

Quantum Jumps of Normal Polytopes

Winfried Bruns¹ · Joseph Gubeladze² ·
Mateusz Michałek^{3,4} 

Received: 8 June 2015 / Revised: 10 February 2016 / Accepted: 25 February 2016 /
Published online: 14 March 2016
© Springer Science+Business Media New York 2016

Abstract We introduce a partial order on the set of all normal polytopes in \mathbb{R}^d . This poset $\text{NPol}(d)$ is a natural discrete counterpart of the continuum of convex compact sets in \mathbb{R}^d , ordered by inclusion, and exhibits a remarkably rich combinatorial structure. We derive various arithmetic bounds on elementary relations in $\text{NPol}(d)$, called *quantum jumps*. The existence of extremal objects in $\text{NPol}(d)$ is a challenge of number theoretical flavor, leading to interesting classes of normal polytopes: minimal, maximal, spherical. Minimal elements in $\text{NPol}(5)$ have played a critical role in disproving various covering conjectures for normal polytopes in the 1990s. Here we report on the first examples of maximal elements in $\text{NPol}(4)$ and $\text{NPol}(5)$, found by a combination of the developed theory, random generation, and extensive computer search.

Keywords Lattice polytope · Normal polytope · Maximal polytope · Quantum jump

Mathematics Subject Classification Primary 52B20 · Secondary 11H06 · 52C07

Editor in Charge: Kenneth Clarkson

Winfried Bruns
wbruns@uos.de

Joseph Gubeladze
soso@sfsu.edu

Mateusz Michałek
wajcha2@poczta.onet.pl

¹ Institut für Mathematik, Universität Osnabrück, 49069 Osnabrück, Germany

² Department of Mathematics, San Francisco State University,
1600 Holloway Ave., San Francisco, CA 94132, USA

³ Institut für Mathematik, Freie Universität, Arnimallee 3, 14195 Berlin, Germany

⁴ Polish Academy of Sciences, ul. Śniadeckich 8, 00-956 Warsaw, Poland

1 Introduction

Normal polytopes are popular objects in combinatorial commutative algebra and toric algebraic geometry: they define the normal homogeneous monoid algebras [5, Chap. 2], [15, Chap. 7] and the projectively normal toric embeddings [5, Chap. 10], [10, Chap. 2.4]. The motivation for normal polytopes in this work is the following more basic observation: these lattice polytopes are natural discrete analogues of (continuous) convex polytopes and, more generally, convex compact sets in \mathbb{R}^d .

Attempts to understand the normality property of lattice polytopes in more intuitive geometric or integer programming terms date back from the late 1980s and 1990s; see Sect. 3.2. The counterexamples in [2, 3, 6] to several conjectures in that direction implicitly used a certain poset $\text{NPol}(d)$ of the normal polytopes in \mathbb{R}^d . We explicitly introduce this poset in Sect. 3.2. If normal polytopes or, rather, the sets of their lattice points are the discrete counterparts of convex compact sets in \mathbb{R}^d , then the poset $\text{NPol}(d)$ is the corresponding discrete analogue of the continuum of all such convex compact sets, ordered by inclusion. Put another way, $\text{NPol}(d)$ provides a formalism for the ‘discrete vs. continuous’ dichotomy in the context of convex geometry.

In this article we focus on the discrete structure of the poset $\text{NPol}(d)$, in particular the existence of maximal elements. In future studies we plan to examine the topological and finer geometric properties of the underlying order complex. One of our motivations is to study global properties of the family of normal polytopes, in analogy to moduli spaces—not only properties of particular polytopes. The present article follows a program that was sketched in [11].

Another aim is to set up a formalism for the search of special normal polytopes (or, equivalently, projective toric varieties) by a random walk on $\text{NPol}(d)$. Motivated from physics, one can consider various measures on the smallest possible changes of the polytope as analogs of potential. Such directed search proved useful in finding maximal polytopes in $\text{NPol}(4)$ and $\text{NPol}(5)$. As it turns out, random search can also hit maximal polytopes, notably in $\text{NPol}(4)$.

A pair (P, Q) of normal polytopes of equal dimension is called a *quantum jump* if $P \subset Q$ and Q has exactly one more lattice point than P . Here the word *quantum* refers to the smallest possible discrete change of a normal polytope and also points to random walks on $\text{NPol}(d)$: among all possible quantum jumps one chooses the ones according to an adopted strategy.

Quantum jumps define a partial order on the set of normal polytopes in which $P < Q$ if and only if there exists an ascending chain of quantum jumps that leads from P to Q . We consider the relation $<$ as the discrete analogue of the set theoretic inclusion between convex compact subsets of \mathbb{R}^d .

The extent of distortion of the continuum in the suggested discretization process is encoded in extremal elements of $\text{NPol}(d)$ and potential topological complexity of the geometric realization of $\text{NPol}(d)$: local and global properties of $\text{NPol}(d)$, respectively. It is not a priori clear that $\text{NPol}(d)$ exhibits any of these irregularities at all.

Explicit nontrivial (i.e., different from unimodular simplices) minimal elements, which were called *tight* polytopes in [3], have been known for quite a while. They exist in all dimensions ≥ 4 and special instances were crucial in disproving the unimodular

covering of normal polytopes or the integral Carathéodory property. The existence of nontrivial minimal elements in $\text{NPol}(3)$ is open.

Finding maximal elements is much more difficult but, as mentioned above, we have been successful in dimensions 4 and 5. There seems to be no way to construct higher dimensional maximal polytopes from a given one, but there is little doubt that maximal polytopes exist in all dimensions ≥ 4 . However, the existence in dimension 3 remains open.

Sections 2 and 3 recall basic notions and results for the study of normal polytopes.

In Sect. 4 we study the height of a lattice point z over a polytope P that, roughly speaking, counts the number of lattice layers between z and the facets of P that are visible from z . It is the natural measure for distance based on the lattice structure. We show that there is no bound on the height of quantum jumps that depends only on dimension. A very precise characterization of quantum jumps in dimension 3 is obtained at the end of Sect. 4.

Section 5 contains a sharp bound for quantum jumps in all dimensions. It is roughly proportional to the product of dimension and the lattice diameter of P that we call *width*. It shows that there are only finitely many jumps (P, Q) for fixed P and allows us to find them efficiently.

Section 6 is devoted to special normal polytopes defined by spheres and, more generally, ellipsoids. In particular we prove that all quantum jumps are infinitesimally close to the initial polytope relative to the size of the latter when the shape approximates a sphere with sufficient precision. The question on normality of the convex hulls of all lattice points in ellipsoids naturally arises. In dimension 3 we always have the normality, and our experiments did not lead to counterexamples in dimensions 4 and 5.

The final Sect. 7 describes our experimental approach to the existence of maximal polytopes. The main difficulty was to find a criterion that lets us choose terminating ascending chains with some positive probability. The maximal polytopes were eventually found by a combination of random generation and directed search. The computational power of Normaliz [7] has been proved invaluable for these experiments.

2 Basic Notions

The sets of nonnegative integer and real numbers are denoted, respectively, by \mathbb{Z}_+ and \mathbb{R}_+ . The Euclidean norm of a vector $v \in \mathbb{R}^d$ is $\|v\|$. We write $\mathbf{e}_1, \dots, \mathbf{e}_d$ for the standard basis vectors of \mathbb{R}^d . A *point configuration* is a finite subset $X \subset \mathbb{Z}^d$. For a subset $X \subset \mathbb{R}^d$ we set $L(X) = X \cap \mathbb{Z}^d$.

For a more detailed account and the proofs for the statements in this section we refer the reader to [5, Chaps. 1 and 2].

2.1 Polytopes

An *affine subspace* of a Euclidean space is a shifted linear subspace. An affine map between two affine spaces is a map that respects *barycentric coordinates*. Equivalently,

an affine map is the restriction of a linear map between the ambient vector spaces followed by a parallel translation.

Two subsets $X, Y \subset \mathbb{R}^d$ are called *unimodularly equivalent* if there is an *integral-affine isomorphism* $X \rightarrow Y$, i.e., there is an affine isomorphism $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$, mapping \mathbb{Z}^d bijectively to itself, such that $f(X) = Y$.

A *closed affine half-space* $H^+ \subset \mathbb{R}^d$ is a subset of the form

$$\{x \in \mathbb{R}^d : h(x) \geq 0\} \subset \mathbb{R}^d,$$

where $h : \mathbb{R}^d \rightarrow \mathbb{R}$ is a non-zero affine map and $H = h^{-1}(0)$ is the bounding hyperplane. When h is a linear map, the half-space H^+ and the hyperplane H will be called *linear* or *homogeneous*.

An affine subspace A is *rational* if it is spanned by points in \mathbb{Q}^d . The bounding hyperplane H of a *rational* half-space H^+ is given in the form

$$H = \{x \in \mathbb{R}^d : \alpha_1 x_1 + \cdots + \alpha_d x_d + \beta \geq 0\} \quad (1)$$

with a rational number β and coprime integers $\alpha_1, \dots, \alpha_d \in \mathbb{Z}$.

The *affine hull* $\text{Aff}(X)$ of a subset $X \subset \mathbb{R}^d$ is the smallest affine subspace of \mathbb{R}^d containing X . The *convex hull* of X will be denoted by $\text{conv}(X)$.

All considered polytopes are assumed to be *convex*, i.e., a polytope is the convex hull of a finite subset $X \subset \mathbb{R}^d$. Equivalently, a polytope P is a bounded intersection of finitely many closed affine half-spaces: $P = \bigcap_{i=1}^n H_i^+$. The *faces* of P are the intersections of the form $P \cap H$ where $H^+ \subset \mathbb{R}^d$ is a closed affine half-space containing P . Also P is a face of itself. The *vertices* of P are the 0-dimensional faces of P , and the $(d-1)$ -dimensional faces of P are called the *facets*. The vertex set of P will be denoted by $\text{vert}(P)$. A *simplex* is a polytope whose number of vertices exceeds the dimension of the polytope by one.

A full-dimensional polytope $P \subset \mathbb{R}^d$ admits a unique representation $P = \bigcap_{i=1}^n H_i^+$, where the $H_i^+ \subset \mathbb{R}^d$ are closed affine half-spaces and $\dim(P \cap H_i) = d-1$, $i = 1, \dots, n$. We call this representation the *irreducible representation* of P .

For every polytope $P \subset \mathbb{R}^d$, its interior and the boundary with respect to $\text{Aff}(P)$ will be denoted, respectively, by $\text{int}(P)$ and $\partial P = P \setminus \text{int}(P)$.

A *lattice polytope* $P \subset \mathbb{R}^d$ is a polytope whose vertices are *lattice points*, i.e., elements of \mathbb{Z}^d . A *rational polytope* has its vertices in \mathbb{Q}^d .

A lattice simplex is called *unimodular* if the edge vectors at some (equivalently, any) vertex define a part of a basis of \mathbb{Z}^d .

Let X be a subset of \mathbb{R}^d such that $\text{Aff}(X) = \text{Aff}(\text{L}(\text{Aff}(X)))$, for example a lattice polytope. Then we can assign to X a *normalized volume*: it is the measure on $\text{Aff}(X)$ in which a unimodular simplex in $\text{Aff}(X)$ has volume 1. If $\text{Aff}(X) = \mathbb{R}^d$, the normalized volume equals $d!$ times the Euclidean d -volume and will be denoted by vol_d . Note that the normalized volume is invariant under integral-affine transformations, but not under Euclidean isometries of \mathbb{R}^d if $\dim \text{Aff}(X) < d$.

For a full-dimensional lattice polytope $P \subset \mathbb{R}^d$, the normalized volume $\text{vol}_d(P)$ is a natural number. Moreover, P is a unimodular simplex if and only if $\text{vol}_d(P) = 1$.

Let $P \subset \mathbb{R}^d$ and $Q \subset \mathbb{R}^e$ be two lattice polytopes. A $(d + e + 1)$ -dimensional lattice polytope R is a *join* of P and Q if it is unimodularly equivalent to the *standard join*, defined by

$$\text{join}(P, Q) = \text{conv}((P, \underbrace{0, \dots, 0}_{e+1}), (\underbrace{0, \dots, 0}_d, 1, Q)) \subset \mathbb{R}^{d+e+1}.$$

Thus unimodular simplices are joins of lattice points.

A d -dimensional polytope will be called a d -polytope. For a lattice d -polytope $P \subset \mathbb{R}^d$ and a facet $F \subset P$, there exists a unique affine map, the *facet-height function*, $\text{ht}_F : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $\text{ht}_F(\mathbb{Z}^d) = \mathbb{Z}$, $\text{ht}_F(F) = 0$, and $\text{ht}_F(P) \subset \mathbb{R}_+$. If $\text{Aff}(F)$ bounds the half-space H^+ , then with the notation introduced in (1),

$$\text{ht}_F(x) = h(x) = \alpha_1 x_1 + \dots + \alpha_d x_d + \beta.$$

A lattice polytope $P \subset \mathbb{R}^d$ is a *unimodular pyramid* over Q if P is a join of the polytope Q and a lattice point. The polytope Q serves as the base and the additional point serves as the apex of the pyramid P . If $\dim P = d$, then P is a unimodular pyramid over Q if Q is a facet of P and $L(P) \setminus L(Q)$ is a single point that has height 1 over Q .

