $\lambda$  runs through the minimal primes of  $\mathbb{T}$  (or, equivalently, all the prime= maximal ideals of  $\mathbb{T}_{\mathbb{Q}_\ell}$ ).

### 5.2 Package for Computing $\ell$ -Adic Decompositions

The second author has developed the MAGMA [4] package PADICALGEBRAS (see [24]) for computing the objects appearing in Proposition 5.5. The package depends on the second author’s earlier MAGMA package ARTINALGEBRAS (see [23]).

The main ingredients are standard linear algebra, especially over finite fields, and the algorithmic idempotent lifting from Corollary 5.4.

### 5.3 Application of the Commutative Algebra to Modular Forms

Let  $S(\mathbb{C})$  be a space of modular forms, e.g.  $S_k(\Gamma_1(N))$ . We only work with spaces that have a basis with coefficients in  $\mathbb{Z}$ . We denote by  $S(R)$  the corresponding space with coefficients in the ring  $R$ . Here the notion  $S(R)$  is the naive one via the standard

$q$ -expansion:  $S(R)$  is the set of  $R$ -linear combinations of the image of the  $\mathbb{Z}$ -basis in  $R[[q]]$  via the standard  $q$ -expansion. The space  $S(R)$  can also be characterised as follows. The Hecke operators  $T_n$  for  $n \in \mathbb{N}$  acting on  $S(\mathbb{C})$  generate a ring (a  $\mathbb{Z}$ -algebra), denoted  $\mathbb{T}$ , and we have the isomorphism

$$S(R) \cong \text{Hom}_{\mathbb{Z}}(\mathbb{T}, R).$$

Concretely, if  $\varphi \in \text{Hom}_{\mathbb{Z}}(\mathbb{T}, R)$ , then  $\sum_{n \geq 1} \varphi(T_n)q^n$  is a cusp form. Thus a  $\mathbb{Z}$ -basis of  $\mathbb{T}$  gives rise to a ‘dual basis’ of  $S(R)$ . We also speak of an ‘echelonised basis’.

By Proposition 5.5, we have the decompositions

$$\mathbb{T}_{\mathbb{Q}} := \mathbb{Q} \otimes_{\mathbb{Z}} \mathbb{T} \cong \prod_{[f]} \mathbb{T}_{[f]} \text{ and } S(\mathbb{Q}) \cong \bigoplus_{[f]} S_{[f]}(\mathbb{Q}),$$

where the product and the sum run over  $G_{\mathbb{Q}}$ -orbits of Hecke eigenforms. If the space  $S(\mathbb{C})$  is a newspace, then  $S_{[f]}(\mathbb{Q})$  is the set of forms with coefficients in  $\mathbb{Q}$  in the  $\mathbb{C}$ -span of all the  $G_{\mathbb{Q}}$ -conjugates of  $f$ . Concretely,  $S_{[f]}(\mathbb{Z})$  is the  $\mathbb{Z}$ -dual of the  $\mathbb{Z}$ -algebra generated by the Hecke operators  $T_n$  in  $\mathbb{T}_{[f]}$ . All Hecke operators acting on  $S_{[f]}(\mathbb{Z})$  are represented as matrices with  $\mathbb{Z}$ -entries.

We now consider  $\mathbb{T}_{\mathbb{Z}_{\ell}} = \mathbb{Z}_{\ell} \otimes_{\mathbb{Z}} \mathbb{T}$ . Then we have  $S(\mathbb{Z}_{\ell}) = \text{Hom}_{\mathbb{Z}_{\ell}}(\mathbb{T}_{\mathbb{Z}_{\ell}}, \mathbb{Z}_{\ell})$ . Importantly, again by Proposition 5.5, we have the decompositions

$$\mathbb{T}_{\mathbb{Z}_{\ell}} \cong \prod_{[\tilde{f}]} \mathbb{T}_{[\tilde{f}]} \text{ and } S(\mathbb{Z}_{\ell}) \cong \bigoplus_{[\tilde{f}]} S_{[\tilde{f}]}(\mathbb{Z}_{\ell}),$$

where the sum and the product run over the  $G_{\mathbb{F}_{\ell}}$ -orbits of Hecke eigenforms in  $S(\overline{\mathbb{F}_{\ell}})$ . These correspond to the maximal ideals of  $\mathbb{T}_{\mathbb{Z}_{\ell}}$ . We refer to the  $S_{[\tilde{f}]}(\mathbb{Z}_{\ell})$  either as  $\mathbb{Z}_{\ell}$ -orbits or as  $G_{\mathbb{F}_{\ell}}$ -orbits.

