Springer INdAM Series 27

Elisabetta Rocca · Ulisse Stefanelli
Lev Truskinovsky · Augusto Visintin
Editors

Trends in Applications of Mathematics to Mechanics

Springer

Springer INdAM Series

Volume 27

More information about this series at http://www.springer.com/series/10283

Elisabetta Rocca • Ulisse Stefanelli •
Lev Truskinovsky • Augusto Visintin
Editors

Trends in Applications of Mathematics to Mechanics

Springer

Editors

Elisabetta Rocca
Dipartimento di Matematica 'F. Casorati'
Università degli Studi di Pavia
Pavia
Pavia, Italy

Ulisse Stefanelli
Fakultät für Mathematik
Universität Wien
Wien, Austria

Lev Truskinovsky
CNRS
Paris
Paris, France

Augusto Visintin
Dipartimento di Matematica
Università degli Studi di Trento
Povo di Trento
Trento, Italy

ISSN 2281-518X ISSN 2281-5198 (electronic)
Springer INdAM Series
ISBN 978-3-319-75939-5 ISBN 978-3-319-75940-1 (eBook)
DOI 10.1007/978-3-319-75940-1

Library of Congress Control Number: 2018939067

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

Mechanics and Mathematics have a long history of mutual development. Across the centuries, mathematical formalism has imposed itself as the natural language of Mechanics. On the other hand, applications to Mechanics have constantly driven the progress of mathematical theories.

This volume originates from the INDAM *Symposium on Trends on Applications of Mathematics to Mechanics* (STAMM), which was held at the INDAM headquarters in Rome on 5–9 September 2016. STAMM is the biennial conference organized by the International Society for the Interaction of Mechanics and Mathematics (ISIMM), and the first meeting of this series dates back to 1975.

The book brings together original contributions at the interface of Mathematics and Mechanics. Consistently with the purpose of ISIMM, the focus is on mathematical models of phenomena issued from various applications. Among others, these include the following themes:

- Functional-analytic theories with applications to the Mechanics of Solids
- Modeling of nematic shells, thin films, dry friction, delamination, and damage
- Phase-field dynamics of Cahn-Hilliard type
- Thermodynamics of gases and continua

The papers in the volume, all of which have been refereed, present novel results and identify possible future developments.

We express our deep gratitude to all the authors and referees for their truly valuable commitment.

Pavia, Italy — Elisabetta Rocca
Wien, Austria — Ulisse Stefanelli
Paris, France — Lev Truskinovsky
Povo di Trento, Italy — Augusto Visintin

Contents

About the Editors

Prof. Elisabetta Rocca graduated with a degree in Mathematics from the University of Pavia in 1999, where she subsequently completed her PhD in 2004. She was a researcher at the University of Milan until 2011, when she became an associate professor. She moved to the WIAS in Berlin in 2013, where she spent 2 years coordinating a research group within the ERC Starting Grant she was awarded as a PI in 2011. She transferred to the University of Pavia in 2016, where she is currently an associate professor. She is the author of more than 80 papers on Mathematical Analysis and Applications.

Prof. Ulisse Stefanelli graduated with a degree in Mathematics and Scientific Computing from the University of Pavia in 2003. Since 2001 he has been working at the National Research Council's Istituto di Matematica Applicata e Tecnologie Informatiche 'E. Magenes'. In 2013 he was appointed Chair of Applied Mathematics and Modeling at the University of Vienna. His research activities focus on Calculus of Variations and partial differential equations, especially in applications to Mechanics and Materials Science.

Prof. Lev Truskinovsky is a CNRS research director at ESPCI PSL, Paris, France. From 1990 to 2004 he served on the faculty at the University of Minnesota. Holding a PhD in Applied Mathematics from the Russian Academy of Sciences, he is the author of more than 120 papers. He served as President of the ISIMM from 2009 to 2014.

Prof. Augusto Visintin graduated with a degree in Mathematics from the University of Pavia in 1975. He has been researcher at the IAN of CNR of Pavia and at SFB 123 in Heidelberg, Germany. He has been a full professor of Mathematical Analysis at the University of Trento since 1987.

Relaxation of *p*-Growth Integral Functionals Under Space-Dependent Differential Constraints

Elisa Davoli and Irene Fonseca

Abstract A representation formula for the relaxation of integral energies

$$(u, v) \mapsto \int_{\Omega} f(x, u(x), v(x))\, dx,$$

is obtained, where f satisfies p-growth assumptions, $1 < p < +\infty$, and the fields v are subjected to space-dependent first order linear differential constraints in the framework of $\mathscr{A}$-quasiconvexity with variable coefficients.

1 Introduction

The analysis of constrained relaxation problems is a central question in materials science. Many applications in continuum mechanics and, in particular, in magnetoelasticity, rely on the characterization of minimizers of non-convex multiple integrals of the type

$$u \mapsto \int_{\Omega} f(x, u(x), \nabla u(x), \dots, \nabla^k u(x))\, dx$$

or

$$(u, v) \mapsto \int_{\Omega} f(x, u(x), v(x))\, dx, \tag{1}$$

E. Davoli (✉)
Faculty of Mathematics, University of Vienna, Vienna, Austria
e-mail: elisa.davoli@univie.ac.at

I. Fonseca
Department of Mathematics, Carnegie Mellon University, Pittsburgh, PA, USA
e-mail: fonseca@andrew.cmu.edu

E. Rocca et al. (eds.), *Trends in Applications of Mathematics to Mechanics*,
Springer INdAM Series 27, https://doi.org/10.1007/978-3-319-75940-1_1

where Ω is an open, bounded subset of $\mathbb{R}^N$, $u : \Omega \to \mathbb{R}^m$, $m \in \mathbb{N}$, and the fields $v : \Omega \to \mathbb{R}^d$, $d \in \mathbb{N}$, satisfy partial differential constraints of the type "$\mathscr{A} v = 0$" other than $\operatorname{curl} v = 0$ (see e.g. [5, 9]).

In this paper we provide a representation formula for the relaxation of nonconvex integral energies of the form (1), in the case in which the energy density f satisfies p-growth assumptions, and the fields v are subjected to linear first-order space-dependent differential constraints.

The natural framework to study this family of relaxation problems is within the theory of $\mathscr{A}$-quasiconvexity with variable coefficients. In order to present this notion, we need to introduce some notation.

For $i = 1 \cdots, N$, let $A^i \in C^\infty(\mathbb{R}^N; \mathbb{M}^{l\times d}) \cap W^{1,\infty}(\mathbb{R}^N; \mathbb{M}^{l\times d})$, let $1 < p < +\infty$, and consider the differential operator

$$\mathscr{A} : L^p(\Omega; \mathbb{R}^d) \to W^{-1,p}(\Omega; \mathbb{R}^l), \quad d, l \in \mathbb{N},$$

defined as

$$\mathscr{A} v := \sum_{i=1}^{N} A^i(x) \frac{\partial v(x)}{\partial x_i} \tag{2}$$

for every $v \in L^p(\Omega; \mathbb{R}^d)$, where (2) is to be interpreted in the sense of distributions. Assume that the symbol $\mathbb{A} : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{M}^{l\times d}$,

$$\mathbb{A}(x, w) := \sum_{i=1}^{N} A^i(x) w_i \quad \text{for } (x, w) \in \mathbb{R}^N \times \mathbb{R}^N,$$

satisfies the uniform constant rank condition (see [22])

$$\operatorname{rank} \mathbb{A}(x, w) = r \quad \text{for every } x \in \mathbb{R}^N \text{and } w \in \mathbb{S}^{N-1}. \tag{3}$$

Let Q be the unit cube in $\mathbb{R}^N$ with sides parallel to the coordinate axis, i.e.,

$$Q := \Big(-\frac{1}{2}, \frac{1}{2}\Big)^N.$$

Denote by $C^\infty_{\text{per}}(\mathbb{R}^N; \mathbb{R}^m)$ the set of $\mathbb{R}^m$-valued smooth maps that are Q-periodic in $\mathbb{R}^N$, and for every $x \in \Omega$ consider the set

$$\mathscr{C}_x := \Big\{ w \in C^\infty_{\text{per}}(\mathbb{R}^N; \mathbb{R}^m) : \int_Q w(y)\, dy = 0, \text{ and } \sum_{i=1}^{N} A^i(x) \frac{\partial w(y)}{\partial y_i} = 0 \Big\}.$$

Let $f : \Omega \times \mathbb{R}^m \times \mathbb{R}^d \to [0, +\infty)$ be a Carathéodory function. The $\mathscr{A}$-quasiconvex envelope of $f(x, u, \cdot)$ for $x \in \Omega$ and $u \in \mathbb{R}^m$ is defined for $\xi \in \mathbb{R}^d$ as

$$Q_{\mathscr{A}(x)} f(x, u, \xi) := \inf \left\{ \int_Q f(x, u, \xi + w(y))\, dy : w \in \mathscr{C}_x \right\}.$$

We say that f is $\mathscr{A}$-quasiconvex if $f(x, u, \xi) = Q_{\mathscr{A}(x)} f(x, u, \xi)$ for a.e. $x \in \Omega$, and for all $u \in \mathbb{R}^m$ and $\xi \in \mathbb{R}^d$.

The notion of $\mathscr{A}$-quasiconvexity was first introduced by B. Dacorogna in [8], and extensively characterized in [17] by I. Fonseca and S. Müller for operators $\mathscr{A}$ defined as in (2), satisfying the constant rank condition (3), and having constant coefficients,

$$A^i(x) \equiv A^i \in \mathbb{M}^{l \times d} \quad \text{for every } x \in \mathbb{R}^N,\ i = 1, \ldots, N.$$

In that paper the authors proved (see [17, Theorems 3.6 and 3.7]) that under p-growth assumptions on the energy density f, $\mathscr{A}$-quasiconvexity is necessary and sufficient for the lower-semicontinuity of integral functionals

$$I(u, v) := \int_\Omega f(x, u(x), v(x))\, dx \quad \text{for every } (u, v) \in L^p(\Omega; \mathbb{R}^m) \times L^p(\Omega; \mathbb{R}^d)$$

along sequences (u^n, v^n) satisfying $u^n \to u$ in measure, $v^n \rightharpoonup v$ in $L^p(\Omega; \mathbb{R}^d)$, and $\mathscr{A} v^n \to 0$ in $W^{-1,p}(\Omega)$. We remark that in the framework $\mathscr{A} = \text{curl}$, i.e., when $v^n = \nabla \phi^n$ for some $\phi^n \in W^{1,p}(\Omega; \mathbb{R}^m)$, $d = n \times m$, $\mathscr{A}$-quasiconvexity reduces to Morrey's notion of quasiconvexity.

The analysis of properties of $\mathscr{A}$-quasiconvexity for operators with constant coefficients was extended in the subsequent paper [6], where A. Braides, I. Fonseca and G. Leoni provided an integral representation formula for relaxation problems under p-growth assumptions on the energy density, and presented (via Γ-convergence) homogenization results for periodic integrands evaluated along $\mathscr{A}$-free fields. These homogenization results were later generalized in [13], where I. Fonseca and S. Krömer worked under weaker assumptions on the energy density f. In [19, 20], simultaneous homogenization and dimension reduction was studied in the framework of $\mathscr{A}$-quasiconvexity with constant coefficients. Oscillations and concentrations generated by $\mathscr{A}$-free mappings are the subject of [14]. Very recently an analysis of the case in which the energy density is nonpositive has been carried out in [18], and applications to the theory of compressible Euler systems have been studied in [7]. A parallel analysis for operators with constant coefficients and under linear growth assumptions for the energy density has been developed in [1, 4, 15, 21]. A very general characterization in this setting has been obtained in [2], following the new insight in [12].

The theory of $\mathscr{A}$-quasiconvexity for operators with variable coefficients has been characterized by P. Santos in [23]. Homogenization results in this setting have been obtained in [10] and [11].

This paper is devoted to proving a representation result for the relaxation of integral energies in the framework of $\mathscr{A}$-quasiconvexity with variable coefficients. To be precise, let $1 < p, q < +\infty$, $d, m, l \in \mathbb{N}$, and consider a Carathéodory function $f : \Omega \times \mathbb{R}^m \times \mathbb{R}^d \to [0, +\infty)$ satisfying

$$\text{(H)} \quad 0 \leq f(x, u, v) \leq C(1 + |u|^p + |v|^q), \quad 1 < p, q < +\infty,$$

for a.e. $x \in \Omega$, and all $(u, v) \in \mathbb{R}^m \times \mathbb{R}^d$, with $C > 0$.

Denoting by $\mathscr{O}(\Omega)$ the collection of open subsets of Ω, for every $D \in \mathscr{O}(\Omega)$, $u \in L^p(\Omega; \mathbb{R}^m)$ and $v \in L^q(\Omega; \mathbb{R}^d)$ with $\mathscr{A}v = 0$, we define

$$\begin{aligned} \mathscr{I}((u, v), D) := \inf \Big\{ \liminf_{n \to +\infty} \int_D f(x, u_n(x), v_n(x)) : u_n \to u \quad \text{strongly in } L^p(\Omega; \mathbb{R}^m), \\ v_n \rightharpoonup v \quad \text{weakly in } L^q(\Omega; \mathbb{R}^d) \text{ and } \mathscr{A}v_n \to 0 \quad \text{strongly in } W^{-1,q}(\Omega; \mathbb{R}^l) \Big\}. \end{aligned} \tag{4}$$

Our main result is the following.

Theorem 1 *Let $\mathscr{A}$ be a first order differential operator with variable coefficients, satisfying* (3). *Let $f : \Omega \times \mathbb{R}^m \times \mathbb{R}^d \to [0, +\infty)$ be a Carathéodory function satisfying* (H). *Then,*

$$\int_D Q_{\mathscr{A}(x)} f(x, u(x), v(x))\, dx = \mathscr{I}((u, v), D)$$

for all $D \in \mathscr{O}(\Omega)$, $u \in L^p(\Omega; \mathbb{R}^m)$ and $v \in L^q(\Omega; \mathbb{R}^d)$ with $\mathscr{A}v = 0$.

Adopting the "blow-up" method introduced in [16], the proof of the theorem consists in showing that the functional $\mathscr{I}((u, v), \cdot)$ is the trace of a Radon measure absolutely continuous with respect to the restriction of the Lebesgue measure $\mathscr{L}^N$ to Ω, and proving that for a.e. $x \in \Omega$ the Radon-Nicodym derivative $\frac{d\mathscr{I}((u,v)\cdot)(x)}{d\mathscr{L}^N}$ coincides with the $\mathscr{A}$-quasiconvex envelope of f.

The arguments used are a combination of the ideas from [6, Theorem 1.1] and from [23]. The main difference with [6, Theorem 1.1], which reduces to our setting in the case in which the operator $\mathscr{A}$ has constant coefficients, is in the fact that while defining the operator $\mathscr{I}$ in (4) we can not work with exact solutions of the PDE, but instead we need to study sequences of asymptotically $\mathscr{A}$-vanishing fields. As pointed out in [23], in the case of variable coefficients the natural framework is the context of pseudo-differential operators. In this setting, we don't know how to project directly onto the kernel of the differential constraint, but we are able to construct an "approximate" projection operator P such that for every field $v \in L^p$, the $W^{-1,p}$ norm of $\mathscr{A}Pv$ is controlled by the $W^{-1,p}$ norm of v itself (we refer to [23, Subsection 2.1] for a detailed explanation of this issue and to the references therein for a treatment of the main properties of pseudo-differential operators). For the same reason, in the proof of the inequality

$$\frac{d\mathscr{I}((u,v)\cdot)(x)}{d\mathscr{L}^N} \leq Q_{\mathscr{A}(x)}f(x,u(x),v(x)) \quad \text{for a.e. } x \in \Omega,$$

an equi-integrability argument is needed (see Proposition 3). We also point out that the representation formula in Theorem 1 was obtained in a simplified setting in [11] as a corollary of the main homogenization result. Here we provide an alternative, direct proof, which does not rely on homogenization techniques.

The paper is organized as follows: in Sect. 2 we establish the main assumptions on the differential operator $\mathscr{A}$ and we recall some preliminary results on $\mathscr{A}$-quasiconvexity with variable coefficients. Section 3 is devoted to the proof of Theorem 1.

Notation Throughout the paper $\Omega \subset \mathbb{R}^N$ is a bounded open set, $1 < p, q < +\infty$, $\mathscr{O}(\Omega)$ is the set of open subsets of Ω, Q denotes the unit cube in $\mathbb{R}^N$, $Q(x_0, r)$ and $B(x_0, r)$ are, respectively, the open cube and the open ball in $\mathbb{R}^N$, with center x_0 and radius r. Given an exponent $1 < q < +\infty$, we denote by q' its conjugate exponent, i.e., $q' \in (1, +\infty)$ is such that

$$\frac{1}{q} + \frac{1}{q'} = 1.$$

Whenever a map $v \in L^q, C^\infty, \cdots$ is Q-periodic, that is

$$v(x + e_i) = v(x) \quad i = 1, \cdots, N,$$

for a.e. $x \in \mathbb{R}^N$, $\{e_1, \cdots, e_N\}$ being the standard basis of $\mathbb{R}^N$, we write $v \in L^q_{\text{per}}, C^\infty_{\text{per}}, \ldots$ We implicitly identify the spaces $L^q(Q)$ and $L^q_{\text{per}}(\mathbb{R}^N)$.

We adopt the convention that C will denote a generic constant, whose value may change from line to line in the same formula.

2 Preliminary Results

In this section we introduce the main assumptions on the differential operator $\mathscr{A}$ and we recall some preliminary results about $\mathscr{A}$-quasiconvexity.

For $i = 1, \cdots, N$, $x \in \mathbb{R}^N$, consider the linear operators $A^i(x) \in \mathbb{M}^{l\times d}$, with $A^i \in C^\infty(\mathbb{R}^N; \mathbb{M}^{l\times d}) \cap W^{1,\infty}(\mathbb{R}^N; \mathbb{M}^{l\times d})$. For every $v \in L^q(\Omega; \mathbb{R}^d)$ we set

$$\mathscr{A}v := \sum_{i=1}^N A^i(x)\frac{\partial v(x)}{\partial x_i} \in W^{-1,q}(\Omega; \mathbb{R}^l).$$

The symbol $\mathbb{A} : \mathbb{R}^N \times \mathbb{R}^N \setminus \{0\} \to \mathbb{M}^{l\times d}$ associated to the differential operator $\mathscr{A}$ is

$$\mathbb{A}(x,\lambda) := \sum_{i=1}^{N} A^i(x)\lambda_i \in \mathbb{M}^{l\times d}$$

for every $x \in \mathbb{R}^N$, $\lambda \in \mathbb{R}^N \setminus \{0\}$. We assume that $\mathscr{A}$ satisfies the following *uniform constant rank condition*:

$$\operatorname{rank}\Big(\sum_{i=1}^{N} A^i(x)\lambda_i\Big) = r \quad \text{for all } x \in \mathbb{R}^N \text{ and } \lambda \in \mathbb{R}^N \setminus \{0\}. \tag{5}$$

For every $x \in \mathbb{R}^N$, $\lambda \in \mathbb{R}^N \setminus \{0\}$, let $\mathbb{P}(x,\lambda) : \mathbb{R}^d \to \mathbb{R}^d$ be the linear projection on Ker $\mathbb{A}(x,\lambda)$, and let $\mathbb{Q}(x,\lambda) : \mathbb{R}^l \to \mathbb{R}^d$ be the linear operator given by

$$\begin{aligned}&\mathbb{Q}(x,\lambda)\mathbb{A}(x,\lambda)v := v - \mathbb{P}(x,\lambda)v \quad \text{for all } v \in \mathbb{R}^d,\\ &\mathbb{Q}(x,\lambda)\xi = 0 \quad \text{if } \xi \notin \text{Range } \mathbb{A}(x,\lambda).\end{aligned}$$

The main properties of $\mathbb{P}(\cdot,\cdot)$ and $\mathbb{Q}(\cdot,\cdot)$ are recalled in the following proposition (see e.g. [23, Subsection 2.1]).

Proposition 1 *Under the constant rank condition* (5)*, for every $x \in \mathbb{R}^N$ the operators $\mathbb{P}(x,\cdot)$ and $\mathbb{Q}(x,\cdot)$ are, respectively,* 0*-homogeneous and* (-1)*-homogeneous. In addition, $\mathbb{P} \in C^\infty(\mathbb{R}^N \times \mathbb{R}^N \setminus \{0\}; \mathbb{M}^{d\times d})$ and $\mathbb{Q} \in C^\infty(\mathbb{R}^N \times \mathbb{R}^N \setminus \{0\}; \mathbb{M}^{d\times l})$.*

Let $\eta \in C_c^\infty(\Omega; [0,1])$, $\eta = 1$ in Ω' for some $\Omega' \subset\subset \Omega$. We denote by $\mathbb{A}_\eta$ the symbol

$$\mathbb{A}_\eta(x,\lambda) := \sum_{i=1}^{N} \eta(x)A^i(x)\lambda_i, \tag{6}$$

for every $x \in \mathbb{R}^N$, $\lambda \in \mathbb{R}^N \setminus \{0\}$, and by $\mathscr{A}_\eta$ the corresponding pseudo-differential operator (see [23, Subsection 2.1] for an overview of the main properties of pseudo-differential operators). Let $\chi \in C^\infty(\mathbb{R}^+; \mathbb{R})$ be such that $\chi(|\lambda|) = 0$ for $|\lambda| < 1$ and $\chi(|\lambda|) = 1$ for $|\lambda| > 2$. Let also P_η be the operator associated to the symbol

$$\mathbb{P}_\eta(x,\lambda) := \eta^2(x)\mathbb{P}(x,\lambda)\chi(|\lambda|) \tag{7}$$

for every $x \in \mathbb{R}^N$, $\lambda \in \mathbb{R}^N \setminus \{0\}$. The following proposition (see [23, Theorem 2.2 and Subsection 2.1]) collects the main properties of the operators P_η and $\mathscr{A}_\eta$.

Proposition 2 *Let $1 < q < +\infty$, and let $\mathscr{A}_\eta$ and P_η be the pseudo-differential operators associated with the symbols* (6) *and* (7)*, respectively. Then there exists* *a constant C, depending on the dimension N, on q, and on the pseudo-differential operators $\mathscr{A}_\eta$ and P_η, such that*

$$\|P_\eta v\|_{L^q(\Omega;\mathbb{R}^d)} \leq C\|v\|_{L^q(\Omega;\mathbb{R}^d)} \tag{8}$$

for every $v \in L^q(\Omega;\mathbb{R}^d)$, and

$$\begin{aligned}
&\|P_\eta v\|_{W^{-1,q}(\Omega;\mathbb{R}^d)} \leq C\|v\|_{W^{-1,q}(\Omega;\mathbb{R}^d)},\\
&\|v - P_\eta v\|_{L^q(\Omega;\mathbb{R}^d)} \leq C\big(\|\mathscr{A}_\eta v\|_{W^{-1,q}(\Omega;\mathbb{R}^l)} + \|v\|_{W^{-1,q}(\Omega;\mathbb{R}^d)}\big),\\
&\|\mathscr{A}_\eta P_\eta v\|_{W^{-1,q}(\Omega;\mathbb{R}^l)} \leq C\|v\|_{W^{-1,q}(\Omega;\mathbb{R}^d)}
\end{aligned}$$

for every $v \in W^{-1,q}(\Omega;\mathbb{R}^d)$.

3 Proof of Theorem 1

Before proving Theorem 1 we state and prove a decomposition lemma, which generalizes [17, Lemma 2.15] to the case of operators with variable coefficients.

Lemma 1 *Let $1 < q < +\infty$. Let $\mathscr{A}$ be a first order differential operator with variable coefficients, satisfying (5). Let $v \in L^q(\Omega;\mathbb{R}^d)$, and let $\{v_n\}$ be a bounded sequence in $L^q(\Omega;\mathbb{R}^d)$ such that*

$$\begin{aligned}
&v_n \rightharpoonup v \quad \textit{weakly in } L^q(\Omega;\mathbb{R}^d),\\
&\mathscr{A}v_n \to 0 \quad \textit{strongly in } W^{-1,q}(\Omega;\mathbb{R}^l).
\end{aligned}$$

Then, there exists a q-equiintegrable sequence $\{\tilde{v}_n\} \subset L^q(\Omega;\mathbb{R}^d)$ such that

$$\mathscr{A}\tilde{v}_n \to 0 \quad \textit{strongly in } W^{-1,s}(\Omega;\mathbb{R}^l) \quad \textit{for every } 1 < s < q, \tag{9}$$

$$\int_\Omega \tilde{v}_n(x)\,dx = \int_\Omega v(x)\,dx,$$

$$\tilde{v}_n - v_n \to 0 \quad \textit{strongly in } L^s(\Omega;\mathbb{R}^d) \quad \textit{for every } 1 < s < q, \tag{10}$$

$$\tilde{v}_n \rightharpoonup v \quad \textit{weakly in } L^q(\Omega;\mathbb{R}^d). \tag{11}$$

In addition, if $\Omega \subset Q$ then we can construct the sequence $\{\tilde{v}^n\}$ so that $\tilde{v}_n - v \in L^q_{per}(\mathbb{R}^N;\mathbb{R}^d)$ for every $n \in \mathbb{N}$.

Proof Arguing as in the first part of [23, Proof of Theorem 1.1], we construct a q-equiintegrable sequence $\{\hat{v}_n\}$ satisfying (9), (10) and (11). The conclusion follows by setting $\tilde{v}_n := \hat{v}_n - \fint_\Omega \hat{v}_n(x)\,dx + \fint_\Omega v(x)\,dx$.

In the case in which $\Omega \subset Q$, let $\{\varphi^i\}$ be a sequence of cut-off functions in Q with $0 \leq \varphi^i \leq 1$ in Q, such that $\varphi^i = 0$ on $Q \setminus \Omega$ and $\varphi^i \to 1$ pointwise in Ω. Define $w_n^i := \varphi^i(\hat{v}_n - v)$. By (11) for every $\psi \in L^{q'}(\Omega; \mathbb{R}^d)$ we have

$$\lim_{i\to+\infty} \lim_{n\to+\infty} \int_\Omega w_n^i(x)\psi(x)\,dx = 0.$$

By (9), (10), and the compact embedding of $L^q(\Omega; \mathbb{R}^d)$ into $W^{-1,q}(\Omega; \mathbb{R}^d)$, there holds

$$\mathscr{A} w_n^i = \varphi^i \mathscr{A} \hat{v}_n + \left(\sum_{j=1}^N A^j \frac{\partial \varphi^i}{\partial x_j}\right)\hat{v}_n \to 0 \quad \text{strongly in } W^{-1,s}(\Omega; \mathbb{R}^l)$$

as $n \to +\infty$, for every $1 < s < q$. Extending the maps w_n^i outside Q by periodicity, by the metrizability of the weak topology on bounded sets and by Attouch's diagonalization lemma (see [3, Lemma 1.15 and Corollary 1.16]), we obtain a sequence

$$w_n := w_n^{i(n)},$$

with $\{w_n\} \subset L^q_{\text{per}}(\mathbb{R}^N; \mathbb{R}^d)$, and such that $w_n + v$ satisfies (9), (10) and (11). The thesis follows by setting

$$\tilde{v}_n := w_n - \int_\Omega w_n(x)\,dx + v.$$

The following proposition will allow us to neglect vanishing perturbations of q-equiintegrable sequences.