2.2 Cones and Hilbert Bases

A *conical set* $C \subset \mathbb{R}^d$ is a subset of \mathbb{R}^d for which $\lambda x + \mu y \in C$ whenever $x, y \in C$ and $\lambda, \mu \in \mathbb{R}_+$. A *cone* means a *finitely generated, rational, and pointed conical set*. That is, a cone $C \subset \mathbb{R}^d$ is a subset such that $C = \mathbb{R}_+ x_1 + \dots + \mathbb{R}_+ x_n$ for some $x_1, \dots, x_n \in \mathbb{Z}^d$ and there is no nonzero element $x \in C$ with $-x \in C$.

For a cone C , the additive submonoid $L(C) \subset \mathbb{Z}^d$ has a unique minimal generating set, which is the set of *indecomposable elements* of the (additive) submonoid $L(C) \subset \mathbb{Z}^d$. This set is called the *Hilbert basis* of the cone C and denoted by $\text{Hilb}(C)$.

A cone C is called *unimodular* if $\text{Hilb}(C)$ is a part of a basis of \mathbb{Z}^d . Equivalently, C is unimodular if $\text{Hilb}(C)$ is a linearly independent set.

The *faces* of a cone $C \subset \mathbb{R}^d$ are the intersections of type $H \cap C$ for a homogeneous half-space $H^+ \subset \mathbb{R}^d$, containing C . Also C is a face of C . Among the faces of C we have the *extremal rays* and *facets*.

A non-zero lattice vector $x \in \mathbb{Z}^d, x \neq 0$, is called *primitive* if it is the generator of the monoid $L(\mathbb{R}_+ v) \cong \mathbb{Z}_+$. This holds if and only if the coordinates of x are coprime. The primitive lattice vectors in the extremal rays of a cone C are called the *extremal generators* of C . They belong to $\text{Hilb}(C)$.

Assume $d > 0$. For a facet F of a d -cone $C \subset \mathbb{R}^d$ there is a unique linear map, the *facet-height function*, $\text{ht}_F : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $\text{ht}_F(F) = 0, \text{ht}_F(\mathbb{Z}^d) = \mathbb{Z}$, and $\text{ht}_F(C) = \mathbb{R}_+$. The last two equalities are equivalent to the condition that $\text{ht}_F(L(C)) = \mathbb{Z}_+$.

Every lattice polytope $P \subset \mathbb{R}^d$ defines the cone $C(P) \subset \mathbb{R}^{d+1}$ as follows. One embeds P into \mathbb{R}^{d+1} by identifying $x \in P$ with the point $(x, 1) \in P \times \{1\}$ and

chooses $C(P)$ as the union of the rays originating from 0 and passing through a point of $P \times \{1\}$. Then $C(P)$ is generated by the vectors $(x, 1)$, $x \in \text{vert}(P)$, and these vectors are the extremal generators of $C(P)$.

We call the process of attaching the $(d + 1)$ st coordinate 1 the *homogenization of coordinates*. In order to facilitate it, we set $x' = (x, 1)$ for $x \in \mathbb{R}^d$. For $y \in \mathbb{R}^{d+1}$ the $(d + 1)$ st coordinate is called its *degree*.

2.3 Simplicial Cones and Simplices

A cone of dimension d is *simplicial* if it has exactly d extremal rays, or, equivalently, its extremal generators are linearly independent.

Let $v_1, \dots, v_d \in \mathbb{Z}^d$ be linearly independent and let $C = \mathbb{R}_+v_1 + \dots + \mathbb{R}_+v_d$ be the simplicial cone spanned by them. Moreover, let $U = \mathbb{Z}v_1 + \dots + \mathbb{Z}v_d$ be the sublattice and $M = \mathbb{Z}_+v_1 + \dots + \mathbb{Z}_+v_d$ be the affine monoid generated by v_1, \dots, v_d . The group U acts on \mathbb{R}^d by translations, and a fundamental domain of this action is

$$\text{par}(v_1, \dots, v_d) = \{a_1v_1 + \dots + a_dv_d : 0 \leq a_i < 1, i = 1, \dots, d\}.$$

The set

$$\text{Lpar}(v_1, \dots, v_d) = \text{par}(v_1, \dots, v_d) \cap \mathbb{Z}^d$$

of the lattice points represents the orbits of U in \mathbb{Z}^d , or, in other words, the residue classes of \mathbb{Z}^d modulo U . Based on these observations one easily proves [5, Prop. 2.43]:

Proposition 2.1 *With the notation just introduced, the following hold:*

- (a) $E = \text{Lpar}(v_1, \dots, v_d)$ is a system of generators of the M -module $\mathbb{L}(C)$ (in the self-explanatory terminology);
- (b) $(x + M) \cap (y + M) = \emptyset$ for $x, y \in E, x \neq y$;
- (c) $\#E = [\mathbb{Z}^d : U]$;
- (d) $\text{Hilb}(C) \subset \{v_1, \dots, v_d\} \cup E$.

Since we are usually interested in the cone C spanned by v_1, \dots, v_d , we can and will assume that v_1, \dots, v_d are the extremal generators of C . The semi-open parallelotope $\text{par}(v_1, \dots, v_d)$ is called the *basic parallelotope* of C , and we call

$$\mu(C) = [\mathbb{Z}^d : U]$$

the *multiplicity* of C . One has $\mu(C) = \text{vol}(S)$ where S is the *basic simplex* with vertices $0, v_1, \dots, v_d$.

Let F be a facet of C and let v be the extremal generator opposite to F . Then we have

$$\mu(C) = \text{ht}_F(v)\mu(F);$$

see [5, Prop. 3.9]. This formula reflects a stratification of $\text{Lpar}(v_1, \dots, v_d)$:

Proposition 2.2 $\text{Lpar}(v_1, \dots, v_d)$ contains exactly $\mu(F)$ lattice points of height j over F , $j = 0, \dots, \text{ht}_F(v) - 1$.

Proof Let $m = \text{ht}_F(v)$. The group homomorphism

$$p : \mathbb{Z}^d \rightarrow \mathbb{Z}/m\mathbb{Z}, \quad p(x) = \text{ht}_F(x) \pmod{m\mathbb{Z}},$$

factors through \mathbb{Z}^d/U . Thus each class in \mathbb{Z}^d/U decomposes into $\mu(C)/m$ classes that have the same height over F modulo m . For each $j = 0, \dots, m - 1$ we must have $\mu(F)$ such classes. □

Let Δ be a lattice d -simplex. Then $C(\Delta)$ is a simplicial cone, and we may write $\text{par}(\Delta)$ for the basic parallelotope of $C(\Delta)$ and $\text{Lpar}(\Delta)$ for its lattice points. Note that

$$\mu(C(\Delta)) = \text{vol}_d(\Delta).$$

The nonzero points in $\text{Lpar}(\Delta)$ are stratified into layers of constant degree. Clearly the maximum degree in $\text{Lpar}(\Delta)$ is at most d and the minimum nonzero degree is at least 1. There seems to be no complete description of this stratification, but in special cases one has more information.

We are particularly interested in the case in which Δ is *empty*: the only lattice points of Δ are its vertices. In this case one can say a little more:

Lemma 2.3 *Suppose Δ is an empty simplex. Then $\text{Lpar}(\Delta)$ has no points in degrees 1 and d . In particular, if $\dim \Delta = 3$, then the nonzero elements of $\text{Lpar}(\Delta)$ live in degree 2.*

Proof That there are no points in degree 1 is the definition of ‘empty’, and that there are no points of degree d follows if one applies the point reflection $\rho : \mathbb{R}^d \rightarrow \mathbb{R}^d$ at the midpoint of $\text{par}(\Delta)$

$$\rho(x) = (v'_1 + \dots + v'_{d+1}) - x,$$

where v_1, \dots, v_{d+1} are the vertices of Δ . □

3 Normal Polytopes

In this section we introduce the class of normal polytopes, recall basic facts and several explicit families. We also define an order structure on the set of normal polytopes in \mathbb{R}^d .

3.1 Normal polytopes

Definition 3.1 A polytope $P \subset \mathbb{R}^d$ is *normal* if, for all $c \in \mathbb{N}$, one has

$$\text{L}(cP) = \{x_1 + \dots + x_c : x_1, \dots, x_c \in \text{L}(P)\}.$$

The continuous version of Definition 3.1 is the equality

$$cX = \underbrace{X + \cdots + X}_c, \quad c \in \mathbb{N},$$

satisfied for any convex subset $X \subset \mathbb{R}^d$, where the left hand side is the c -th dilation and the right hand side the Minkowski sum of c copies of X .

This version of ‘normal’ is used in many sources. It is called ‘integrally closed’ in [5, Chap. 2] where ‘normal’ was used for a weaker property, namely the normality of the monoid $M(P)$ defined below. Further, in [9, p. 4] the authors distinguish the *Integral Decomposition Property* (IDP) from normality. The first one is referring to the ambient lattice, as we do, the second one to the lattice generated by integral points of the polytope. In this paper we prefer the more succinct ‘normal’ to the more algebraically oriented ‘integrally closed’.

By Pick’s 19th century theorem [5, Cor. 2.54] all lattice polygons are normal. But in high dimensions, starting with 3, the normal polytopes form a small portion of all lattice polytopes.

The next theorem encapsulates some basic facts about normal polytopes, the parts (a), (b), (c) and (d) being direct consequences of Definition 3.1 and the parts (e), (f), (g), (h) and (i) being proved in [3, 3.1], [5, 2.81, 2.57], [12], and [16], respectively.

Theorem 3.2 (a) *A lattice polytope that is unimodularly equivalent to a normal polytope, is normal.*

- (b) *If P is a union of normal polytopes then P is normal.*
- (c) *If P is normal then every face of P is normal.*
- (d) *Cartesian products and joins of normal polytopes are normal. Unimodular pyramids over normal polytopes are normal; in particular, unimodular simplices are normal.*
- (e) *If P is normal then for every complete flag of faces*

$$\mathbb{F} : F_0 \subset F_1 \subset \cdots \subset F_{d-1} \subset P, \quad d = \dim P,$$

there exists an \mathbb{F} -incident unimodular d -dimensional simplex $\Delta \subset P$, i.e.,

$$\dim(\Delta \cap F_i) = i, \quad i = 1, \dots, d - 1.$$

- (f) *For any lattice polytope P , the dilated polytopes cP are normal as soon as $c \geq \dim P - 1$.*
- (g) *Lattice parallelotopes (not necessarily rectangular) of any dimension are normal.*
- (h) *A full dimensional lattice polytope $P \subset \mathbb{R}^d$ is normal if the primitive normal vectors to the facets of P form a subset of a root system, whose irreducible summands are of type A, B, C, or D.*
- (i) *A lattice polytope P is normal if each of its edges contains at least $4d(d + 1) + 1$ lattice points. When P is a lattice simplex this bound can be lowered to $d(d + 1) + 1$.*

Call a lattice polytope $P \subset \mathbb{R}^d$ *smooth* if the primitive edge vectors at every vertex $v \in P$ form a part of a basis of \mathbb{Z}^d . The terminology is explained by the observation

that the projective toric variety of a lattice polytope is smooth if and only if P is smooth; see [5, p. 371]. *Oda’s question* asks whether smooth lattice polytopes are normal; see, for instance, [13] and the references therein.

The recent extensive treatment [14] of unimodular triangulations for various classes of normal polytopes presents the state of the art in the field.

For every lattice polytope R the set $L(R)' = \{x' = (x, 1) : x \in L(P)\}$ generates an affine submonoid $M(R)$ of \mathbb{Z}^{d+1} , and the normality of R (as used in this paper) is equivalent to $\text{Hilb}(C(R)) = L(R)'$, or, in other words, to the equality

$$\sum_{x \in L(R)} \mathbb{Z}_+(x, 1) = L(C(R)) \quad (\subset \mathbb{Z}^{d+1}),$$

of graded affine monoids, where the degree is chosen as introduced in Sect. 2.2. We set $\overline{M}(R) = L(C(R))$. In algebraic terms, it is the integral closure of $M(R)$ in \mathbb{Z}^{d+1} (and in general is larger than the normalization of $M(R)$, the integral closure in the sublattice generated by $M(R)$). See [5, Chap. 2] for an extensive discussion. The k -th degree components of the monoids just introduced will be denoted by $M(R)_k$ and $\overline{M}(R)_k$, respectively.

As a consequence of Lemma 2.3 one can show (see [5, Thm. 2.52]):

Lemma 3.3 *Let P be a lattice polytope. Then all elements of $\text{Hilb}(P)$ have degree $\leq d - 1$.*

3.2 The poset $\text{NPol}(d)$

Definition 3.4 The partially ordered set $\text{NPol}(d)$ is the set of normal polytopes in \mathbb{R}^d , ordered as follows: $P < Q$ if and only if there exists a finite sequence of normal polytopes of the form

$$\begin{aligned} P &= P_0 \subset \cdots \subset P_{n-1} \subset P_n = Q, \\ \#L(P_i) &= \#L(P_{i-1}) + 1, \quad i = 1, \dots, n. \end{aligned} \tag{2}$$

One easily observes that, in the sequence above, if $\dim(P_i) = \dim(P_{i-1}) + 1$ then P_i is a unimodular pyramid over P_{i-1} .