We are also interested in  $\mathbb{Q}_{\ell}$ -orbits of eigenforms inside a  $\mathbb{Z}_{\ell}$ -orbit. By Proposition 5.5,  $\mathbb{Q}_{\ell} \otimes_{\mathbb{Z}_{\ell}} S(\mathbb{Z}_{\ell}) = S(\mathbb{Q}_{\ell})$  breaks as a direct sum

$$S(\mathbb{Q}_{\ell}) \cong \bigoplus_{[\tilde{f}]} S_{[\tilde{f}]}(\mathbb{Q}_{\ell}),$$

where the sum runs over the  $\overline{\mathbb{Q}_{\ell}}$ -valued eigenforms up to  $G_{\mathbb{Q}_{\ell}}$ -conjugation. The fact that these  $G_{\mathbb{Q}_{\ell}}$ -orbits lie in a single  $\mathbb{Z}_{\ell}$ -orbit simply means that they are all congruent modulo a uniformiser.

### 5.4 Testing Weak Congruences

The second author has developed the MAGMA package WEAKCONG (see [25]), which has the purpose to compute whether Hecke eigenforms over  $\overline{\mathbb{Q}_{\ell}}$  belong to

given  $\mathbb{Z}_\ell$ -orbits of Hecke eigenforms modulo powers of  $\ell$  (or uniformisers). Here we briefly describe how it functions.

Let  $n_1, \dots, n_r$  be indices such that  $T_{n_1}, \dots, T_{n_r}$  form a basis of the Hecke algebra  $\mathbb{T}_{\mathbb{Z}_\ell}$  (which we may assume to be local by using the MAGMA package PADICALGEBRAS, see above). We speak of *basis indices*. These indices are computed via Nakayama’s lemma, i.e. by reducing the matrices to  $\mathbb{F}_\ell$ .

For any  $n$ , we have  $T_n = \sum_{i=1}^r a_{n,i} T_{n_i}$ ; in particular,  $a_{n_j,i} = \delta_{i,j}$ . For each  $i \in \{1, \dots, r\}$ , we define a cusp form  $f_i$  by specifying its coefficients as follows:

$$a_n(f_i) := a_{n,i}.$$

Then  $f_1, \dots, f_r$  form an  $R$ -basis of  $\text{Hom}_{\mathbb{Z}_\ell}(\mathbb{T}_{\mathbb{Z}_\ell}, R)$  for any  $\mathbb{Z}_\ell$ -algebra  $R$ . We call this basis *echelonised* because it is at the coefficients  $n_1, \dots, n_r$ . It is the dual basis with respect to the basis  $T_{n_1}, \dots, T_{n_r}$  of  $\mathbb{T}_{\mathbb{Z}_\ell}$ .

Furthermore, we compute one  $\overline{\mathbb{Q}_\ell}$ -eigenform for each  $\mathbb{Q}_\ell$ -orbit inside the given  $\mathbb{Z}_\ell$ -orbit. This is done via standard linear algebra over local fields, using both the new MAGMA command LocalField and the older implementation. If we find that a system of linear equations which mathematically must have a solution does not seem to have any, then we lower the precision until the desired solution exists. Thus, in this procedure generally some precision is lost.

Let  $g = \sum_{n \geq 1} b_n q^n \in S(\overline{\mathbb{Q}_\ell})$  be an eigenform in some level and weight. Let  $\mathcal{O}$  be the valuation ring of some finite extension of  $\mathbb{Q}_\ell$  that contains all coefficients  $b_n$  of  $g$ , and let  $\lambda$  be a uniformiser of  $\mathcal{O}$ . The main purpose of this package is to compute the maximum integer  $m$  such that  $g$  lies in a given  $\mathbb{Z}_\ell$ -orbit (some level and some weight) modulo  $\lambda^m$ .