Proposition 3 *For every $n \in \mathbb{N}$, let $f_n : Q \times \mathbb{R}^d \to [0, +\infty)$ be a continuous function. Assume that there exists a constant $C > 0$ such that, for $q > 1$,*

$$\sup_{n\in\mathbb{N}} f_n(y, \xi) \leq C(1 + |\xi|^q) \quad \textit{for every } y \in Q \textit{ and } \xi \in \mathbb{R}^d, \tag{12}$$

and that the sequence $\{f_n(y, \cdot)\}$ is equicontinuous in $\mathbb{R}^d$, uniformly in y. Let $\{w_n\}$ be a q-equiintegrable sequence in $L^q(Q; \mathbb{R}^d)$, and let $\{v_n\} \subset L^q(Q; \mathbb{R}^d)$ be such that

$$v_n \to 0 \quad \textit{strongly in } L^q(Q; \mathbb{R}^d). \tag{13}$$

Then

$$\lim_{n\to+\infty} \left| \int_Q f_n\big(y, w_n(y)\big)\,dy - \int_Q f_n\big(y, v_n(y) + w_n(y)\big)\,dy \right| = 0.$$

Proof Fix $\eta > 0$. In view of (13), the sequence $\{C(1 + |v_n|^q + |w_n|^q)\}$ is equiintegrable in Q, thus there exists $0 < \varepsilon < \frac{\eta}{3}$ such that

$$\sup_{n\in\mathbb{N}} \int_A C\big(1 + |v_n(y)|^q + |w_n(y)|^q\big)\, dy < \frac{\eta}{3} \tag{14}$$

for every $A \subset Q$ with $|A| < \varepsilon$. By the q-equiintegrability of $\{w_n\}$ and $\{v_n\}$, and by Chebyshev's inequality there holds

$$\big|Q \cap \big(\{|w_n| > M\} \cup \{|v_n| > M\}\big)\big| \le \frac{1}{M^q} \int_Q (|w_n(y)|^q + |v_n(y)|^q)\, dy \le \frac{C}{M^q}$$

for every $n \in \mathbb{N}$. Therefore, there exists M_0 satisfying

$$\sup_{n\in\mathbb{N}} \big|Q \cap \big(\{|w_n| > M_0\} \cup \{|v_n| > M_0\}\big)\big| \le \frac{\varepsilon}{2}. \tag{15}$$

By the uniform equicontinuity of the sequence $\{f_n(y,\cdot)\}$, there exists $\delta > 0$ such that, for every $\xi_1, \xi_2 \in \overline{B(0, M_0)}$, with $|\xi_1 - \xi_2| < \delta$, we have

$$\sup_{y\in Q} |f_n(y, \xi_1) - f_n(y, \xi_2)| < \varepsilon \tag{16}$$

for every $n \in \mathbb{N}$. By (13) and Egoroff's theorem, there exists a set $E_\varepsilon \subset Q$, $|E_\varepsilon| < \frac{\varepsilon}{2}$, such that

$$v_n \to 0 \quad \text{uniformly in } Q \setminus E_\varepsilon,$$

and, in particular,

$$|v_n(x)| < \delta \quad \text{for a.e. } x \in Q \setminus E_\varepsilon, \tag{17}$$

for every $n \ge n_0$, for some $n_0 \in \mathbb{N}$.

We observe that

$$\begin{aligned}\int_Q f_n(y, v_n(y) + w_n(y))\, dy = &\int_{Q\cap\{|w_n|\le M_0\}\cap\{|v_n|\le M_0\}} f_n(y, v_n(y) + w_n(y))\, dy \\ &+ \int_{Q\cap(\{|w_n|>M_0\}\cup\{|v_n|>M_0\})} f_n(y, v_n(y) + w_n(y))\, dy.\end{aligned} \tag{18}$$

The first term in the right-hand side of (18) can be further decomposed as

$$\begin{aligned}
&\int_{Q\cap\{|w_n|\leq M_0\}\cap\{|v_n|\leq M_0\}} f_n(y, v_n(y)+w_n(y))\,dy\\
&\quad=\int_{(Q\setminus E_\varepsilon)\cap\{|w_n|\leq M_0\}\cap\{|v_n|\leq M_0\}} f_n(y, v_n(y)+w_n(y))\,dy\\
&\qquad+\int_{E_\varepsilon\cap\{|w_n|\leq M_0\}\cap\{|v_n|\leq M_0\}} f_n(y, v_n(y)+w_n(y))\,dy\\
&\quad=\int_{(Q\setminus E_\varepsilon)\cap\{|w_n|\leq M_0\}\cap\{|v_n|\leq M_0\}} f_n(y, w_n(y))\,dy\\
&\qquad+\int_{(Q\setminus E_\varepsilon)\cap\{|w_n|\leq M_0\}\cap\{|v_n|\leq M_0\}} \big(f_n(y, v_n(y)+w_n(y))-f_n(y, w_n(y))\big)\,dy\\
&\qquad+\int_{E_\varepsilon\cap\{|w_n|\leq M_0\}\cap\{|v_n|\leq M_0\}} f_n(y, v_n(y)+w_n(y))\,dy\\
&\quad=\int_Q f_n(y, w_n(y))\,dy-\int_{E_\varepsilon\cap\{|w_n|\leq M_0\}\cap\{|v_n|\leq M_0\}} f_n(y, w_n(y))\,dy\\
&\qquad-\int_{Q\cap(\{|w_n|>M_0\}\cup\{|v_n|>M_0\})} f_n(y, w_n(y))\,dy\\
&\qquad+\int_{(Q\setminus E_\varepsilon)\cap\{|w_n|\leq M_0\}\cap\{|v_n|\leq M_0\}} \big(f_n(y, v_n(y)+w_n(y))-f_n(y, w_n(y))\big)\,dy\\
&\qquad+\int_{E_\varepsilon\cap\{|w_n|\leq M_0\}\cap\{|v_n|\leq M_0\}} f_n(y, v_n(y)+w_n(y))\,dy.
\end{aligned}$$

We observe that by (15)

$$|E_\varepsilon\cup(\{|w_n|>M_0\}\cup\{|v_n|>M_0\})|<\varepsilon.$$

Hence, for $n\geq n_0$, by (12), (14), (16), and (17) we deduce the estimate

$$\begin{aligned}
&\Big|\int_Q f_n(y, w_n(y))\,dy-\int_Q f_n(y, v_n(y)+w_n(y))\,dy\Big| \qquad (19)\\
&\quad\leq\varepsilon+\int_{E_\varepsilon\cup(\{|w_n|>M_0\}\cup\{|v_n|>M_0\})} 2C(1+|w_n(y)|^p+|v_n(y)|^p)\,dy\leq\varepsilon+\frac{2\eta}{3}.
\end{aligned}$$

The thesis follows by the arbitrariness of η.

We now prove our main result.

Proof (Proof of Theorem 1) The proof is subdivided into 4 steps. Steps 1 and 2 follow along the lines of [6, Proof of Theorem 1.1]. Step 3 is obtained by modifying

[6, Lemma 3.5], whereas Step 4 follows by adapting an argument in [23, Proof of Theorem 1.2]. We only outline the main ideas of Steps 1 and 2 for convenience of the reader, whilst we provide more details for Steps 3 and 4.

Step 1:

The first step consists in showing that

$$\mathscr{I}((u,v),D)=\inf\Big\{\liminf_{n\to+\infty}\int_D f(x,u(x),v_n(x))\,dx\,:\,\{v_n\}\text{ is }q-\text{equiintegrable},\\ \mathscr{A}v_n\to 0\text{ strongly in }W^{-1,s}(D;\mathbb{R}^l)\text{ for every }1<s<q\\ \text{and }v_n\rightharpoonup v\text{ weakly in }L^q(D;\mathbb{R}^d)\Big\}.$$

This identification is proved by adapting [6, Proof of Lemma 3.1]. The only difference is the application of Lemma 1 instead of [6, Proposition 2.3 (i)].

Step 2:

The second step is the proof that $\mathscr{I}((u,v),\cdot)$ is the trace of a Radon measure absolutely continuous with respect to $\mathscr{L}^N\lfloor\Omega$. This follows as a straightforward adaptation of [6, Lemma 3.4]. The only modifications are due to the fact that [6, Proposition 2.3 (i)] and [6, Lemma 3.1] are now replaced by Lemma 1 and Step 1.

Step 3:

We claim that

$$\frac{d\mathscr{I}((u,v),\cdot)}{d\mathscr{L}^N}(x_0)\geq Q_{\mathscr{A}(x_0)}f(x_0,u(x_0),v(x_0))\quad\text{for a.e. }x_0\in\Omega. \tag{20}$$

Indeed, since $g(x,\xi):=f(x,u(x),\xi)$ is a Carathéodory function, by the Scorza-Dragoni Theorem there exists a sequence of compact sets $K_j\subset\Omega$ such that

$$|\Omega\setminus K_j|\leq\tfrac{1}{j}$$

and the restriction of g to $K_j\times\mathbb{R}^d$ is continuous. Hence, the set

$$\omega:=\bigcup_{j=1}^{+\infty}(K_j\cap K_j^*)\cap\mathscr{L}(u,v), \tag{21}$$

where K_j^* is the set of Lebesgue point for the characteristic function of K_j and $\mathscr{L}(u,v)$ is the set of Lebesgue points of u and v, is such that

$$|\Omega\setminus\omega|\leq|\Omega\setminus K_j|\leq\frac{1}{j}\quad\text{for every }j,$$

and so $|\Omega \setminus \omega| = 0$. Let $x_0 \in \omega$ be such that

$$\lim_{r\to 0^+} \frac{1}{r^N} \int_{Q(x_0,r)} |u(x) - u(x_0)|^p \, dx = \lim_{r\to 0^+} \frac{1}{r^N} \int_{Q(x_0,r)} |v(x) - v(x_0)|^q \, dx = 0, \tag{22}$$

and

$$\frac{d\mathscr{I}((u,v),\cdot)}{d\mathscr{L}^N}(x_0) = \lim_{r\to 0^+} \frac{\mathscr{I}((u,v), Q(x_0,r))}{r^N} < +\infty, \tag{23}$$

where the sequence of radii r is such that $\mathscr{I}((u,v), \partial Q(x_0,r)) = 0$ for every r. (Such a choice of the sequence is possible due to Step 2).

By Step 1, for every r there exists a q-equiintegrable sequence $\{v_{n,r}\}$ such that

$$\begin{aligned} &v_{n,r} \rightharpoonup v \quad \text{weakly in } L^q(Q(x_0,r);\mathbb{R}^d), \\ &\mathscr{A} v_{n,r} \to 0 \quad \text{strongly in } W^{-1,s}(Q(x_0,r);\mathbb{R}^l) \text{ for every } 1 < s < q \end{aligned} \tag{24}$$

as $n \to +\infty$, and

$$\lim_{n\to+\infty} \int_{Q(x_0,r)} g(x, v_{n,r}(x)) \, dx \leq \mathscr{I}((u,v), Q(x_0,r)) + r^{N+1}.$$

A change of variables yields

$$\frac{d\mathscr{I}((u,v),\cdot)}{d\mathscr{L}^N}(x_0) \geq \liminf_{r\to 0^+} \lim_{n\to+\infty} \int_Q g(x_0 + ry, v(x_0) + w_{n,r}(y)) \, dy,$$

where

$$w_{n,r}(y) := v_{n,r}(x_0 + ry) - v(x_0) \quad \text{for a.e. } y \in Q.$$

Arguing as in [6, Proof of Lemma 3.5], Hölder's inequality and a change of variables imply

$$w_{n,r} \rightharpoonup 0 \quad \text{weakly in } L^q(Q;\mathbb{R}^d) \tag{25}$$

as $n \to +\infty$ and $r \to 0^+$, in this order. We claim that

$$\mathscr{A}(x_0 + r\cdot) w_{n,r} \to 0 \quad \text{strongly in } W^{-1,s}(Q;\mathbb{R}^l), \tag{26}$$

as $n \to +\infty$, for every r and every $1 < s < q$.

Indeed, let $\varphi \in W_0^{1,s'}(Q;\mathbb{R}^d)$. There holds

$$\begin{aligned}
&\langle \mathscr{A}(x_0+r\cdot)w_{n,r},\ \varphi\rangle_{W^{-1,s}(Q;\mathbb{R}^l),W_0^{1,s'}(Q;\mathbb{R}^l)}\\
&\quad= -\sum_{i=1}^N \left\{ r\int_Q \frac{\partial A^i(x_0+ry)}{\partial x_i} v_{n,r}(x_0+ry)\cdot\varphi(y)\,dy \right.\\
&\qquad \left. + \int_Q A^i(x_0+ry)v_{n,r}(x_0+ry)\cdot\ \frac{\partial\varphi(y)}{\partial y_i}\,dy\right\}\\
&\quad= -\sum_{i=1}^N \left\{ \frac{1}{r^{N-1}}\int_{Q(x_0,r)} \frac{\partial A^i(x)}{\partial x_i} v_{n,r}(x)\cdot\psi_r(x)\,dx \right.\\
&\qquad \left. + \frac{1}{r^{N-1}}\int_{Q(x_0,r)} A^i(x)v_{n,r}(x)\cdot \frac{\partial\psi_r(x)}{\partial x_i}\,dx\right\}\\
&\quad= \frac{1}{r^{N-1}}\langle \mathscr{A}v_{n,r},\ \psi_r\rangle_{W^{-1,s}(Q(x_0,r);\mathbb{R}^l),W_0^{1,s'}(Q(x_0,r);\mathbb{R}^l)},
\end{aligned}$$

where $\psi_r(x) := \varphi\big(\frac{x-x_0}{r}\big)$ for a.e. $x \in Q(x_0,r)$. Since $\psi_r \in W_0^{1,s'}(Q(x_0,r);\mathbb{R}^d)$ and

$$\|\psi_r\|_{W_0^{1,s'}(Q(x_0,r);\mathbb{R}^d)} \le C(r)\|\varphi\|_{W_0^{1,s'}(Q;\mathbb{R}^d)},$$

we obtain the estimate

$$\|\mathscr{A}(x_0+r\cdot)w_{n,r}\|_{W^{-1,s}(Q;\mathbb{R}^l)} \le C(r)\|\mathscr{A}v_{n,r}\|_{W^{-1,s}(Q(x_0,r);\mathbb{R}^l)}.$$

Claim (26) follows by (24).

In view of (25) and (26), a diagonalization procedure yields a q-equiintegrable sequence $\{\hat{w}_k\} \subset L^q(Q;\mathbb{R}^d)$ satisfying

$$\hat{w}_k \rightharpoonup 0 \quad \text{weakly in } L^q(Q;\mathbb{R}^d), \tag{27}$$

$$\mathscr{A}(x_0+r_k\cdot)\hat{w}_k \to 0 \quad \text{strongly in } W^{-1,s}(Q;\mathbb{R}^l) \quad \text{for every } 1<s<q, \tag{28}$$

and

$$\frac{d\mathscr{I}((u,v),\cdot)}{d\mathscr{L}^N}(x_0) \ge \liminf_{k\to+\infty}\int_Q g(x_0+r_ky, v(x_0)+\hat{w}_k(y))\,dy. \tag{29}$$

For every $\varphi \in W_0^{1,s'}(Q;\mathbb{R}^l)$, $1 < s < q$, there holds

$$\begin{aligned}&\langle (\mathscr{A}(x_0 + r_k\cdot) - \mathscr{A}(x_0))\hat{w}_k,\ \varphi\rangle_{W^{-1,s}(Q;\mathbb{R}^l),W_0^{1,s'}(Q;\mathbb{R}^l)}\\ &= -\sum_{i=1}^{N}\Bigg[r_k\int_Q \frac{\partial A^i(x_0 + r_k y)}{\partial x_i}\hat{w}_k(y)\cdot\varphi(y)\,dy\\ &\quad + \int_Q (A^i(x_0 + r_k y) - A^i(x_0))\hat{w}_k(y)\cdot\frac{\partial\varphi(y)}{\partial y_i}\,dy\Bigg].\end{aligned}$$

Thus,

$$\|(\mathscr{A}(x_0 + r_k\cdot) - \mathscr{A}(x_0))\hat{w}_k\|_{W^{-1,s}(Q;\mathbb{R}^l)} \leq r_k\sum_{i=1}^{N}\|A^i\|_{W^{1,\infty}(\mathbb{R}^N;\mathbb{R}^{l\times d})}\|\hat{w}_k\|_{L^q(Q;\mathbb{R}^d)}$$

for every $1 < s < q$. By (27) and (28) we conclude that

$$\mathscr{A}(x_0)\hat{w}_k \to 0 \quad \text{strongly in } W^{-1,s}(Q;\mathbb{R}^l) \quad \text{for every } 1 < s < q. \tag{30}$$

In view of (27) and (30), an adaptation of [6, Corollary 3.3] yields a q-equiintegrable sequence $\{w_k\}$ such that

$$\begin{aligned}&w_k \rightharpoonup 0 \quad \text{weakly in } L^q(Q;\mathbb{R}^d),\\ &\int_Q w_k(y)\,dy = 0 \quad \text{for every } k,\\ &\mathscr{A}(x_0)w_k = 0 \quad \text{for every } k,\end{aligned} \tag{31}$$

and

$$\liminf_{k\to+\infty}\int_Q g(x_0, v(x_0) + w_k(y))\,dy \leq \liminf_{k\to+\infty}\int_Q g(x_0 + r_k y, v(x_0) + \hat{w}_k(y))\,dy. \tag{32}$$

Finally, by combining (29), (31), and (32), and by the definition of $\mathscr{A}$-quasiconvex envelope for operators with constant coefficients, we obtain

$$\begin{aligned}\frac{d\mathscr{I}((u,v),\cdot)}{d\mathscr{L}^N}(x_0) &\geq \liminf_{k\to+\infty}\int_Q g(x_0, v(x_0) + w_k(y))\,dy\\ &= \liminf_{k\to+\infty}\int_Q f(x_0, u(x_0), v(x_0) + w_k(y))\,dy\\ &\geq Q_{\mathscr{A}(x_0)}f(x_0, u(x_0), v(x_0))\end{aligned}$$

for a.e. $x_0 \in \Omega$. This concludes the proof of Claim (20).

Step 4:

To complete the proof of the theorem we need to show that

$$\frac{d\mathscr{I}((u,v),\cdot)}{d\mathscr{L}^N}(x_0) \leq Q_{\mathscr{A}(x_0)} f(x_0, u(x_0), v(x_0)) \quad \text{for a.e. } x_0 \in \Omega. \tag{33}$$

To this aim, let $\mu > 0$, and $x_0 \in \omega$ be such that (22) and (23) hold. Let $w \in C^\infty_{\text{per}}(\mathbb{R}^N; \mathbb{R}^d)$ be such that

$$\int_Q w(y)\,dy = 0, \quad \mathscr{A}(x_0)w = 0, \tag{34}$$

and

$$\int_Q f(x_0, u(x_0), v(x_0) + w(y))\,dy \leq Q_{\mathscr{A}(x_0)} f(x_0, u(x_0), v(x_0)) + \mu. \tag{35}$$

Let $\eta \in C^\infty_c(\Omega; [0,1])$ be such that $\eta \equiv 1$ in a neighborhood of x_0 and let r be small enough so that

$$Q(x_0, r) \subset \{x : \eta(x) = 1\} \quad \text{and } Q(x_0, 2r) \subset\subset \Omega. \tag{36}$$

Consider a map $\varphi \in C^\infty_c(Q(x_0, r); [0,1])$ satisfying

$$\mathscr{L}^N(Q(x_0, r) \cap \{\varphi \neq 1\}) < \mu r^N, \tag{37}$$

and define

$$z^r_m(x) := \varphi(x) w\Big(\frac{m(x - x_0)}{r}\Big) \quad \text{for } x \in \mathbb{R}^N. \tag{38}$$

We observe that $z^r_m \in L^q(\Omega; \mathbb{R}^d)$, and for $\psi \in L^{q'}(\Omega; \mathbb{R}^d)$ we have

$$\begin{aligned}\int_\Omega z^r_m(x) \cdot \psi(x)\,dx &= \int_\Omega \varphi(x) w\Big(\frac{m(x - x_0)}{r}\Big) \cdot \psi(x)\,dx \\ &= r^N \int_Q \varphi(x_0 + ry) w(my) \cdot \psi(x_0 + ry)\,dy.\end{aligned}$$

By (34) and by the Riemann-Lebesgue lemma we have

$$z^r_m \rightharpoonup 0 \quad \text{weakly in } L^q(\Omega; \mathbb{R}^d) \tag{39}$$

as $m \to +\infty$. We claim that

$$\limsup_{m \to +\infty} \|\mathscr{A}_\eta z^r_m\|_{W^{-1,q}(\Omega; \mathbb{R}^l)} \leq C r^{\frac{N}{q} + 1}, \tag{40}$$

where $\mathscr{A}_\eta$ is the pseudo-differential operator defined in (6). Indeed, by (36) we obtain

$$\begin{aligned}\mathscr{A}_\eta z_m^r &= \mathscr{A} z_m^r - \mathscr{A}(x_0) z_m^r + \mathscr{A}(x_0) z_m^r \\ &= \sum_{i=1}^N \frac{\partial((A^i(x) - A^i(x_0)) z_m^r(x))}{\partial x_i} + \sum_{i=1}^N A^i(x_0) \frac{\partial z_m^r(x)}{\partial x_i} - \sum_{i=1}^N \frac{\partial A^i(x)}{\partial x_i} z_m^r(x).\end{aligned} \tag{41}$$

By the regularity of the operators A^i and by a change of variables, the first term in the right-hand side of (41) is estimated as

$$\begin{aligned}&\Big\| \sum_{i=1}^N \frac{\partial((A^i(x) - A^i(x_0)) z_m^r(x))}{\partial x_i} \Big\|_{W^{-1,q}(\Omega;\mathbb{R}^l)} \\ &\leq \sum_{i=1}^N \Big\| (A^i(x) - A^i(x_0)) \varphi(x) w\Big(\frac{m(x-x_0)}{r}\Big) \Big\|_{L^q(Q(x_0,r);\mathbb{R}^l)} \\ &\leq \sum_{i=1}^N \|A^i\|_{W^{1,\infty}(\mathbb{R}^N;\mathbb{R}^{l\times d})} \|\varphi\|_{L^\infty(Q(x_0,r))} \|w(m\cdot)\|_{L^q(Q;\mathbb{R}^d)} r^{\frac{N}{q}+1} \leq C r^{\frac{N}{q}+1}.\end{aligned} \tag{42}$$

In view of (34) the second term in the right-hand side of (41) becomes

$$\sum_{i=1}^N A^i(x_0) \frac{\partial z_m^r(x)}{\partial x_i} = \sum_{i=1}^N A^i(x_0) \frac{\partial \varphi(x)}{\partial x_i} w\Big(\frac{m(x-x_0)}{r}\Big),$$

and thus converges to zero weakly in $L^q(\Omega;\mathbb{R}^l)$, as $m \to +\infty$, due to (34) and by the Riemann-Lebesgue lemma. Hence,

$$\Big\| \sum_{i=1}^N A^i(x_0) \frac{\partial z_m^r(x)}{\partial x_i} \Big\|_{W^{-1,q}(\Omega;\mathbb{R}^l)} \to 0 \quad \text{as } m \to +\infty \tag{43}$$

by the compact embedding of $L^q(\Omega;\mathbb{R}^l)$ into $W^{-1,q}(\Omega;\mathbb{R}^l)$. Finally, the third term in the right-hand side of (41) satisfies

$$\sum_{i=1}^N \frac{\partial A^i(x)}{\partial x_i} z_m^r(x) = \sum_{i=1}^N \frac{\partial A^i(x)}{\partial x_i} \varphi(x) w\Big(\frac{m(x-x_0)}{r}\Big),$$

which again converges to zero weakly in $L^q(\Omega;\mathbb{R}^l)$, as $m \to +\infty$, owing again to (34) and the Riemann-Lebesgue lemma. Therefore,

$$\Big\| \sum_{i=1}^N \frac{\partial A^i(x)}{\partial x_i} z_m^r(x) \Big\|_{W^{-1,q}(\Omega;\mathbb{R}^l)} \to 0 \quad \text{as } m \to +\infty. \tag{44}$$

Claim (40) follows by combining (42)–(44).

Consider the maps

$$v_m^r := P_\eta z_m^r,$$

where P_η is the projection operator introduced in (7). By Proposition 2 we have

$$\|v_m^r\|_{L^q(Q(x_0,r);\mathbb{R}^d)} \leq C\|z_m^r\|_{L^q(\Omega;\mathbb{R}^d)}, \tag{45}$$

$$\|v_m^r\|_{W^{-1,q}(Q(x_0,r);\mathbb{R}^d)} \leq C\|z_m^r\|_{W^{-1,q}(\Omega;\mathbb{R}^d)}, \tag{46}$$

$$\|\mathscr{A}_\eta v_m^r\|_{W^{-1,q}(Q(x_0,r);\mathbb{R}^l)} \leq C\|z_m^r\|_{W^{-1,q}(\Omega;\mathbb{R}^d)}, \tag{47}$$

$$\|v_m^r - z_m^r\|_{L^q(Q(x_0,r);\mathbb{R}^d)} \leq C(\|\mathscr{A}_\eta z_m^r\|_{W^{-1,q}(\Omega;\mathbb{R}^l)} + \|z_m^r\|_{W^{-1,q}(\Omega;\mathbb{R}^d)}). \tag{48}$$

By (39) and (45), the sequence $\{v_m^r\}$ is uniformly bounded in $L^q(Q(x_0,r);\mathbb{R}^d)$. Thus, there exists a map $v^r \in L^q(Q(x_0,r);\mathbb{R}^d)$ such that, up to the extraction of a (not relabelled) subsequence,

$$v_m^r \rightharpoonup v^r \quad \text{weakly in } L^q(Q(x_0,r);\mathbb{R}^d) \tag{49}$$

as $m \to +\infty$. Again by (39), and by the compact embedding of L^q into $W^{-1,q}$, we deduce that

$$z_m^r \to 0 \quad \text{strongy in } W^{-1,q}(\Omega;\mathbb{R}^d) \tag{50}$$

as $m \to +\infty$. Therefore, by combining (46) and (49), we conclude that

$$v_m^r \rightharpoonup 0 \quad \text{weakly in } L^q(Q(x_0,r);\mathbb{R}^d)$$

as $m \to +\infty$, and the convergence holds for the entire sequence. Additionally, by (36), (47), and (50), we obtain

$$\mathscr{A}v_m^r = \mathscr{A}_\eta v_m^r \to 0 \quad \text{strongly in } W^{-1,q}(Q(x_0,r);\mathbb{R}^l)$$

as $m \to +\infty$. Finally, by (40), (48), and (50), there holds

$$\lim_{r\to 0}\lim_{m\to+\infty} r^{-\frac{N}{q}}\|v_m^r - z_m^r\|_{L^q(Q(x_0,r);\mathbb{R}^d)} = 0. \tag{51}$$

We recall that, since x_0 satisfies (23), Step 1 yields

$$\begin{aligned}\frac{d\mathscr{I}(u,v)}{d\mathscr{L}^N}(x_0) &= \lim_{r\to 0^+}\frac{\mathscr{I}((u,v);Q(x_0,r))}{r^N}\\ &\leq \liminf_{r\to 0^+}\liminf_{m\to+\infty}\frac{1}{r^N}\int_{Q(x_0,r)} f(x,u(x),v(x)+v_m^r(x))\,dx.\end{aligned} \tag{52}$$

We claim that

$$\begin{aligned}\frac{d\mathscr{I}(u,v)}{d\mathscr{L}^N}(x_0) &= \lim_{r\to 0^+}\frac{\mathscr{I}((u,v);\,Q(x_0,r))}{r^N}\\ &\leq \liminf_{r\to 0^+}\liminf_{m\to+\infty}\frac{1}{r^N}\int_{Q(x_0,r)} g(x,v(x)+z_m^r(x))\,dx,\end{aligned} \tag{53}$$

where g is the function introduced in Step 3. Indeed, for every $r\in\mathbb{R}$, consider the function $g^r : Q\times\mathbb{R}^d \to [0,+\infty)$ defined as

$$g^r(y,\xi) := g(x_0+ry,\xi) \quad \text{for every } y\in Q,\, \xi\in\mathbb{R}^d.$$

Since $x_0\in\omega$, by (21) there exists K_j such that $x_0\in K_j$. In particular, this yields the existence of $r_0>0$ such that for $r\leq r_0$, the maps g^r are continuous on $Q\times\mathbb{R}^d$, and the family $\{g^r(y,\cdot)\}$ is equicontinuous in $\mathbb{R}^d$, uniformly with respect to y. A change of variables yields

$$\begin{aligned}&\frac{1}{r^N}\Big|\int_{Q(x_0,r)} f(x,u(x),v(x)+v_m^r(x))\,dx - \int_{Q(x_0,r)} f(x,u(x),v(x)+z_m^r(x))\,dx\Big|\\ &= \Big|\int_Q g^r(y,v(x_0+ry)+v_m^r(x_0+ry))\,dy - \int_Q g^r(y,v(x_0+ry)+z_m^r(x_0+ry))\,dy\Big|.\end{aligned}$$

On the other hand, by (51) we have

$$\begin{aligned}&\lim_{r\to 0}\lim_{m\to+\infty}\|z_m^r(x_0+r\cdot)-v_m^r(x_0+r\cdot)\|_{L^q(Q;\mathbb{R}^d)}\\ &\quad= \lim_{r\to 0}\lim_{m\to+\infty} r^{-\frac{N}{q}}\|z_m^r - v_m^r\|_{L^q(Q(x_0,r);\mathbb{R}^d)} = 0.\end{aligned}$$

Therefore, by a diagonal procedure we extract a subsequence $\{m_r\}$ such that

$$\begin{aligned}&\limsup_{r\to 0}\limsup_{m\to+\infty}\Big|\int_Q g^r(y,v(x_0+ry)+v_m^r(x_0+ry))\,dy\\ &\quad -\int_Q g^r(y,v(x_0+ry)+z_m^r(x_0+ry))\,dy\Big|\\ &\quad= \lim_{r\to 0}\Big|\int_Q g^r(y,v(x_0+ry)+v_{m_r}^r(x_0+ry))\,dy\\ &\quad -\int_Q g^r(y,v(x_0+ry)+z_{m_r}^r(x_0+ry))\,dy\Big|,\end{aligned} \tag{54}$$

and

$$z_{m_r}^r(x_0+r\cdot) - v_{m_r}^r(x_0+r\cdot) \to 0 \quad \text{strongly in } L^q(Q;\mathbb{R}^d).$$

In view of (22), (38) and the Riemann-Lebesgue lemma, the sequence $\{v(x_0+r\cdot)+z^r_{m_r}(x_0+r\cdot)\}$ is q-equiintegrable in Q. Hence, by (H) we are under the assumptions of Proposition 3, and we conclude that

$$\lim_{r\to 0}\left|\int_Q g^r(y, v(x_0+ry)+v^r_{m_r}(x_0+ry))\,dy-\int_Q g^r(y, v(x_0+ry)+z^r_{m_r}(x_0+ry))\,dy\right|=0. \tag{55}$$

Claim (53) follows by combining (54) with (55).