The importance of the poset $\text{NPol}(d)$ is explained as follows. In the late 1980s, in an attempt to give a more succinct characterization of the normal point configurations, the following two distinguished conjectures were proposed in [17], of which the second one had been already asked as a question in [8]:

Unimodular cover (UC) A lattice polytope P is normal if and only if P is the union of unimodular simplices.

Integral carathéodory property (ICP) A lattice polytope $P \subset \mathbb{R}^d$ is normal if and only if for an arbitrary natural number $c \in \mathbb{N}$ and an arbitrary integer point $z \in L(cP)$ there exist integer points $x_1, \dots, x_{d+1} \in L(P)$ and integer numbers $a_1, \dots, a_{d+1} \in \mathbb{Z}_+$ with $z = a_1x_1 + \cdots + a_{d+1}x_{d+1}$ and $a_1 + \cdots + a_{d+1} = c$.

Informally, (UC) says that the continuity, modeled by normal polytopes, is piecewise by nature, resulting from the constituent unimodular simplices. The (ICP) is an arithmetic version of (UC). (The original conjectures were formulated for general cones C that are not necessarily of the form $C(P)$ for normal P , using $\text{Hilb}(C)$ instead of $L(P)$.)

The first indication of the relevance of the poset $\text{NPol}(d)$ was the following observation in [3], without introducing the poset structure in $\text{NPol}(d)$ explicitly: for both conjectures (UC) and (ICP) it is critical to check their validity on the *minimal elements* of the poset $\text{NPol}(d)$. Minimal polytopes are called *tight polytopes* in [3]. The very existence of tight polytopes is not quite intuitive: computational evidence shows that many descending sequences of the type (2) reversed lead to complete erasure of the initial normal polytope. However, tight polytopes have popped up in dimensions 4 and higher, and the larger the dimension the more frequently so. A counterexample to (UC) was finally found in [3]. In [6] it was shown that the example also disproves (ICP). That (ICP) is strictly weaker than (UC) was shown in [2], where the same strategy of shrinking normal polytope was used with the following important refinement: if a shrinking process halts at a minimal counterexample to (UC) then chances are that, somewhere along the descent path, the stronger property (UC) is lost before (ICP). This can be viewed as the second indication of the relevance of the poset $\text{NPol}(d)$ in understanding the normality property.

Every normal 3-dimensional polytope that is comparable with a unimodular simplex within $\text{NPol}(3)$, is covered by unimodular simplices, as follows immediately from [3, Lem. 2.2]. In particular, the lack of nontrivial minimal elements, if true, would imply that (UC) holds for normal polytopes of dimension 3. (UC) is open in dimension 4 as well, but there are nontrivial minimal elements in $\text{NPol}(4)$ so that the same argument cannot work.

One easily generates infinitely many higher dimensional minimal normal polytopes from a single one. In fact, for any minimal element $P \in \text{NPol}(d)$ and any element $Q \in \text{NPol}(e)$ the product polytope $P \times Q$ is a minimal element of $\text{NPol}(d+e)$. However, as this paper shows, the situation is very different for maximal normal polytopes – so far we have been able to find only a handful (up to unimodular equivalence) maximal normal polytopes in $\text{NPol}(4)$ and $\text{NPol}(5)$.

4 Lattice Stratifications and Quantum Jumps

In this section we single out the elementary relations in the poset $\text{NPol}(d)$ between two full dimensional polytopes as the main object of our study and show that, already in dimension 3, their arithmetic picture is quite involved.

4.1 Large Empty Layers Around Polytopes

In the following it will be very convenient to say that a facet F of a d -polytope $P \subset \mathbb{R}^d$ is *visible* from $x \in \mathbb{R}^d \setminus P$ if for every $y \in F$ the line segment $[x, y]$ intersects P exactly in y . Note that $\text{ht}_F(x) < 0$ if and only if F is visible from x because the points in P have nonnegative height by convention.

Definition 4.1 Let $P \subset \mathbb{R}^d$ be a full-dimensional lattice polytope.

(a) For a point $z \in \mathbb{Z}^d \setminus P$ the *height* $\text{ht}_P(z)$ of z over P is defined by

$$\text{ht}_P(z) = \max(-\text{ht}_F(z) : F \subset P \text{ a facet, visible from } z).$$

The points in P have height 0 over P .

(b) For $j \in \mathbb{Z}_+$, the polytope P^{-j} is defined by

$$P^{-j} = \{x : \text{ht}_P(x) \leq j\}.$$

(c) The *lattice stratum of height j* around P is the subset $\partial P^{-j} \cap \mathbb{Z}^d$.

(d) The *width* of P with respect to a facet $F \subset P$ and the *absolute width* are defined as follows

$$\begin{aligned} \text{width}_F P &= \max(\text{ht}_F(x) : x \in P), \\ \text{width } P &= \max(\text{width}_F P : F \subset P \text{ a facet}). \end{aligned}$$

The term ‘stratum’ above is justified: one has the stratification

$$\mathbb{Z}^d \setminus P = \bigcup_{j=1}^{\infty} (\mathbb{L}(\partial P^{-j})). \tag{3}$$

Informally, $\mathbb{L}(\partial P^{-j})$ consists of lattice points outside of the polytope P on ‘lattice distance’ j from P . The polytopes P^{-j} are rational polytopes, but usually *not* lattice polytopes. In fact, as we will see below, $\mathbb{L}(\partial P^{-j})$ can very well be empty.

Remark 4.2 If P is a normal polytope then it defines a normal projective toric variety X together with a very ample line bundle \mathcal{L} providing a projectively normal embedding. The points $\mathbb{L}(P)$ correspond to a basis of global sections $H^0(X, \mathcal{L})$ [5, Chap. 10.B][10, Chap. 4.3]. Let K be the canonical (Weil) divisor on X . Points of $\mathbb{L}(P^{-j})$ correspond to a basis of global sections $H^0(X, \mathcal{L} - jK)$. In particular, if some strata are empty this is equivalent to the fact that by adding the (effective) divisor $-K$ we do not obtain any new global sections.

Many of our results may be interpreted in this language. For example, Theorem 4.3 below implies that there is no lower bound on j , even for normal toric threefolds X with a very ample line bundle \mathcal{L} , guaranteeing the inequality $h^0(X, \mathcal{L}) < h^0(X, \mathcal{L} - jK)$.

Two easy observations:

(i) If a point $z \in \mathbb{Z}^d \setminus P$ is at the smallest possible positive height above P , then

$$\mathbb{L}(\text{conv}(P, z)) = \mathbb{L}(P) \cup \{z\}.$$

(ii) $\text{width}_F P$ is always attained at a vertex of P not in F . It is positive since $P \neq F$.

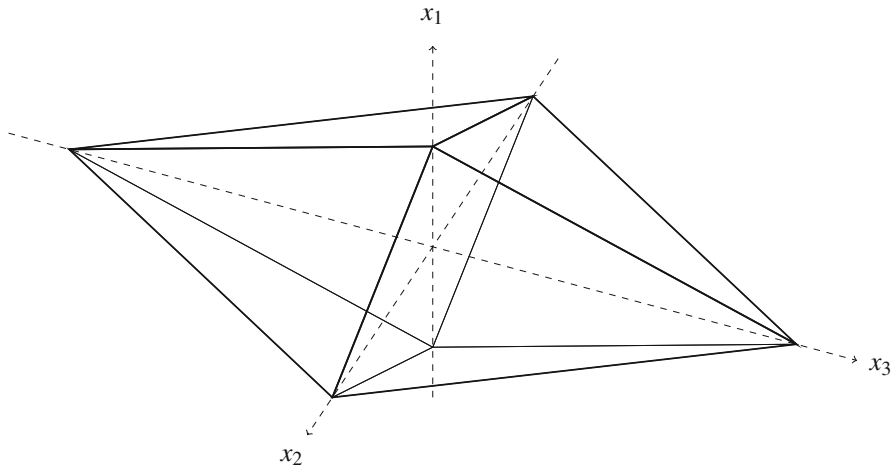


Fig. 1 The polytope of Theorem 4.3

Without any constraints on the lattice point z and the lattice polytope P , except the requirement $L(\text{conv}(P, z)) = L(P) \cup \{z\}$, there is no upper bound for the heights $\text{ht}_P(z)$, not even for normal 3-polytopes P . The simplest such example is the unit tetrahedron $P = \text{conv}(0, \mathbf{e}_1, \mathbf{e}_2, -\mathbf{e}_3)$ and the points $z_k = \mathbf{e}_1 + \mathbf{e}_2 + k\mathbf{e}_3$: we have $\text{ht}_P(z_k) = k$. Although one should note that none of the lattice strata around a unimodular simplex of any dimension is empty.

Our next result shows that there is no dimensionally uniform upper bound for the height of the *lowest* lattice points above lattice polytopes, not even in the class of normal polytopes, and not even in dimension 3.

Theorem 4.3 *There is a sequence of normal 3-polytopes $P_k \subset \mathbb{R}^3$ and lattice points $z_k \in \mathbb{Z}^d \setminus P_k, k \in \mathbb{N}$, such that for every index k we have:*

- (a) width $P_k = 2k(k + 1)(k^2 + k + 1)$,
- (b) the lattice strata around P_k up to height $k - 1$ ($\approx \sqrt[4]{\frac{1}{2}}$ width P_k as $k \rightarrow \infty$) are all empty.
- (c) $\text{ht}_P(z_k) = k$ and $\text{conv}(P_k, z_k) \in \text{NPol}(d)$.

As mentioned in observation (i) above, in part (c) we also have $L(\text{conv}(P_k, z_k) \setminus P_k) = \{z_k\}$.

Proof Consider the cross-polytopes (Fig. 1)

$$P_k = \text{conv}(\pm k\mathbf{e}_1, \pm(k + 1)\mathbf{e}_2, \pm(k^2 + k + 1)\mathbf{e}_3) \subset \mathbb{R}^3, \quad k \in \mathbb{N}.$$

Each P_k is the union of eight congruent copies of the rectangular tetrahedron

$$\Delta_k = \text{conv}(0, k\mathbf{e}_1, (k + 1)\mathbf{e}_2, (k^2 + k + 1)\mathbf{e}_3) \subset \mathbb{R}^3.$$

For every k the tetrahedron

$$\text{conv}(0, k\mathbf{e}_1, (k + 1)\mathbf{e}_2, \mathbf{e}_3), \quad k \in \mathbb{N},$$

is a unimodular pyramid over the right triangle $\text{conv}(0, k\mathbf{e}_1, (k + 1)\mathbf{e}_2)$ and so is normal. Therefore, by [4, Thm. 1.6], the tetrahedra

$$\Delta_k = \text{conv}(0, k\mathbf{e}_1, (k + 1)\mathbf{e}_2, (k(k + 1) + 1)\mathbf{e}_3)$$

are normal for all $k \in \mathbb{N}$. By Theorem 3.2(a), the cross-polytopes P_k are normal for all $k \in \mathbb{N}$.

To complete the proof of (b), because of reasons of symmetry between the coordinate orthants in \mathbb{R}^3 , it is enough to show that

$$\min(-\text{ht}_{F_k}(z) : z \in \mathbb{Z}_+^d \setminus \Delta_k) \geq k, \quad k \in \mathbb{N}, \tag{4}$$

where F_k is the facet of Δ_k opposite to 0. The corresponding height function $\text{ht}_{F_k} : \mathbb{Z}^3 \rightarrow \mathbb{Z}$ is given by

$$\begin{aligned} (\xi_1, \xi_2, \xi_3) \mapsto & -(k + 1)(k^2 + k + 1)\xi_1 - k(k^2 + k + 1)\xi_2 \\ & - k(k + 1)\xi_3 + k(k + 1)(k^2 + k + 1). \end{aligned}$$

The lattice point $z_k = (k - 1, 1, 1)$ belongs to Δ_k and satisfies $\text{ht}_{F_k}(z_k) = 1$. Because $(1, -1, -1) = k\mathbf{e}_1 - z_k$ and $k\mathbf{e}_1 \in \text{vert}(\Delta_k)$, the parallel translates

$$(j, -j, -j) + F_k \subset \mathbb{R}^3, \quad j \in \mathbb{N},$$

of the triangle F_k live in the planes that are defined correspondingly by $\text{ht}_{F_k}(-) = -j$. Since these are unimodular triangles, we have

$$\begin{aligned} & ((j, -j, -j) + \text{Aff}(F_k)) \cap \mathbb{Z}^3 \\ & = (k + j, -j, -j) + \mathbb{Z}(-k, k + 1, 0) + \mathbb{Z}(-k, 0, k^2 + k + 1). \end{aligned}$$

Therefore, for every natural number k , the inequality (4) is equivalent to the system of equalities

$$\begin{aligned} & ((k + j, -j, -j) + \mathbb{Z}(-k, k + 1, 0) + \mathbb{Z}(-k, 0, k^2 + k + 1)) \cap \mathbb{R}_+^3 = \emptyset, \\ & j = 1, \dots, k - 1. \end{aligned} \tag{5}$$

For the mentioned range of j , if the components of the triple

$$(k + j, -j, -j) + a(-k, k + 1, 0) + b(-k, 0, k^2 + k + 1)$$

are positive for some $a, b \in \mathbb{Z}$ then $-j < 0$ implies $a, b > 0$ and $k + j \leq 2k - 1$ implies $a + b < 2$. This contradiction proves (5) and, hence, (b).