Put  $h := g - \sum_{i=1}^r b_{n_i} f_i$ . We then have:

$$h \equiv 0 \pmod{\lambda^m} \Leftrightarrow \exists s_1, \dots, s_r \in \mathcal{O} : g \equiv \sum_{i=1}^r s_i f_i \pmod{\lambda^m}.$$

This equivalence is clear as the basis is echelonised, whence automatically  $s_i \equiv b_{n_i} \pmod{\lambda^m}$  for all  $i = 1, \dots, r$ . The desired highest exponent  $m$  can thus be computed as the minimum of the valuations of the coefficients of  $h$  up to the Sturm bound.

## 6 Database of Modular Form Orbits and Higher Congruences

The first author has created a PostgreSQL database containing data on  $\mathbb{Q}$ -,  $\mathbb{Q}_\ell$ - and  $\mathbb{Z}_\ell$ -orbits, as well as information on congruences modulo powers of  $\ell$ . We are currently planning to integrate parts of the database into the LMFDB.<sup>1</sup>

---

<sup>1</sup><http://lmfdb.org>.

## 6.1 Technical Features

In this section we describe the way our database is organised and what kind of data it contains. This will also highlight two important aspects of our approach:

- We do our best to avoid computing again data that are used more than once. This aims to speed up the process of computing the  $G_{\mathbb{Q}_\ell}$ -orbits. In order to do this we store a lot of useful information, even intermediate results, e.g. congruences with forms other than those that provide an optimal weight or level, even congruence of individual coefficients.
- We try to parallelise as much of the problem as possible. This also aims at speeding up the computation of congruences. This becomes especially handy when the coefficient fields of the forms that are compared become large.

We will come back to both of these features after the description of the database tables. We list them together with a brief description of the data each one holds.

1. **Modular form spaces over  $\mathbb{Q}$** : For every level and weight we store some useful information: The dimension of its Eisenstein subspace, old cuspidal subspace, new cuspidal subspace as well as the number of new Eisenstein  $\mathbb{Q}$ -Galois orbits and the number of newform  $\mathbb{Q}$ -Galois orbits.
2. **Bases of modular form spaces over  $\mathbb{Q}$** : Here we store the basis in terms of modular symbols for every space in the previous table. This in Magma readable format.
3. **Eigenforms over  $\mathbb{Q}$** : For every space over  $\mathbb{Q}$ , we store an entry for every Eisenstein and newform  $\mathbb{Q}$ -Galois orbit uniquely determined by its level, weight and orbit number.
4. **Hecke matrices over  $\mathbb{Z}$** : For each of the newform orbits in the previous table we store a list (up to a bound that can be increased as needed) of all the Hecke matrices acting on the  $\mathbb{Q}$ -subspace spanned by this orbit.
5. **Lattices**: For each of the newform orbits in the  $\mathbb{Q}$ -eigenforms table, we store a list of base change matrices that ensure the matrices in the table above, after base change, are with respect to the same basis.
6.  **$\ell$ -adic idempotents**: Given a newform from the  $\mathbb{Q}$ -eigenforms list and a prime number  $\ell$ , we store a list of idempotents which provide the decomposition of the corresponding  $\ell$ -adic Hecke algebra into local factors (see Proposition 5.5), their number and the  $\ell$ -adic precision that they were computed in.
7.  **$\mathbb{F}_\ell$ -Galois orbits**: For each entry in the table above (i.e. a list of idempotents), we store an  $\mathbb{Z}$ -integral basis for each of the components (indexed by the idempotents in this list) that the parent  $\mathbb{Q}$ -Galois orbit of newforms breaks into.
8.  **$\mathbb{Q}_\ell$ -Galois orbits of newforms**: For each  $\mathbb{Q}$ -Galois orbit of newforms and the prime  $\ell$ , we store the  $\mathbb{Q}_\ell$ -Galois orbit of newforms it decomposes into, along with the  $\ell$ -adic precision they were computed in.

These are the tables that provide a hierarchical organisation of the objects involved in the database and we tried to present it in a top to bottom fashion were an

entry in one of these table will be associated with many entries in the ones mentioned after it.

There are some auxiliary tables where all the congruence information is stored. We store everything down to congruences of individual pairs of coefficients. These are detailed catalogs of all meaningful congruences when it comes to level or weight lowering, weak or strong.

It is obvious that the comparison of two eigenvalues at a prime  $p$  is independent from the comparison of the ones at some other prime  $q$ . We thus run a multi-threaded application utilising as many CPU cores as possible where all threads compare a specific pair of eigenvalues each simultaneously. Let us stress here that the design of the database and the multi-threaded application is such that it allows us to utilise more than one server and/or personal computers to compute even more congruences simultaneously. Extra care has been taken to avoid overlapping of threads, i.e. two of those computing the same congruence, but we choose not to elaborate on these technical matters.