Arguing as in [6, Proof of Lemma 3.5], for every $x_0\in\omega$ (where ω is the set defined in (21)) we have

$$\begin{aligned}&\liminf_{r\to 0^+}\liminf_{m\to+\infty}\frac{1}{r^N}\int_{Q(x_0,r)} f(x,u(x),v(x)+z^r_m(x))\,dx\\&\quad\le \liminf_{r\to 0^+}\liminf_{m\to+\infty}\frac{1}{r^N}\int_{Q(x_0,r)} f(x_0,u(x_0),v(x_0)+z^r_m(x))\,dx,\end{aligned}$$

hence by (53) we deduce that

$$\frac{d\mathscr{I}(u,v)}{d\mathscr{L}^N}(x_0)\le \liminf_{r\to 0^+}\liminf_{m\to+\infty}\frac{1}{r^N}\int_{Q(x_0,r)} f(x_0,u(x_0),v(x_0)+z^r_m(x))\,dx.$$

By (38) we obtain

$$\begin{aligned}\frac{d\mathscr{I}(u,v)}{d\mathscr{L}^N}(x_0)&\le \liminf_{r\to 0^+}\liminf_{m\to+\infty}\frac{1}{r^N}\int_{Q(x_0,r)} f(x_0,u(x_0),v(x_0)+z^r_m(x))\,dx\\&\le \liminf_{r\to 0^+}\liminf_{m\to+\infty}\frac{1}{r^N}\left\{\int_{Q(x_0,r)} f\Big(x_0,u(x_0),v(x_0)+w\Big(\frac{m(x-x_0)}{r}\Big)\Big)\,dx\right.\\&\qquad\left.+\int_{Q(x_0,r)\cap\{\varphi\neq 1\}} f\Big(x_0,u(x_0),v(x_0)+\varphi(x)w\Big(\frac{m(x-x_0)}{r}\Big)\Big)\,dx\right\}.\end{aligned}$$

The growth assumption (H) and estimate (37) yield

$$\begin{aligned}&\int_{Q(x_0,r)\cap\{\varphi\neq 1\}} f\Big(x_0,u(x_0),v(x_0)+\varphi(x)w\Big(\frac{m(x-x_0)}{r}\Big)\Big)\,dx \qquad (56)\\&\quad\le C\int_{Q(x_0,r)\cap\{\varphi\neq 1\}}\Big(1+\Big|w\Big(\frac{m(x-x_0)}{r}\Big)\Big|^q\Big)\,dx\\&\quad\le C(1+\|w\|^q_{L^\infty(\mathbb{R}^N;\mathbb{R}^d)})\mathscr{L}^N(Q(x_0,r)\cap\{\varphi\neq 1\})\le C\mu r^N.\end{aligned}$$

Thus, by (56), the periodicity of w, and the Riemann-Lebesgue lemma, we deduce

$$\begin{aligned}\frac{d\mathscr{I}(u,v)}{d\mathscr{L}^N}(x_0) &\leq C\mu \\ &\quad + \liminf_{r\to 0^+}\liminf_{m\to+\infty}\frac{1}{r^N}\int_{Q(x_0,r)} f\Big(x_0, u(x_0), v(x_0) + w\Big(\frac{m(x-x_0)}{r}\Big)\Big)\,dx \\ &= C\mu + \liminf_{m\to+\infty}\int_Q f(x_0, u(x_0), v(x_0) + w(my))\,dy \\ &= C\mu + \int_Q f(x_0, u(x_0), v(x_0) + w(y))\,dy \\ &\leq C\mu + Q_{\mathscr{A}(x_0)} f(x_0, u(x_0), v(x_0)),\end{aligned}$$

where the last inequality is due to (35). Letting $\mu \to 0^+$ we conclude (33).

Acknowledgements The authors thank the Center for Nonlinear Analysis (NSF Grant No. DMS-0635983), where this research was carried out, and also acknowledge support of the National Science Foundation under the PIRE Grant No. OISE-0967140. The research of I. Fonseca and E. Davoli was funded by the National Science Foundation under Grant No. DMS- 0905778. E. Davoli acknowledges the support of the Austrian Science Fund (FWF) projects P 27052 and I 2375. The research of I. Fonseca was further partially supported by the National Science Foundation under Grant No. DMS-1411646.

References

1. Arroyo-Rabasa, A.: Relaxation and optimization for linear-growth convex integral functionals under PDE constraints. J. Funct. Anal. **273**, 2388–2427 (2017)
2. Arroyo-Rabasa, A., De Philippis, G., Rindler, F.: Lower semicontinuity and relaxation of linear-growth convex integral functionals under PDE constraints (To appear Adv. Calc. Var.) (2018). https://doi.org/10.1515/acv-2017-0003
3. Attouch, H.: Variational Convergence for Functions and Operators. Pitman (Advanced Publishing Program), Boston (1984)
4. Baía, M., Chermisi, M., Matias, J., Santos, P. M.: Lower semicontinuity and relaxation of signed functionals with linear growth in the context of $\mathscr{A}$-quasiconvexity. Calc. Var. Partial Differ. Equ. **47**, 465–498 (2013)
5. Benešová, B., Kružík, M.: Weak lower semicontinuity of integral functionals and applications. SIAM Rev. **59**, 703–766 (2017)
6. Braides, A., Fonseca, I., Leoni, G.: $\mathscr{A}$-quasiconvexity: relaxation and homogenization. ESAIM Control Optim. Calc. Var. **5**, 539–577 (2000)
7. Chiodaroli, E., Feireisl, E., Kreml, O., Wiedemann, E.: $\mathscr{A}$-free rigidity and applications to the compressible Euler system. Ann. Mat. Pura Appl. **4**, 1–16 (2017)
8. Dacorogna, B.: Weak Continuity and Weak Lower Semicontinuity of Nonlinear Functionals. Springer, Berlin-New York (1982)
9. Dacorogna, B., Fonseca, I.: A-B quasiconvexity and implicit partial differential equations. Calc. Var. Partial Differ. Equ. **14**, 115–149 (2002)
10. Davoli, E., Fonseca,I.: Homogenization of integral energies under periodically oscillating differential constraints. Calc. Var. Partial Differ. Equ. **55**, 1–60 (2016)

11. Davoli, E., Fonseca, I.: Periodic homogenization of integral energies under space-dependent differential constraints. Port. Math. **73**, 279–317 (2016)
12. De Philippis, G., Rindler, F.: On the structure of $\mathscr{A}$-free measures and applications. Ann. Math. **2**, 1017–1039 (2016)
13. Fonseca, I., Krömer, S.: Multiple integrals under differential constraints: two-scale convergence and homogenization. Indiana Univ. Math. J. **59**, 427–457 (2010)
14. Fonseca, I., Kružík, M.: Oscillations and concentrations generated by $\mathscr{A}$-free mappings and weak lower semicontinuity of integral functionals. ESAIM Control Optim. Calc. Var. **16**, 472–502 (2010)
15. Fonseca, I., Leoni, G., Müller, S.: $\mathscr{A}$-quasiconvexity: weak-star convergence and the gap. Ann. Inst. H. Poincaré Anal. Non Linéaire **21**, 209–236 (2004)
16. Fonseca, I., Müller, S.: Relaxation of quasiconvex functionals in $\mathrm{BV}(\Omega, \mathbf{R}^p)$ for integrands $f(x, u, \nabla u)$. Arch. Ration. Mech. Anal. **123**, 1–49 (1993)
17. Fonseca, I., Müller, S.: $\mathscr{A}$-quasiconvexity, lower semicontinuity, and Young measures. SIAM J. Math. Anal. **30**, 1355–1390 (1999)
18. Krämer, J., Krömer, S., Kružík, M., Pathó, G.: $\mathscr{A}$-quasiconvexity at the boundary and weak lower semicontinuity of integral functionals. Adv. Calc. Var. **10**, 49–67 (2017).
19. Kreisbeck, C., Krömer, S.: Heterogeneous thin films: combining homogenization and dimension reduction with directors. SIAM J. Math. Anal. **48**, 785–820 (2016)
20. Kreisbeck, C., Rindler, F.: Thin-film limits of functionals on $\mathscr{A}$-free vector fields. Indiana Univ. Math. J. **64**, 1383–1423 (2015)
21. Matias, J., Morandotti, M., Santos, P M.: Homogenization of functionals with linear growth in the context of $\mathscr{A}$-quasiconvexity. Appl. Math. Optim. **72**, 523–547 (2015)
22. Murat, F.: Compacité par compensation: condition nécessaire et suffisante de continuité faible sous une hypothèse de rang constant. Ann. Scuola Norm. Sup. Pisa Cl. Sci. **4**, 69–102 (1981)
23. Santos, P. M.: $\mathscr{A}$-quasi-convexity with variable coefficients. Proc. Roy. Soc. Edinburgh Sect. A **134**, 1219–1237 (2004)

Weak Lower Semicontinuity by Means of Anisotropic Parametrized Measures

Agnieszka Kałamajska, Stefan Krömer, and Martin Kružík

Abstract It is well known that besides oscillations, sequences bounded only in L^1 can also develop concentrations, and if the latter occurs, we can at most hope for weak* convergence in the sense of measures. Here we derive a new tool to handle mutual interferences of an oscillating and concentrating sequence with another weakly converging sequence. We introduce a couple of explicit examples showing a variety of possible kinds of behavior and outline some applications in Sobolev spaces.

1 Introduction

Mutual interactions of oscillations and concentrations appears in many problems of optimal control and calculus of variations. We refer, for example, to [7, 25] for optimal control of dynamical systems with oscillations and concentrations, to [26] for a model of mechanical debonding, or to analysis of mechanical problems [29, 30]. Analytical problems related to these phenomena in the calculus of variations are described in detail in [6]. Moreover, oscillations, concentrations, and discontinuities naturally appear in problems of the variational calculus where one is interested in weak lower semicontinuity in the Sobolev space $W^{1,p}(\Omega;\mathbb{R}^m)$ for a sufficiently regular domain $\Omega \subset \mathbb{R}^n$ and $m, n \geq 1$. Indeed, consider

$$I(u) := \int_\Omega h(x, u(x), \nabla u(x))\,\mathrm{d}x \ , \tag{1}$$

A. Kałamajska
Institute of Mathematics, University of Warsaw, Warsaw, Poland
e-mail: Agnieszka.Kalamajska@mimuw.edu.pl

S. Krömer · M. Kružík (✉)
The Czech Academy of Sciences, Institute of Information Theory and Automation, Praha 8, Czech Republic
e-mail: skroemer@utia.cas.cz; kruzik@utia.cas.cz

E. Rocca et al. (eds.), *Trends in Applications of Mathematics to Mechanics*,
Springer INdAM Series 27, https://doi.org/10.1007/978-3-319-75940-1_2

where $h : \bar{\Omega} \times \mathbb{R}^m \times \mathbb{R}^{m\times n} \to \mathbb{R}$ is continuous and such that $|h(x,r,s)| \leq C(1+|r|^q+|s|^p)$ for some $C > 0$, $p > 1$, and $q \geq 1$ so small that $W^{1,p}(\Omega;\mathbb{R}^m)$ compactly embeds into $L^q(\Omega;\mathbb{R}^m)$. If one wants to investigate lower semicontinuity of I with respect to the weak topology in $W^{1,p}(\Omega;\mathbb{R}^m)$, a usual way is to show first that

$$\lim_{k\to\infty}\int_\Omega h(x,u(x),\nabla u_k(x))\,\mathrm{d}x = \lim_{k\to\infty}\int_\Omega h(x,u_k(x),\nabla u_k(x))\,\mathrm{d}x\,. \tag{2}$$

for a suitable sequence $u_k \rightharpoonup u$ in $W^{1,p}(\Omega;\mathbb{R}^m)$, and then to prove that the left-hand side of (2) is bounded from below by $\int_\Omega h(x,u(x),\nabla u(x))\,\mathrm{d}x$. That, however, is not possible without some additional assumptions on h or $\{u_k\}$. We refer to [1] or [5] for such cases. Indeed, if $p \leq n$ then u and u_k, $k \in \mathbb{N}$, are not necessarily continuous and if $\{|\nabla u_k|^p\}$ is not equi-integrable then concentrations can interact with $\{u_k\}_{k\in\mathbb{N}}$. This phenomenon is clearly visible in the following example.

Example 1 Consider $\Omega = B(0,1)$, the unit ball in $\mathbb{R}^n$ centered at the origin, a mapping $w \in W_0^{1,p}(B(0,1);\mathbb{R}^m)$, $p > 1$, extended by zero to the whole space and $u_k(x) := k^{n/p-1}w(kx)$. Hence $u_k \rightharpoonup u := 0$ in $W^{1,p}(B(0,1);\mathbb{R}^m)$ as $k \to \infty$. Assume that h as above is positively p-homogeneous in the last variable, i.e., $h(x,r,\alpha s) = \alpha^p h(x,r,s)$, for all (x,r,s) admissible and all $\alpha \geq 0$. Then a simple calculation yields

$$\begin{aligned}\liminf_{k\to\infty}\int_{B(0,1)} h(x,u_k(x),\nabla u_k(x))\,\mathrm{d}x &= \liminf_{k\to\infty}\int_{B(0,1)} k^n h(x,k^{n/p-1}w(kx),\nabla w(kx))\,\mathrm{d}x\\ &= \liminf_{k\to\infty}\int_{B(0,1)} h(\frac{y}{k},k^{n/p-1}w(y),\nabla w(y))\,\mathrm{d}y\\ &= \begin{cases}\int_{B(0,1)} h(0,w(y),\nabla w(y))\,\mathrm{d}y & \text{if } p=n,\\ \int_{B(0,1)} h(0,0,\nabla w(y))\,\mathrm{d}y & \text{if } p>n,\\ \liminf_{k\to\infty}\int_{B(0,1)} h(y/k,k^{n/p-1}w(y),\nabla w(y))\,\mathrm{d}y & \text{if } p<n.\end{cases}\end{aligned} \tag{3}$$

We see that if $p > n$ then (2) really holds. On the other hand, if $p = n$ the map u appears in the limit besides its gradient and the most complex case is $p < n$ where the limit cannot be calculated explicitly. Notice that the sequence $\{|\nabla u_k|^p\}_{k\in\mathbb{N}} \subset L^1(\Omega)$ is uniformly bounded in this space and concentrates at $x = 0$, i.e., $|\nabla u_k|^p \overset{*}{\rightharpoonup} \|\nabla w\|^p_{L^p(\Omega;\mathbb{R}^m)}\delta_0$ in $\mathscr{M}(\overline{B(0,1)})$ as $k \to \infty$. Here δ_0 denotes the Dirac measure supported at the origin and $\mathscr{M}(\overline{B(0,1)})$ denotes the set of Radon measures on $\overline{B(0,1)}$.

If $p = 1$, concentrations of the gradient can even interact with jump discontinuities.

Example 2 Consider $\Omega = (-1, 1)$ and a sequence $\{u_k\}_{k\in\mathbb{N}} \subset W^{1,1}(-1, 1)$ such that $u_k \to u$ in $L^q(-1, 1)$ for every $1 \leq q < +\infty$. We are interested in

$$\lim_{k\to\infty} \int_{-1}^{1} f(u_k(x))\psi(u_k'(x))\,\mathrm{d}x$$

for continuous function ψ such that $|\psi| \leq C(1+|\cdot|)$ with some constant $C > 0$ and continuous $f : \mathbb{R} \to \mathbb{R}$. If ψ is the identity map then the calculation is easy, namely the limit equals $\liminf_{k\to\infty}(F(u_k(1)) - F(u_k(-1)))$ where F is the primitive of f. In case of more general ψ, the situation is more involved. Let

$$u_k(x) := \begin{cases} 0 & \text{if } -1 \leq x \leq 0, \\ kx & \text{if } 0 \leq x \leq 1/k, \\ 1 & \text{if } 1/k \leq x \leq 1. \end{cases}$$

Assume further that $\lim_{t\to\infty} \psi(t)/t$ exists. Then it is easy to see that

$$\lim_{k\to\infty} \int_{-1}^{1} f(u_k(x))\psi(u_k'(x))\,\mathrm{d}x = (f(0)+f(1))\psi(0) + \Big(\int_0^1 f(r)\,\mathrm{d}r\Big) \lim_{k\to\infty} \frac{\psi(k)}{k}\,. \tag{4}$$

The sequence of $\{u_k'\}_{k\in\mathbb{N}}$ concentrates at zero which is exactly the point of discontinuity of the pointwise limit of $\{u_k\}_{k\in\mathbb{N}}$ which we denote by u. Also notice that $u_k' \overset{*}{\rightharpoonup} \delta_0$ in $\mathscr{M}([-1, 1])$ for $k \to \infty$. Hence, the second term on the right-hand side of (4) suggests that we should refine the definition of u at zero by saying that $u(0)$ is the Lebesgue measure supported on the interval of the jump of u, i.e., on the interval $(0, 1)$.

In this contribution, we introduce a new tool which allows us to describe limits of nonlinear maps along sequences that oscillate, concentrate, and concentrations possibly interfere with discontinuities. While oscillations are successfully treated by Young measures [36] or [4], to handle oscillations and concentrations require finer tools as in, e.g., Young measures and varifolds [2] or DiPerna-Majda measures [9]. We also refer to [24] for an explicit characterization of the DiPerna-Majda measures and to [14, 17] for characterization of those measures which are generated by sequences of gradients, as well as to [21] and [3] for related results in case $p = 1$.

1.1 Basic Notation

Let us start with a few definitions and with the explanation of our notation. If not said otherwise, we will assume throughout this article that $\Omega \subset \mathbb{R}^n$ is a bounded domain with a Lipschitz boundary. Furthermore, $C(\Omega; \mathbb{R}^m)$ (respectively $C(\bar{\Omega}; \mathbb{R}^m)$) is the space of continuous functions defined on Ω (respectively $\bar{\Omega}$) with values in $\mathbb{R}^m$. Here, as well as in similar notation for other function spaces, if the dimension of the target space is $m = 1$, then $\mathbb{R}^m$ is omitted and we only write $C(\Omega)$. In what follows $\mathcal{M}(S)$ denotes the set of regular countably additive set functions on the Borel σ-algebra on a metrizable set S (cf. [10]), its subset, $\mathcal{M}_1^+(S)$, denotes regular probability measures on a set S. We write "γ-almost all" or "γ-a.e." if we mean "up to a set with the γ-measure zero". If γ is the n-dimensional Lebesgue measure we omit writing γ in the notation. The support of a measure $\sigma \in \mathcal{M}(\Omega)$ is the smallest closed set S such that $\sigma(A) = 0$ if $S \cap A = \emptyset$. If $\sigma \in \mathcal{M}(\bar{\Omega})$ we write σ_s and d_σ for the singular part and density of σ defined by the Lebesgue decomposition (with respect to the Lebesgue measure), respectively. By $L^p(\Omega; \mathbb{R}^m)$ we denote the usual Lebesgue space of $\mathbb{R}^m$-valued maps. Further, $W^{1,p}(\Omega; \mathbb{R}^m)$ where $1 \le p \le +\infty$ denotes the usual Sobolev space (of $\mathbb{R}^m$-valued functions) and $W_0^{1,p}(\Omega; \mathbb{R}^m)$ denotes the completion of $C_0^\infty(\Omega, \mathbb{R}^m)$ (smooth functions with support in Ω) in $W^{1,p}(\Omega; \mathbb{R}^m)$. We say that Ω has the extension property in $W^{1,p}$ if every function $u \in W^{1,p}(\Omega)$ can be extended outside Ω to $\tilde{u} \in W^{1,p}(\mathbb{R}^n)$ and the extension operator is linear and bounded. If Ω is an arbitrary domain and $u, w \in W^{1,p}(\Omega, \mathbb{R}^m)$ we say that $u = w$ on $\partial\Omega$ if $u - w \in W_0^{1,p}(\Omega; \mathbb{R}^m)$. We denote by 'w-lim' or by $\rightharpoonup$ the weak limit. Analogously we indicate weak* limits by $\overset{*}{\rightharpoonup}$.

1.2 Quasiconvex Functions

Let $\Omega \subset \mathbb{R}^n$ be a bounded domain. We say that a function $\psi : \mathbb{R}^{m\times n} \to \mathbb{R}$ is quasiconvex (cf. [28]) if for any $s_0 \in \mathbb{R}^{m\times n}$ and any $\varphi \in W_0^{1,\infty}(\Omega; \mathbb{R}^m)$

$$\psi(s_0)|\Omega| \le \int_\Omega \psi(s_0 + \nabla\varphi(x))\,\mathrm{d}x\ .$$

If $\psi : \mathbb{R}^{m\times n} \to \mathbb{R}$ is not quasiconvex we define its quasiconvex envelope $Q\psi : \mathbb{R}^{m\times n} \to \mathbb{R}$ as

$$Q\psi(s) = \sup\left\{h(s);\ h \le \psi;\ h : \mathbb{R}^{m\times n} \to \mathbb{R} \text{ quasiconvex}\right\} \tag{5}$$

and we put $Q\psi = -\infty$ if the set on the right-hand side of (5) is empty. If ψ is locally bounded and Borel measurable then for any $s_0 \in \mathbb{R}^{m\times n}$ (see [8])

$$Q\psi(s_0) = \inf_{\varphi\in W_0^{1,\infty}(\Omega;\mathbb{R}^m)} \frac{1}{|\Omega|}\int_\Omega \psi(s_0 + \nabla\varphi(x))\,\mathrm{d}x\ . \tag{6}$$

1.3 Young Measures

For $p \geq 0$ we define the following subspace of the space $C(\mathbb{R}^{m\times n})$ of all continuous functions on $\mathbb{R}^{m\times n}$:

$$C_p(\mathbb{R}^{m\times n}) = \{\psi \in C(\mathbb{R}^{m\times n});\ \psi(s) = o(|s|^p) text for |s| \to \infty\}\ ,$$

with the obvious modification for any Euclidean space instead of $\mathbb{R}^{m\times n}$. The Young measures on a measurable set $\Lambda \subset \mathbb{R}^l$ are weakly* measurable mappings $x \mapsto \nu_x : \Lambda \to \mathscr{M}(\mathbb{R}^{m\times n})$ with values in probability measures; and the adjective "weakly* measurable" means that, for any $\psi \in C_0(\mathbb{R}^{m\times n})$, the mapping $\Lambda \to \mathbb{R} : x \mapsto \langle \nu_x, \psi\rangle = \int_{\mathbb{R}^{m\times n}} \psi(s)\nu_x(\mathrm{d}s)$ is measurable in the usual sense. Let us remind that, by the Riesz theorem the space $\mathscr{M}(\mathbb{R}^{m\times n})$, normed by the total variation, is a Banach space which is isometrically isomorphic with $C_0(\mathbb{R}^{m\times n})^*$. Let us denote the set of all Young measures by $\mathscr{Y}(\Lambda; \mathbb{R}^{m\times n})$.

Below, we are mostly interested in the case $\Lambda = \Omega$, i.e., a bounded domain. It is known that $\mathscr{Y}(\Omega; \mathbb{R}^{m\times n})$ is a convex subset of $L^\infty_{\mathrm{w}*}(\Omega; \mathscr{M}(\mathbb{R}^{m\times n})) \cong L^1(\Omega; C_0(\mathbb{R}^{m\times n}))^*$, where the index "$w*$" indicates the property "weakly* measurable". A classical result [36] is that, for every sequence $\{y_k\}_{k\in\mathbb{N}}$ bounded in $L^\infty(\Omega; \mathbb{R}^{m\times n})$, there exists its subsequence (denoted by the same indices for notational simplicity) and a Young measure $\nu = \{\nu_x\}_{x\in\Omega} \in \mathscr{Y}(\Omega; \mathbb{R}^{m\times n})$ such that

$$\forall \psi \in C_0(\mathbb{R}^{m\times n}) : \quad \lim_{k\to\infty} \psi \circ y_k = \psi_\nu \qquad \text{weakly* in } L^\infty(\Omega)\ , \tag{7}$$

where $[\psi \circ y_k](x) = \psi(y_k(x))$ and

$$\psi_\nu(x) = \int_{\mathbb{R}^{m\times n}} \psi(s)\nu_x(\mathrm{d}s)\ . \tag{8}$$

Let us denote by $\mathscr{Y}^\infty(\Omega; \mathbb{R}^{m\times n})$ the set of all Young measures which are created by this way, i.e. by taking all bounded sequences in $L^\infty(\Omega; \mathbb{R}^{m\times n})$. Note that (7) actually holds for any $\psi : \mathbb{R}^{m\times n} \to \mathbb{R}$ continuous.

A generalization of this result was formulated by Schonbek [34] (cf. also [4]): if $1 \leq p < +\infty$: for every sequence $\{y_k\}_{k\in\mathbb{N}}$ bounded in $L^p(\Omega; \mathbb{R}^{m\times n})$ there exists its subsequence (denoted by the same indices) and a Young measure $\nu = \{\nu_x\}_{x\in\Omega} \in \mathscr{Y}(\Omega; \mathbb{R}^{m\times n})$ such that

$$\forall \psi \in C_p(\mathbb{R}^{m\times n}) : \quad \lim_{k\to\infty} \psi \circ y_k = \psi_\nu \qquad \text{weakly in } L^1(\Omega)\ . \tag{9}$$

We say that $\{y_k\}$ generates ν if (9) holds. Let us denote by $\mathscr{Y}^p(\Omega; \mathbb{R}^{m\times n})$ the set of all Young measures which are created by this way, i.e. by taking all bounded sequences in $L^p(\Omega; \mathbb{R}^{m\times n})$. The subset of $\mathscr{Y}^p(\Omega; \mathbb{R}^{m\times n})$ containing Young measures generated by gradients of $W^{1,p}(\Omega; \mathbb{R}^m)$ maps will be denoted by

$\mathscr{G}\mathscr{Y}^p(\Omega;\mathbb{R}^{m\times n})$. An explicit characterization of this set is due to Kinderlehrer and Pedregal [18, 19].

1.4 DiPerna-Majda Measures

1.4.1 Definition and Basic Properties

Let $\mathscr{R}$ be a complete (i.e. containing constants, separating points from closed subsets and closed with respect to the Chebyshev norm) separable ring of continuous bounded functions $\mathbb{R}^{m\times n}\to\mathbb{R}$. It is known [11, Sect. 3.12.21] that there is a one-to-one correspondence $\mathscr{R}\leftrightarrow\beta_{\mathscr{R}}\mathbb{R}^{m\times n}$ between such rings and metrizable compactifications of $\mathbb{R}^{m\times n}$ (also see [20] concerning the metrizability); by a compactification we mean here a compact set, denoted by $\beta_{\mathscr{R}}\mathbb{R}^{m\times n}$, into which $\mathbb{R}^{m\times n}$ is embedded homeomorphically and densely. For simplicity, we will not distinguish between $\mathbb{R}^{m\times n}$ and its image in $\beta_{\mathscr{R}}\mathbb{R}^{m\times n}$. Similarly, we will not distinguish between elements of $\mathscr{R}$ and their unique continuous extensions defined on $\beta_{\mathscr{R}}\mathbb{R}^{m\times n}$. This means that if $i:\mathbb{R}^{m\times n}\to\beta_{\mathscr{R}}\mathbb{R}^{m\times n}$ is the homeomorphic embedding and $\psi_0\in\mathscr{R}$ then the same notation is used also for $\psi_0\circ i^{-1}:i(\mathbb{R}^{m\times n})\to\mathbb{R}$ and for its unique continuous extension to $\beta_{\mathscr{R}}\mathbb{R}^{m\times n}$.

Let $\sigma\in\mathscr{M}(\bar{\Omega})$ be a positive Radon measure on a closure of a bounded domain $\Omega\subset\mathbb{R}^n$. A mapping $\hat{\nu}:x\mapsto\hat{\nu}_x$ belongs to the space $L^\infty_{\mathrm{w}*}(\bar{\Omega},\sigma;\mathscr{M}(\beta_{\mathscr{R}}\mathbb{R}^{m\times n}))$ if it is weakly* σ-measurable (i.e., for any $\psi_0\in C_0(\mathbb{R}^{m\times n})$, the mapping $\bar{\Omega}\to\mathbb{R}:x\mapsto\int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}}\psi_0(s)\hat{\nu}_x(\mathrm{d}s)$ is σ-measurable in the usual sense). If additionally $\hat{\nu}_x\in\mathscr{M}_1^+(\beta_{\mathscr{R}}\mathbb{R}^{m\times n})$ for σ-a.a. $x\in\bar{\Omega}$ the collection $\{\hat{\nu}_x\}_{x\in\bar{\Omega}}$ is the so-called Young measure on $(\bar{\Omega},\sigma)$ [36, see also [4, 33]].