Because we know the height functions of the facets, (a) follows easily.

To prove (c), we put $z_k = (0, 1, k^2 + 1)$. One can immediately check that z_k has height k and that only the facets in the orthants of \mathbb{R}^3 with sign patterns $+++$ and $-++$ are visible from z_k . By symmetry it is enough to prove that $\text{conv}(\Delta_k, z_k)$ is normal.

According to Theorem 4.11 below, it is enough to exhibit lattice points in Δ_k that have heights $1, \dots, k - 1$ over $F: y_j = (k - j, j, j), j = 1, \dots, k - 1$, is in Δ_k and has height j over F . □

While there is no upper bound on the number of strata around P that do not contain lattice points, we have the following uniform bound depending only on width P .

Proposition 4.4 *For all natural numbers d and every lattice polytope $P \subset \mathbb{R}^d$ there is a lattice point $z \notin P$ such that $\text{ht}_P(z) \leq \text{width } P$. In particular, there is a point $z \notin P$ such that $L(\text{conv}(P, z)) = L(P) \cup \{z\}$ and $\text{ht}_P(z) \leq \text{width } P$.*

Proof Choose an edge of P and consider its two endpoints u, v . Then $z = u + 2(v - u) \notin P$ since it lies on the straight line through u and v and does not belong to the edge.

For any face F of P one has

$$\text{ht}_F(z) = \text{ht}_F(2v - u) = 2 \text{ht}_F(v) - \text{ht}_F(u) \geq - \text{ht}_F(u) \geq - \text{width}_F P.$$

(Note that the coefficients in $2v - u$ sum to 1.) □

4.2 Quantum Jumps

Definition 4.5 (a) A *minimal (resp. maximal) polytope* is a minimal (resp. maximal) element of $\text{NPol}(d)$.

- (b) A pair of d -polytopes (P, Q) in $\text{NPol}(d)$ with $P < Q$ and $\#L(Q) = \#L(P) + 1$ will be called a *quantum jump from P* , or simply a *jump (of dimension d)*. If z is the additional lattice point in Q , we will say that z is a *quantum jump over P* .
- (c) The *height* of a jump (P, Q) is defined to be the height over P of the only lattice point in $Q \setminus P$; we denote this number by $\text{ht}(P, Q)$.

Remark 4.6 There are several natural measures one can associate to a jump (P, Q) , of which the height is one. Examples include the *volume* $\mathbf{v}(P, Q)$, equal to $\text{vol}_d(Q \setminus P)$, and the *base* $\mathbf{b}(P, Q)$, equal to the sum of $(d - 1)$ -volumes of the facets $F \subset P$, visible from the vertex of Q outside P , normalized correspondingly with respect to the lattices $\mathbb{Z}^d \cap \text{Aff}(F)$. Both these measures are natural numbers. If $z = \text{vert}(Q) \setminus P$ then the equality

$$\mathbf{v}(P, Q) = \mathbf{b}(P, Q) \text{ht}(P, Q)$$

is equivalent to the condition that z is on same height with respect to any facet $F \subset P$, visible from z .

The chains in $\text{NPol}(d)$, consisting of jumps that maximize the volumes at each step, lead to normal polytopes in which the lattice points are relatively rarefied. One

may think that such ascending chains have potential to lead to maximal elements in $\text{NPol}(d)$: after all, the lattice points in a normal polytope are meant to be regularly distributed. The reality is not as simple though; see Sect. 7.

Example 4.7 (Dark vertices of polygons) The order in $\text{NPol}(2)$ coincides with the inclusion order on the lattice polygons in \mathbb{R}^d and, consequently, the order complex of $\text{NPol}(2)$ is topologically trivial, i.e., contractible. In fact, all lattice polygons are normal (Theorem 3.2(f)) and if $P \subsetneq Q$ in $\text{NPol}(2)$ and v is a vertex of Q , not in P , then (Q_1, Q) is a jump, where $Q_1 = \text{conv}(L(Q) \setminus \{v\})$. Iterating the process, we find a finite descending sequence

$$\begin{aligned} Q &= Q_0 \supset Q_1 \supset \dots \supset Q_n = P, \\ (Q_{i+1}, Q_i) &\text{ a jump for every } i, \\ n &= \#L(Q) - \#L(P). \end{aligned}$$

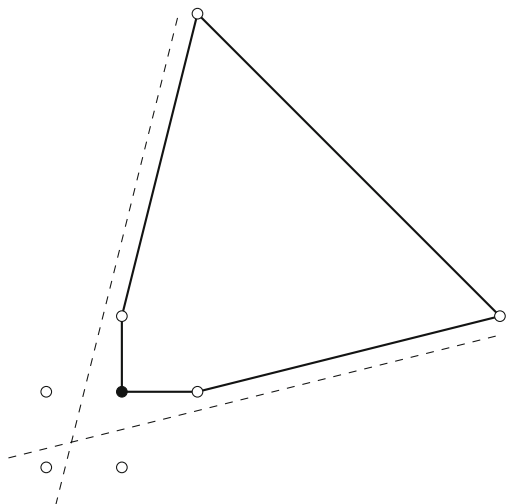
Although no polygon can be maximal, constructing jumps from a given polygon is not quite straightforward. Let us say that a vertex v of P is *dark* if there is no jump z over P such that v is visible (or ‘illuminated’) from z . The origin is a dark vertex of the polygon with vertices

$$(0, 0), (0, 1), (1, 0), (5, 1), (1, 5).$$

In fact, every jump z from which $(0, 0)$ is visible must have one coordinate equal to -1 . But each of these points has height > 1 over one of the other facets. See Fig. 2; the dashed lines are the lines of height -1 over the facets parallel to them.

If we add $(20, 5)$ as a further vertex, then $(0, 0)$ and $(1, 0)$ will become dark. This construction can be continued an arbitrary number of steps: if $v_{-2}, v_{-1}, v_0, \dots, v_{n+2}$

Fig. 2 A polygon with a dark vertex



have been constructed such that v_0, \dots, v_n are dark, choose the next vertex v_{n+3} at height 5 over $[v_n, v_{n+1}]$ and height 1 over $[v_{n+1}, v_{n+2}]$. This will lead to a polygon with an arbitrary number of adjacent dark vertices at which the corner cones are unimodular. By the standard technique of toric desingularization [10, Chap. 11], we can change the polygon to a smooth one keeping the adjacent dark vertices untouched and still dark.

If the construction is continued infinitely many times, it yields an unbounded polygon P with all dark vertices and unimodular corner cones. Equivalently, all lattice points outside P have infinite height over it.

Example 4.8 (The poset $\text{NPol}(3)$) As shown above, the order in $\text{NPol}(2)$ is simply the inclusion order. Although it is open whether extreme elements apart from unimodular simplices exist in $\text{NPol}(3)$, the example below, found by computer search, shows that the inclusion order is finer than the one induced by jumps.

Consider the 3-polytope P with vertices:

$$(0, 0, 2), (0, 0, 1), (0, 1, 3), (1, 0, 0), (2, 1, 2), (1, 2, 1).$$

It is a normal lattice polytope with two additional lattice points: $(1, 1, 2), (1, 1, 1)$.

Removing either the first *or* the second vertex and taking the convex hull of the other lattice points in P yields a nonnormal polytope. However, if Q is the convex hull of all lattice points in P apart from the first *and* the second vertex, then Q is a normal polytope. Clearly Q is inside P , but $Q \not\subset P$. More examples similar to the one above can be found.

Example 4.9 (Unimodular simplices) Any two unimodular d -simplices in \mathbb{R}^d belong to same connected component of $\text{NPol}(d)$. In fact, let Δ_1 and $\Delta_2 \subset \mathbb{R}^d$ be two unimodular simplices and $v \in \Delta_1$ and $w \in \Delta_2$ be vertices. Choose a lattice broken line $[v_1, v_2, \dots, v_k]$ in \mathbb{R}^d , where $v = v_1, w = v_k$, and $v_{j+1} - v_j \in \mathbb{Z}^d$ is a primitive vector for every $j = 1, \dots, k - 1$. Then we have $v < \Delta_1, w < \Delta_2$, and $v_j, v_{j+1} < [v_j, v_{j+1}]$.

If, in addition, $\dim \Delta_1 = \dim \Delta_2 = d$, then the two simplices can be even connected by quantum jumps. To this end, we first reduce the general case to the case when Δ_1 and Δ_2 share a vertex. Let $[v_1, \dots, v_k]$ be a broken line as above. There exist unimodular d -simplices T_1, \dots, T_{k-1} such that $v_j, v_{j+1} \in \text{vert}(T_j)$ for every $j = 1, \dots, k - 1$. In particular, it is enough to connect by quantum jumps the simplices in each of the doublets

$$\{\Delta_1, T_1\}, \{T_1, T_2\}, \dots, \{T_{k-1}, \Delta_2\}.$$

But the simplices in each of these pairs share a vertex. At this point without loss of generality we can assume that 0 is a vertex of Δ_1 and Δ_2 . Let $x_1, \dots, x_d \in \Delta_1$ and $y_1, \dots, y_d \in \Delta_2$ be the other vertices. Consider the following two matrices in $GL_d(\mathbb{Z})$: $A = [x_1 \dots x_d]$ and $B = [y_1 \dots y_d]$. By an appropriate enumeration of the nonzero vertices, we can further assume $\det A = \det B = 1$. Then, because \mathbb{Z} is a Euclidean domain, every integer matrix with determinant 1 is a product of elementary matrices: $SL_d(\mathbb{Z}) = E_d(\mathbb{Z})$. Equivalently, we can transform $\{x_1, \dots, x_d\}$ into $\{y_1, \dots, y_d\}$ by

a series of successive elementary transformations of the following two types:

$$\begin{aligned} \{z_1, \dots, z_p, \dots, z_q, \dots, z_d\} &\longrightarrow \{z_1, \dots, z_p, \dots, z_q + z_p, \dots, z_d\}, \\ \{z_1, \dots, z_p, \dots, z_q, \dots, z_d\} &\longrightarrow \{z_1, \dots, z_p, \dots, z_q - z_p, \dots, z_d\}. \end{aligned}$$

In particular, it is enough to show that, for a basis $\{z_1, \dots, z_d\} \subset \mathbb{Z}^d$ and two natural numbers $1 \leq p \neq q \leq d$, the unimodular simplices in each of the pairs

$$\begin{aligned} &\{\text{conv}(0, z_1, \dots, z_d), \text{conv}(0, z_1, \dots, z_p, \dots, z_q + z_p, \dots, z_d)\}, \\ &\{\text{conv}(0, z_1, \dots, z_d), \text{conv}(0, z_1, \dots, z_p, \dots, z_q - z_p, \dots, z_d)\} \end{aligned}$$

can be connected by quantum jumps. For simplicity of notation we can assume $p = 1$ and $q = 2$. Now the desired jumps are provided by:

$$\begin{aligned} \text{conv}(0, z_1, z_2, \dots, z_d) < \text{conv}(0, z_1, z_1 + z_2, z_2, \dots, z_d) > \text{conv}(0, z_1, z_1 + z_2, \dots, z_d), \\ \text{conv}(0, z_1, z_2, \dots, z_d) < \text{conv}(0, z_1, z_1 - z_2, z_2, \dots, z_d) > \text{conv}(0, z_1, z_1 - z_2, \dots, z_d), \end{aligned}$$

where the middle polytopes are normal, each being the union of two unimodular simplices:

$$\begin{aligned} &\text{conv}(0, z_1, z_1 + z_2, z_2, \dots, z_d) \\ &\quad = \text{conv}(0, z_1, z_2, \dots, z_d) \cup \text{conv}(z_1, z_1 + z_2, z_2, \dots, z_d), \\ &\text{conv}(0, z_1, z_1 - z_2, z_2, \dots, z_d) \\ &\quad = \text{conv}(0, z_1, z_2, \dots, z_d) \cup \text{conv}(0, z_2 - z_1, z_2, \dots, z_d). \end{aligned}$$

Below, in Theorems 4.11, 5.1, and 6.1, we will give useful criteria for a pair of lattice polytopes to be a quantum jump. In dimension 2 the situation is very simple.