The current size of the database is 488GB. It contains 3906  $\mathbb{Q}$ -eigenforms, of level and weight up to 361 and 298 respectively (not of all possible combinations of course).

## 6.2 Accessibility

We have designed a basic web interface<sup>2</sup> for the database which currently allows one to query the database about the following:

1. Given a  $G_{\mathbb{Q}}$ -orbit  $[f]$  and a prime  $\ell$ , return  $G_{\mathbb{Q}_{\ell}}$ -orbits appearing in it.
2. Given a  $G_{\mathbb{Q}}$  orbit  $[f]$ , a prime  $\ell$  and a positive integer  $n$ , return the  $G_{\mathbb{Q}_{\ell}}$ -orbits that are congruent to the ones corresponding to  $[f]$  and  $\ell$  modulo  $\ell^n$  and are of the smallest weight possible, i.e. the answer to the strong weight lowering modulo  $\ell^n$  problem for  $[f]$ .
3. Given a  $G_{\mathbb{Q}}$ -orbit  $[f]$  and a prime  $\ell$ , return a list of downloadable files (one for each  $G_{\mathbb{Q}_{\ell}}$ -orbit) containing all the  $\ell$ -adic, prime-indexed Hecke polynomials (that are stored in the database) for each  $G_{\mathbb{Q}_{\ell}}$ -orbit.

## 6.3 Some Remarks on the Algorithms Used

We now describe how we computed the various orbits. Our algorithm is implemented in the MAGMA computer algebra system [4]. Assume as input a given level  $N$ , weight  $k$  and prime  $\ell$ .

---

<sup>2</sup>[http://math.uni.lu/~tsaknias/elladicdatabase\\_2.php](http://math.uni.lu/~tsaknias/elladicdatabase_2.php).

1. Compute the newspace of the cuspidal subspace of the modular symbols of level  $N$  and weight  $k$ . Decompose this subspace into irreducible Hecke modules. These correspond to  $G_{\mathbb{Q}}$ -orbits. This is done with standard MAGMA commands.
2. For a given irreducible Hecke module of the previous decomposition, compute the matrices for all operators  $T_n$  acting on it up to a sufficient bound  $B$ .
3. Use the package PADICALGEBRAS [24] to factor the completion of the Hecke algebra at  $\ell$  into local factors over  $\mathbb{Z}_{\ell}$ . Each of these factors corresponds to a  $G_{\mathbb{F}_{\ell}}$ -orbit. Project the matrices representing the  $T_n$ 's onto each of these local factors.
4. After tensoring with  $\mathbb{Q}_{\ell}$ , each of these  $G_{\mathbb{F}_{\ell}}$ -orbits is the sum of all the  $G_{\mathbb{Q}_{\ell}}$ -orbits admitting the same reduction mod  $\ell$ . For each such orbit, take the collection of projections of the Hecke matrices onto it computed in the previous step and decompose the corresponding  $\mathbb{Q}_{\ell}$ -vector space into simultaneous generalised eigenspaces by applying each operator successively. The resulting decomposition is the breaking of the corresponding  $G_{\mathbb{F}_{\ell}}$ -orbit into the  $G_{\mathbb{Q}_{\ell}}$ -ones that coincide mod  $\ell$ .

**Acknowledgements** The authors would like to thank Rajender Adibhatla, Sara Arias-de-Reyna, Gebhard Böckle, Frank Calegari, Imin Chen, Shaunak Deo, Frazer Jarvis, Ian Kiming, Ariel Pacetti, Nadim Ruston and many others for various discussions about topics on modular Galois representations modulo prime powers. They also thank Ken Ribet for having pointed out an inaccuracy in a previous version. Thanks are also due to the referee for a careful reading and useful suggestions. The second author thanks Gabi Nebe for having explained the simple algorithmic idempotent lifting (Eq. (4)) to him a long time ago.

This project was supported by the Luxembourg Research Fund (Fonds National de la Recherche Luxembourg) INTER/DFG/12/10/COMFGREP in the framework of the priority program 1489 of the Deutsche Forschungsgemeinschaft.