DiPerna and Majda [9] showed that given a bounded sequence in $L^p(\Omega;\mathbb{R}^{m\times n})$ with $1\le p<+\infty$ defined on an open domain $\Omega\subseteq\mathbb{R}^n$, there exist a subsequence (denoted by the same indices), a positive Radon measure $\sigma\in\mathscr{M}(\bar{\Omega})$ and a Young measure $\hat{\nu}:x\mapsto\hat{\nu}_x$ on $(\bar{\Omega},\sigma)$ such that $(\sigma,\hat{\nu})$ is attainable by a sequence $\{y_k\}_{k\in\mathbb{N}}\subset L^p(\Omega;\mathbb{R}^{m\times n})$ in the sense that $\forall g\in C(\bar{\Omega})$ and $\forall\psi_0\in\mathscr{R}$:

$$\lim_{k\to\infty}\int_\Omega g(x)\psi(y_k(x))\mathrm{d}x=\int_{\bar{\Omega}}g(x)\int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}}\psi_0(s)\hat{\nu}_x(\mathrm{d}s)\sigma(\mathrm{d}x)\ , \tag{10}$$

where

$$\psi\in\Upsilon^p_{\mathscr{R}}(\mathbb{R}^{m\times n}):=\{\psi_0(1+|\cdot|^p);\ \psi_0\in\mathscr{R}\}. \tag{11}$$

In particular, putting $\psi_0\equiv 1\in\mathscr{R}$ in (10) we can see that

$$\lim_{k\to\infty}(1+|y_k|^p)\ =\ \sigma\qquad\text{weakly* in }\ \mathscr{M}(\bar{\Omega})\ . \tag{12}$$

If (10) holds, we say that $\{y_k\}_{\in\mathbb{N}}$ generates $(\sigma,\hat\nu)$. Let us denote by $\mathscr{DM}^p_{\mathscr{R}}(\Omega;\mathbb{R}^{m\times n})$ the set of all pairs $(\sigma,\hat\nu)\in\mathscr{M}(\bar\Omega)\times L^\infty_{\mathrm{w}*}(\bar\Omega,\sigma;\mathscr{M}(\beta_{\mathscr{R}}\mathbb{R}^{m\times n}))$ attainable by sequences from $L^p(\Omega;\mathbb{R}^{m\times n})$; note that, taking $\psi_0=1$ in (10), one can see that these sequences must be inevitably bounded in $L^p(\Omega;\mathbb{R}^{m\times n})$.

It is well known [33] that (10) can also be rewritten with the help of classical Young measures as

$$\lim_{k\to\infty}\int_\Omega g(x)\psi(y_k(x))\mathrm{d}x=\int_\Omega\int_{\mathbb{R}^{m\times n}}g(x)\psi(s)\nu_x(\mathrm{d}s)\mathrm{d}x \\ +\int_{\bar\Omega}g(x)\int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}\setminus\mathbb{R}^{m\times n}}\psi_0(s)\hat\nu_x(\mathrm{d}s)\sigma(\mathrm{d}x), \qquad (13)$$

where $\{\nu_x\}_{x\in\Omega}\in\mathscr{Y}^\infty(\Omega,\mathbb{R}^{m\times n})$ and $\{\nu_x\}_{x\in\Omega}$ are as in (10).

Formula (13) clarifies connections between Young measures and DiPerna-Majda measures. Namely, the latter ones provide us with more details about behavior of $\{y_k\}$. If $\{|y_k|^p\}\subset L^1(\Omega)$ is uniformly integrable then the second term on the right-hand side of (13) vanishes and $\{y_k\}$ exhibits only oscillations. On the other hand, if for almost all $x\in\Omega$ it holds that $\nu_x=\delta_{y(x)}$ for some $y\in L^p(\Omega;\mathbb{R}^{m\times n})$ then $y_k\to y$ in measure, $\{y_k\}$ does not oscillate but it still can concentrate. Concentrations are then recorded in $(\sigma,\hat\nu)$. We refer to formula (21) below which defines the Young measure given a DiPerna-Majda one. See also [33] for more details.

There are two prominent examples of compactifications of $\mathbb{R}^{m\times n}$. The simplest example is the so-called one point compactification which corresponds to the ring of continuous bounded functions which have limits if the norm of its argument tends to infinity, i.e., we denote $\psi_0(\infty):=\lim_{|s|\to+\infty}\psi_0(s)$.

A richer compactification is the one by the sphere. In that case, we consider the following ring of continuous bounded functions:

$$\mathscr{S}:=\Big\{\psi_0\in C(\mathbb{R}^{m\times n}):\ \text{there exist } c\in\mathbb{R},\ \psi_{0,0}\in C_0(\mathbb{R}^{m\times n}),\ \text{and } \psi_{0,1}\in C(S^{(m\times n)-1})\text{ s.t.} \\ \psi_0(s)=c+\psi_{0,0}(s)+\psi_{0,1}\left(\frac{s}{|s|}\right)\frac{|s|^p}{1+|s|^p}\text{ if } s\neq 0 \text{ and } \psi_0(0)=\psi_{0,0}(0)\Big\}, \qquad (14)$$

where $S^{m\times n-1}$ denotes the $(mn-1)$-dimensional unit sphere in $\mathbb{R}^{m\times n}$. Then $\beta_{\mathscr{R}}\mathbb{R}^{m\times n}$ is homeomorphic to the unit ball $\overline{B(0,1)}\subset\mathbb{R}^{m\times n}$ via the mapping $d:\mathbb{R}^{m\times n}\to B(0,1)$, $d(s):=s/(1+|s|)$ for all $s\in\mathbb{R}^{m\times n}$. Note that $d(\mathbb{R}^{m\times n})$ is dense in $\overline{B(0,1)}$.

The following proposition from [24] explicitly characterizes the set of DiPerna-Majda measures $\mathscr{DM}^p_{\mathscr{R}}(\Omega;\mathbb{R}^{m\times n})$.

Proposition 1 *Let $\Omega\subset\mathbb{R}^n$ be a bounded open domain such that $|\partial\Omega|=0$, $\mathscr{R}$ be a separable complete subring of the ring of all continuous bounded functions on $\mathbb{R}^{m\times n}$ and $(\sigma,\hat\nu)\in\mathscr{M}(\bar\Omega)\times L^\infty_{\mathrm{w}}(\bar\Omega,\sigma;\mathscr{M}(\beta_{\mathscr{R}}\mathbb{R}^{m\times n}))$ and $1\le p<+\infty$. Then the following two statements are equivalent with each other:*

(i) *the pair* $(\sigma, \hat{\nu})$ *is the DiPerna-Majda measure, i.e.* $(\sigma, \hat{\nu}) \in \mathscr{DM}^p_{\mathscr{R}}(\Omega; \mathbb{R}^{m\times n})$,

(ii) *The following properties are satisfied simultaneously:*

1. σ *is positive,*
2. $\sigma_{\hat{\nu}} \in \mathscr{M}(\bar{\Omega})$ *defined by* $\sigma_{\hat{\nu}}(\mathrm{d}x) = (\int_{\mathbb{R}^{m\times n}} \hat{\nu}_x(\mathrm{d}s))\sigma(\mathrm{d}x)$ *is absolutely continuous with respect to the Lebesgue measure* ($d_{\sigma_{\hat{\nu}}}$ *will denote its density),*
3. *for a.a.* $x \in \Omega$ *it holds*

$$\int_{\mathbb{R}^{m\times n}} \hat{\nu}_x(\mathrm{d}s) > 0, \qquad d_{\sigma_{\hat{\nu}}}(x) = \left(\int_{\mathbb{R}^{m\times n}} \frac{\hat{\nu}_x(\mathrm{d}s)}{1+|s|^p}\right)^{-1} \int_{\mathbb{R}^{m\times n}} \hat{\nu}_x(\mathrm{d}s)\ ,$$

4. *for* σ*-a.a.* $x \in \bar{\Omega}$ *it holds*

$$\hat{\nu}_x \geq 0, \qquad \int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}} \hat{\nu}_x(\mathrm{d}s) = 1\ .$$

Remark 1 Consider a metrizable compactification $\beta_{\mathscr{R}}\mathbb{R}^{m\times n}$ of $\mathbb{R}^{m\times n}$ and the corresponding separable complete closed ring $\mathscr{R}$ with its dense subset $\{\psi_k\}_{k\in\mathbb{N}}$. We take a bounded continuous function $\psi : \mathbb{R}^{m\times n} \to \mathbb{R}$, $\psi \notin \mathscr{R}$ and take a closure (in the Chebyshev norm) of all the products of elements from $\{\psi\}\cup\{\psi_k\}_{k\in\mathbb{N}}$. The corresponding ring is again separable and the corresponding compactification is metrizable but strictly finer than $\beta_{\mathscr{R}}\mathbb{R}^{m\times n}$.

2 Anisotropic Parametrized Measures Generated by Pairs of Sequences

This section is devoted to a new tools which might be seen as a multiscale oscillation/concentration measures. It is a generalization of the approach introduced in [32] where only oscillations were taken into account. We also wish to mention that if $\{u_k\}_{k\in\mathbb{N}}$ is bounded in $W^{1,p}(\Omega;\mathbb{R}^m)$ for $1 < p < \infty$ then (at least for a nonrelabeled subsequence) the Young measure generated by the pair $\{(u_k, \nabla u_k)\}$ is $\xi_x(\mathrm{d}(r,s)) = \delta_{u(x)}(\mathrm{d}r)\nu_x(\mathrm{d}s)$ for almost all $x \in \Omega$. Here u is the weak limit of $\{u_k\}_{k\in\mathbb{N}}$ in $W^{1,p}(\Omega;\mathbb{R}^m)$ and $\{\nu_x\}_{x\in\Omega}$ is the Young measure generated by $\{\nabla u_k\}$. We refer to [31] for the proof of this statement. If we are interested also in concentrations of $\{|\nabla u_k|^p\}$ and in their interactions with $\{u_k\}$ the situation is more involved.

As before, let $\mathscr{R}$ be a complete separable ring of continuous bounded functions $\mathbb{R}^{m\times n} \to \mathbb{R}$. Similarly, we take a complete separable ring $\mathscr{U}$ of continuous bounded real-valued functions on $\mathbb{R}^m$, and denote the corresponding metrizable compactification of $\mathbb{R}^m$ by $\beta_{\mathscr{U}}\mathbb{R}^m$. We will consider the ring $C(\bar{\Omega}) \otimes \mathscr{U} \otimes \mathscr{R}$, the subset of bounded continuous functions on $\Omega \times \mathbb{R}^m \times \mathbb{R}^{m\times n}$ spanned by $\{(x,s,r) \mapsto g(x)f_0(r)\psi_0(s) : g \in C(\bar{\Omega}),\ \ f_0 \in \mathscr{U},\ \ \psi_0 \in \mathscr{R}\}$.

Remark 2 Notice that:

1) $\beta_{\mathscr{U}}\mathbb{R}^m \times \beta_{\mathscr{R}}\mathbb{R}^{m\times n} = \beta_{\mathscr{U}\otimes\mathscr{R}}(\mathbb{R}^m \times \mathbb{R}^{m\times n})$;
2) the linear hull of $\{g \otimes f_0 \otimes \psi_0 :\ g \in C(\bar{\Omega})\,,\, f_0 \in C(\beta_{\mathscr{U}})\,,\, \psi_0 \in C(\beta_{\mathscr{R}}\mathbb{R}^{m\times n})\}$ is dense in $C(\bar{\Omega} \times \beta_{\mathscr{U}}\mathbb{R}^m \times \beta_{\mathscr{R}}\mathbb{R}^{m\times n})$ due to the Stone-Weierstrass theorem, where $[g \otimes f_0 \otimes \psi_0](x, r, s) := g(x)f_0(r)\psi_0(s)$ for all $x \in \bar{\Omega}$, $r \in \mathbb{R}^m$, and all $s \in \mathbb{R}^{m\times n}$.

Remark 3 There always exists a separable ring into which a given continuous bounded function f_0 belongs. Indeed, consider a ring $\mathscr{U}_0$ of continuous functions which possess limits if the norm of their argument tends to infinity. This ring to the one-point compactification of $\mathbb{R}^m$. If f_0 does not belong to $\mathscr{U}_0$ we construct a larger ring from f_0 and $\mathscr{U}$ by taking the closure (in the maximum norm) of all products of $\{f_0\} \cup \mathscr{U}$.

2.1 Representation of Limits Using Parametrized Measures

The following statement is rather standard generalization of the DiPerna-Majda Theorem to the anisotropic case. It can be obtained using a special case of the representation theorem in [16].[1]

Theorem 1 *Let* $1 \le q \le +\infty$, $1 \le p < +\infty$ *and*

$$Y^{q,p}(\Omega, \mathscr{U}, \mathscr{R}) = \{h_0(r, s)(1 + |r|^q + |s|^p) : h_0 \in C(\bar{\Omega} \times \beta_{\mathscr{U}}\mathbb{R}^m \times \beta_{\mathscr{R}}\mathbb{R}^{m\times n})\}.$$

Moreover, let $\{u_k\}_{k\in\mathbb{N}}$ *be bounded sequence in* $L^q(\Omega; \mathbb{R}^m)$ *and* $\{w_k\}$ *a bounded sequence in* $L^p(\Omega; \mathbb{R}^{m\times n})$*. Then there is a subsequence* $\{(u_k, w_k)\}$ *(denoted by the same indices), a measure* $\hat{\sigma}(dx)$ *such that*

$$(1 + |u_k|^q + |w_k|^p)dx \overset{*}{\rightharpoonup} \hat{\sigma},$$

and probability measures $\{\hat{\gamma}_x\}_{x\in\bar{\Omega}} \in L^\infty_{\mathrm{w}*}(\bar{\Omega}, \mathscr{M}(\beta_{\mathscr{U}}\mathbb{R}^m \times \beta_{\mathscr{R}}\mathbb{R}^{m\times n}); \hat{\sigma})$ *such that for any* $h \in Y^{q,p}(\Omega, \mathscr{U}, \mathscr{R})$ *and any* $g \in C(\bar{\Omega})$ *we have*

$$\lim_{k\to\infty} \int_\Omega g(x)h_0(u_k(x), w_k(x))(1 + |u_k(x)|^q + |w_k(x)|^p)dx \to \int_{\bar{\Omega}} g(x) \int_{\beta_{\mathscr{U}}\mathbb{R}^m\times\beta_{\mathscr{R}}\mathbb{R}^{m\times n}} h_0(r, s)\hat{\gamma}_x(dr, ds)\hat{\sigma}(\mathrm{d}x).$$

[1] In [16] it is assumed that the compactification of the entire space $\mathbb{R}^m \times \mathbb{R}^{m\times n}$ is a subset in $\mathbb{R}^N$ for some $N \in \mathbb{N}$. This however is not required for the proof in [16] which only uses separability of the compactification.

Remark 4 In a sense, the pair $(\hat{\sigma}, \hat{\gamma})$ is an anisotropic (q, p) DiPerna-Majda measure generated by the sequence $\{(u_k, w_k)\}$, generalizing the isotropic case $p = q$. However, while this approach is a rather intuitive generalization of standard DiPerna-Majda measures, it has a drawback: Several extremely simple and often prototypical choices for the integrands which we would like to use in applications are not admissible. For instance, $h(x, r, s) := |s|^p$ *never* is an element of $Y^{q,p}(\Omega, \mathscr{U}, \mathscr{R})$. Indeed, the limit of $h_0(x, r, s) := |s|^p(1 + |r|^q + |s|^p)^{-1}$ is 1 as $|s| \to \infty$ for fixed r, and it is 0 as $|s| \to \infty$ for fixed r. If h_0 was continuous on product of compactifications, for any $(r, s) \in \beta_{\mathscr{U}} \setminus \mathbb{R}^m \times \beta_{\mathscr{R}} \setminus \mathbb{R}^{m\times n}$ we would have $h_0(r, s) = \lim_{r_n \to r, r_n \in \mathbb{R}^m} h_0(r_n, s) = 0$ and $h_0(r, s) = \lim_{s_n \to s, s_n \to \mathbb{R}^{m\times n}} h_0(r, s_n) = 1$, a contradiction. Hence, this function h_0 does not have a continuous extension to the compactification $\beta_{\mathscr{U}} \times \beta_{\mathscr{R}}$ of $\mathbb{R}^m \times \mathbb{R}^{m\times n}$. Similarly, $h(x, r, s) := |r|^q$ is not admissible, either. Note that this problem is completely independent of the choice of compactifications.

In view of the issue pointed out in Remark 4, we will not use Theorem 1 and its class of anisotropic DiPerna-Majda measures below. Instead, our next statement provides an alternative approach which in particular does allow integrands of the form $h(x, r, s) := |s|^p$.

Theorem 2 *Let $1 \le q \le +\infty$ and $1 \le p < +\infty$. Let $\{u_k\}_{k\in\mathbb{N}}$ be bounded sequence in $L^q(\Omega; \mathbb{R}^m)$ and $\{w_k\}$ a bounded sequence in $L^p(\Omega; \mathbb{R}^{m\times n})$. Then there is a (non-relabeled) subsequence $\{(u_k, w_k)\}$, a DiPerna-Majda measure $(\sigma, \hat{\nu}) \in \mathscr{D}\mathscr{M}^p_{\mathscr{R}}(\Omega; \mathbb{R}^{m\times n})$ and $\hat{\mu} \in \mathscr{Y}(\bar{\Omega} \times \beta_{\mathscr{R}}\mathbb{R}^{m\times n}; \beta_{\mathscr{U}}\mathbb{R}^m)$, such that for every $f_0 \in \mathscr{U}$, every $\psi_0 \in \mathscr{R}$ and every $g \in C(\bar{\Omega})$*

$$\begin{aligned}\lim_{k\to\infty} \int_\Omega & g(x) f_0(u_k(x))\psi(w_k(x))\,\mathrm{d}x \\ &= \int_{\bar{\Omega}} \int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}} \int_{\beta_{\mathscr{U}}\mathbb{R}^m} g(x) f_0(r)\psi_0(s)\hat{\mu}_{s,x}(\mathrm{d}r)\hat{\nu}_x(\mathrm{d}s)\sigma(\mathrm{d}x)\ ,\end{aligned} \tag{15}$$

where $\psi(s) := \psi_0(s)(1 + |s|^p)$. Moreover, measure $(\sigma, \hat{\nu})$ is generated by $\{w_k\}$.

Proof Due to separability of $\mathscr{U}$, $\mathscr{R}$ and of $C(\bar{\Omega})$ there is a (non-relabeled) subsequence of $\{(u_k, w_k)\}$ such that for all $[g \otimes f_0 \otimes \psi_0] \in C(\bar{\Omega}) \times C(\beta_{\mathscr{U}}\mathbb{R}^m) \times C(\beta_{\mathscr{R}}\mathbb{R}^{m\times n})$ and $\psi(s) := \psi_0(s)(1 + |s|^p)$

$$\lim_{k\to\infty} \int_\Omega g(x) f_0(u_k(x))\psi(w_k(x))\,\mathrm{d}x = \langle \Lambda, g \otimes f_0 \otimes \psi_0\rangle\ , \tag{16}$$

for some $\Lambda \in \mathscr{M}(\bar{\Omega} \times \beta_{\mathscr{U}}\mathbb{R}^m \times \beta_{\mathscr{R}}\mathbb{R}^{m\times n})$.

We further define $\hat{T}_\Lambda : \mathscr{U} \times \mathscr{R} \to C(\bar{\Omega})^* = \mathscr{M}(\bar{\Omega})$ by $\left\langle \hat{T}_\Lambda(f_0, \psi_0), g\right\rangle := \langle \Lambda, g \otimes f_0 \otimes \psi_0\rangle$. Let $\sigma \in \mathscr{M}(\bar{\Omega})$ be the weak* limit of $\{1 + |w_k|^p)\}$. Then we see that due to (16)

$$\left|\left\langle \hat{T}_\Lambda(f_0, \psi_0), g\right\rangle\right| = |\langle \Lambda, g \otimes f_0 \otimes \psi_0\rangle| \le \|f_0\|_{C(\mathbb{R}^m)}\|\psi_0\|_{C(\mathbb{R}^{m\times n})}\int_{\bar\Omega} g(x)\,\sigma(\mathrm{d}x)\,. \tag{17}$$

This means that $\hat{T}_\Lambda(f_0, \psi_0)$ is absolutely continuous with respect to σ and by the Radon-Nikodým theorem there is $T_\Lambda : \mathscr{U} \times \mathscr{R} \to L^1(\bar\Omega; \sigma)$ such that for any Borel subset $\omega \subset \bar\Omega$ we get $\hat{T}_\Lambda(f_0, \psi_0)(\omega) = \int_\omega T_\Lambda(f_0, \psi_0)(x)\sigma(\mathrm{d}x)$. Consequently, the right-hand side of (16) can be written as $\int_{\bar\Omega} T_\Lambda(f_0, \psi_0)(x)g(x)\sigma(\mathrm{d}x)$.

As $\mathscr{U} \times \mathscr{R}$ is separable, $\beta_{\mathscr{U}}\mathbb{R}^m \times \beta_{\mathscr{R}}\mathbb{R}^{m\times n}$ is metrizable and separable (with $\mathbb{R}^m \times \mathbb{R}^{m\times n}$ a dense subset) and σ is a regular measure. Hence, the linear span of $C(\bar\Omega) \otimes C(\beta_{\mathscr{U}}\mathbb{R}^m) \otimes C(\beta_{\mathscr{R}}\mathbb{R}^{m\times n})$ is dense in $L^1(\bar\Omega, \sigma; C(\beta_{\mathscr{U}}\mathbb{R}^m \times \beta_{\mathscr{R}}\mathbb{R}^{m\times n}))$ [35, Thm. 1.5.25]. Because of this and (17), Λ can be continuously extended to a continuous linear functional on the space $L^1(\bar\Omega, \sigma; C(\beta_{\mathscr{U}}\mathbb{R}^m \times \beta_{\mathscr{R}}\mathbb{R}^{m\times n}))$; however, the dual of this space is isometrically isomorphic to $L^\infty_{\mathrm{w}}(\bar\Omega, \sigma; \mathscr{M}(\beta_{\mathscr{U}}\mathbb{R}^m \times \beta_{\mathscr{R}}\mathbb{R}^{m\times n}))$. Arguing as in [33, p. 133] we get that there is a family $\lambda := \{\lambda_x\}_{x\in\bar\Omega}$ of probability measures on $\beta_{\mathscr{U}}\mathbb{R}^m \times \beta_{\mathscr{R}}\mathbb{R}^{m\times n}$ which is σ-weak* measurable, that is to say, for any $z \in C(\beta_{\mathscr{U}}\mathbb{R}^m \times \beta_{\mathscr{R}}\mathbb{R}^{m\times n})$, the mapping $\bar\Omega \to \mathbb{R} : x \mapsto \int_{\beta_{\mathscr{U}}\mathbb{R}^m\times\beta_{\mathscr{R}}\mathbb{R}^{m\times n}} z(r, s)\lambda_x(\mathrm{d}r\mathrm{d}s)$ is σ-measurable in the usual sense. Moreover, for σ-almost all $x \in \bar\Omega$ it holds that

$$T_\Lambda(f_0, \psi_0)(x) = \int_{\beta_{\mathscr{U}}\mathbb{R}^m\times\beta_{\mathscr{R}}\mathbb{R}^{m\times n}} f_0(r)\psi_0(s)\lambda_x(\mathrm{d}r\mathrm{d}s)\,. \tag{18}$$

Altogether, we see that (16) can be rewritten as

$$\begin{aligned}\lim_{k\to\infty}\int_\Omega g(x) f_0(u_k(x))\psi(w_k(x))\,\mathrm{d}x \\ = \int_{\bar\Omega} g(x)\int_{\beta_{\mathscr{U}}\mathbb{R}^m\times\beta_{\mathscr{R}}\mathbb{R}^{m\times n}} f_0(r)\psi_0(s)\lambda_x(\mathrm{d}r\mathrm{d}s)\sigma(\mathrm{d}x)\,.\end{aligned} \tag{19}$$

Applying the slicing-measure decomposition (e.g., [13, Theorem 1.45]) to λ_x, we write $\lambda_x(\mathrm{d}r\mathrm{d}s) = \hat\mu_{s,x}(\mathrm{d}r)\hat\nu_x(\mathrm{d}s)$, with a probability measure $\hat\mu_{s,x}$ on $\beta_{\mathscr{U}}\mathbb{R}^m$ for each pair (s, x) and a probability measure $\hat\nu_x$ on $\beta_{\mathscr{R}}\mathbb{R}^{m\times n}$ for each x.

Plugging this decomposition into (19) and testing it with $f_0 := 1$, we get

$$\lim_{k\to\infty}\int_\Omega g(x)\psi(w_k(x))\,\mathrm{d}x = \int_{\bar\Omega} g(x)\int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}} \psi_0(s)\hat\nu_x(\mathrm{d}s)\sigma(\mathrm{d}x)\,. \tag{20}$$

This means that $(\sigma, \hat\nu)$ is the DiPerna-Majda measure *generated by* $\{w_k\}$ [9]. □

In the situation of Theorem 2, passing to a subsequence (not relabeled) if necessary, we may assume in addition that $\{(u_k, w_k)\}$ generates the (classical) Young measure ξ_x. Using the slicing-measure decomposition [12, Thm. 1.5.1] as before, we can always decompose $\xi_x(\mathrm{d}(r,s)) = \mu_{x,s}(\mathrm{d}r)\nu_x(\mathrm{d}s)$, so that

$$\int_\Omega g(x) f_0(u_k)\psi_0(w_k)\,\mathrm{d}x \to \int_\Omega \int_{\mathbb{R}^m\times\mathbb{R}^{m\times n}} g(x) f_0(r)\psi_0(s)\,\xi_x(\mathrm{d}(r,s))\mathrm{d}x$$
$$= \int_\Omega \int_{\mathbb{R}^{m\times n}} \int_{\mathbb{R}^m} g(x) f_0(r)\psi_0(s)\,\mu_{x,s}(\mathrm{d}r)\nu_x(\mathrm{d}s)\mathrm{d}x,$$

in particular for every $f_0 \in \mathscr{U}$, every $\psi_0 \in \mathscr{R}$ and every $g \in C(\bar{\Omega})$. The link between (μ,ν) and $(\hat{\mu},\hat{\nu})$ is the following:

Corollary 1 *In the situation of Theorem 2, let* $\xi_x(\mathrm{d}(r,s)) = \mu_{x,s}(\mathrm{d}r)\nu_x(\mathrm{d}s)$ *be the Young measure generated by* $\{(u_k,w_k)\}$. *Then* $\mathrm{d}x = \left(\int_{\mathbb{R}^{m\times n}} \frac{1}{1+|t|^p}\hat{\nu}_x(\mathrm{d}t)\right)\sigma(\mathrm{d}x)$, *and for a.e.* $x \in \Omega$,

$$\nu_x(\mathrm{d}s) = \left(\int_{\mathbb{R}^{m\times n}} \frac{1}{1+|t|^p}\hat{\nu}_x(\mathrm{d}t)\right)^{-1} \frac{\hat{\nu}_x(\mathrm{d}s)}{1+|s|^p} \tag{21}$$

(this is actually the well known connection between the DiPerna-Majda-measure and the associated Young measure) and

$$\mu_{x,s} = \hat{\mu}_{x,s} \text{ for } \hat{\nu}_x\text{-a.e. } s \in \mathbb{R}^{m\times n} \tag{22}$$

Proof In the following, let $\psi_0 \in C_0(\mathbb{R}^{m\times n})$, i.e., $\psi_0 \in \mathscr{R}$ with the added property that $\psi_0(s) = 0$ for every $s \in \beta_{\mathscr{R}}\mathbb{R}^{m\times n} \setminus \mathbb{R}^{m\times n}$. Consequently, $\psi(s) := \psi_0(s)(1 + |s|^p)$ satisfies $(1+|s|^p)^{-1}\psi(s) \to 0$ as $|s| \to \infty$ ($s \in \mathbb{R}^{m\times n}$) and $\frac{\psi(s)}{1+|s|^p} = 0$ for $s \in \beta_{\mathscr{R}}\mathbb{R}^{m\times n} \setminus \mathbb{R}^{m\times n}$. In addition, let $g \in C(\bar{\Omega})$ and $f_0 \in \mathscr{U}$. From (19), also using the decomposition $\lambda_x(\mathrm{d}r\mathrm{d}s) = \hat{\mu}_{s,x}(\mathrm{d}r)\hat{\nu}_x(\mathrm{d}s)$, we get that

$$\lim_{k\to\infty} \int_\Omega g(x) f_0(u_k(x))\psi(w_k(x))\,\mathrm{d}x$$
$$= \int_{\bar{\Omega}} g(x) \int_{\mathbb{R}^{m\times n}} \int_{\beta_{\mathscr{U}}\mathbb{R}^m} f_0(r)\hat{\mu}_{s,x}(\mathrm{d}r)\frac{\psi(s)\hat{\nu}_x(\mathrm{d}s)}{1+|s|^p}\sigma(\mathrm{d}x)\,. \tag{23}$$

Moreover, since f_0 is bounded, $\{w_k\}$ is bounded in L^p and ψ has less than p-growth, the left hand side can be expressed using the Young measure $\xi_x(\mathrm{d}(r,s)) = \mu_{x,s}(\mathrm{d}r)\nu_x(\mathrm{d}s)$ generated by $\{(u_k,w_k)\}$:

$$\lim_{k\to\infty} \int_\Omega g(x) f_0(u_k(x))\psi(w_k(x))\,\mathrm{d}x = \int_\Omega g(x) \int_{\mathbb{R}^{m\times n}} \int_{\mathbb{R}^m} f_0(r)\mu_{s,x}(\mathrm{d}r)\psi(s)\nu_x(\mathrm{d}s)\mathrm{d}x\,. \tag{24}$$

Since $(\sigma,\hat{\nu})$ is a DiPerna-Majda measure (the one generated by $\{w_k\}$), we in particular know that the density of the Lebesgue measure with respect to σ is given by

$$\frac{\mathrm{d}\mathscr{L}^n}{\mathrm{d}\sigma}(x) = \int_{\mathbb{R}^{m\times n}} \frac{\hat{\nu}_x(\mathrm{d}s)}{1+|s|^p} ,$$

cf. Proposition 1 (ii). Hence, we can also write the outer integral on right hand side of (24) as an integral with respect to σ, and then compare it to the right hand side of (23). Since g is arbitrary, this implies that for σ-a.e. $x \in \Omega$,

$$\begin{aligned}\Big(\int_{\mathbb{R}^{m\times n}} \int_{\mathbb{R}^m} f_0(r)\mu_{s,x}(\mathrm{d}r)\psi(s)\nu_x(\mathrm{d}s)\Big)\Big(\int_{\mathbb{R}^{m\times n}} \frac{\hat{\nu}_x(\mathrm{d}t)}{1+|t|^p}\Big)\\ = \int_{\mathbb{R}^{m\times n}} \int_{\beta_{\mathscr{U}}\mathbb{R}^m} f_0(r)\hat{\mu}_{s,x}(\mathrm{d}r)\frac{\psi(s)\hat{\nu}_x(\mathrm{d}s)}{1+|s|^p} .\end{aligned} \tag{25}$$

Here, also notice that it is enough to state (25) for a.e. $x \in \Omega$, because $\mathscr{L}^n$ is absolutely continuous with respect to σ and $\int_{\mathbb{R}^{m\times n}} \frac{\hat{\nu}_x(\mathrm{d}t)}{1+|t|^p} = 0$ for σ^s-a.e. $x \in \bar{\Omega}$.