Proposition 4.10 *Let P be a normal polytope.*

- (a) *If z is a height 1 lattice point over P , it is a quantum jump. In particular, the first lattice stratum around any maximal polytope is empty.*
- (b) *If $\dim P \leq 2$ then every quantum jump over P has height 1.*

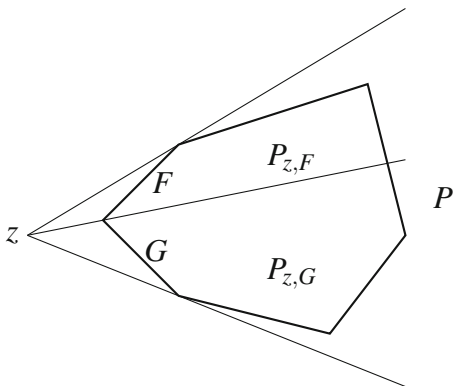
Proof (a) Clearly there are no additional lattice points in $Q = \text{conv}(P, z)$. Let F be a facet of P that is visible from z . Then F is normal by Theorem 3.2(c), and $\text{conv}(F, z)$, being a unimodular pyramid over F , is normal by Theorem 3.2(d). Thus Q is normal by Theorem 3.2(b).

(b) This is obvious in dimension 1. In dimension 2, let F be a facet of P that is visible from z , and let Δ be a unimodular line segment in F . Then $\text{conv}(\Delta, z)$ is an empty triangle and therefore unimodular. But this implies $\text{ht}_F(z) = -1$. □

4.3 Quantum Jumps in Dimension 3

In dimension 3 we have a rather precise description of quantum jumps $(P, \text{conv}(P, z))$, which uses the subdivision of P according to the rays emerging from z and passing

Fig. 3 The subdivision of P by facets visible from z



through a facet F visible from z : we set

$$P_{z,F} = \{x \in P : [x, z] \cap F \neq \emptyset\};$$

see Fig. 3.

Theorem 4.11 *Let $P \subset Q$ be lattice 3-polytopes such that P is normal and $\#\mathbf{L}(Q) = \#\mathbf{L}(P) + 1$. Let z be the additional lattice point in Q . Then the following are equivalent:*

- (a) z is a quantum jump over P .
- (b) For each facet F of P that is visible from z , the polytope $P_{z,F}$ contains at least (equivalently: exactly) $\mu(F)$ lattice points y such that $\text{ht}_F(y) = j$, $j = 1, \dots, \text{ht}_F(z) - 1$.

Proof Let F be a facet of P that is visible from z . Since F has dimension 2 it has a unimodular triangulation Σ . Let Δ be a triangle in Σ . For (b) \implies (a) it is enough to show that all degree 2 lattice points in $C(Q) \setminus C(P)$ are reducible (Lemma 3.3).

Let y be such a point. Since the tetrahedra $\text{conv}(\Delta, z)$ form a triangulation of $\text{conv}(F, z)$ there are two possibilities for y :

- (i) y is in the boundary of one (or more) cones $C(\Delta, z)$;
- (ii) y is in the interior of exactly one such cone.

In case (i) y is reducible since the facets of $\text{conv}(\Delta, z)$ are unimodular: they are empty triangles. Thus the Hilbert basis of such a facet lives in degree 1, and the degree 1 lattice points different from z' are of the form x' with $x \in P$.

In case (ii) we have $y \in \text{Lpar}(\Delta, z)$. Observe that there is exactly one point in $\text{Lpar}(\Delta, z)$ that has height j , $j = 1, \dots, m - 1$, over the facet $\Delta \subset \text{conv}(\Delta, z)$, where $m = \text{ht}_F(z)$ (Proposition 2.2). We want to show that $y = u' + z'$ for a lattice point u of height $m - j$ in $P_{z,F}$.

It is enough to show that

$$\bigcup_{\Sigma} \text{Lpar}(\Delta, z) = \{w' + z' : w \text{ a lattice point of height } m - j \text{ in } P_{z,F}\}. \quad (6)$$

Equality (6) follows if we show that every point $w' + z'$, where w' is as in (6), is in the interior of $\text{par}(\Delta, z)$ for one of the triangles Δ .

Clearly $w' + z'$ is outside $C(P)$ since it has negative height over F (considered as a facet of P). But it is in $C(Q)$. So one of the alternatives (i) or (ii) applies. Since the lattice points satisfying (i) are sums $v' + z'$ with v a vertex of Δ for some Δ and $w \notin F$, the alternative (i) is excluded. So (ii) applies, and we get indeed the desired $\mu(F)$ points — one in each $\text{Lpar}(\Delta, z)$.

Similarly, for (a) \implies (b), for each Δ we consider the point of height j and degree 2 in $\text{Lpar}(\Delta, z)$. As Q is normal, each such point must be a sum of two homogenized points in $\text{L}(Q)$, one of which has to be equal to z' . All the other points must be different, belong to $P_{z,F}$, and have height $m - j$ over F . □

Remark 4.12 Theorem 4.11 can of course be used to analyze the jumps over specific polytopes. For example, let $P = 2\Delta_{pq}$ where Δ_{pq} is the empty 3-simplex spanned by $0, \mathbf{e}_1, \mathbf{e}_3, q\mathbf{e}_1 + p\mathbf{e}_2 + \mathbf{e}_3, 1 \leq q \leq p - 1, p, q$ coprime. Then P has facets of multiplicity 4, but for each facet F only a single lattice point of height 1 over F . Thus a quantum jump over P must have height 1 (and such exists). It is an old result of White [18] that all empty 3-simplices are unimodularly equivalent to the Δ_{pq} .

Remark 4.13 All normal 3-polytopes P that have been encountered in our experiments, millions of them, have the following remarkable property: every point in the lowest nonempty stratum over P is a jump. On the other hand, the jumps in $\text{NPol}(3)$ need not be confined to the lowest nonempty stratum.

This changes completely in dimension 4. There exist 4-polytopes over which there is no jump at all (see Sect. 7), but there are examples where jumps exist and none of them belongs to the lowest nonempty stratum.

Despite all the information on $\text{NPol}(3)$ at hand, we do not know whether there are maximal normal 3-polytopes (or nontrivial minimal elements). In one special case we can provide the answer.

Proposition 4.14 *There are no simplices that are maximal elements of $\text{NPol}(3)$.*

Proof Let $S \in \text{NPol}(3)$ be a simplex with vertices v_0, v_1, v_2, v_3 and $F = \text{conv}(v_1, v_2, v_3)$. Among the lattice points not in S , from which the only visible facet is F , let z minimize $|\text{ht}_F(z)|$. Say $\text{ht}_F(z) = -k$. We claim that z is a quantum jump. We have $\text{L}(\text{conv}(S, z)) = \text{L}(S) \cup \{z\}$. Because F is the only visible facet, we have $S_{z,F} = S$, using notation as in Theorem 4.11. Hence, by the mentioned theorem, we only have to check that S contains $\mu(F)$ lattice points of height j over F for $j = 1, \dots, k - 1$.

Note that there are no lattice points in S with height $\text{ht}_F(v_0) - 1, \dots, \text{ht}_F(v_0) - k + 1$. Indeed, if such a point existed, we could consider the ray starting at v_0 passing through that point. The first point on that ray outside S would contradict the choice of z . In particular $\text{ht}_F(v_0) \geq k$. Moreover, in view of the inversion map

$$\text{Lpar}(S) \rightarrow \mathbb{Z}^4, \quad m \rightarrow v'_0 + v'_1 + v'_2 + v'_3 - m,$$

this implies that there are no points of degree three and of height j over F for $j = 1, \dots, k - 1$ in $\text{Lpar}(S)$. Hence, all height j points must appear in degree two and one.

However, if such a point q of degree two existed, then the only facet visible from the point

$$w = v'_1 + v'_2 + v'_3 - q = (v'_1 + v'_2 + v'_3 + v'_0 - q) - v'_0$$

would be F , and $\text{ht}_F(w) = -j$, which would contradict the choice of z . We thus conclude that the height j points must appear in degree one. But there are exactly $\mu(F)$ such points by Proposition 2.2. \square

As it turns out, already in dimension 4 there are maximal normal simplices, see Sect. 7.

5 Bounding Quantum Jumps in $\text{NPol}(d)$

In this section we derive a bound for the heights of quantum jumps in all dimensions and show that this bound is sharp.

We begin with a criterion for a quantum jump.

Theorem 5.1 *Let $P \subset Q$ be lattice d -polytopes such that P is normal. Suppose that $L(Q) = L(P) \cup \{z\}$. For every facet F of P that is visible from z , let Σ_F be a triangulation of F . Then the following are equivalent:*

- (a) Q is normal.
- (b) For each facet F of P that is visible from z and every $(d - 1)$ -simplex $\Delta \in \Sigma_F$ one has

$$y - z' \in C(P)$$

for every $y \in \text{Lpar}(\Delta, z)$ with $\text{ht}_F(y) < 0$.

Proof Suppose that Q is normal and let y be one of the points as in (b). Then $y \in C(Q) \setminus C(P)$. On the other hand, the Hilbert basis of $C(Q)$ is given by the vectors x' , $x \in L(P)$, and z' . So z' must appear in a representation of y as a sum of Hilbert basis elements. The ray from z' towards y leaves the simplicial cone $C(\Delta, z)$ through the facet $C(\Delta)$ and thus passes through $C(F)$. Since $y - z' \in C(Q)$ lies on this ray and has positive height over F , it must be in $C(P)$.

For the converse we observe that the simplices $\text{conv}(\Delta, z)$, $\Delta \in \Sigma_F$, are a triangulation of $\text{conv}(F, z)$. Then the union of $\text{Hilb}(C(P))$, z' and the $\text{Lpar}(\Delta, z) \setminus C(P)$, where Δ ranges over the $(d - 1)$ -simplices in Σ_F , contains a generating set of the monoid $L(C(Q))$. But (b) implies that all the lattice points in $\text{Lpar}(\Delta, z) \setminus C(P)$ are reducible. \square

The bound on all quantum jumps over a polytope $P \in \text{NPol}(d)$ in Theorem 5.3 below can be also derived from Theorem 5.1. However, we present an independent proof which uses a weaker condition than normality, related to the s.c. very ampleness of lattice polytopes.

Definition 5.2 A lattice polytope $R \subset \mathbb{R}^d$ is *very ample* if $\text{Hilb}(\mathbb{R}_+(R - v)) \subset R - v$ for every vertex v of R .

Normality implies very ampleness but not conversely; very ample polytopes define the normal projective toric varieties and very ample line bundles on them, which also explains the name; a lattice polytope $R \subset \mathbb{R}^d$ is very ample if and only if the complement $\overline{M}(R) \setminus M(R)$ for the monoids introduced in Sect. 3.1 is finite. For these and other generalities see [1, Sect. 2].

We have already seen in Theorem 4.3 that the situation drastically changes from dimension 2 to 3: there is no uniform limit on the number of empty strata for all $P \in \text{NPol}(3)$. For fixed dimension and width there is however such a bound (even after relaxing the normality condition).

Theorem 5.3 *Let $P \subset \mathbb{R}^d$ be a (not necessarily very ample) lattice d -polytope and let z be a point in \mathbb{Z}^d outside P . If $\text{Hilb}(\mathbb{R}_+(P - z)) \subset P - z$ then*

$$|\text{ht}_F(z)| \leq 1 + (d - 2) \text{width}_F P \tag{7}$$

for every facet F of P that is visible from z . In particular, if the polytope $\text{conv}(P, z)$ is very ample and $\mathbb{L}(\text{conv}(P, z)) = \mathbb{L}(P) \cup \{z\}$, then $|\text{ht}_F(z)|$ satisfies the bound (7).

Proof By applying the parallel translation by $-z$, we can assume $z = 0$. Set $R = \text{conv}(P, 0)$. Let F be a facet of P , visible from 0. By Theorem 3.2(f), the dilated polytope $(d - 1)\text{conv}(F, 0)$ is normal. Hence, by Theorem 3.2(e), there exists a lattice point $x \in (d - 1)\text{conv}(F, 0)$ on lattice height 1 above the facet $(d - 1)F \subset (d - 1)\text{conv}(F, 0)$.

Because $\text{Hilb}(\mathbb{R}_+P) \subset P$, the point x is a positive integral linear combination of lattice points of P . However, x cannot be the sum of $(d - 1)$ or more such points because the ht_F -value of the sum will be at least $(d - 2)|\text{ht}_F(0)|$, whereas $\text{ht}_F(x) = (d - 2)\text{ht}_F(0) - 1$.