## References

1. R. Adibhatla, J. Manoharmayum, Higher congruence companion forms. *Acta Arith.* **156**(2), 159–175 (2012)
2. R. Adibhatla, P. Tsaknias, A characterisation of ordinary modular eigenforms with CM, in *Arithmetic and Geometry*. London Mathematical Society Lecture Note Series, vol. 420 (Cambridge University Press, Cambridge, 2015), pp. 24–35
3. G. Böckle, On the density of modular points in universal deformation spaces. *Am. J. Math.* **123**(5), 985–1007 (2001)
4. W. Bosma, J. Cannon, C. Playoust, The Magma algebra system. I. The user language. *J. Symb. Comput.* **24**(3–4), 235–265 (1997). *Computational algebra and number theory (London, 1993)*
5. K. Buzzard, Questions about slopes of modular forms. *Astérisque* **298**, 1–15 (2005). *Automorphic forms. I*
6. M. Camporino, A. Pacetti, Congruences between modular forms modulo prime powers (2013). arXiv:1312.4925
7. I. Chen, I. Kiming, G. Wiese, On modular Galois representations modulo prime powers. *Int. J. Number Theory* **9**(1), 91–113 (2013)
8. S.V. Deo, Structure of Hecke algebras of modular forms modulo  $p$ . *Algebra Number Theory* **11**(1), 1–38 (2017)



9. F. Diamond, M. Flach, L. Guo, The Tamagawa number conjecture of adjoint motives of modular forms. *Ann. Sci. Éc. Norm. Supér. (4)* **37**(5), 663–727 (2004)
10. N. Dummigan, Level-lowering for higher congruences of modular forms (2015). <http://neil-dummigan.staff.shef.ac.uk/levell8.pdf>
11. B. Edixhoven, The weight in Serre’s conjectures on modular forms. *Invent. Math.* **109**(3), 563–594 (1992)
12. M. Emerton, Local-global compatibility in the  $p$ -adic Langlands programme for  $GL_2/\mathbb{Q}$  (2011). <http://www.math.uchicago.edu/~emerton/pdffiles/lg.pdf>
13. C. Khare, R. Ramakrishna, Lifting torsion Galois representations. *Forum Math. Sigma* **3**, e14, 37 (2015)
14. C. Khare, J.P. Wintenberger, Serre’s modularity conjecture. I. *Invent. Math.* **178**(3), 485–504 (2009)
15. I. Kiming, N. Rustom, G. Wiese, On certain finiteness questions in the arithmetic of modular forms. *J. Lond. Math. Soc.* **94**(2), 479–502 (2016)
16. M. Kisin, The Fontaine-Mazur conjecture for  $GL_2$ . *J. Am. Math. Soc.* **22**(3), 641–690 (2009)
17. K.A. Ribet, On modular representations of  $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$  arising from modular forms. *Invent. Math.* **100**(2), 431–476 (1990)
18. K.A. Ribet, Raising the levels of modular representations, in *Séminaire de Théorie des Nombres, Paris 1987–88*. Progress in Mathematics, vol. 81 (Birkhäuser Boston, Boston, MA, 1990), pp. 259–271
19. J. Tilouine, Hecke algebras and the Gorenstein property, in *Modular Forms and Fermat’s Last Theorem (Boston, MA, 1995)* (Springer, New York, 1997), pp. 327–342
20. P. Tsaknias, On higher congruences of modular Galois representations, Ph.D. thesis, University of Sheffield, 2009
21. X.T.i. Ventosa, G. Wiese, Computing congruences of modular forms and Galois representations modulo prime powers, in *Arithmetic, Geometry, Cryptography and Coding Theory 2009*. Contemporary Mathematics, vol. 521 (American Mathematical Society, Providence, RI, 2010), pp. 145–166
22. G. Wiese, Multiplicities of Galois representations of weight one. *Algebra Number Theory* **1**(1), 67–85 (2007). With an appendix by Niko Naumann
23. G. Wiese, Magma package `ArtinAlgebras` (2008). <http://math.uni.lu/~wiese/programs/ArtinAlgebras>
24. G. Wiese, Magma package `pAdicAlgebras` (2014). <http://math.uni.lu/~wiese/programs/pAdicAlgebras>
25. G. Wiese, Magma package `WeakCong` (2016). <http://math.uni.lu/~wiese/programs/WeakCong>