Using the probability measure given by the right hand side of (21), i.e.,

$$\nu_x(\mathrm{d}s) := \Big(\int_{\mathbb{R}^{m\times n}} \frac{\hat{\nu}_x(\mathrm{d}t)}{1+|t|^p}\Big)^{-1} \frac{\hat{\nu}_x(\mathrm{d}s)}{1+|s|^p},$$

we see that (25) is equivalent to

$$\int_{\mathbb{R}^{m\times n}} \int_{\mathbb{R}^m} f_0(r)\mu_{s,x}(\mathrm{d}r)\psi(s)\nu_x(\mathrm{d}s)= \int_{\mathbb{R}^{m\times n}} \int_{\beta_{\mathscr{U}}\mathbb{R}^m} f_0(r)\hat{\mu}_{s,x}(\mathrm{d}r)\psi(s)\tilde{\nu}_x(\mathrm{d}s) . \tag{26}$$

Since (26) holds for all $\psi_0 \in C_0(\mathbb{R}^{m\times n})$ (and therefore all ψ with less than p-growth, in particular all bounded ψ) and $\mu_{s,x}$ and $\hat{\mu}_{s,x}$ are probability measures, choosing $f_0 \equiv 1 \in \mathscr{U}$ in (26) yields that $\nu_x = \tilde{\nu}_x$, i.e., (21). Finally, replacing $\tilde{\nu}_x$ by ν_x in (26), and using that the latter holds in particular for all bounded $\psi \in C(\mathbb{R}^{m\times n})$ and all $f_0 \in C_0(\mathbb{R}^m) \subset \mathscr{U}$, we infer (22).

Remark 5 In the situation of Corollary 1, suppose in addition that $u_k \to u$ in L^q for some $q \geq 1$ (for instance by compact embedding, if $\{u_k\}$ is bounded in $W^{1,p}$). We recall that in this case, for the Young measure $\xi_x(\mathrm{d}(r,s)) = \mu_{x,s}(\mathrm{d}r)\nu_x(\mathrm{d}s)$ generated by $\{(u_k, w_k\}$ we have $\mu_{x,s} = \delta_{u(x)}$ for a.e. $x \in \Omega$ (in particular independent of s, cf. [31, Proposition 6.13], e.g.). Consequently, (22) implies that

$$\hat{\mu}_{x,s} = \delta_{u(x)} \text{ for a.e. } x \in \Omega \text{ and } \hat{\nu}_x\text{-a.e. } s \in \mathbb{R}^{m\times n} \tag{27}$$

Remark 6 It is left to the interested reader to show that if $u_k \to u$ in $C(\bar{\Omega};\mathbb{R}^m)$ for $k \to \infty$ then $\hat{\mu}_{s,x} = \delta_{u(x)}$ for σ-a.e. $x \in \bar{\Omega}$. Also, $\hat{\mu}_{s,x}$ is then supported only on $\mathbb{R}^m$, so it is independent of the choice of the compactification $\beta_{\mathscr{U}}\mathbb{R}^m$.

The next statement is similar to Theorem 2, but now we consider the limits of the sequence $\int_\Omega f_0(u_k)\psi_0(w_k)(1+|u_k|^q)\,\mathrm{d}x$ where $f_0 \in \mathscr{U}$, $\psi_0 \in \mathscr{R}$. In particular, the integrand $|u_k|^q$ will thus be admissible. Its proof can easily be deduced by adapting the proof of Theorem 2, essentially interchanging the role of the two sequences.

Theorem 3 *Let $1 \le q < +\infty$ and $1 \le p \le +\infty$. Let $\{u_k\}_{k\in\mathbb{N}}$ be bounded sequence in $L^q(\Omega;\mathbb{R}^m)$ and $\{w_k\}$ a bounded sequence in $L^p(\Omega;\mathbb{R}^{m\times n})$. Then there is a (non-relabeled) subsequence $\{(u_k, w_k)\}$, a positive measure $\sigma^* \in \mathscr{M}(\bar{\Omega})$ and parametrized probability measures $\hat{\nu}^* \in \mathscr{Y}(\bar{\Omega};\beta_{\mathscr{R}}\mathbb{R}^{m\times n})$ (defined σ^*-a.e.) and $\hat{\mu}^* \in \mathscr{Y}(\bar{\Omega}\times\beta_{\mathscr{R}}\mathbb{R}^{m\times n};\beta_{\mathscr{U}}\mathbb{R}^m)$ (defined $\sigma^*\otimes\hat{\nu}^*_x$-a.e.) such that for every $f_0 \in \mathscr{U}$, every $\psi_0 \in \mathscr{R}$ and every $g \in C(\bar{\Omega})$*

$$\begin{aligned}&\lim_{k\to\infty}\int_\Omega g(x)f(u_k(x))\psi_0(w_k(x))\,\mathrm{d}x\\ &\quad=\int_{\bar{\Omega}}\int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}}\int_{\beta_{\mathscr{U}}\mathbb{R}^m} g(x)f_0(r)\psi_0(s)\hat{\mu}^*_{s,x}(\mathrm{d}r)\hat{\nu}^*_x(\mathrm{d}s)\sigma^*(\mathrm{d}x)\;,\end{aligned}\tag{28}$$

where $f(r) := f_0(r)(1+|r|^q)$. Moreover, $(\sigma^*,\overline{\hat{\mu}^*}_x) \in \mathscr{DM}^q_{\mathscr{U}}(\bar{\Omega};\mathbb{R}^m)$ is the the DiPerna-Majda measure generated by $\{u_k\}$, where $\overline{\hat{\mu}^*}_x$ is given as follows:

$$\int_{\beta_{\mathscr{U}}\mathbb{R}^m} f_0(r)\overline{\hat{\mu}^*}_x(\mathrm{d}r) = \int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}}\int_{\beta_{\mathscr{U}}\mathbb{R}^m} f_0(r)\hat{\mu}^*_{s,x}(\mathrm{d}r)\hat{\nu}^*_x(\mathrm{d}s)\tag{29}$$

for all $f_0 \in \mathscr{U}$ and σ^*-a.e. $x \in \bar{\Omega}$.

Analogously to Corollary 1, we have

Corollary 2 *In the situation of Theorem 3, let $\xi_x(\mathrm{d}(r,s)) = \mu_{x,s}(\mathrm{d}r)\nu_x(\mathrm{d}s)$ be the Young measure generated by $\{(u_k, w_k)\}$. Then*

$$\mathrm{d}x = \left(\int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}}\int_{\mathbb{R}^m}\frac{1}{1+|z|^q}\hat{\mu}^*_{x,t}(\mathrm{d}z)\;\hat{\nu}^*_x(\mathrm{d}t)\right)\sigma^*(\mathrm{d}x),$$

and for a.e. $x \in \Omega$,

$$\nu_x(\mathrm{d}s) = \left(\int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}}\int_{\mathbb{R}^m}\frac{1}{1+|z|^q}\hat{\mu}^*_{x,t}(\mathrm{d}z)\hat{\nu}^*_x(\mathrm{d}t)\right)^{-1}\left(\int_{\mathbb{R}^m}\frac{1}{1+|z|^q}\hat{\mu}^*_{x,s}(\mathrm{d}z)\right)\hat{\nu}^*_x(\mathrm{d}s),\tag{30}$$

$$\mu_{x,s}(\mathrm{d}r) = \left(\int_{\mathbb{R}^m}\frac{1}{1+|z|^q}\hat{\mu}^*_{x,s}(\mathrm{d}z)\right)^{-1}\frac{\hat{\mu}_{x,s}(\mathrm{d}r)}{1+|r|^q}\;\textit{ for } \hat{\nu}_x\textit{-a.e. } s \in \beta_{\mathscr{R}}\mathbb{R}^{m\times n}.\tag{31}$$

Analogous to the case of Young measures or DiPerna-Majda-measures, we say that $(\sigma, \hat{\nu}, \hat{\mu})$ [or $(\sigma^*, \hat{\nu}^*, \hat{\mu}^*)$, respectively] *is generated by* $\{(u_k, w_k)\}$ whenever (15) [(28)] holds for all $(g, f_0, \psi_0) \in C(\bar{\Omega}) \times \mathscr{U} \times \mathscr{R}$.

Theorems 2 and 3 can be combined, leading to the following statement. It provides a representation for limits of rather general nonlinear functionals along a given sequence. The suitable class of integrands is

$$\mathbb{H}^{q,p}(\Omega, \mathscr{U}, \mathscr{R}) = \left\{ h \;\middle|\; \begin{array}{l} h(x,r,s) = h_0^{(1)}(x,r,s)(1+|s|^p) + h_0^{(2)}(x,r,s)(1+|r|^q) \\ h_0^{(1)}, h_0^{(2)} \in C(\bar{\Omega} \times \beta_{\mathscr{U}}\mathbb{R}^m \times \beta_{\mathscr{R}}\mathbb{R}^{m\times n}) \end{array} \right\}. \tag{32}$$

According to Remark 2 the linear hull of $\{g \otimes f_0 \otimes \psi_0 \,:\, g \in C(\bar{\Omega}),\, f_0 \in C(\beta_{\mathscr{U}}),\, \psi_0 \in C(\beta_{\mathscr{R}}\mathbb{R}^{m\times n})\}$ is dense in $C(\bar{\Omega} \times \beta_{\mathscr{U}}\mathbb{R}^m \times \beta_{\mathscr{R}}\mathbb{R}^{m\times n})$ and we have the following statement.

Theorem 4 (Representation Theorem) *Let* $1 \le q < +\infty$ *and* $1 \le p < +\infty$. *Let* $\{u_k\}_{k\in\mathbb{N}}$ *be bounded sequence in* $L^q(\Omega; \mathbb{R}^m)$ *and* $\{w_k\}$ *a bounded sequence in* $L^p(\Omega; \mathbb{R}^{m\times n})$. *Then there is a (non-relabeled) subsequence* $\{(u_k, w_k)\}$ *generating the measures* $(\sigma, \hat{\nu}, \hat{\mu})$ *and* $(\sigma^*, \hat{\nu}^*, \hat{\mu}^*)$ *(in the sense of* (15) *and* (28)*, respectively), and in addition, for every* $h_0^{(1)}, h_0^{(2)} \in C(\bar{\Omega} \times \beta_{\mathscr{U}}\mathbb{R}^m \times \beta_{\mathscr{R}}\mathbb{R}^{m\times n})$*,*

$$\begin{aligned} \lim_{k\to\infty} \int_\Omega \big(h_0^{(1)}(x,u_k,w_k)(1+|w_k|^p) + h_0^{(2)}(x,u_k,w_k)(1+|u_k|^q))\big)\,\mathrm{d}x \\ = \int_{\bar{\Omega}} \int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}} \int_{\beta_{\mathscr{U}}\mathbb{R}^m} h_0^{(1)}(x,r,s)\hat{\mu}_{s,x}(\mathrm{d}r)\hat{\nu}_x(\mathrm{d}s)\sigma(\mathrm{d}x) \\ + \int_{\bar{\Omega}} \int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}} \int_{\beta_{\mathscr{U}}\mathbb{R}^m} h_0^{(2)}(x,r,s)\hat{\mu}^*_{s,x}(\mathrm{d}r)\hat{\nu}^*_x(\mathrm{d}s)\sigma^*(\mathrm{d}x)\,. \end{aligned} \tag{33}$$

Remark 7 The cases $p = \infty$ and $q = \infty$ are excluded in Theorem 4, but as long as only one of the two is infinite, either Theorems 2 or 3 can then be used instead.

Remark 8 As a special case, we recover a representation of the limit for functionals with integrands in $Y^{q,p}(\Omega; \mathscr{U}; \mathscr{R})$ as in Theorem 1, since

$$\tilde{h}_0(x,r,s)(1+|r|^q+|s|^p) = h_0(x,r,s)(1+|r|^q) + h_0(x,r,s)(1+|s|^p),$$

where

$$h_0(x,r,s) := \frac{1+|r|^q+|s|^p}{2+|r|^q+|s|^p}\tilde{h}_0(x,r,s)$$

The quotient which appears here does not matter, because $(r, s) \mapsto \frac{1+|r|^q+|s|^p}{2+|r|^q+|s|^p}$ converges to the constant 1 as $|(r, s)| \to \infty$, and therefore it is an element of $\overline{\mathscr{U} \otimes \mathscr{R}} = C(\beta_{\mathscr{U}}\mathbb{R}^m \times \beta_{\mathscr{R}}\mathbb{R}^{m\times n})$.

Remark 9 Given a sequence $\{(u_k, w_k)\}$, the generated measure $(\sigma, \hat{\nu}, \hat{\mu})$ is uniquely determined by (15) in the following sense:

1) σ is unique as measure on $\bar{\Omega}$;
2) $\hat{\nu}_x$ is uniquely defined for σ-almost every $x \in \bar{\Omega}$;
3) there exists a set $E \subset \bar{\Omega}$ with full σ-measure such that for every $x \in E$ and $\hat{\nu}_x$-a.e. $s \in \beta_{\mathscr{R}}\mathbb{R}^{m\times n}$, $\mu_{s,x} \in (C(\beta_{\mathscr{U}}\mathbb{R}^m))^*$ is uniquely defined.

A proof can be obtained by checking that if the right hand side of (15) coincides for two measure triples and all admissible test functions, the measure triples already must be equal in the sense outlined above. In particular, the uniqueness relies on the fact that both $\hat{\nu}_x$ and $\mu_{s,x}$ are probability measures, which also makes them unique as the slicing decomposition of λ_x in the proof of Theorem 2 (otherwise, arbitrary constants factors could be moved from one to the other and only their "product" λ_x would be unique). We omit the details.

The same holds when we deal with $(\sigma^*, \hat{\nu}^*, \hat{\mu}^*)$ instead of $(\sigma, \hat{\nu}, \hat{\mu})$ with obvious modifications.

Remark 10 Notice that $(\sigma, \hat{\nu}, \hat{\mu})$ and $(\sigma^*, \hat{\nu}^*, \hat{\mu}^*)$ are not independent, because they share the same underlying Young measure $\xi_x(\mathrm{d}(r, s)) = \mu_{x,s}(\mathrm{d}r)\nu_x(\mathrm{d}s)$, see Corollary 1 and Corollary 2. Using that, we get yet another representation: For $h \in \mathbb{H}^{q,p}(\Omega, \mathscr{U}, \mathscr{R})$ (cf. (32)),

$$\begin{aligned}
&\lim_{k\to\infty} \int_\Omega h(x, u_k, w_k)\,\mathrm{d}x \\
&= \int_{\bar{\Omega}} \int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}\setminus\mathbb{R}^{m\times n}} \int_{\beta_{\mathscr{U}}\mathbb{R}^m} h_0^{(1)}(x, r, s)\hat{\mu}_{s,x}(\mathrm{d}r)\hat{\nu}_x(\mathrm{d}s)\sigma(\mathrm{d}x) \\
&\quad + \int_{\bar{\Omega}} \int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}} \int_{\beta_{\mathscr{U}}\mathbb{R}^m\setminus\mathbb{R}^m} h_0^{(2)}(x, r, s)\hat{\mu}^*_{s,x}(\mathrm{d}r)\hat{\nu}^*_x(\mathrm{d}s)\sigma^*(\mathrm{d}x) \\
&\quad + \int_\Omega \int_{\mathbb{R}^{m\times n}} \int_{\mathbb{R}^m} h(x, r, s)\mu_{s,x}(\mathrm{d}r)\nu_x(\mathrm{d}s)\mathrm{d}x.
\end{aligned} \tag{34}$$

Remark 11 If either $\{|u_k|^q\}$ or $\{|w_k|^q\}$ is equi-integrable, then (34) can be further simplified. For instance, if $\{u_k\}$ is bounded in $L^{\tilde{q}}$ for some $\tilde{q} > q$), then $\{|u_k|^q\}$ is equi-integrable, and it that case, it is known (e.g., see [33, Lemma 3.2.14]) that for the associated DiPerna-Majda measure $(\sigma^*, \overline{\hat{\mu}^*}_x)$, we have that σ^* is absolutely continuous with respect to $\mathscr{L}^n$ and $\overline{\hat{\mu}^*}_x(\beta_{\mathscr{U}}\mathbb{R}^m \setminus \mathbb{R}^m) = 0$ for a.e. $x \in \Omega$. Due to (29), the latter implies that $\hat{\mu}^*_{x,s}(\beta_{\mathscr{U}}\mathbb{R}^m \setminus \mathbb{R}^m) = 0$ for a.e. $x \in \Omega$ and

$\hat{\nu}^*_x$-a.e. $s \in \beta_{\mathscr{R}}\mathbb{R}^{m\times n}$. Accordingly, for $h \in \mathbb{H}^{q,p}(\Omega, \mathscr{U}, \mathscr{R})$ (cf. (32)),

$$\lim_{k\to\infty}\int_\Omega h(x,u_k,w_k)\,\mathrm{d}x = \int_{\bar\Omega}\int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}\setminus\mathbb{R}^{m\times n}}\int_{\beta_{\mathscr{U}}\mathbb{R}^m} h_0^{(1)}(x,r,s)\hat{\mu}_{s,x}(\mathrm{d}r)\hat{\nu}_x(\mathrm{d}s)\sigma(\mathrm{d}x)$$
$$+\int_\Omega\int_{\mathbb{R}^{m\times n}}\int_{\mathbb{R}^m} h(x,r,s)\mu_{s,x}(\mathrm{d}r)\nu_x(\mathrm{d}s)\mathrm{d}x. \tag{35}$$

2.2 Analysis for Couples $\{(u_k, \nabla u_k)\}$

For the rest of the article, we are mainly interested in sequences of the form $(u_k, w_k) = (u_k, \nabla u_k)$, with a bounded sequence $\{u_k\} \subset W^{1,p}(\Omega;\mathbb{R}^m)$, $1 \le p < \infty$, and integrands $h \in \mathbb{H}^{q,p}(\Omega, \mathscr{U}, \mathscr{R})$ (cf. (32)) for some $q < p^*$. Here, p^* is the exponent of the Sobolev embedding, i.e.,

$$p^* := \begin{cases} pn/(n-p) & \text{if } 1 \le p < n, \\ +\infty & \text{otherwise.}\end{cases}$$

In particular, such integrands satisfy

$$|h(x,r,s)| \le C(1+|r|^q+|s|^p) \quad \text{for all } x \in \bar\Omega,\ r \in \mathbb{R}^m,\ s \in \mathbb{R}^{m\times n}, \tag{36}$$

with a constant $C \ge 0$.

Since we assume that $q < p^*$, we have that $u_k \to u$ strongly in $L^q(\Omega;\mathbb{R}^m)$. In this case, the Young measure generated by $\{u_k\}_{k\in\mathbb{N}}$ is just $\{\delta_{u(x)}\}_{x\in\Omega}$; cf. e.g [31]. Hence, we can represent limits using (35) and $\mu_{x,s} = \delta_{u(x)}$ for all s. This gives the following result.

Theorem 5 *Let $(u_k, w_k) := (u_k, \nabla u_k)$, with a bounded sequence $\{u_k\} \subset W^{1,p}(\Omega;\mathbb{R}^m)$, $1 \le p < \infty$, such that $u_k \rightharpoonup u$ in $W^{1,p}(\Omega;\mathbb{R}^m)$, $\{(\nabla u_k)\}$ generates the (classical) Young measure ν_x in the sense of* (9) *and $\{(u_k, \nabla u_k)\}$ generates the measure $(\sigma, \hat\nu, \hat\mu)$ in the sense of* (15)*. Then for every $h \in \mathbb{H}^{q,p}(\Omega, \mathscr{U}, \mathscr{R})$ (cf.* (32)*),*

$$\lim_{k\to\infty}\int_\Omega h(x,u_k(x),\nabla u_k(x))\,\mathrm{d}x$$
$$= \int_\Omega\int_{\mathbb{R}^{m\times n}} h(x,u(x),s)\nu_x(\mathrm{d}s)\,\mathrm{d}x$$
$$+\int_{\bar\Omega}\int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}\setminus\mathbb{R}^{m\times n}}\int_{\beta_{\mathscr{U}}\mathbb{R}^m} h_0^{(1)}(x,r,s)\hat{\mu}_{s,x}(\mathrm{d}r)\hat{\nu}_x(\mathrm{d}s)\sigma(\mathrm{d}x)\,. \tag{37}$$

Remark 12 If $h(x,u(x),\cdot)$ is quasiconvex, we can further calculate in (37) as follows:

$$\int_{\Omega}\int_{\mathbb{R}^{m\times n}} h(x,u(x),s)\nu_x(\mathrm{d}s)\,\mathrm{d}x \geq \int_{\Omega} h(x,u(x),\nabla u(x))\,\mathrm{d}x\ . \tag{38}$$

Here we exploit the fact that ν is generated by the sequence of gradients $\{\nabla u_k\}$ and therefore it fulfills the mentioned Jensen-like inequality due to the well-known characterization of gradient Young measures by Kinderlehrer and Pedregal [18, 19, 31].

Remark 13 If $p > n$, $W^{1,p}(\Omega;\mathbb{R}^m)$ is compactly embedded in $C(\bar{\Omega};\mathbb{R}^m)$, and therefore $u_k \to u$ uniformly on $\bar{\Omega}$. In view of Remark 6, we then have that $\hat{\mu}_{s,x} = \delta_{u(x)}$ for σ-a.e. $x \in \bar{\Omega}$, for $\hat{\nu}_x$-a.e. $s \in \beta_{\mathscr{R}}\mathbb{R}^{m\times n}$. Hence,

$$\int_{\beta_{\mathscr{U}}\mathbb{R}^m} h_0^{(1)}(x,r,s)\hat{\mu}_{s,x}(\mathrm{d}r) = h_0^{(1)}(x,u(x),s)$$

in the right hand side of (37).

Although an explicit characterization of measures $(\sigma,\hat{\nu},\hat{\mu})$ generated by a sequence of pairs $\{(u_k,\nabla u_k)\}_{k\in\mathbb{N}}$ is not currently available, we can at least characterize DiPerna-Majda measures generated by gradients. The following result can be found in [17] and its extension in [23]. Here and in the sequel d_σ denotes density of the absolutely continuous part of σ with respect to the Lebesgue measure $\mathscr{L}^n$.

Theorem 6 *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with the extension property in $W^{1,p}$, $1 < p < +\infty$ and $(\sigma,\hat{\nu}) \in \mathscr{DM}^p_{\mathscr{R}}(\Omega;\mathbb{R}^{m\times n})$. Then then there is a bounded sequence $\{u_k\}_{k\in\mathbb{N}} \subset W^{1,p}(\Omega;\mathbb{R}^m)$ such that $u_k = u_j$ on $\partial\Omega$ for any $j,k \in \mathbb{N}$ and $\{\nabla u_k\}_{k\in\mathbb{N}}$ generates $(\sigma,\hat{\nu})$ if and only if the following three conditions hold:*

$$\exists u \in W^{1,p}(\Omega;\mathbb{R}^m)\ :\ \textit{for a.a. } x \in \Omega\textit{: } \nabla u(x) = d_\sigma(x)\int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}} \frac{s}{1+|s|^p}\hat{\nu}_x(\mathrm{d}s)\ , \tag{39}$$

for almost all $x \in \Omega$ and for all $\psi_0 \in \mathbb{R}$ and $\psi(s) := (1+|s|^p)\psi_0(s)$,

$$Q\psi(\nabla u(x)) \leq d_\sigma(x)\int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}} \psi_0(s)\hat{\nu}_x(\mathrm{d}s)\ , \tag{40}$$

for σ-almost all $x \in \bar{\Omega}$ and all $\psi_0 \in \mathbb{R}$ with $Q\psi > -\infty$, where $\psi(s) := (1+|s|^p)\psi_0(s)$,

$$0 \leq \int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}\setminus\mathbb{R}^{m\times n}} \psi_0(s)\hat{\nu}_x(\mathrm{d}s)\ . \tag{41}$$

Remark 14 Inequality (40) can be written in terms of $\nu = \{\nu_x\}$, the Young measure generated by $\{u_k\}$, as follows [18]: There exists a zero-measure set $\omega \subset \Omega$ such that for every $x \in \Omega \setminus \omega$

$$\psi(\nabla u(x)) \leq \int_{\mathbb{R}^{m\times n}} \psi(s)\nu_x(\mathrm{d}s) \ , \tag{42}$$

for all $\psi : \mathbb{R}^{m\times n} \to \mathbb{R}$ quasiconvex and such that $|\psi| \leq C(1 + |\cdot|^p)$ for some $C > 0$.

2.3 Examples

Below, we give a couple of examples of sequences and measures from Theorem 2 generated by them (Fig. 1).