In particular, x is the sum of at most $(d - 2)$ points from $\mathbb{L}(P)$. The largest ht_F -value of such a sum is $\text{width}_F P + (d - 3)(\text{width}_F P + |\text{ht}_F(0)|)$, forcing

$$\begin{aligned} \text{ht}_F(x) = (d - 2)|\text{ht}_F(0)| - 1 &\leq \text{width}_F P + (d - 3)(\text{width}_F P + |\text{ht}_F(0)|) \\ \implies |\text{ht}_F(0)| &\leq 1 + (d - 2) \text{width}_F P. \end{aligned}$$

□

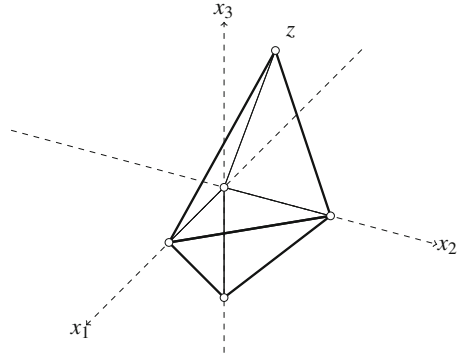
Remark 5.4 One should note that in the special case when $(P, \text{conv}(P, z))$ is a jump, Theorem 5.1 contains information beyond the bound in Theorem 5.3: the multiplicity of F also plays an essential role. We have already observed this in Remark 4.12. Furthermore, if $\text{conv}(P, z)$ is normal but P is not, then one can show based on Theorem 5.1 that the bound in Theorem 5.3 can be improved to $|\text{ht}_F(z)| \leq (d - 2) \text{width}_F P$.

We will see below that the bound in Theorem 5.3 cannot be improved, not even for quantum jumps of any dimension d .

As a consequence the number of lattice points that are candidates for quantum jumps over a polytope P is bounded, and the set of candidates can be efficiently described: the candidates are contained in the set

$$\mathbb{L}(P^{-1-(d-2)\text{width } P}) \setminus P.$$

Fig. 4 The polytope of Theorem 5.5 for $d = 3, w = 1$



This is the basis of our experiments with quantum jumps that helped us to find maximal elements in $\text{NPol}(d)$ for $d = 4$ and $d = 5$.

Our next theorem shows that the bound in Theorem 5.3 is sharp even for normal polytopes (Fig. 4).

Theorem 5.5 *For every natural number $d \geq 2$ and $w \geq 1$ there exists a jump (P, Q) of dimension d satisfying the following conditions:*

- (a) *The vertex of Q , not in P , is visible from exactly one facet $F \subset P$,*
- (b) *$\text{width}_F P = w$,*
- (c) *$\text{ht}(P, Q) = (d - 2)w + 1$.*

Proof There is nothing to show for $d = 2$. Therefore we assume $d \geq 3$.

We choose the polytope P to be spanned by the vertices $0, \mathbf{e}_1, \dots, \mathbf{e}_{d-1}$ and $-w\mathbf{e}_d$. It is the top element of the unique chain of length $w - 1$ in $\text{NPol}(d)$, starting with the unimodular simplex $\text{conv}(0, \mathbf{e}_1, \dots, \mathbf{e}_{d-1}, -\mathbf{e}_d)$ and finishing with P . In particular, P is normal. Over the ‘horizontal’ facet F spanned by 0 and the $\mathbf{e}_i, i \leq d - 1$, it has width w .

Let

$$z = (1, \dots, 1, (d - 2)w + 1) = \mathbf{e}_1 + \dots + \mathbf{e}_{d-1} + ((d - 2)w + 1)\mathbf{e}_d.$$

It is easy to check that z is the only additional lattice point in $Q = \text{conv}(P, z)$, that it has height $(d - 2)w + 1$ over F , and that F is the only facet of P that is visible from F .

The critical issue is the normality of Q . For the application of Theorem 5.1 it is advisable to use homogenized coordinates in \mathbb{R}^{d+1} , as usual indicated by ‘.

We claim that the nonzero points in $\text{Lpar}(F, z)$ are given by

$$y_k = u(\mathbf{e}'_1 + \dots + \mathbf{e}'_{d-1}) + vz' + t0',$$

where

$$\begin{aligned}
 v &= \frac{k}{w(d-2)+1}, & k &= 1, \dots, w(d-2), \\
 u &= 1 - v, \\
 t &= \lceil s \rceil - s, & s &= (d-1)u + v.
 \end{aligned}$$

In fact, since F is unimodular and $\text{vol}_d(\text{conv}(F, z)) = (d-2)w + 1$, there are exactly $w(d-2)$ points in $\text{Lpar}(F, z)$, one on height k above $C(F)$ for each $k = 1, \dots, w(d-2)$ (Proposition 2.2). Hence the indicated values of v are as above. The sum $u + v$ must be integer and $0 \leq u < 1$, which motivates the value of u . The last coordinates of the points y_k are integers and, simultaneously, $0 \leq t < 1$, yielding the indicated values for t .

We have

$$y_k = (1, \dots, 1, k, h_k), \quad k = 1, \dots, (d-2)w.$$

For the difference $y_{(d-2)w+1-k} - z'$ one obtains

$$\begin{aligned}
 (0, \dots, 0, -k, 1), & \quad k = 1, \dots, w, \\
 (0, \dots, 0, -k, 2), & \quad k = w + 1, \dots, 2w, \\
 & \quad \vdots \\
 (0, \dots, 0, -k, d-2), & \quad k = (d-3)w + 1, \dots, (d-2)w,
 \end{aligned}$$

where the $(d+1)$ -st coordinates on the left are computed by the formula

$$h_k - 1 = \lceil s \rceil - 1 = \left\lceil \frac{(d-2)(w+k)+1}{(d-2)w+1} \right\rceil - 1.$$

All these points lie in $C(P)$, i.e., after dehomogenization with respect to the last coordinate we get points in P . □

Remark 5.6 Theorems 4.11 and 5.5 rely on the exact knowledge of the distribution of the numbers $\text{ht}_F(x)$ in the critical areas relative to $\text{deg } x$.

In the proof of Theorem 5.5 there is only a single facet F visible from z and all facets of $\text{conv}(F, z)$ are not only empty, but even unimodular. This follows from the fact that all nonzero elements of $\text{Lpar}(F, z)$ are in the interior. But even under these ‘optimal’ conditions it seems difficult to find a transparent generalization of Theorem 4.11 to higher dimensions.

It is instructive to compute the heights of the elements of $\text{Lpar}(F, z)$ over the other facets of $\text{conv}(F, z)$ from the proof of Theorem 5.5 for $d = 4, w = 2$. Over F the degree 2 elements have heights 3 and 4, and the degree 3 elements have heights 1 and 2 (and the height 1 element lets us reach the upper bound). Over the other facets the height distributions are 1, 2 in degree 2, vs. 3, 4 in degree 3 (three facets) and 3, 1 in degree 2 vs. 2, 4 in degree 3.

This example shows that one cannot predict a priori the distribution of heights over a facet of an empty 4-simplex, not even if all facets are unimodular.

6 Spherical Polytopes

Throughout this section we fix a natural number $d \geq 2$.

Under certain constrains on the shapes of the normal polytopes we can derive more stringent bounds on the heights of quantum jumps than in Theorem 5.3. More precisely, in this section we show that for asymptotically spherical polytopes the heights of jumps become infinitesimally small compared to the widths. This also naturally leads to interesting number theoretical questions.

6.1 Asymptotically Infinitesimal Jumps

Below we will need the following criterion for quantum jumps, which is reminiscent of a dehomogenized version of Theorem 5.1:

Theorem 6.1 *Let $P \in \text{NPol}(d)$ with $0 \notin P$ and $Q = \text{conv}(P, 0)$. Then the following conditions are equivalent:*

- (a) (P, Q) is a jump,
- (b) $L(kQ \setminus ((k - 1)Q \cup kP)) = \emptyset$ for all $k \in \mathbb{N}$,
- (c) $L(kQ \setminus ((k - 1)Q \cup kP)) = \emptyset$ for $k = 1, \dots, d - 1$.

Proof In the following we use the monoids $M(Q)$ and $\overline{M}(Q)$ and their k -th degree components $M(Q)_k$ and $\overline{M}(Q)_k$, introduced in Section 3.1.

(a) \implies (b) Consider $x \in L(kQ)$. By normality, $x = \sum_{i=1}^k q_i$ for $q_i \in L(Q)$. If there exists $q_i = 0$, we may omit it in the sum, hence $x \in L((k - 1)Q)$. Otherwise all $q_i \in L(P)$, hence $x \in L(kP)$.

(b) \implies (c) is obvious.

(c) \implies (a) Assume (P, Q) is not a jump. Let k be the smallest natural number for which $M(Q)_k \subsetneq \overline{M}(Q)_k$ (notation as above). Since P is normal and $k \geq 2$, we must have

$$L(kQ \setminus ((k - 1)Q \cup kP)) \neq \emptyset.$$

Therefore, (c) implies that the monoids $M(Q)$ and $\overline{M}(Q)$ coincide up to degrees $d - 1$. But then, in view of Lemma 3.3, the two monoids are equal—a contradiction. \square

The closed d -ball in \mathbb{R}^d with radius r and centered at $z \in \mathbb{R}^d$ will be denoted by $B(z, r)$.

Theorem 6.2 *Let $P_i \in \text{NPol}(d)$, $z_i \in \mathbb{R}^d$, and r_i, ε_i be positive real numbers, where $i \in \mathbb{N}$. Assume $\lim_{i \rightarrow \infty} r_i = \infty$ and $B(z_i, r_i - \varepsilon_i) \subset P_i \subset B(z_i, r_i + \varepsilon_i)$ for all $i \gg 0$.*

(a) *If $\lim_{i \rightarrow \infty} \frac{\varepsilon_i}{r_i} = 0$ then*

$$\lim_{i \rightarrow \infty} \max \left(\frac{\text{ht}(P_i, Q)}{\text{width } P_i} : (P_i, Q) \text{ a jump} \right) = 0.$$

(b) If $\varepsilon = \varepsilon_1 = \varepsilon_2 = \dots$ then for all $i \gg 0$ and all jumps (P_i, Q) we have

$$\min(\|v - x\| : x \in P_i) < 47\varepsilon + 12,$$

where $Q = \text{conv}(P_i, v)$.

Remark 6.3 Theorem 6.2(a) and its proof straightforwardly extend to the more general families of polytopes when instead of spheres one uses ellipsoids – with fixed eccentricities in a family. We present the argument only in the spherical case in order to avoid cumbersome notation. Ellipsoids will appear explicitly in the next subsection in a more number theoretical context. Also, the proof below uses the following weaker condition than normality: the lattice points in the 2nd multiples of the polytopes in question are the sums of pairs of lattice points in the original polytopes.

Remark 6.4 The strong metric bound in Theorem 6.2(b) does not necessarily translate into a strong bound for the corresponding heights. In fact, as i gets larger, some facets of P_i get increasingly *sloped*, i.e., the ratio of the lattice and metric widths with respect to facets of P_i can be made arbitrarily small as $i \rightarrow \infty$. In fact, the normal unit vectors to the facets of P_i define increasingly dense subsets of the unit sphere S^{d-1} as $i \rightarrow \infty$. In particular, any neighborhood of any integer nonzero point $w \in \mathbb{Z}^d$ meets a hyperplane of the form $\text{Aff}(F) - v$ for some $F \in \mathbb{F}(P_i)$ and $v \in \text{vert}(F)$ when $i \gg 0$, depending on the neighborhood, so that w is not in the hyperplane. Actually, the same argument shows that, when $i \rightarrow \infty$, the absolute majority of the facets of P_i get increasingly sloped.

Proof of Theorem 6.2 (a) First we observe that the claim follows from the following equality

$$\lim_{i \rightarrow \infty} \max_v \left(\frac{\|z_i - v\|}{r_i} : (P_i, \text{conv}(P_i, v)) \text{ a jump} \right) = 1, \quad i \in \mathbb{N}. \tag{8}$$

In fact, the equality (8) is equivalent to the claim that the ratios

$$\frac{\|z_i - v\| - r_i}{r_i},$$

where $(P_i, \text{conv}(P_i, v))$ is a jump, can be made arbitrarily close to 0 by choosing i sufficiently large. For an index i and a jump $(P_i, \text{conv}(P_i, v))$ pick a facet $F_i \subset P_i$, visible from v and such that

$$\text{ht}(P_i, \text{conv}(P_i, v)) = \text{ht}_{P_i}(v) = -\text{ht}_{F_i}(v).$$

For all $i \gg 0$ and all jumps $(P_i, \text{conv}(P_i, v))$ we have the inequalities

$$\begin{aligned} \frac{\text{ht}(P_i, \text{conv}(P_i, v))}{\text{width } P_i} &\leq \frac{\text{ht}_{P_i}(v)}{\text{width}_{F_i} P_i} \\ &\leq \frac{\inf_{\text{Aff}(F_i)} \|\text{Aff}(F_i) - v\|}{2(r_i - \varepsilon_i)} \leq \left| \frac{\|z_i - v\| - r_i + \varepsilon_i}{2(r_i - \varepsilon_i)} \right|. \end{aligned}$$

So Theorem (6.2)(a) follows from (8).

Next we prove (8). Assume to the contrary that the considered limit is not 1. This means that infinitely many of the considered ratios exceed 1 by some real number $\theta > 0$. After picking the corresponding subsequence and re-indexing, we can assume that there are jumps $Q_i = (P_i, \text{conv}(P_i, v_i))$ such that

$$\theta_i := \frac{\|z_i - v_i\| - r_i}{r_i} \geq \theta, \quad i \in \mathbb{N}. \tag{9}$$

Applying the parallel translations by the vectors $-v_i$, we can further assume that $v_i = 0$ for all i . By Theorem 6.1, we have

$$L(2Q_i \setminus (Q_i \cup 2P_i)) = \emptyset, \quad i \in \mathbb{N}. \tag{10}$$

As $i \rightarrow \infty$, the subsets

$$T_i := \text{conv}(\text{B}((2 + 2\theta_i)\mathbf{e}_1, 2) \cup \{0\}) \setminus (\text{conv}(\text{B}((1 + \theta_i)\mathbf{e}_1, 1) \cup \{0\}) \cup \text{B}((2 + 2\theta_i)\mathbf{e}_1, 2)) \subset \mathbb{R}^d$$

become approximately congruent with increased precision to the rescaled subsets

$$\frac{1}{r_i}(2Q_i \setminus (Q_i \cup 2P_i)) \subset \mathbb{R}^d.$$

The following can be said on the geometry of the closure $\bar{T}_i \subset \mathbb{R}^d$ of T_i in the Euclidean topology:

- (i) \bar{T}_i is homeomorphic to a d -torus;
- (ii) \bar{T}_i is invariant under rotation of \mathbb{R}^d about the axis $\mathbb{R}\mathbf{e}_1$.

These properties, together with the inequalities (9), imply the existence of a real number $\rho > 0$ such that for every index i the set \bar{T}_i contains a ball B'_i of radius ρ . The easiest way to see this is by induction over d : the case $d = 2$ is obvious and every such ball $B_{d'}$ of radius ρ' in dimension $d' < d$ gives rise by revolution about the axis $\mathbb{R}_+\mathbf{e}_1$ to a torus of dimension $d' + 1$, which in its turn contains a $(d' + 1)$ -ball of radius ρ only depends on the radius of $B_{d'}$; see Fig. 5. (It is an exercise to show that, actually, we can take $\rho' = \rho$.)

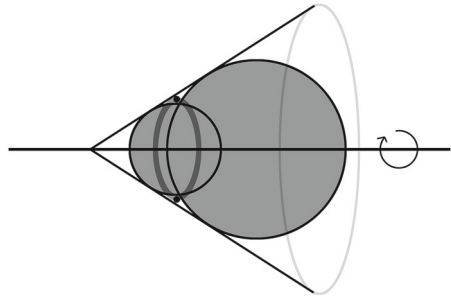
We see that, for any real number $0 < \kappa < 1$, the subsets

$$2Q_i \setminus (Q_i \cup 2P_i) \subset \mathbb{R}^d$$

contain balls of radius $\kappa\rho r_i$ whenever $i \gg 0$. But this contradicts (10) because $\kappa\rho r_i \rightarrow \infty$ as $i \rightarrow \infty$.

(b) Pick a sequence of jumps (P_i, Q_i) , $i \in \mathbb{N}$. Without loss of generality we can assume $L(Q_i) \setminus P_i = \{0\}$. By Theorem 6.1, we have the same equality (10) as in the proof of the part (a).

Fig. 5 Generating a d -torus in \bar{T}_i by revolving a $(d - 1)$ -ball



For simplicity of notation, put $B_i = B(z_i, r_i)$. Consider the subset

$$R_i = \text{conv}(2B_i \cup \{0\}) \setminus (\text{conv}(B_i \cup \{0\}) \cup 2B_i) \subset \mathbb{R}^d$$

and consider the closure $\bar{R}_i \subset \mathbb{R}^d$ of R_i in the Euclidean topology. The set \bar{R}_i is homeomorphic to a d -torus, invariant under rotation of \mathbb{R}^d about the axis $\mathbb{R}z_i$.

Let $C_i \subset \mathbb{R}^d$ be the cone $\mathbb{R}_+ B_i$. Then the boundary ∂C_i is tangent to the ball B_i . Let δ_i denote the distance from 0 to any point $B_i \cap \partial C_i$.

First we observe that the part (a) implies

$$\lim_{i \rightarrow \infty} \frac{\delta_i}{r_i} = 0. \tag{11}$$

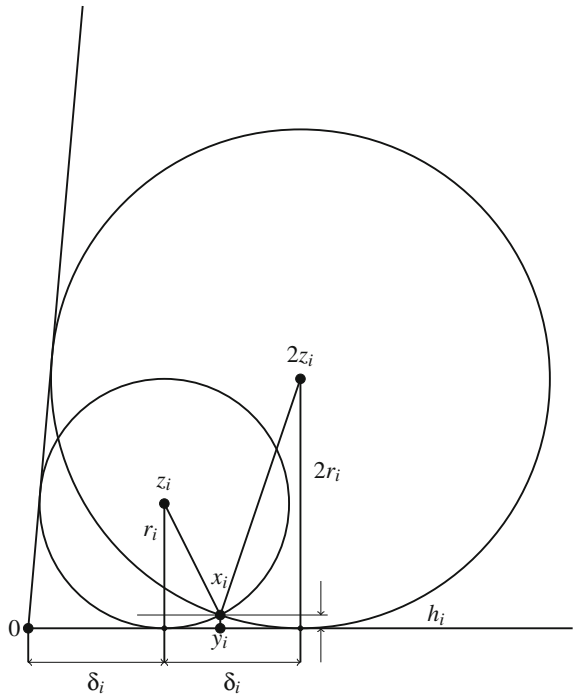
For every index i , the set of farthest points of \bar{R}_i from ∂C_i , which can be connected to ∂C_i by segments perpendicular to ∂C_i and entirely inside \bar{R}_i , forms a circle S_i with center in the line $\mathbb{R}_+ z_i$. Assume the distance from S_i to ∂C_i is h_i . Then

$$\delta_i = \sqrt{4r_i^2 - (2r_i - h_i)^2} + \sqrt{r_i^2 - (r_i - h_i)^2}, \tag{12}$$

as follows from Fig. 6, representing a section by any 2-dimensional plane in \mathbb{R}^d through 0 and z_i .

For every index i , pick a point $x_i \in S_i$ and let $[x_i, y_i]$ be the segment inside \bar{R}_i , perpendicular to ∂C_i and with $y_i \in \partial C_i$. (In particular, $\|x_i - y_i\| = h_i$). Because $r_i \rightarrow \infty$ and (11), the boundary $\partial \bar{R}_i$ close to the point x_i and y_i becomes increasingly close to the two $(d - 1)$ -dimensional affine hyperplanes through x_i and y_i , perpendicular to $[x_i, y_i]$. In fact, $r_i \rightarrow \infty$ implies that the cones C_i become increasingly obtuse, flattening $\partial \bar{R}_i$ close to y_i as $i \rightarrow \infty$, and (11) implies that the balls B_i and $2B_i$ have increasingly large radii but they stay close to each other relative to the radii, flattening $\partial \bar{R}_i$ close to x_i as $i \rightarrow \infty$. Consequently, as $i \rightarrow \infty$, the tori \bar{R}_i contain right cylinders of arbitrarily large radius around the axes $x_i + \mathbb{R}(y_i - x_i)$ with heights arbitrarily close to h_i . Fix a system of such cylinders $\Pi_i \subset \bar{R}_i$. We can assume that the heights of the Π_i are more than $\frac{1}{2}h_i$.

Fig. 6 Planar cross section



We have

$$2Q_i \setminus (Q_i \cup 2P_i) \supset \text{conv}(B(2z_i, 2r_i - 2\varepsilon) \cup \{0\}) \setminus (\text{conv}(B_i(z_i, r_i + \varepsilon) \cup \{0\}) \cup 2B_i(2z_i, 2z_i + 2\varepsilon)). \tag{13}$$

Using the same flatness of $\partial \bar{R}_i$ close to x_i and y_i , one concludes that the intersection of Π_i with the right side of (13) contains a coaxial right sub-cylinder of height $>$ (the height of $\Pi_i - 4\varepsilon) > \frac{1}{2}h_i - 4\varepsilon$ and the same radius as Π_i , provided $i \gg 0$.

If $\frac{1}{2}h_i - 4\varepsilon > 1$ for infinitely many indices i then the mentioned cylinders contain lattice points for $i \gg 0$, contradicting (10) in view of the containments (13).

Since the functions $f_i(x) = 4r_i^2 - (2r_i - x)^2$ and $g_i(x) = r_i r - (r_i - x)^2$ are increasing over the segment $[0, r_i]$, $\lim_{i \rightarrow \infty} r_i = \infty$, and $\frac{1}{2}h_i - 4\varepsilon \leq 1$ for $i \gg 0$, the equalities (12) imply

$$\delta_i \leq \sqrt{4r_i^2 - (2r_i - 2 - 8\varepsilon)^2} + \sqrt{r_i^2 - (r_i - 2 - 8\varepsilon)^2} < \sqrt{r_i}(2 + \sqrt{2})\sqrt{2 + 8\varepsilon},$$

provided $i \gg 0$.

Finally, for all $i \gg 0$ we have

$$\begin{aligned} \min(\|x\| : x \in P_i) &\leq \|z_i\| - r_i + \varepsilon \\ &= \sqrt{r_i^2 + \delta_i^2} - r_i + \varepsilon \\ &= \frac{\delta_i^2}{\sqrt{r_i^2 + \delta_i^2} + r_i} + \varepsilon \\ &< \frac{r_i(2 + \sqrt{2})^2(2 + 8\varepsilon)}{2r_i} + \varepsilon < 47\varepsilon + 12. \end{aligned}$$

□

A more careful choice in the cylinders inside \bar{R}_i in the argument above leads to a better estimate in Theorem 6.2(b), but in view of Remark 6.4 such an improvement is not worth pursuing.

6.2 Convex Hulls of All Lattice Points in Spheres

There is a ubiquity of sequences $\{P_i\}_{i \in \mathbb{N}}$, satisfying the stronger condition in Theorem 6.2(b), which one could call *rapidly spherical families*. Here is one recipe for deriving such a sequence. Choose any divergent series of real numbers $0 < r'_1 < r'_2 < \dots$ and put $P'_i = \text{conv}(L(B(r'_i, 0)))$. Because every unit d -cube in \mathbb{R}^d contains a lattice point and every d -ball $B \subset \mathbb{R}^d$ of radius $\frac{\sqrt{d}}{2}$ contains a unit cube, we have

$$\text{vert}(P'_i) \subset B(0, r'_i) \setminus B(0, r'_i - \sqrt{d}/2), \quad i \in \mathbb{N}.$$

Fix an arbitrary real number $\theta > 0$. The inclusions above imply

$$B(0, r'_i - (1 + \theta)\sqrt{d}/2) \subset P'_i \subset B(0, r'_i), \quad i \gg 0.$$

By Theorem 3.2(f), the polytopes $P_i = (d - 1)P'_i$ are normal for all i and we also have

$$B(0, r_i - \varepsilon) \subset P_i \subset B(0, r_i + \varepsilon), \quad i \gg 0,$$

where

$$r_i = (d - 1)(r'_i - (1 + \theta)\sqrt{d}/4) \quad \text{and} \quad \varepsilon = (d - 1)(1 + \theta)\sqrt{d}/4.$$

Similar examples can be derived when instead of balls one uses ellipsoids of fixed eccentricities per a family, not necessarily centered at 0.

Dilated lattice polytopes usually have non-empty first lattice strata around them. In particular the proposed recipe for deriving rapidly ellipsoidal families are unlikely to represent maximal elements in $\text{NPol}(d)$. This observation motivates the interest in

studying the normality of the convex hulls of *all* lattice points in spheres or, more generally, ellipsoids. We have the following partial result.

Theorem 6.5 *Let l_1, \dots, l_d be linearly independent real linear d -forms and $(z_1, \dots, z_d) \in \mathbb{R}^d$. Consider the ellipsoid*

$$E = \{ \xi = (\xi_1, \dots, \xi_d) : (l_1(\xi) - z_1)^2 + \dots + (l_d(\xi) - z_d)^2 \leq 1 \} \subset \mathbb{R}^d,$$

and the polytope $P = \text{conv}(\mathbb{L}(E))$.