Example 3 Let $u_k \in W^{1,1}(0, 2)$ be such that

$$u_k(x) := \begin{cases} 0 & \text{if } 0 \leq x \leq 1 - 1/k, \\ kx - k + 1 & \text{if } 1 - 1/k \leq x \leq 1, \\ -2kx + 2k + 1 & \text{if } 1 \leq x \leq 1 + 1/k, \\ -1 & \text{if } 1 + 1/k \leq x \leq 2. \end{cases}$$

Let $w_k := u_k'$, i.e.,

$$w_k(x) := \begin{cases} 0 & \text{if } 0 \leq x \leq 1 - 1/k, \\ k & \text{if } 1 - 1/k \leq x \leq 1, \\ -2k & \text{if } 1 \leq x \leq 1 + 1/k, \\ 0 & \text{if } 1 + 1/k \leq x \leq 2. \end{cases}$$

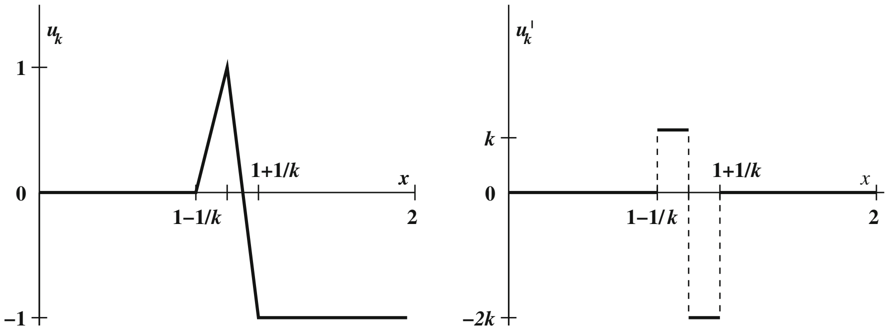

Fig. 1 Sequence $\{u_k, u_k'\}_{k\in\mathbb{N}}$ from Example 3

Let $f_0 \in C(\mathbb{R})$ be bounded with its primitive denoted by F, i.e., $F' = f_0$, $g \in C(\bar{\Omega})$, and let $\psi = \psi_0(1 + |\cdot|)$ where $\psi_0 \in \mathscr{R}$ corresponding to the two-point (or sphere) compactification $\beta_{\mathscr{R}}\mathbb{R} = \mathbb{R} \cup \{\pm\infty\}$, i.e., $\psi_0 \in C(\mathbb{R})$ is such that $\lim_{s\to\pm\infty} \psi_0(s) =: \psi_0(\pm\infty) \in \mathbb{R}$. Then

$$
\begin{aligned}
&\lim_{k\to\infty} \int_0^2 f_0(u_k(x))\psi(w_k(x))g(x)\,\mathrm{d}x \\
&= \lim_{k\to\infty} \Big(\int_0^{1-1/k} f_0(0)\psi_0(0)g(x)\,\mathrm{d}x + \int_{1+1/k}^2 f_0(-1)\psi_0(0)g(x)\,\mathrm{d}x \Big) \\
&\quad + \lim_{k\to\infty} \Big(\int_{1-1/k}^1 f_0(kx-k+1)\psi_0(k)(1+k)g(x)\,\mathrm{d}x \\
&\qquad\qquad + \int_1^{1+1/k} f_0(-2kx+2k+1)\psi_0(-2k)(1+2k)g(x)\,\mathrm{d}x \Big) \\
&= \psi_0(0)(f_0(0) \int_0^1 g(x)\,\mathrm{d}x + f_0(-1) \int_1^2 g(x)\,\mathrm{d}x) \\
&\quad + \lim_{k\to\infty} \Big(\int_{1-1/k}^1 [F(kx-k+1)]'\psi_0(k)\frac{(1+k)}{k}g(x)\,\mathrm{d}x \\
&\qquad\qquad + \int_1^{1+1/k} [F(-2kx+2k+1)]'\psi_0(-2k)\frac{(1+2k)}{-2k}g(x)\,\mathrm{d}x \Big) \\
&= f_0(0)\psi_0(0) \int_0^1 g(x)\,\mathrm{d}x + f_0(-1)\psi_0(0) \int_1^2 g(x)\,\mathrm{d}x \\
&\quad + g(1)(F(1) - F(0))\psi_0(+\infty) + g(1)(F(1) - F(-1))\psi_0(-\infty) \\
&= \int_0^2 \int_{\beta_{\mathscr{R}}\mathbb{R}} \int_{\beta_{\mathscr{U}}\mathbb{R}} g(x)f_0(r)\psi_0(s)\hat{\mu}_{s,x}(\mathrm{d}r)\hat{\nu}_x(\mathrm{d}s)\sigma(\mathrm{d}x)\ ,
\end{aligned}
$$

where $\sigma = \mathscr{L}^1 + 3\delta_1$,

$$
\hat{\nu}_x = \begin{cases} \delta_0 & \text{if } x \in [0,1) \cup (1;2], \\ \frac{1}{3}\delta_\infty + \frac{2}{3}\delta_{-\infty} & \text{if } x = 1, \end{cases}
$$

and

$$
\hat{\mu}_{s,x} = \begin{cases} \delta_0 & \text{if } 0 \le x < 1, \\ \delta_{-1} & \text{if } 1 < x \le 2, \\ \mathscr{L}^1 \llcorner_{(0,1)} & \text{if } s = +\infty \text{ and } x = 1, \\ \frac{1}{2}\mathscr{L}^1 \llcorner_{(-1,1)} & \text{if } s = -\infty \text{ and } x = 1. \end{cases}
$$

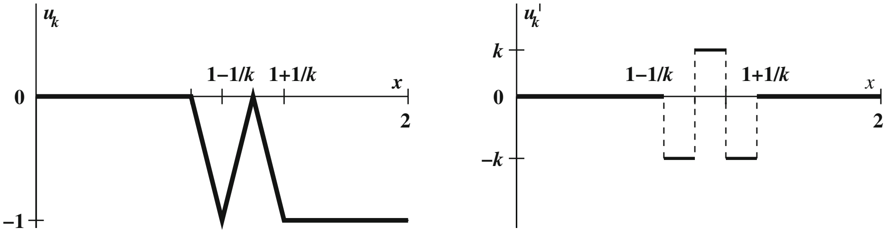

Fig. 2 Sequence $\{u_k, u_k'\}_{k\in\mathbb{N}}$ from Example 4

Changing the previous sequence slightly we get the same measure $(\sigma, \hat{\nu})$, the same limit of $\{u_k\}$ but a different measure $\hat{\mu}$.

Example 4 Let $u_k \in W^{1,1}(0,2)$ be such that (see also Fig. 2)

$$u_k(x) := \begin{cases} 0 & \text{if } 0 \le x \le 1 - 2/k, \\ -kx + k - 2 & \text{if } 1 - 2/k \le x \le 1 - 1/k, \\ kx - k & \text{if } 1 - 1/k \le x \le 1, \\ -kx + k & \text{if } 1 \le x \le 1 + 1/k, \\ -1 & \text{if } 1 + 1/k \le x \le 2. \end{cases}$$

Let $w_k := u_k'$, i.e.,

$$w_k(x) := \begin{cases} 0 & \text{if } 0 \le x \le 1 - 2/k, \\ -k & \text{if } 1 - 2/k \le x \le 1 - 1/k, \\ k & \text{if } 1 - 1/k \le x \le 1, \\ -k & \text{if } 1 \le x \le 1 + 1/k, \\ 0 & \text{if } 1 + 1/k \le x \le 2. \end{cases}$$

Then a computation analogous to the one above shows that
$\sigma = \mathscr{L}^1 + 3\delta_1$,

$$\hat{\nu}_x = \begin{cases} \delta_0 & \text{if } x \in [0,1) \cup (1;2], \\ \frac{1}{3}\delta_{\infty} + \frac{2}{3}\delta_{-\infty} & \text{if } x = 1, \end{cases}$$

and

$$\hat{\mu}_{s,x} = \begin{cases} \delta_0 & \text{if } 0 \le x < 1, \\ \delta_{-1} & \text{if } 1 < x \le 2, \\ \mathscr{L}^1 \llcorner_{(-1,0)} & \text{if } s = -\infty \text{ and } x = 1, \\ \mathscr{L}^1 \llcorner_{(-1,0)} & \text{if } s = +\infty \text{ and } x = 1, \end{cases}$$

These two examples show that $\hat{\mu}$ captures behavior of $\{u_k\}$ and cannot be read off either from $(\sigma, \hat{\nu})$ and/or from u.

Example 5 In the next example, we just set $u_k := u$, where $u(x) := 0$ if $x \in [0, 1]$ and $u(x) = -1$ if $x \in (1; 2]$, and $\{w_k\}_{k\in\mathbb{N}}$ for all $k \in \mathbb{N}$ as before. This gives us

$$
\begin{aligned}
&\lim_{k\to\infty}\int_0^2 f_0(u(x))\psi(w_k(x))g(x)\,\mathrm{d}x\\
&= \lim_{k\to\infty}\Big(\int_0^{1-1/k} f_0(0)\psi_0(0)g(x)\,\mathrm{d}x + \int_{1+1/k}^2 f_0(-1)\psi_0(0)g(x)\,\mathrm{d}x\Big)\\
&\quad + \lim_{k\to\infty}\Big(\int_{1-1/k}^1 f_0(0)\psi_0(k)(1+k)g(x)\,\mathrm{d}x + \int_1^{1+1/k} f_0(-1)\psi_0(-2k)(1+2k)g(x)\,\mathrm{d}x\Big)\\
&= f_0(0)\psi_0(0)\int_0^1 g(x)\,\mathrm{d}x + f_0(-1)\psi_0(0)\int_1^2 g(x)\,\mathrm{d}x\\
&\quad + \lim_{k\to\infty}\Big(k\int_{1-1/k}^1 f_0(0)\psi_0(k)\frac{1+k}{k}g(x)\,\mathrm{d}x + k\int_1^{1+1/k} f_0(-1)\psi_0(-2k)\frac{1+2k}{k}g(x)\,\mathrm{d}x\Big)\\
&= f_0(0)\psi_0(0)\int_0^1 g(x)\,\mathrm{d}x + f_0(-1)\psi_0(0)\int_1^2 g(x)\,\mathrm{d}x\\
&\quad + g(1)f_0(0)\psi_0(+\infty) + 2g(1)f_0(-1)\psi_0(-\infty))\\
&= \int_0^2\int_{\beta_{\mathscr{R}}\mathbb{R}}\int_{\beta_{\mathscr{U}}} g(x)f_0(r)\psi_0(s)\nu_{s,x}(\mathrm{d}r)\hat{\nu}_x(\mathrm{d}s)\sigma(\mathrm{d}x)\ ,
\end{aligned}
$$

where $\sigma = \mathscr{L}^1 + 3\delta_1$,

$$
\hat{\nu}_x = \begin{cases} \delta_0 & \text{if } x \in [0, 1) \cup (1; 2],\\ \frac{1}{3}\delta_\infty + \frac{2}{3}\delta_{-\infty} & \text{if } x = 1, \end{cases}
$$

and

$$
\hat{\mu}_{s,x} = \begin{cases} \delta_0 & \text{if } 0 \le x < 1,\\ \delta_0 & \text{if } x = 1 \text{ and } s = +\infty,\\ \delta_{-1} & \text{if } x = 1 \text{ and } s = -\infty,\\ \delta_{-1} & \text{if } 1 < x \le 2. \end{cases}
$$

In the example below, we calculate the measure $\hat{\mu}$ of a strongly converging sequence.

Example 6 Let $p = 1$, consider the one-point compactification $\beta_{\mathscr{R}}\mathbb{R} = \mathbb{R} \cup \{\infty\}$ of $\mathbb{R}$, and let $\{u_k\}_{k\in\mathbb{N}} \subset W^{1,1}(0, 2)$, $u_k \rightharpoonup u$, be a sequence of nondecreasing functions such that $u_k(0) = 0$ and $u_k(2) = 1$ for all $k \in \mathbb{N}$. In addition, suppose that

$\{u'_k\}_{k\in\mathbb{N}} \subset L^1(0,2)$ converges to zero in measure and it concentrates at $x = 1$, i.e., $\{u'_k\}$ generates $(\sigma, \hat{\nu}) \in \mathscr{DM}^p_{\mathscr{R}}(\Omega; \mathbb{R}^{m\times n})$ given by

$$\sigma = \mathscr{L}^1 + \delta_1, \quad \hat{\nu}_x = \begin{cases} \delta_0 & \text{if } x \in [0,1) \cup (1,2], \\ \delta_\infty & \text{if } x = 1. \end{cases}$$

Moreover, let $\alpha \geq 0$, let $f_0(r) \in C_0(\mathbb{R})$ be such that

$$f_0(r) = \begin{cases} r^\alpha & \text{if } 0 \leq r \leq 1 \\ 1 & \text{for } r \geq 1 \end{cases}$$

and let $\psi(s) := |s|$. As u_k is nondecreasing it must always satisfy $u_k \in [0,1]$, so that $f_0(u_k) = u_k^\alpha$, and $u'_k \geq 0$. Consequently, in view of Theorem 2

$$\begin{aligned} \lim_{k\to\infty} \int_0^2 f_0(u_k(x))\psi(u'_k(x))\,\mathrm{d}x &= \int_0^2 \int_{\beta_{\mathscr{R}}\mathbb{R}} \int_{\beta_{\mathscr{U}}\mathbb{R}} r^\alpha \hat{\mu}_{s,x}(\mathrm{d}r) \frac{s}{1+|s|} \hat{\nu}_x(\mathrm{d}s)\sigma(\mathrm{d}x) \\ &= \int_{\beta_{\mathscr{U}}\mathbb{R}} r^\alpha \hat{\mu}_{\infty,1}(\mathrm{d}r)\ . \end{aligned}$$

On the other hand,

$$\begin{aligned} \lim_{k\to\infty} \frac{1}{\alpha+1}(u_k^{\alpha+1}(2) - u_k^{\alpha+1}(0)) &= \lim_{k\to\infty} \int_0^2 \frac{1}{\alpha+1}(u_k^{\alpha+1}(x))'\,\mathrm{d}x \\ &= \lim_{k\to\infty} \int_0^2 u_k^\alpha(x) u'_k(x)\,\mathrm{d}x \\ &= \int_{\beta_{\mathscr{U}}\mathbb{R}} r^\alpha \hat{\mu}_{\infty,1}(\mathrm{d}r) \\ &= \lim_{k\to\infty} \int_{u_k(0)}^{u_k(2)} r^\alpha\,\mathrm{d}r = \int_0^1 r^\alpha\,\mathrm{d}r\ . \end{aligned}$$

Since $\alpha \geq 0$ is arbitrary and the polynomials are dense in the continuous functions on all compact subsets of $\mathbb{R}$, we infer that

$$\hat{\mu}_{s,x} = \begin{cases} \delta_{u(x)} & \text{if } x \in [0,1) \cup (1,2], \\ \mathscr{L}^1 \llcorner_{(0,1)} & \text{if } x = 1 \text{ and } s = \infty. \end{cases}$$

The measure $\hat{\mu}_{\infty,1}$ is supported on $(0,1)$ because u jumps between zero and one at $x = 1$. However, notice that particular behavior of $\{u_k\}$ in the vicinity of $x = 1$ is not important for the measure.

Let us finally identify the measure generated by the $\{(u_k, u_k')\}$ from Example 2 in case of the one-point compactification for the DiPerna-majda measure. Here we get $\sigma = \mathscr{L} + \delta_0$ and

$$\hat{\nu}_x = \begin{cases} \delta_0 & \text{if } x \in [-1, 0) \cup (0; 1], \\ \delta_\infty & \text{if } x = 1, \end{cases}$$

and

$$\hat{\mu}_{s,x} = \begin{cases} \delta_0 & \text{if } -1 \leq x < 0, \\ \delta_1 & \text{if } 0 < x \leq 1, \\ \mathscr{L}^1 \llcorner_{(0,1)} & \text{if } s = \infty \text{ and } x = 1. \end{cases}$$

3 Applications to Weak Lower Semicontinuity in Sobolev Spaces

We here focus on weak lower semicontinuity of "signed" integral functionals in $W^{1,p}$, i.e., functional whose integrand may have a negative part which has p-growth in the gradient variable. The case of non-negative integrands (or weaker growth in the negative direction) is well-known, see e.g. [1].

Throughout this section, let $\mathscr{U}$ and $\mathscr{R}$ denote rings of bounded continuous functions corresponding to suitable metrizable compactifications $\beta_{\mathscr{U}}\mathbb{R}^m$ and $\beta_{\mathscr{R}}\mathbb{R}^{m\times n}$ of $\mathbb{R}^m$ and $\mathbb{R}^{m\times n}$, respectively, as before. The choice of these rings can be adapted to the particular integrand h at hand in the results presented below. Compactifications by the sphere are sufficiently rich for most practical purposes.

If $p > n$, we can exploit the embedding of $W^{1,p}(\Omega; \mathbb{R}^m)$ into continuous functions on $\bar{\Omega}$. Still, even for quasiconvex integrands concentration effects near the boundary of the domain can prevent lower semicontinuity. However, as it turns out this is the only remaining obstacle. Unlike in the related result of Ball and Zhang [5] where small measurable (but otherwise pretty unknown) sets are removed from the domain, for us it is enough to "peel" away a layer near $\partial\Omega$:

Lemma 1 (Peeling Lemma for $p > n$**)** *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with a boundary of class C^1, let $\infty > p > n$ and let $h \in \mathbb{H}^{q,p}(\Omega, \mathscr{U}, \mathscr{R})$ (cf. (32)). Moreover, assume that $h(x, r, \cdot)$ is quasiconvex for a.e. $x \in \Omega$ (and therefore all $x \in \bar{\Omega}$, by continuity) and every $r \in \mathbb{R}^m$, and let $\{u_k\} \subset W^{1,p}(\Omega; \mathbb{R}^m)$ be a bounded sequence with $u_k \rightharpoonup u$ in $W^{1,p}(\Omega; \mathbb{R}^m)$. Then there exists an increasing sequence of open set Ω_j (possibly depending on the subsequence of $\{u_k\}$) with boundary of class C^∞, $\bar{\Omega}_j \subset \Omega$ and $\bigcup_j \Omega_j = \Omega$ such that*

$$\liminf_{k\to\infty} \int_{\Omega_j} h(x, u_k(x), \nabla u_k(x))\,\mathrm{d}x \geq \int_{\Omega_j} h(x, u(x), \nabla u(x))\,\mathrm{d}x.$$

Proof We select a subsequence of $\{u_k\}$ so that "$\liminf = \lim$" and such that $\{(u_k)\}$ generates a Young measure ν, and $\{(u_k, \nabla u_k)\}$ generates a measure $(\sigma, \hat{\nu}, \hat{\mu})$ in the sense of (15). Now let $\Omega_0 := \emptyset$. For each j, we choose an open set Ω_j with smooth boundary such that

$$K_j := \bar{\Omega}_{j-1} \cup \{x \in \Omega : \operatorname{dist}(x; \partial\Omega) \geq \tfrac{1}{j}\} \subset \Omega_j \subset \bar{\Omega}_j \subset \Omega$$

and

$$\sigma(\partial\Omega_j) = 0 \tag{43}$$

Here, notice that since the distance of the compact set K_j to $\partial\Omega$ is positive, we can find uncountably many pairwise disjoint candidates for Ω_j. Since σ is a finite measure, all but countably many of them must satisfy (43). Clearly, the measure generated by $\{(u_k, \nabla u_k)\}$ on Ω_j coincides with $(\sigma, \hat{\nu}, \hat{\mu})$ on the open set Ω_j, and due to (43) even on $\bar{\Omega}_j$. Hence, by Theorem 5 and Remark 13,

$$\begin{aligned}
\lim_{k\to\infty} \int_{\Omega_j} h(x, u_k(x), \nabla u_k(x))\,\mathrm{d}x &= \int_{\Omega_j} \int_{\mathbb{R}^{m\times n}} h(x, u(x), s)\nu_x(\mathrm{d}s)\,\mathrm{d}x \\
&\quad + \int_{\bar{\Omega}_j} \int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}\setminus\mathbb{R}^{m\times n}} h_0^{(1)}(x, u(x), s)\hat{\nu}_x(\mathrm{d}s)\sigma(\mathrm{d}x) \\
&\geq \int_{\Omega_j} h(x, u(x), \nabla u(x))\,\mathrm{d}x.
\end{aligned}$$

Here, the inequality above is due to Remark 12 and (41) with $\psi_0(s) := h_0^{(1)}(x, u(x), s)$ (separately applied for each x); for $\psi(s) := (1 + |s|^p)\psi_0(s)$ and its quasiconvex hull $Q\psi$ we have $Q\psi > -\infty$ because $h(x, u(x), \cdot)$ is quasiconvex and $\psi(s) - h(x, u(x), s) = h_0^{(2)}(x, u(x), s)(1 + |u(x)|^q)$ is bounded.

To get lower semicontinuity for all sequences and on the whole domain, we need additional assumptions. Theorem 6 can be used to obtain weak lower semicontinuity results along sequences with prescribed boundary data [17]. If we do not control boundary conditions the situation is more complicated. To the best of our knowledge, the first results in this direction are due to Meyers [27] who also deals with higher-order variational problems. However, his condition is stated in terms of sequences. A refinement was proved in [22], showing that even near the boundary, the necessary and sufficient conditions for weak lower semicontinuity in terms of the integrand can be expressed in terms of localized test functions, similar to quasiconvexity:

Theorem 7 ([22, Thm. 1.6]) *Let $1 < p < \infty$, $\Omega \subset \mathbb{R}^n$ be a bounded domain with the C^1-boundary. Let $\tilde{h} : \bar{\Omega} \times \mathbb{R}^{m\times n} \to \mathbb{R}$ be continuous and such that $\tilde{h}(\cdot, s)/(1 + |s|^p)$ is bounded and continuous in $\bar{\Omega}$, uniformly in s. Then*

$J(u) := \int_\Omega \tilde{h}(x, \nabla u(x))\,\mathrm{d}x$ is weakly lower semicontinuous in $W^{1,p}(\Omega; \mathbb{R}^m)$ if and only if the following two conditions hold simultaneously:

(i) $\tilde{h}(x, \cdot)$ is quasiconvex for all $x \in \Omega$;
(ii) for every $x_0 \in \partial\Omega$ and for every $\epsilon > 0$, there exists $C_\epsilon \geq 0$ such that

$$\int_{D_\varrho} \tilde{h}(x_0, \nabla\varphi(x))\,\mathrm{d}x \geq -\epsilon \int_{D_\varrho} |\nabla\varphi(x)|^p\,\mathrm{d}x - C_\epsilon \text{ for every } \varphi \in C_c^\infty(B(0,1); \mathbb{R}^m).$$

Here, $D_\varrho := \{x \in B(0,1);\ x \cdot \varrho < 0\}$ where ϱ denotes the outer unit normal to $\partial\Omega$ at x_0.

Definition 1 (p-Quasisubcritical Growth from Below) If $\tilde{h}$ satisfies (ii) in Theorem 7, we say that it has p-quasisubcritical growth from below (p-qscb) at x_0.

With the help of the results of Sect. 2, we can provide an extension of this result to integrands that also depend on u, at least if $p > n$:

Theorem 8 *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with a boundary of class C^1, let $\infty > p > n$ and let $h \in \mathbb{H}^{q,p}(\Omega, \mathscr{U}, \mathscr{R})$ (cf. (32)). Then, if $h(x, r, \cdot)$ is quasiconvex for a.e. $x \in \Omega$ (and therefore all $x \in \bar{\Omega}$, by continuity) and all $r \in \mathbb{R}^m$ and $\tilde{h}(x, s) := h(x, u(x), s)$ has p-quasisubcritical growth from below for all $x \in \partial\Omega$ and all $u \in W^{1,p}(\Omega; \mathbb{R}^m)$, $w \mapsto \int_\Omega h(x, w(x), \nabla w(x))\,\mathrm{d}x$ is weakly lower semicontinuous in $W^{1,p}(\Omega; \mathbb{R}^m)$.*

Proof Let $u_k \rightharpoonup u$ weakly in $W^{1,p}(\Omega; \mathbb{R}^m)$. In view of Remark 13, the measures generated by (subsequences of) $\{(u, \nabla u_k)\}$ and $\{(u_k, \nabla u_k)\}$ in the sense of (15) always coincide. As a consequence of (35) and (37), it therefore suffices to show that for each $u \in W^{1,p}(\Omega; \mathbb{R}^m) \subset C(\bar{\Omega}; \mathbb{R}^m)$, $w \mapsto \int_\Omega h(x, u(x), \nabla w(x))\,\mathrm{d}x$ is weakly lower semicontinuous. The latter follows from Theorem 5.

Remark 15 In Theorem 8, quasiconvexity of $h(x, u(x), \cdot)$ in Ω and p-qscb of $h(x, u(x), \cdot)$ at every $x \in \partial\Omega$ are also necessary for weak lower semicontinuity. We omit the details.

As already briefly pointed out in the introduction, the situation becomes significantly more complicated if $p \leq n$. Using our measures to express the limit as in Theorem 5, we can at least reduce the problem to a property of an integrand without explicit dependence on u, for each given sequence:

Proposition 2 *Let $p \leq n$, suppose that $h(x, r, \cdot)$ is quasiconvex, $h \in \mathbb{H}^{q,p}(\Omega, \mathscr{U}, \mathscr{R})$, and let $\{u_k\} \subset W^{1,p}(\Omega; \mathbb{R}^m)$ be a bounded sequence such that $u_k \rightharpoonup u$ and $\{(u_k, \nabla u_k)\}$ generates a measure $(\sigma, \hat{\nu}, \hat{\mu})$ in the sense of (15). Then*

$$\liminf_{k\to\infty} \int_\Omega h(x, u_k, \nabla u_k)\,\mathrm{d}x \geq \int_\Omega h(x, u, \nabla u)\,\mathrm{d}x,$$

provided that for σ-a.e. $x \in \bar{\Omega}$,

$$\int_{\bar{\Omega}} \int_{\beta_{\mathscr{R}}\mathbb{R}^{m\times n}\setminus\mathbb{R}^{m\times n}} \tilde{h}(x,s)\,\hat{\nu}_x(\mathrm{d}s)\sigma(\mathrm{d}x) \geq 0, \tag{44}$$

where $\tilde{h}(x,s) := \int_{\beta_{\mathscr{U}}} h_0^{(1)}(x,r,s)\,\hat{\mu}_{x,s}(\mathrm{d}r)$. *Here, recall that* $h(x,r,s) = h_0^{(1)}(x,r,s)(1+|s|^p) + h_0^{(2)}(x,r,s)(1+|r|^q)$, *cf.* (32).

Proof This is a straightforward consequence of Theorem 5 and Remark 12.

Remark 16 Given $h \in \mathbb{H}^{q,p}(\Omega,\mathscr{U},\mathscr{R})$, $h_0^{(1)}(x,r,s)$ is uniquely determined for $s \in \beta_{\mathscr{R}}\mathbb{R}^{m\times n} \setminus \mathbb{R}^{m\times n}$, but not for $s \in \mathbb{R}^{m\times n}$. Of course, (44) actually is only a condition on the restriction of $h_0^{(1)}$ to $\bar{\Omega} \times \beta_{\mathscr{U}}\mathbb{R}^m \times (\beta_{\mathscr{R}}\mathbb{R}^{m\times n} \setminus \mathbb{R}^{m\times n})$.

4 Concluding Remarks

We have seen that generalized DiPerna-Majda measures introduced here can be helpful in proofs of weak lower semicontinuity. Other applications are, for example, in impulsive control problems where the concentration of controls typically results in discontinuity of the state variable [15]. An open challenging problem is to find some explicit characterization of generalized Diperna-Majda measures generated by pairs of functions and their gradients, namely $\{(u_k, \nabla u_k)\} \subset W^{1,p}(\Omega;\mathbb{R}^m) \times L^p(\Omega;\mathbb{R}^{m\times n})$. This could then help us to find necessary and sufficient conditions for weak lower semicontinuity of $u \mapsto \int_\Omega h(x,u(x),\nabla u(x))\,\mathrm{d}x$ in $W^{1,p}(\Omega;\mathbb{R}^m)$ for $1 < p < +\infty$ and for $h \in \mathbb{H}^p$.

Acknowledgements This work was partly done during MK's visiting Giovanni-Prodi professorship at the University of Würzburg, Germany. The hospitality and support of the Institute of Mathematics is gratefully acknowledged. This work was also supported by GAČR through projects 16-34894L and 17-04301S. The work of AK was supported by NCN grant 2011/03/N/ST1/00111.