- (a) *For any integer $k \geq 2$ and any point $y \in kP$ there exists a point $w \in \mathbb{L}(P)$ such that $y - w \in (k - 1)E$.*
- (b) *For any $y \in \mathbb{L}(2P)$ there exist $w_1, w_2 \in \mathbb{L}(P)$ such that $y = w_1 + w_2$.*
- (c) *If $d = 3$ then P is a normal polytope.*

Proof (a) For simplicity of notation, put $\sum = \sum_{i=1}^d$. Consider the (potentially 0) linear form

$$l(\xi) = \sum l_i(\xi)(l_i(y) - kz_i).$$

As $y/k \in P$ there must exist a vertex $w \in P$ such that $l(w) \geq l(y/k)$, i.e.,

$$\sum l_i(w)(l_i(y) - kz_i) \geq \sum \frac{l_i(y)}{k}(l_i(y) - kz_i).$$

This is equivalent to

$$\sum (l_i(w) - z_i)(l_i(y) - kz_i) \geq \sum \left(\frac{l_i(y)}{k} - z_i \right) (l_i(y) - kz_i)$$

and, therefore, to

$$\sum 2(l_i(y) - kz_i)(l_i(w) - z_i) \geq \frac{2}{k} \sum (l_i(y) - kz_i)^2. \tag{14}$$

We have

$$\begin{aligned} \sum (l_i(y - w) - (k - 1)z_i)^2 &= \sum ((l_i(y) - kz_i) - (l_i(w) - z_i))^2 \\ &= \sum ((l_i(y) - kz_i)^2 - 2(l_i(y) - kz_i)(l_i(w) - z_i) + (l_i(w) - z_i)^2), \end{aligned}$$

which, in view of (14), implies

$$\sum (l_i(y - w) - (k - 1)z_i)^2 \leq \frac{k - 2}{k} \sum (l_i(y) - kz_i)^2 + \sum (l_i(w) - z_i)^2.$$

As $y \in kE$ and $w \in E$ we obtain:

$$\sum (l_i(y - w) - (k - 1)z_i)^2 \leq \frac{k - 2}{k} \cdot k^2 + 1 = (k - 1)^2,$$

i.e., $y - w \in (k - 1)E$.

(b) We choose w as in (a). Then $y - w \in L(E) = L(P)$.

(c) Follows from (b) because of Lemma 3.3. \square

Remark 6.6 (a) We have tested several dozens of polytopes defined by ellipsoids with axes parallel to the coordinate axes in dimensions four and five, all of which turned out to be normal.

(b) The standard three dimensional balls $B(0, r)$, $r = 1, 2, \dots, 21$, define nonmaximal polytopes: all of them have height 1 jumps. The maximal height of jumps over them varies in an irregular manner: the smallest is 2 for $r = 2$, the largest is 11 for $r = 13$, and for $r = 21$ it is 9. Despite of its irregular behavior, the maximal height of jumps seems to grow slowly with r .

7 Explicit maximal polytopes

We have found maximal polytopes of dimension 4 and 5. This leaves little doubt that there exist maximal polytopes of any dimension ≥ 4 , but dimension 3 remains open. The experiments described in this section are based on a computer program written in C++ that makes heavy use of the library interface of Normaliz [7].

We want to emphasize that the experiments described below have not only produced maximal polytopes, but have also motivated several central results of the preceding sections.

7.1 The extension approach

The basic search strategy for finding maximal elements by successive extension is very simple:

- (1) Choose a normal start polytope P .
- (2) If $\#L(P)$ exceeds a preset bound, go to (1).
- (3) Find a jump Q over P .
- (4) If none exists, stop and save the maximal polytope P .
- (5) Replace P by Q and go to (2).

In addition to special constructions, like the cross-polytopes, we have implemented two methods for finding a start polytope:

- (U) Take the unimodular d -simplex and extend it by a random number of random height 1 jumps. The polytope thus reached is considered the start polytope.
- (S) We start from a lattice parallelotope and ‘shrink’ it successively by removing a vertex and taking the convex hull of the remaining vertices until no vertex can be removed without losing normality or the full dimension. The reached polytope serves as the starting point for subsequent extensions.

When we say ‘random’, we mean the choice of a random integer or vector within a certain range that can be modified via parameters of the search program.

At first glance, the shrinking technique (S) seems paradoxical: we shrink a parallelotope and then extend the shrunk polytope in order to reach a maximal one. However, (S) has proved very successful. Also (U) has led to maximal polytopes.

We can apply various strategies for finding quantum jumps over P . There are two major variants:

- (1) Choose a height 1 jump at random, provided such exists.
- (P) Choose a jump which maximizes a certain parameter, meant to lead to some sort of irregular normal polytopes.

If (1) is applied, one needs to compute only the points in the first stratum around the given polytope, and this is usually quite fast. Moreover, there is no need to test if the candidates are really jumps. For (P) we compute all candidate points according to Theorem 5.3 in dimension 3, but use a lower bound in dimensions ≥ 4 for the search phase, applying the full bound in the verification phase only.

The polytopes containing the candidates are highly rational. Nevertheless their lattice points can be computed very fast via the approximation algorithm of Normaliz.

It might seem most promising to always apply strategy (P), for example with the volume of the jump. But in pure form it has two drawbacks: (i) it tends to create successive jumps along straight lines that are not limited, and (ii) it is rather time consuming to test all candidate points in decreasing order of volume.

The following mixed strategy for step (3) of the basic algorithm has led us to the maximal polytopes P_4 and P_5 described below (and many others):

- (3a) Extend P according to (1) if a height 1 jump exists.
- (3b) Otherwise apply (P).

The two parameters for (P) that have proved successful are

- (V) the volume of the jump, see Remark 4.6;
- (A) the average multiplicity (or normalized $(d - 1)$ -volume) of the facets of Q .

In fact, the larger the multiplicity of a facet F , the more lattice points of low height over F in $P, \dots, (d - 2)P$ are necessary to guarantee normality of the extension; see Theorem 5.1. It is not surprising that the facet multiplicities of the maximal polytopes are quite large; see Table 2.

7.2 The Random Generation Approach

In this approach we

- (1) Choose a normal polytope at random and
- (2) Check it for maximality.

Creating a normal polytope by randomly choosing vertices becomes more and more difficult with growing dimension and number of vertices. According to our experience it works very well in dimension 4 if we limit ourselves to simplices.

The main advantage of this brute force approach is the enormous number of candidate polytopes that can be scanned if one gives up the idea of successive extension, and one can say that even in mathematics mass production may beat sophistication.

The random generation approach has produced the simplex P_4^1 below, many others in $\text{NPol}(4)$ and two in $\text{NPol}(5)$, of which one has only 21 lattice points.

The frequency of hitting maximal elements of $\text{NPol}(d)$ in our computations so far has been more or less the same for the two methods.

7.3 Some Maximal Polytopes

Table 1 contains the vertices of some maximal polytopes. The numbers of lattice points are 41 in P_4 , 42 in P_5 and only 22 in P'_4 . Note that P'_4 is a simplex with 22 lattice points. These numbers are small in relation to the widths of the polytopes over their facets that we have listed in Table 2 together with the multiplicities of the facets. Although we have no analogue of Theorem 4.11 in higher dimensions, one can expect that a maximal polytope has few lattice points relative to its facet widths and multiplicities.

By now, more than 40 maximal polytopes have emerged in dimension 4 and 6 in dimension 5. Despite of millions of attempts with varying strategies, our search has been futile in dimension 3.

For the three maximal polytopes the second lattice stratum is nonempty. In other words there exist height 2 points over P_4 , P_5 and P'_5 . There also exist maximal polytopes whose first two strata are empty.

We add a few data of the computations for P_4 and P_5 . The number of lattice points satisfying the height bound of Theorem 5.3 are 196, 697 for P_4 and 13, 525, 003 for P_5 . The computation of these candidate points takes <2 s for P_4 and <7 min for P_5 .

In order to verify that a candidate point is not a quantum jump, we first check whether $\#L(\text{conv}(P, z)) = \#L(P) + 1$. Only few candidates survive, namely 84 for P_4 and 980 for P_5 . For these we compute the Hilbert bases of the extended polytope and look for Hilbert basis elements of degree >1 . The computation times for the verifications are <2 min for P_4 and <2.5 h for P_5 . The verifications are documented in log files that list every candidate together with a ‘witness’, namely an extra element of $\text{Hilb}(\text{conv}(P, z))$. (The computations were done on a system with an Intel Xeon

Table 1 Vertices of maximal polytopes in dimensions 4 and 5

P_4 : (0,0,0,0)	P_5 : (4-13,-2,-1,1)	P'_4 : (0,3,2,0)
(3,0,2,0)	(4,12,13,4,-2)	(1,1,3,2)
(-2,-3,3,-1)	(-2,0,-8,-2,1)	(2,3,0,4)
(10,3,-3,-1)	(0,-2,0,0,0)	(4,0,0,2)
(0,-3,1,-2)	(27-26-15,-6,3)	(4,4,4,2)
(2,-2,0,-2)	(10,-1-11,-4,1)	
(-9,4,10,4)	(10-13,-2,-1,1)	

Table 2 Widths and facet multiplicities of maximal polytopes

P_4	Width	29	180	66	8	20	116	40	91	32	80	160
	Mult	4	1	4	10	4	1	2	4	10	4	2
P_5	Width	27	105	24	24	105	105	48	105	27	105	
	Mult	18	9	18	18	9	9	9	9	18	9	
P'_4	Width	24	48	48	48	48						
	Mult	8	4	4	4	4						

CPU E5-2660 0 at 2.20 GHz in strictly serial mode, the data available on request from the authors.)

Remark 7.1 (a) We have checked that the 4-dimensional maximal polytopes P_4 and P'_4 remain maximal if we consider very ample polytopes instead of normal ones.

(b) It is obvious that our findings rely crucially on the correctness of Normaliz. In order to enhance our confidence we have verified the maximality of P_4 with the dual algorithm of Normaliz. It takes considerable more time than the primal algorithm.

(c) P_5 and P'_4 have nontrivial symmetries: their automorphism groups are isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$ and \mathbb{Z}_2 , respectively.

The techniques employed in our experiments, apart of random generation, follow descending and ascending chains in $\text{NPol}(d)$. They can hardly find polytopes that are simultaneously minimal and maximal.

We end with the following question to which we were naturally led in this paper and which seems very difficult at present.

Question 7.2 (a) Does $\text{NPol}(3)$ have maximal elements? Does it have nontrivial minimal elements?

(b) Is the convex hull of all lattice points in every ellipsoid normal?

(c) Does there exist a normal polytope that is both a minimal and maximal element of the poset $\text{NPol}(d)$?

Acknowledgments We thank B. van Fraassen for his comments in the early stages of this project. We are grateful to anonymous reviewers for their helpful comments and spotting several inaccuracies. Supported by Grants DFG BR 688/22-1 (Bruns), NSF DMS-1301487 and GNSF DI/16/5-103/12 (Gubeladze), Polish National Science Center Grant No. 2012/05/D/ST1/01063 (Michalek)

References

1. Beck, M., Delgado, J., Gubeladze, J., Michalek, M.: Very ample and Koszul segmental fibrations. *J. Algebr. Comb.* **42**, 165–182 (2015)
2. Bruns, W.: On the integral Carathéodory property. *Exp. Math.* **16**, 359–365 (2007)
3. Bruns, W., Gubeladze, J.: Normality and covering properties of affine semigroups. *J. Reine Angew. Math.* **510**, 161–178 (1999)
4. Bruns, W., Gubeladze, J.: Rectangular simplicial semigroups. In: *Commutative Algebra, Algebraic Geometry, and Computational Methods* (Hanoi, 1996), pp. 201–213. Springer, Singapore (1999)
5. Bruns, W., Gubeladze, J.: *Polytopes, Rings, and K-theory*. Springer Monographs in Mathematics. Springer, New York (2009)
6. Bruns, W., Gubeladze, J., Henk, M., Martin, A., Weismantel, R.: A counterexample to an integer analogue of Carathéodory's theorem. *J. Reine Angew. Math.* **510**, 179–185 (1999)
7. Bruns, W., Ichim, B., Tim, R., Christof, S.: Normaliz. <http://www.math.uos.de/normaliz/>
8. Cook, W., Fonlupt, J., Schrijver, A.: An integer analogue of Carathéodory's theorem. *J. Comb. Theory, Ser. B* **40**(1), 63–70 (1986)
9. Cox, D.A., Haase, C., Hibi, T., Higashitani, A.: Integer decomposition property of dilated polytopes. *Electron. J. Comb.* **21**(4) (2014)
10. Cox, D.A., Little, J.B., Schenck, H.K.: *Toric Varieties*. Graduate Studies in Mathematics, vol. 124. American Mathematical Society, Providence, RI (2011)
11. Gubeladze, J.: Normal polytopes. In: *Proceedings of 22nd International Conference on Formal Power Series and Algebraic Combinatorics*, San Francisco, pp. 4–8 (2010). http://math.sfsu.edu/fpsac/local_proceedings.pdf

12. Gubeladze, J.: Convex normality of rational polytopes with long edges. *Adv. Math.* **230**, 372–389 (2012)
13. Haase, C., Maclagan, D., Hibi, T.: Miniworkshop: Projective Normality of Smooth Toric Varieties (2007). Oberwolfach Reports, 39/2007
14. Haase, C., Paffenholz, A., Piechnik, L., Santos, F.: Existence of unimodular triangulations—positive results. Preprint. <http://arxiv.org/abs/1405.1687>
15. Miller, E., Sturmfels, B.: *Combinatorial Commutative Algebra*. Graduate Texts in Mathematics, vol. 227. Springer, New York (2005)
16. Payne, S.: Lattice polytopes cut out by root systems and the Koszul property. *Adv. Math.* **220**, 926–935 (2009)
17. Sebö, A.: Hilbert bases, Carathéodory’s theorem and combinatorial optimization. In: Proceedings of the IPCO Conference, Waterloo, pp. 431–455 (1990)
18. White, G.K.: Lattice tetrahedra. *Can. J. Math.* **16**, 389–396 (1964)