References

1. Acerbi, E., Fusco, N.: Semicontinuity problems in the calculus of variations. Arch. Ration. Mech. Anal. **86**, 125–145 (1984)
2. Alibert, J., Bouchitté, G.: Non-uniform integrability and generalized Young measures. J. Convex Anal. **4**, 125–145 (1997)
3. Baía, M., Krömer, S., Kružík, M.: Generalized $\mathbf{W^{1,1}}$-Young measures and relaxation of problems with linear growth. Preprint arXiv:1611.04160v1, submitted (2016)
4. Ball, J.M.: A version of the fundamental theorem for Young measures. In: Rascle, M., Serre, D., Slemrod, M. (eds.) PDEs and Continuum Models of Phase Transition. Lecture Notes in Physics, vol. 344, pp. 207–215. Springer, Berlin (1989)

5. Ball, J.M., Zhang K.-W.: Lower semicontinuity of multiple integrals and the biting lemma. Proc. Roy. Soc. Edinburgh **114A**, 67–379 (1990)
6. Benešová, B., Kružík, M.: Weak lower semicontinuity of integral functionals and applications. SIAM Rev. **59**, 703–766 (2017)
7. Claeys, M., Henrion, D., Kružík, M.: Semi-definite relaxations for optimal control problems with oscillations and concentration effects. ESAIM Control Optim. Calc. Var. **23**, 95–117 (2017)
8. Dacorogna, B.: Direct Methods in the Calculus of Variations, 2nd edn. Springer, Berlin (2008)
9. DiPerna, R.J., Majda, A.J.: Oscillations and concentrations in weak solutions of the incompressible fluid equations. Commun. Math. Phys. **108**, 667–689 (1987)
10. Dunford, N., Schwartz, J.T.: Linear Operators, Part I. Interscience, New York (1967)
11. Engelking, R.: General topology. Translated from the Polish by the author, 2nd edn. Heldermann Verlag, Berlin (1989)
12. Evans, L.C.: Weak Convergence Methods for Nonlinear Partial Differential Equations. AMS, Providence (1990)
13. Evans, L.C., Gariepy, R.F.: Measure Theory and Fine Properties of Functions. CRC Press, Boca Raton (1992)
14. Fonseca, I., Müller, S., Pedregal, P.: Analysis of concentration and oscillation effects generated by gradients. SIAM J. Math. Anal. **29**, 736–756 (1998)
15. Henrion, D., Kružík, M., Weisser, T.: Optimal control problems with oscillations, concentrations, and discontinuities. In preparation (2017)
16. Kałamajska, A.: On Young measures controlling discontinuous functions. J. Conv. Anal. **13**(1), 177–192 (2006)
17. Kałamajska, A., Kružík, M.: Oscillations and concentrations in sequences of gradients. ESAIM Control Optim. Calc. Var. **14**, 71–104 (2008)
18. Kinderlehrer, D., Pedregal, P.: Characterization of Young measures generated by gradients. Arch. Ration. Mech. Anal. **115**, 329–365 (1991)
19. Kinderlehrer, D., Pedregal, P.: Gradient Young measures generated by sequences in Sobolev spaces. J. Geom. Anal. **4**, 59–90 (1994)
20. Kozarzewski, P.: On certain compactifcation of an arbitrary subset of $\mathbb{R}^n$ and its applications to DiPerna-Majda measures theory. In preparation
21. Kristensen, J., Rindler, F.: Characterization of generalized gradient Young measures generated by sequences in $W^{1,1}$ and BV. Arch. Ration. Mech. Anal. **197**, 539–598 (2010); Erratum **203**, 693–700 (2012)
22. Krömer, S.: On the role of lower bounds in characterizations of weak lower semicontinuity of multiple integrals. Adv. Calc. Var. **3**, 387–408 (2010)
23. Krömer, S., Kružík, M.: Oscillations and concentrations in sequences of gradients up to the boundary. J. Convex Anal. **20**, 723–752 (2013)
24. Kružík, M., Roubíček, T.: On the measures of DiPerna and Majda. Mathematica Bohemica **122**, 383–399 (1997)
25. Kružík, M., Roubíček, T.: Optimization problems with concentration and oscillation effects: relaxation theory and numerical approximation. Numer. Funct. Anal. Optim. **20**, 511–530 (1999)
26. Licht, C., Michaille, G., Pagano, S.: A model of elastic adhesive bonded joints through oscillation-concentration measures. J. Math. Pures Appl. **87**, 343–365 (2007)
27. Meyers, N.G.: Quasi-convexity and lower semicontinuity of multiple integrals of any order. Trans. Am. Math. Soc. **119**, 125–149 (1965)
28. Morrey, C.B.: Multiple Integrals in the Calculus of Variations. Springer, Berlin (1966)
29. Paroni, R., Tomassetti, G.: A variational justification of linear elasticity with residual stress. J. Elasticity **97**, 189–206 (2009)
30. Paroni, R., Tomassetti, G.: From non-linear elasticity to linear elasticity with initial stress via Γ-convergence. Cont. Mech. Thermodyn. **23**, 347–361 (2011)
31. Pedregal, P.: Parametrized Measures and Variational Principles. Birkäuser, Basel (1997)
32. Pedregal, P.: Multiscale Young measures. Trans. Am. Math. Soc. **358**, 591–602 (2005)

33. Roubíček, T.: Relaxation in Optimization Theory and Variational Calculus. W. de Gruyter, Berlin (1997)
34. Schonbek, M.E.: Convergence of solutions to nonlinear dispersive equations. Comm. Partial Differ. Equ. **7**, 959–1000 (1982)
35. Warga, J.: Optimal Control of Differential and Functional Equations. Academic Press, New York (1972)
36. Young, L.C.: Generalized curves and the existence of an attained absolute minimum in the calculus of variations. Comptes Rendus de la Société des Sciences et des Lettres de Varsovie, Classe III **30**, 212–234 (1937)

What Does Rank-One Convexity Have to Do with Viscosity Solutions?

Pablo Pedregal

Abstract Relying on Hilbert's classical theorem for non-negative polynomials as a main tool, we show that rank-one convex functions for 2 × 2-matrices admit a decomposition as a sum of a multiple of the determinant and a viscosity solution of a certain equation.

1 Introduction

The paradigmatic problem in the Calculus of Variations for vector problems is that of minimizing an integral cost functional of the form

$$\int_{\Omega} \phi(\nabla \mathbf{u}(\mathbf{x}))\, d\mathbf{x}, \quad \mathbf{u}(\mathbf{x}) : \Omega \subset \mathbb{R}^N \to \mathbb{R}^m, N, m > 1, \tag{1}$$

where structural properties of the density

$$\phi(\mathbf{F}) : \mathbb{R}^{m\times N} \to \mathbb{R}$$

determine fundamental properties of the corresponding cost functional. Such variational problems are of paramount importance in non-linear elasticity [1, 3, 5] where they represent non-quadratic internal energies associated with deformations **u** of the (hyper)elastic body under consideration, characterized by its own internal energy density ϕ. In particular, minimizers of the integral energy in (1) represent stable states of the body, and so a basic fundamental problem is to understand under which sets of assumptions, the existence of such equilibrium configurations may be shown. This job depends on the properties of the energy density ϕ.

P. Pedregal (✉)
INEI, U. Castilla-La Mancha, Ciudad Real, Spain
e-mail: pablo.pedregal@uclm.es

E. Rocca et al. (eds.), *Trends in Applications of Mathematics to Mechanics*,
Springer INdAM Series 27, https://doi.org/10.1007/978-3-319-75940-1_3

The way in which one can try to understand the existence of minimizers in (1) under competing deformations $\mathbf{u} : \Omega \subset \mathbb{R}^N \to \mathbb{R}^m$ complying with standard Dirichlet-type boundary conditions $\mathbf{u} - \mathbf{u}_0 \in W_0^{1,p}(\Omega; \mathbb{R}^m)$ proceeds through the direct methods of the Calculus of Variations [7], whose main ingredient is the (sequential) weak lower semicontinuity properties ensuring that the weak convergence $\mathbf{u}_j \rightharpoonup \mathbf{u}$ in $W^{1,p}(\Omega; \mathbb{R}^m)$ implies

$$\int_\Omega \phi(\nabla \mathbf{u}(\mathbf{x}))\, d\mathbf{x} \le \liminf_{j\to\infty} \int_\Omega \phi(\nabla \mathbf{u}_j(\mathbf{x}))\, d\mathbf{x}.$$

What are the properties of ϕ guaranteeing this weak lower semicontinuity? This has been a main concern since the beginning of the discipline. For the scalar case when either of the two dimensions N or m is unity, it was very well-understood since the time of Tonelli [17] that convexity of ϕ was the necessary and sufficient condition for the weak lower semicontinuity of the corresponding functional. However, it was Morrey [12, 13], who in the 50's, realized that for vector problems, when both dimension N and m are greater than one, the situation could be much more involved. Convexity was definitely a sufficient condition for weak lower semicontinuity, but given that this property was incompatible with other physical requirements in non-linear elasticity [5], more general conditions were to be found.

It was Morrey himself who introduced the concept of quasi convexity which, in this context, means

$$\int_Q \phi(\mathbf{F} + \nabla \mathbf{u}(\mathbf{x}))\, d\mathbf{x} \ge \phi(\mathbf{F}), \qquad \text{for all } \mathbf{u}\text{, } Q\text{-periodic, and all } \mathbf{F}. \tag{2}$$

Here Q is the unit cube in $\mathbb{R}^N$. This is not the form in which Morrey introduced quasi convexity , but it can easily be proved to be equivalent to this form. The issue is, however, far from being settled because, in practice, (2) is almost impossible to check. Necessary conditions were first sought, and rank-one convexity was shown (by Morrey) to be the main such condition. A function ϕ like the integrand in (1) is said to be rank-one convex if the sections

$$t \mapsto \phi(\mathbf{F} + t\mathbf{a} \otimes \mathbf{n})$$

are convex functions of the single variable t for all $\mathbf{F} \in \mathbb{R}^{m\times N}$, $\mathbf{a} \in \mathbb{R}^m$, $\mathbf{n} \in \mathbb{R}^N$. Then important sufficient conditions were given [3] in the form of polyconvexity. An integrand like $\phi(\mathbf{F})$ is said to be polyconvex if

$$\phi(\mathbf{F}) = \Phi(\mathbf{F}, M(\mathbf{F})), \qquad M(\mathbf{F}), \text{ vector of all subdeterminants of } \mathbf{F},$$

and Φ is a convex (in the usual sense) function of all its arguments. This is the main structural condition that allows for existence theorems in non-linear elasticity [3].

It was then clear that

$$convexity \Longrightarrow polyconvexity \Longrightarrow quasiconvexity \Longrightarrow rank{-}one\,convexity,$$

and a lot of collective effort was devoted to distinguishing among all these convexity notions. As it turns out, (almost) all reverse implications are false. In particular the last one (whether rank-one convexity is equivalent to quasi convexity) stood longer as unsolved, until the remarkable counterexample by V. Sverak [16] who constructed a fourth degree, rank-one convex polynomial that is not quasi convex. That kind of examples are valid for $m \geq 3$, and several attempts to extend it to $m = 2$ failed [15], so that it is still a main open problem in the field to prove or disprove if rank-one convexity implies quasi convexity for two-dimensional deformations.

This is the situation in which we would like to place ourselves for this contribution. Our densities $\phi : \mathbf{M}^{2\times 2} \to \mathbb{R}$ correspond to the case of mappings $\mathbf{u} : Q \subset \mathbb{R}^2 \to \mathbb{R}^2$. To state our main result, we take into account the following notation. The letter $\mathbf{F}$ is an independent variable with four components corresponding to its four entries as a 2×2-matrix. We focus on the function

$$\lambda_1(\mathbf{X}) : \mathbb{R}^{2\times 2} \to \mathbb{R}$$

providing the smallest eigenvalue of the symmetric matrix $\mathbf{X}$. The negative of this function $-\lambda_1$ is degenerate elliptic according to Example 1.8 in the celebrated reference [6]. Our main result is the following.

Theorem 1 *The $\mathscr{C}^2$-smooth function $\phi : \mathbf{M}^{2\times 2} \to \mathbb{R}$ is rank-one convex if and only if there is a function $\alpha : \mathbf{M}^{2\times 2} \to \mathbb{R}$, such that $\phi(\mathbf{F}) = \psi(\mathbf{F}) + \alpha(\mathbf{F})\,\mathrm{Det}\,\mathbf{F}$ with ψ a viscosity sub-solution of*

$$-\lambda_1[\nabla^2\psi(\mathbf{F}) + \mathrm{Det}\,\mathbf{F}\nabla^2\alpha(\mathbf{F}) + \nabla\alpha(\mathbf{F}) \otimes \mathbf{DF} + \mathbf{DF} \otimes \nabla\alpha(\mathbf{F})] = 0,$$

i.e.

$$\lambda_1[\nabla^2\psi(\mathbf{F}) + \mathrm{Det}\,\mathbf{F}\nabla^2\alpha(\mathbf{F}) + \nabla\alpha(\mathbf{F}) \otimes \mathbf{DF} + \mathbf{DF} \otimes \nabla\alpha(\mathbf{F})] \geq 0,$$

in every domain where α is smooth.

One may have an impression that those rank-one convex functions for which the ψ's in this result are in fact viscosity solutions, instead of just sub-solutions as this theorem states, of the equation might play a special role. This is something to be further investigated but, under a suitable set of assumptions, there are such rank-one convex functions.

Theorem 2 *Let $\alpha : \mathbf{M}^{2\times 2} \to \mathbb{R}$ be $\mathscr{C}^2$. Suppose that for a bounded, open subset $\Omega \subset \mathbf{M}^{2\times 2}$ two functions $\psi^+(\mathbf{F})$, $\psi^-(\mathbf{F})$, can be found so that $\psi^+ = \psi^-$ on $\partial\Omega$, and if we put*

$$\phi^\pm(\mathbf{F}) = \psi^\pm(\mathbf{F}) + \alpha(\mathbf{F})\,\mathrm{Det}\,\mathbf{F},$$

α is smooth ($\mathscr{C}^2$) in Ω, ϕ^+ is rank-one convex in Ω, and $\lambda_1(\nabla^2\phi^-) \leq 0$ in Ω. Then:

1. *there are viscosity solutions ψ for the problem*

$$-\lambda_1[\nabla^2\psi(\mathbf{F}) + \operatorname{Det}\mathbf{F}\nabla^2\alpha(\mathbf{F}) + \nabla\alpha(\mathbf{F}) \otimes \mathbf{DF} + \mathbf{DF} \otimes \nabla\alpha(\mathbf{F})] = 0 \text{ in } \Omega,$$
$$\psi(\mathbf{F}) = \psi^+(\mathbf{F}) = \psi^-(\mathbf{F}) \text{ on } \partial\Omega.$$

2. *the function*

$$\phi : \mathbf{M}^{2\times 2} \to \mathbb{R}, \quad \phi(\mathbf{F}) = \psi(\mathbf{F}) + \alpha(\mathbf{F}) \operatorname{Det}\mathbf{F}$$

is rank-one convex.

This result is a direct application of Perron's method (Theorem 4.1 in [6]).

The material in this contribution builds in a fundamental way upon [4].

2 Non-Negative Polynomials

To motivate our claim that quasi convexity can be directly related to the general issue of non-negativeness of polynomials, let us go back to the basic definition of a quasi convex function (2)

$$\int_Q \phi(\mathbf{F} + \nabla\mathbf{u}(\mathbf{x}))\, d\mathbf{x} \geq \phi(\mathbf{F}), \qquad \text{for all } \mathbf{u},\ Q\text{-periodic, and all } \mathbf{F},$$

and rewrite it in the form

$$\Phi(\mathbf{F}, \mathbf{u}) \equiv \int_Q [\phi(\mathbf{F} + \nabla\mathbf{u}(\mathbf{x})) - \phi(\mathbf{F})]\, d\mathbf{x} \geq 0, \quad \mathbf{F} \in \mathbb{M}^{m\times N}, \mathbf{u} : Q \to \mathbb{R}^m.$$

Because of the periodicity requirement for test fields $\mathbf{u}$, we can use Fourier series to have

$$\mathbf{u}(\mathbf{x}) = \frac{1}{2\pi} \sum_{\mathbf{n}\in\mathbb{Z}^N} \sin(2\pi\mathbf{n}\cdot\mathbf{x})\, \mathbf{a_n}, \quad \mathbf{a_n} \in \mathbb{R}^m,$$
$$\nabla\mathbf{u}(\mathbf{x}) = \sum_{\mathbf{n}\in\mathbb{Z}^N} \cos(2\pi\mathbf{n}\cdot\mathbf{x})\, \mathbf{a_n} \otimes \mathbf{n}.$$

The quasi convexity condition for ϕ can be recast as the inequality

$$\Phi(\mathbf{F}, \{\mathbf{a_n}\}) = \int_Q \left[\phi \left(\mathbf{F} + \sum_{\mathbf{n} \in \mathbb{Z}^N} \cos(2\pi \mathbf{n} \cdot \mathbf{x})\, \mathbf{a_n} \otimes \mathbf{n} \right) - \phi(\mathbf{F}) \right] d\mathbf{x} \geq 0.$$

If ϕ is a polynomial of a certain degree of all its variables, then the functional Φ will also be a polynomial of the same degree, possibly in an infinite number of variables, or in an arbitrary large number of variables, but still a polynomial. Hence, the quasi convexity condition for a polynomial ϕ is equivalent to the non-negativeness of a certain, more sophisticated polynomial, and this realization brings us to the subject of non-negativeness of polynomials.

This is a main area in Algebraic Geometry of considerable relevance for global optimization problems [10, 14]. The issue of the non-negativeness of polynomials is a difficult question not fully understood or solved. Throughout the years, since the time of Hilbert, researchers have been looking for efficient tests or certificates for such non-negativeness.

2.1 Hilbert's Theorem

The first, elementary condition to ensure the non-negativeness of polynomials is the sum-of-squares test. If for a given polynomials $p(\mathbf{x})$ of several variables $\mathbf{x}$, we have that

$$p(\mathbf{x}) = \sum_i p_i(\mathbf{x})^2, \quad \text{each } p_i, \text{ a polynomial,}$$

we immediately have that $p \geq 0$. The sum-of-squares test was deeply studied by D. Hilbert [8], leading to his celebrated theorem on the equivalence between the non-negativity of polynomials and the sum-of-squares condition. Put n for the degree of the polynomial, and d for the number of variables.

Theorem 3 (D. Hilbert) *Non-negative polynomials coincide with sums-of squares polynomials, in the following three cases:*

1. $d = 1$*: polynomials of arbitrary degree in one variable;*
2. $n = 2$*: second degree polynomials in any number of variables;*
3. $n = 4$, $d = 2$*: quartic polynomials in two variables.*

In all other cases, there are non-negative polynomials which are not sums of squares.

Hilbert later, and motivated by his result, proposed his 17th problem in his famous list [9]: Does every non-negative polynomial have a representation as a sum of squares of "rational" functions? He was essentially asking about the existence of a polynomial $q(\mathbf{x})$ so that $q(\mathbf{x})^2 p(\mathbf{x})$ is a sum of squares of polynomials. Artin proved in 1927 that this is so [2].

We would like to focus on the case $n = 4$, $d = 2$ of Hilbert's theorem to see if it can be utilized to show something interesting concerning our vector variational problems. In particular, we want to work with rank-one convexity. This convexity condition for smooth ($\mathscr{C}^2$) functions is equivalent to the Legendre-Hadamard condition demanding

$$\nabla^2\phi(\mathbf{F}) : (\mathbf{a} \otimes \mathbf{n}) \otimes (\mathbf{a} \otimes \mathbf{n}) \geq 0 \tag{3}$$

for every matrix $\mathbf{F}$, and vectors $\mathbf{a}$, $\mathbf{n}$ of the appropriate dimensions.

3 The Fundamental Lemma

We first fix notation to avoid misunderstandings. We put

$$\mathbf{F} = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix} \mapsto \mathbf{F} = (F_{11}, F_{12}, F_{21}, F_{22}), \tag{4}$$

$$\mathbf{D} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{D} : \mathbf{F} \otimes \mathbf{F} = \mathbf{F}^T \mathbf{D} \mathbf{F} = 2 \det \mathbf{F},$$

$$\mathbf{Q}, \text{ a } 4 \times 4\text{-symmetric matrix}, \quad \phi(\mathbf{F}) = \mathbf{Q} : \mathbf{F} \otimes \mathbf{F} = \mathbf{F}^T \mathbf{Q} \mathbf{F}.$$

Notice that we are here restricting attention to quadratic forms so that our ϕ is just the function $\phi(\mathbf{F}) = \mathbf{Q} : \mathbf{F} \otimes \mathbf{F}$ for a constant, 4×4-matrix $\mathbf{Q}$. Note also that we identify, in all of our formulas, the matrix $\mathbf{F}$, a 2×2-tensor, with the four-vector as in (4).

Our main result is the following lemma which was already shown by Marcellini [11] many years ago in a straightforward way.

Lemma 1 *The quadratic form* $\mathbf{Q}$ *is rank-one convex if and only if there is a number* α *such that* $\mathbf{Q} = \mathbf{S} + \alpha\mathbf{D}$, *and* $\mathbf{S}$ *is non-negative definite.*

Proof If $\mathbf{Q}$ is of the form $\mathbf{S} + \alpha\mathbf{D}$ for some number α, and a non-negative definite matrix $\mathbf{S}$, it is elementary to check that it is rank-one convex.

Suppose $\mathbf{Q}$ is rank-one convex. According to the Legendre-Hadamard condition (3),

$$\mathbf{Q} : (\mathbf{x} \otimes \mathbf{y}) \otimes (\mathbf{x} \otimes \mathbf{y}) \geq 0, \quad \mathbf{x} = (x_1, x_2), \mathbf{y} = (y_1, y_2).$$

By homogeneity, it is clear that this inequality is equivalent to

$$\mathbf{Q} : (\tilde{\mathbf{x}} \otimes \tilde{\mathbf{y}}) \otimes (\tilde{\mathbf{x}} \otimes \tilde{\mathbf{y}}) \geq 0, \quad \tilde{\mathbf{x}} = (x, 1), x = x_1/x_2, \tilde{\mathbf{y}} = (y, 1), y = y_1/y_2,$$

and so, the positivity condition determining rank-one convexity becomes the nonnegativeness of the polynomial

$$P_4(x, y) = \mathbf{Q} : [(x, 1) \otimes (y, 1)] \otimes [(x, 1) \otimes (y, 1)] \geq 0,$$
$$P_4(x, y) = \mathbf{Q} : \mathbf{X} \otimes \mathbf{X} \geq 0, \quad \mathbf{X} = (xy, x, y, 1).$$

This is precisely the situation of Hilbert's theorem for the case of quartic polynomials in two variables, and so there must be a representation of $P_4(x, y)$ as a sum of squares. Given that

$$\mathbf{A} : \mathbf{X} \otimes \mathbf{X} = 0 \iff \mathbf{A} = \lambda \mathbf{D}$$

all possible representations of P_4 are of the form

$$P_4(x, y) = (\mathbf{Q} - \alpha \mathbf{D}) : \mathbf{X} \otimes \mathbf{X}, \quad \alpha \in \mathbb{R}.$$

Hilbert's theorem then implies that there must be, at least one real number α, such that

$$(\mathbf{Q} - \alpha \mathbf{D}) : \mathbf{X} \otimes \mathbf{X} = (\mathbf{C}\mathbf{X}) \otimes (\mathbf{C}\mathbf{X}), \quad (\mathbf{Q} - \alpha \mathbf{D}) = \mathbf{C}^T \mathbf{C},$$

and this finishes the proof.

4 Some Consequences for Rank-One Convexity

We can now use Lemma 1 to show some interesting characterization of rank-one convexity of functions defined on 2 × 2-matrices.

An immediate consequence follows.

Corollary 1 *A smooth function $\phi : \mathbf{M}^{2\times 2} \to \mathbb{R}$ is rank-one convex if and only if there is a scalar function $\alpha : \mathbf{M}^{2\times 2} \to \mathbb{R}$ and a symmetric, non-negative definite matrix field $\mathbf{S} : \mathbf{M}^{2\times 2} \to \mathbf{M}^{4\times 4}$ such that*

$$\nabla^2\phi(\mathbf{F}) = \mathbf{S}(\mathbf{F}) + \alpha(\mathbf{F})\mathbf{D}.$$

One can elaborate on the fact that the difference

$$\nabla^2\phi(\mathbf{F}) - \alpha(\mathbf{F})\mathbf{D}$$

is a positive definite matrix to find a more explicit characterization. Namely, if we set

$$\phi^-(\mathbf{F}) = \sup_{\mathbf{G}}\{-\nabla^2\phi(\mathbf{F}) : \mathbf{G}\otimes\mathbf{G} : \det \mathbf{G} = -1\},$$
$$\phi^+(\mathbf{F}) = \inf_{\mathbf{G}}\{\nabla^2\phi(\mathbf{F}) : \mathbf{G}\otimes\mathbf{G} : \det \mathbf{G} = 1\},$$

then we have the following statement.

Theorem 4 *Such ϕ is rank-one convex if and only if, for each matrix $\mathbf{F}$,*

$$\phi^-(\mathbf{F}) \le \phi^+(\mathbf{F}).$$

Moreover, for every function $\alpha(\mathbf{F})$ such that

$$\phi^-(\mathbf{F}) \le 2\alpha(\mathbf{F}) \le \phi^+(\mathbf{F}),$$

we have that

$$\nabla^2\phi(\mathbf{F}) - \alpha(\mathbf{F})\mathbf{D}$$

is non-negative definite.

Proof We start by exploring the fact that the combination

$$\nabla^2\phi(\mathbf{F}) - \alpha(\mathbf{F})\mathbf{D}$$

ought to be positive definite for an appropriate function $\alpha(\mathbf{F})$. This means that

$$(\nabla^2\phi(\mathbf{F}) - \alpha(\mathbf{F})\mathbf{D}) : \mathbf{G}\otimes\mathbf{G} \ge 0$$

for every matrix $\mathbf{G}$, that is to say

$$\nabla^2\phi(\mathbf{F}) : \mathbf{G}\otimes\mathbf{G} \ge 2\alpha(\mathbf{F}) \det \mathbf{G}.$$

If $\det \mathbf{G} = 1$, then

$$\alpha(\mathbf{F}) \le \frac{1}{2}\nabla^2\phi(\mathbf{F}) : \mathbf{G} \otimes \mathbf{G},$$

whereas if $\det \mathbf{G} = -1$,

$$\alpha(\mathbf{F}) \ge -\frac{1}{2}\nabla^2\phi(\mathbf{F}) : \mathbf{G} \otimes \mathbf{G}.$$

The arbitrariness of $\mathbf{G}$ in both cases leads to the statement in the theorem.

It is worth exploring a bit the two optimization problems defining $\phi^{\pm}(\mathbf{F})$. For a given fixed matrix $\mathbf{F}$, put $\mathbf{A} = \nabla^2\phi(\mathbf{F})$, and consider jointly the two quadratic mathematical programming problems

$$a_- = \sup_{\mathbf{G}}\{-\frac{1}{2}\mathbf{A} : \mathbf{G} \otimes \mathbf{G} : \det \mathbf{G} = -1\},$$

$$a_+ = \inf_{\mathbf{G}}\{\frac{1}{2}\mathbf{A} : \mathbf{G} \otimes \mathbf{G} : \det \mathbf{G} = 1\}.$$

Note that the rank-one convexity of the quadratic form determined by a given, constant matrix $\mathbf{A}$ amounts to having

$$a_0 = \min_{\mathbf{G}}\{\frac{1}{2}\mathbf{A} : \mathbf{G} \otimes \mathbf{G} : \det \mathbf{G} = 0, |\mathbf{G}|^2 = 1\} \ge 0.$$

However, the calculation of this minimum is not that elementary because there are two quadratic constraints, one of which is not convex. We would like to relate a_+ and a_- to a_0.

Lemma 2 *If either the infimum determining a_+ or the supremum determining a_- is achieved for matrices going to infinity, then $a_0 \le 0$.*

Proof We prove it by contradiction for the first case, the second one being completely parallel. Take a sequence $\mathbf{G}_j$ with

$$|\mathbf{G}_j| \to \infty, \quad \det \mathbf{G}_j = 1.$$

Without loss of generality, we can assume that there is some $\mathbf{G}$ such that

$$\frac{1}{|\mathbf{G}_j|}\mathbf{G}_j \to \mathbf{G}, \quad \det \mathbf{G} = 0, \quad |\mathbf{G}|^2 = 1.$$

For instance, one can take

$$\mathbf{G}_j = \begin{pmatrix} j+1 & j \\ 1 & 1 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ 0 & 0 \end{pmatrix}.$$

If $a_0 > 0$, then

$$\frac{1}{2}\mathbf{A} : \mathbf{G} \otimes \mathbf{G} \geq a_0 > 0,$$

and

$$\frac{1}{2}\mathbf{A} : \mathbf{G}_j \otimes \mathbf{G}_j = |\mathbf{G}_j|^2 \frac{1}{2}\mathbf{A} : \frac{\mathbf{G}_j}{|\mathbf{G}_j|} \otimes \frac{\mathbf{G}_j}{|\mathbf{G}_j|}$$

has to converge to $+\infty$, a contradiction because a_+ can never be $+\infty$.

Lemma 3 *Let λ_+ be the least eigenvalue of the matrix* **AD** *among those having eigenvectors with positive determinant. If the infimum determining a_+ is achieved in finite matrices, then $a_+ = \lambda_+$. Similarly, if λ_- is the greatest of the eigenvalues of* **AD** *among those having eigenvectors with negative determinant, and the supremum determining a_- is achieved for finite matrices, then $a_- = \lambda_-$.*

This is elementary. Use optimality, and notice that $\mathbf{D}^2$ is the identity matrix.

We finally identify circumstances when the equality $a_+ = a_-$ may take place.

Proposition 1 *Suppose $a_+ = a_-$. Then, this common value must vanish.*

Proof If $a_+ = a_-$, one of the two optimization problems defining a_+ and a_- ought to be attained at infinity. Indeed if they both were attained by finite matrices, Lemma 3 clearly implies that $a_\pm$ are taken from disjoint groups of (eigen)values, and so they cannot match. In that case, Lemma 2 leads to $a_0 \leq 0$. But if $a_+ = a_-$, the quadratic form must be rank-one convex, and then $a_0 \geq 0$. We conclude that $a_0 = 0$. Once we know this, by continuity, both a_+ and a_- must vanish. Notice that there cannot be a real gap between the infimum defining a_+, if it is taken on at infinity, and the minimum determining a_0.

Applying the previous conclusion to a non-quadratic rank-one convex function, we find the following remarkable corollary.

Corollary 2 *Let ϕ be smooth and rank-one convex. Suppose there is a subset* $\mathbf{S} \subset \mathbf{M}^{2\times 2}$ *of matrices with non-empty interior where $\phi^+ = \phi^-$. Then*

$$\phi^+\big|_{\mathbf{S}} = \phi^-\big|_{\mathbf{S}} = 0.$$

5 Proof of Theorem 1

For the proof of Theorem 1, let us start with the condition in Corollary 1

$$\nabla^2\phi(\mathbf{F}) = \mathbf{S}(\mathbf{F}) + \alpha(\mathbf{F})\mathbf{D}. \tag{5}$$

By looking at this identity, it is quite natural to consider the function

$$\psi(\mathbf{F}) = \phi(\mathbf{F}) - \alpha(\mathbf{F})\,\mathrm{Det}\,\mathbf{F}.$$

If we compute $\nabla^2\psi(\mathbf{F})$, we find

$$\nabla^2\phi(\mathbf{F}) - \mathrm{Det}\,\mathbf{F}\nabla^2\alpha(\mathbf{F}) - \nabla\alpha(\mathbf{F}) \otimes \mathbf{DF} - \mathbf{DF} \otimes \nabla\alpha(\mathbf{F}) - \alpha(\mathbf{F})\mathbf{D}.$$

Conclude by comparison with (5) that

$$\mathbf{S}(\mathbf{F}) = \nabla^2\psi(\mathbf{F}) + \mathrm{Det}\,\mathbf{F}\nabla^2\alpha(\mathbf{F}) + \nabla\alpha(\mathbf{F}) \otimes \mathbf{DF} + \mathbf{DF} \otimes \nabla\alpha(\mathbf{F}) \geq 0.$$

But this condition is saying that

$$-\lambda_1[\nabla^2\psi(\mathbf{F}) + \mathrm{Det}\,\mathbf{F}\nabla^2\alpha(\mathbf{F}) + \nabla\alpha(\mathbf{F}) \otimes \mathbf{DF} + \mathbf{DF} \otimes \nabla\alpha(\mathbf{F})] \leq 0.$$

Acknowledgements Partially supported by MINECO/FEDER grant MTM2013-47053-P, by PEII-2014-010-P of the Conserjería de Cultura (JCCM), and by grant GI20152919 of UCLM.

References

1. Antman, S.S.:, Nonlinear Problems of Elasticity, 2nd edn. Applied Mathematical Sciences, vol 107. Springer, New York (2005)
2. Artin, E.: Über die Zerlegung definiter Funktionen in Quadrate. Abh. Math. Sem. Univ. Hamburg **5**(1), 100–115 (1927)
3. Ball, J.M.: Convexity conditions and existence theorems in nonlinear elasticity. Arch. Ration. Mech. Anal. **63**, 337–403 (1977)
4. Bandeira, L., Pedregal, P.: The role of non-negative polynomials for rank-one convexity and quasi convexity. J. Ellipic Parab. Equ. **2**(2), 27–36 (2016)
5. Ciarlet, P.G.: Mathematical Elasticity. Vol. I. Three-Dimensional Elasticity. Studies in Mathematics and Its Applications, vol. 20. North-Holland, Amsterdam (1988)
6. Crandall, M.G., Ishii, H., Lions, P.L.: User's guide to viscosity solutions of second order partial differential equations. Bull. Am. Math. Soc. (N.S.) **27**(1), 1–67 (1992)
7. Dacorogna, B.: Direct Methods in the Calculus of Variations, 2nd edn. Springer, New York (2008)
8. Hilbert, D.: Über die Darstellung Definiter Formen als Summe von Formenquadraten. Mathematische Annalen **32**, 342–250 (1888)
9. Hilbert, D.: Mathematische Probleme, Lecture, Second Internat. Congr. Math. (Paris, 1900), Nachr. Ges. Wiss. Göttingen Math. Phys. KL., 253–297 (1900); English transl., Bull. Am. Math. Soc. **8**, 437–479 (1902); Bull. (New Series) Am. Math. Soc. **37**, 407–436 (2000)

10. Laserre, J.B.: Moments, Positive Polynomials and Their Applications. Imperial College Press, London (2010)
11. Marcellini, P.: Quasi convex quadratic forms in two dimensions. Appl. Math. Optim. **11**(2), 183–189 (1984)
12. Morrey, C.B.: Quasiconvexity and the lower semicontinuity of multiple integrals. Pac. J. Math. **2**, 25–53 (1952)
13. Morrey, C.B.: Multiple Integrals in the Calculus of Variations. Springer, New York (1966)
14. Nie, J.: Discriminants and nonnegative polynomials. J. Symb. Comp. **47**, 167–191 (2012)
15. Pedregal, P., Šverák, V.: A note on quasiconvexity and rank-one convexity for 2x2 matrices. J. Convex Anal. **5**(1), 107–117 (1998)
16. Šverák, V.: Rank-one convexity does not imply quasiconvexity. Proc. Roy. Soc. Edinburgh Sect. **120 A**, 293–300 (1992)
17. Tonelli, L.: Fondamenti di calcolo delle variazioni. Zanichelli, Bologna (1921)

On Friedrichs Inequality, Helmholtz Decomposition, Vector Potentials, and the div-curl Lemma

Ben Schweizer

Abstract We study connections between four different types of results that are concerned with vector-valued functions $u : \Omega \to \mathbb{R}^3$ of class $L^2(\Omega)$ on a domain $\Omega \subset \mathbb{R}^3$: Coercivity results in $H^1(\Omega)$ relying on div and curl, the Helmholtz decomposition, the construction of vector potentials, and the global div-curl lemma.

1 Introduction

The original motivation of this text was to derive a variant of the div-curl lemma. This important lemma treats the convergence properties of products of two weakly convergent sequences of functions. Besides other applications, the lemma plays an important role in homogenization theory, in particular in non-periodic homogenization problems. At some places, the name “compensated compactness” is used to refer to the div-curl lemma. We will use below the name “global” div-curl lemma to indicate that we are not satisfied with the distributional convergence of the product of functions, but that we want to obtain the convergence of the integral of the product.

The usual proof of the div-curl lemma is based on the construction of vector potentials, see e.g. [7]. In the *global* div-curl lemma, the construction of potentials must be performed taking special care of appropriate boundary conditions. The proof of the div-curl lemma becomes shorter if it is based on a Helmholtz decomposition result. Once more, the *global* div-curl lemma requires a careful analysis of the boundary conditions.

Both, the construction of vector potentials and the proof of the Helmholtz decomposition, can be obtained from coercivity results involving divergence and curl of a function $u : \mathbb{R}^3 \supset \Omega \to \mathbb{R}^3$. We have not been able to find a clear description of this connection in the literature. Moreover, the literature discusses the

B. Schweizer (✉)
Fakultät für Mathematik, Technische Universität Dortmund, Dortmund, Germany
e-mail: ben.schweizer@tu-dortmund.de

E. Rocca et al. (eds.), *Trends in Applications of Mathematics to Mechanics*,
Springer INdAM Series 27, https://doi.org/10.1007/978-3-319-75940-1_4

coercivity results usually in a form that is not strong enough to obtain the above-mentioned consequences.

In this text we present coercivity results in various forms and provide sketches of their proofs. We demonstrate how the other results can be obtained quite directly from the coercivity estimates. Moreover, we obtain these other results in a strong form, i.e. with a good control of boundary data. To summarize, we treat the following closely connected subjects and describe their most relevant connections:

(1) Coercivity results relying on div and curl
(2) The Helmholtz decomposition
(3) Construction of vector potentials
(4) The global div-curl lemma

More specifically, we will show the following: Let Ω be a domain for which the coercivity estimate of Item (1) holds. Then Ω permits statements as in Items (2)–(4) in a strong form.

Let us describe more clearly what is meant by the above items (1)–(4).

(1) The coercivity regards inequalities that allow to estimate, in the space $L^2(\Omega)$, all derivatives of a field $u : \Omega \to \mathbb{R}^3$ in terms of its divergence $\operatorname{div} u : \Omega \to \mathbb{R}$ and its rotation curl $u : \Omega \to \mathbb{R}^3$.
(2) In the Helmholtz decomposition we are interested in constructing, given $f \in L^2(\Omega, \mathbb{R}^3)$, two functions ϕ and w such that $f = \nabla\phi + w$ with $\operatorname{div} w = 0$. In *strong* Helmholtz decomposition results, we want to write curl ψ instead of w and impose boundary conditions on ψ.
(3) Construction of vector potentials: Given a field $f : \Omega \to \mathbb{R}^3$ with $\operatorname{div} f = 0$, we want to find a potential $\psi : \Omega \to \mathbb{R}^3$ such that $f = \operatorname{curl} \psi$, again imposing boundary conditions on ψ.
(4) In the div-curl lemma one considers sequences $f_k \rightharpoonup f$ and $p_k \rightharpoonup p$ as $k \to \infty$ in $L^2(\Omega, \mathbb{R}^3)$. The additional information is that both $\|\operatorname{div} f_k\|_{L^2(\Omega)}$ and $\|\operatorname{curl} p_k\|_{L^2(\Omega)}$ are bounded sequences. One is interested in the product $f_k \cdot p_k$. In the standard div-curl lemma, one obtains the distributional convergence $f_k \cdot p_k \to f \cdot p$ as $k \to \infty$. We are interested in the *global* div-curl lemma, which provides $\int_\Omega f_k \cdot p_k \to \int_\Omega f \cdot p$ as $k \to \infty$.

We note that the coercivity result (1) requires quite strong assumptions on $\Omega \subset \mathbb{R}^3$ (the regularity $\mathscr{C}^{1,1}$ or convexity of Ω, simple connectedness of Ω and connectedness of the boundary $\partial\Omega$). On the other hand, given (1), the results (2)–(4) can be derived easily in strong forms. In particular, we obtain these results with a control of the boundary data and with natural estimates.

1.1 Disclaimer

All the results that are presented in this note are known to the experts in the field. Moreover, the results are not stated with optimized assumptions. In particular, we oftentimes make assumptions on the domain that are not necessary. Our goal in this note is to show simple proofs for non-optimized results and to highlight connections between the different results.

At this point we would like to express our gratitude to D. Pauly for useful discussions and for pointing out an error in a first version of these notes.

1.2 Applications

As already mentioned, the div-curl lemma plays a crucial role in the derivation of homogenization limits, in particular if one follows the Russian approach, which is well adapted to perform stochastic homogenization limits, see [7]. A recent application is the non-periodic homogenization of plasticity equations. For such nonlinear non-periodic problems, the so-called "needle-problem approach" was developed in [12]. The crucial step in this approach is to find, given a sequence of functions u^ε on a domain Ω, a triangulation of Ω such that the global div-curl lemma can be applied on every simplex of the triangulation. The method was applied to perform the homogenization of plasticity equation in [5, 6]. The global div-curl lemma is also needed in a recent existence result for plasticity equations with curl-contribution, see [10].

1.3 An Observation

Since our proofs are based on coercivity estimates, we start with an observation regarding divergence and curl of functions.

Remark 1 Let $\Omega \subset \mathbb{R}^3$ be a bounded domain. We consider functions $u : \Omega \to \mathbb{R}^3$ with vanishing boundary values, $u \in H_0^1(\Omega)$. For such functions, the control of curl u and div u in $L^2(\Omega)$ is equivalent to the control of the full gradient in $L^2(\Omega)$. Indeed, for $u \in H_0^1(\Omega)$ (i.e.: all components of u vanish along the boundary), the following calculation is valid

$$\begin{aligned}\int_\Omega \left\{ |\nabla \cdot u|^2 + |\operatorname{curl} u|^2 \right\} &= \int_\Omega [-\nabla(\nabla \cdot u) + \operatorname{curl}\operatorname{curl} u] \cdot u \\ &= \int_\Omega [-\Delta u] \cdot u = \int_\Omega |\nabla u|^2 .\end{aligned} \tag{1}$$

Remark 1 indicates that all derivatives of u are controlled by $\nabla \cdot u$ and curl u—at least up to contributions from boundary integrals.

We note that Remark 1 has some similarity with the trivial Korn's inequality (Korn's inequality for functions with vanishing boundary values): The integral over squared gradients is (up to a factor 2) identical to the integral over squared *symmetrized* gradients. This is similar to equation (1). Korn's inequality (the non-trivial version) shows that, indeed, the full gradient of u can be estimated in terms of the symmetrized gradient of u.

2 Notation

In the following, $\Omega \subset \mathbb{R}^3$ always denotes a bounded open set, further properties will be specified when needed. For Lipschitz domains Ω, we denote the exterior normal by $\nu : \partial\Omega \to \mathbb{R}^3$ (ν is defined almost everywhere on the boundary).

We use the space $H(\Omega, \mathrm{curl}) := \{u \in L^2(\Omega, \mathbb{R}^3) \mid \mathrm{curl}\ u \in L^2(\Omega, \mathbb{R}^3)\}$, where curl u is understood in the distributional sense. The norm on this space is $\|u\|_{L^2} + \|\mathrm{curl}\ u\|_{L^2}$. The subspace of functions with vanishing boundary condition is defined as $H_0(\Omega, \mathrm{curl}) = \{u \in H(\Omega, \mathrm{curl}) \mid \nu \times u|_{\partial\Omega} = 0\}$. We emphasize that, since only the curl of u is controlled, only tangential boundary data can be evaluated in the sense of traces. Since trace estimates require Lipschitz boundaries, we define the space $H_0(\Omega, \mathrm{curl})$ with a weak formulation as follows:

$$H_0(\Omega, \mathrm{curl}) := \left\{ u \in H(\Omega, \mathrm{curl}) \middle| \int_\Omega \mathrm{curl}\ u \cdot \eta = \int_\Omega u \cdot \mathrm{curl}\ \eta \ \forall \eta \in H^1(\Omega, \mathbb{R}^3) \right\} . \tag{2}$$

Similarly, the space of functions with divergence in $L^2(\Omega)$ can be defined as: $H(\Omega, \mathrm{div}) := \{u \in L^2(\Omega, \mathbb{R}^3) \mid \mathrm{div}\, u \in L^2(\Omega, \mathbb{R}^3)\}$ and the corresponding space with vanishing boundary data is $H_0(\Omega, \mathrm{div}) = \{u \in H(\Omega, \mathrm{div}) \mid \nu \cdot u|_{\partial\Omega} = 0\}$, defined as

$$H_0(\Omega, \mathrm{div}) := \left\{ u \in H(\Omega, \mathrm{div}) \middle| \int_\Omega (\mathrm{div}\, u)\ \eta = -\int_\Omega u \cdot \nabla\eta \ \forall \eta \in H^1(\Omega, \mathbb{R}) \right\} . \tag{3}$$

We emphasize that the index 0 enforces in both cases that certain components of the vector field vanish on the boundary; these are tangential components in the case of $H_0(\Omega, \mathrm{curl})$ and normal components in the case of $H_0(\Omega, \mathrm{div})$.

Following [1], we use the space $X(\Omega) := H(\Omega, \mathrm{curl}) \cap H(\Omega, \mathrm{div})$ and the two subspaces

$$X_N(\Omega) := H_0(\Omega, \mathrm{curl}) \cap H(\Omega, \mathrm{div}) = \{u \in X(\Omega) \mid \nu \times u|_{\partial\Omega} = 0\} , \tag{4}$$

$$X_T(\Omega) := H(\Omega, \mathrm{curl}) \cap H_0(\Omega, \mathrm{div}) = \{u \in X(\Omega) \mid \nu \cdot u|_{\partial\Omega} = 0\} . \tag{5}$$

We note that the boundary values $\nu \times u|_{\partial\Omega}$ are well defined in the sense of distributions for functions $u \in H(\Omega, \text{curl})$. Similarly, $\nu \cdot u|_{\partial\Omega}$ is well defined in the sense of distributions for functions $u \in H(\Omega, \text{div})$.

3 Friedrichs Inequality

Many references are available for the following coercivity estimate. See e.g. (11) in [9] or Corollary 2.2 in [3], or Theorems 2.9 and 2.12 in [1].

We emphasize that the coercivity estimate of Theorem 1 remains valid on convex Lipschitz domains (the regularity $\partial\Omega \in \mathscr{C}^{1,1}$ is replaced by the convexity requirement), see Theorem 2.17 in [1]. It is also known as Gaffney-inequality.

Theorem 1 (Coercivity Estimate) *Let Ω be a bounded Lipschitz domain with $\partial\Omega \in \mathscr{C}^{1,1}$. Then there exists a coercivity constant $C_C > 0$ such that*

$$\|u\|_{H^1}^2 \le C_C \int_\Omega \left\{ |\nabla \cdot u|^2 + |\operatorname{curl} u|^2 + |u|^2 \right\} \tag{6}$$

for every $u \in X_T(\Omega)$. The constant C_C can be chosen such that (6) holds also for every $u \in X_N(\Omega)$.

Proof (Sketch) The proof of (6) relies on the fact that for u in either $X_T(\Omega)$ or $X_N(\Omega)$ critical boundary terms in the calculation (1) cancel. The remaining terms are products containing the curvature of the boundary and squares of values of u on the boundary. It is important that, in the boundary integrals, no terms containing derivatives of u remain. Moreover, for convex domains, the remaining terms have the good sign (see Lemma 2.11 in [1]). Combining the calculation (1) with a trace estimate for u and an interpolation, one obtains (6). The full proof requires density results in the spaces $X_T(\Omega)$ and $X_N(\Omega)$.

Our next step is to improve inequality (6) so that the $L^2(\Omega)$-norm of u does not appear on the right hand side. We call the result a Friedrichs inequality.

Let us describe why we call the result a Friedrichs inequality: The above sketch of proof (more precisely, the positivity of boundary contributions for convex domains) suggests that the following inequality holds on convex domains with $C_F = 1$:

$$\|\nabla u\|_{L^2(\Omega,\mathbb{R}^3)}^2 \le C_F \int_\Omega \left\{ |\nabla \cdot u|^2 + |\operatorname{curl} u|^2 \right\} \tag{7}$$

Inequality (7) is known as Friedrichs second inequality, see e.g. Theorem 3.1 in [11]. Using a general constant C_F in (7) is necessary for non-convex domains. We note that (7) for $L^p(\Omega)$-spaces is treated in [13] with methods from potential theory. Regarding the result in space dimension 2 we refer to [8], Theorem 4.3.

We want to improve (7) and estimate the full H^1-norm.

Corollary 1 (Friedrichs Inequality) *Let Ω be a simply connected bounded Lipschitz domain. We assume that (6) holds (we recall that $\partial\Omega \in \mathscr{C}^{1,1}$ or convexity of Ω is sufficient). Then there exists a constant $C_F > 0$ such that*

$$\|u\|^2_{H^1(\Omega,\mathbb{R}^3)} \leq C_F \int_\Omega \left\{ |\nabla \cdot u|^2 + |\operatorname{curl} u|^2 \right\} \tag{8}$$

holds for all functions u in the space $X_T(\Omega)$. If the boundary $\partial\Omega$ of the domain is connected, the estimate (8) holds also for every $u \in X_N(\Omega)$.

Proof We argue by contradiction. Let $(u_k)_k$ be a sequence with $\|u_k\|_{H^1(\Omega)} = 1$ for every k and with $\nabla \cdot u_k \to 0$ and curl $u_k \to 0$ in $L^2(\Omega)$. Rellich compactness allows to extract a subsequence and to find $u \in H^1(\Omega)$ such that $u_k \rightharpoonup u$ in $H^1(\Omega)$ and $u_k \to u$ in $L^2(\Omega)$. Weak limits coincide with distributional limits, hence $\nabla \cdot u = 0$ and curl $u = 0$.

The curl-free function u has a potential, $u = \nabla\Phi$ for some $\Phi \in H^1(\Omega)$. This fact is known as Poincaré lemma, we use at this point that Ω is simply connected. The potential Φ can be constructed for smooth curl-free functions $u : \Omega \to \mathbb{R}^3$ with the help of line integrals. The extension of the map $u \mapsto \Phi$ to functions $u \in L^2(\Omega, \mathbb{R}^3)$ is straightforward using the density of smooth functions and the fact that the gradient of Φ is controlled (it is u). The potential Φ solves $\Delta\Phi = 0$ because of $\nabla \cdot u = 0$. Furthermore, the boundary condition $u \in X_T(\Omega)$ implies that Φ has a vanishing normal derivative on $\partial\Omega$. The boundary condition $u \in X_N(\Omega)$ implies that tangential components of $\nabla\Phi$ vanish on the boundary, hence Φ can be chosen in $H^1_0(\Omega)$. In both cases, due to $\Delta\Phi = 0$, the potential Φ is a constant function and u vanishes.

The fact $u_k \to u = 0$ in $L^2(\Omega)$ implies that the three terms on the right hand side of (6) vanish in the limit $k \to \infty$ for the sequence u_k. Inequality (6) yields $\|u_k\|_{H^1} \to 0$, which is the desired contradiction.

4 Helmholtz Decomposition

We formulate a strong Helmholtz decomposition result in Theorem 2. In order to explain why we call Theorem 2 a strong Helmholtz decomposition result, let us first state and prove an elementary version.

Proposition 1 (Elementary Helmholtz Decomposition) *Let $\Omega \subset \mathbb{R}^3$ be a bounded Lipschitz domain. Then there exists a constant $C_H > 0$ such that, for every vector field $f \in L^2(\Omega, \mathbb{R}^3)$, the following holds:*

1. **Imposing a boundary condition for** w. *There exist $\phi : \Omega \to \mathbb{R}$ and $w : \Omega \to \mathbb{R}^3$ such that*

$$f = \nabla\phi + w\,, \quad \phi \in H^1(\Omega, \mathbb{R})\,, \tag{9}$$

$$w \in W_0 := \left\{ w \in L^2(\Omega) \,\middle|\, \int_\Omega w \cdot \nabla\varphi = 0\; \forall\varphi \in H^1(\Omega) \right\}\,. \tag{10}$$

2. **Imposing a boundary condition for ϕ.** *There exist $\phi : \Omega \to \mathbb{R}$ and $w : \Omega \to \mathbb{R}^3$ such that*

$$f = \nabla\phi + w\,, \quad \phi \in H_0^1(\Omega, \mathbb{R})\,, \tag{11}$$

$$w \in W := \left\{ w \in L^2(\Omega) \,\middle|\, \int_\Omega w \cdot \nabla\varphi = 0 \,\forall\varphi \in H_0^1(\Omega) \right\}. \tag{12}$$

Both decompositions are valid with the estimate

$$\|\phi\|_{H^1(\Omega)} + \|w\|_{L^2(\Omega)} \le C_H \|f\|_{L^2(\Omega)}\,. \tag{13}$$

Proof For Item (1), we define $\phi \in H^1(\Omega)$ as the solution of the Neumann problem

$$\int_\Omega \nabla\phi \cdot \nabla\varphi = \int_\Omega f \cdot \nabla\varphi \quad \forall\varphi \in H^1(\Omega)\,. \tag{14}$$

The solution exists by the Lax-Milgram theorem in the space of H^1-functions with vanishing mean value. With this choice of ϕ, the remainder $w := f - \nabla\phi$ satisfies $w \in W_0$ by definition.

For Item (2), we define $\phi \in H_0^1(\Omega)$ as the solution of the Dirichlet problem

$$\int_\Omega \nabla\phi \cdot \nabla\varphi = \int_\Omega f \cdot \nabla\varphi \quad \forall\varphi \in H_0^1(\Omega)\,. \tag{15}$$

The solution exists by the Lax-Milgram theorem in $H_0^1(\Omega)$. With this choice of ϕ, the remainder $w := f - \nabla\phi$ satisfies $w \in W$ by definition.

In both cases, due to the solution estimate of the Lax-Milgram theorem, the norm of ϕ in $H^1(\Omega)$ and, hence, the norm of w in $L^2(\Omega)$ are controlled by the norm of f in $L^2(\Omega)$.

We next show a stronger Helmholtz decomposition result. Here, we write the solenoidal function w as the curl of a vector potential ψ. Furthermore, we can prescribe a boundary condition for the vector potential. Again, all norms are controlled by the datum f.

Theorem 2 (Helmholtz Decomposition with Vector Potential) *Let $\Omega \subset \mathbb{R}^3$ be a simply connected bounded Lipschitz domain with a connected boundary $\partial\Omega$ of class $\mathscr{C}^{1,1}$. Then there exists a constant $C_H > 0$ such that, for every vector field $f \in L^2(\Omega, \mathbb{R}^3)$, we have:*

1. **Imposing a boundary condition for ψ.** *There exist $\phi : \Omega \to \mathbb{R}$ and $\psi : \Omega \to \mathbb{R}^3$ such that*

$$f = \nabla\phi + \operatorname{curl}\psi\,, \quad \phi \in H^1(\Omega, \mathbb{R})\,, \quad \nabla \cdot \psi = 0\,, \quad \psi \in X_N(\Omega)\,. \tag{16}$$

2. **Imposing a boundary condition for ϕ.** *There exist $\phi : \Omega \to \mathbb{R}$ and $\psi : \Omega \to \mathbb{R}^3$ such that*

$$f = \nabla\phi + \operatorname{curl}\psi\,, \quad \phi \in H_0^1(\Omega, \mathbb{R})\,, \quad \nabla \cdot \psi = 0\,, \quad \psi \in X_T(\Omega)\,. \tag{17}$$

In both cases, the decomposition satisfies the estimate

$$\|\phi\|_{H^1(\Omega)} + \|\psi\|_{H^1(\Omega)} \le C_H \|f\|_{L^2(\Omega)}\,. \tag{18}$$

Remark 2 Many parts of Theorem 2 remain valid under the following weaker assumption (A_1) on Ω:

(A_1) Let Ω be a bounded Lipschitz domain such that the Friedrichs inequality (8) holds.

Item (1) of the Theorem remains valid without any changes in the proof. Instead, our proof of Item (2) makes use of the $\mathscr{C}^{1,1}$-regularity of the boundary.

In order to clarify the connection with Proposition 1, we note the following: With ψ as in Item (1) above, there holds $w := \operatorname{curl}\psi \in W_0$ (as in Item (1) of Proposition 1). Indeed, for $\varphi \in H^1(\Omega)$,

$$\int_\Omega w \cdot \nabla\varphi = \int_\Omega \operatorname{curl}\psi \cdot \nabla\varphi = \int_\Omega \psi \cdot \operatorname{curl}\nabla\varphi = 0\,. \tag{19}$$

Proof **Proof of Item (1).**

Step 1. Construction of ϕ. In this first step, we compensate the divergence $\nabla \cdot f$ and the normal boundary data $f \cdot \nu$ with a scalar potential ϕ (as in (14) in the proof of Item (1) of Proposition 1). We define $\phi \in H^1(\Omega, \mathbb{R})$ as the solution with vanishing average of the Neumann problem

$$\int_\Omega \nabla\phi \cdot \nabla\varphi = \int_\Omega f \cdot \nabla\varphi \qquad \forall\varphi \in H^1(\Omega, \mathbb{R})\,. \tag{20}$$

The solution ϕ satisfies the estimate (18). In the rest of the proof our aim is to write the function $\tilde{f} := f - \nabla\phi \in W_0$ as the curl of a vector potential.

Step 2. Construction of ψ. We introduce the bilinear form

$$b(u, v) := \int_\Omega \{(\nabla \cdot u)(\nabla \cdot v) + (\operatorname{curl} u) \cdot (\operatorname{curl} v)\} \tag{21}$$

on the space $X_N(\Omega)$ of (4). We consider the following auxiliary problem: Find $\psi \in X_N(\Omega)$ such that

$$b(\psi, \varphi) = \int_\Omega \tilde{f} \cdot \operatorname{curl}\varphi \qquad \forall\varphi \in X_N(\Omega)\,. \tag{22}$$

The bilinear form b is coercive on $X_N(\Omega)$ by the Friedrichs coercivity estimate (8). This implies the solvability of problem (22) by some $\psi \in X_N(\Omega)$. We note that the solution $\psi \in X_N(\Omega)$ satisfies the estimate (18).

Step 3. The divergence of ψ. We claim that ψ satisfies $\nabla \cdot \psi = 0$.

To verify this claim, we solve, for arbitrary $\eta \in L^2(\Omega, \mathbb{R})$, the Dirichlet problem $\Delta\Phi = \eta$ with $\Phi \in H_0^1(\Omega)$. We want to use $\varphi := \nabla\Phi$ as a test-function in (22). The construction and the regularity $\Phi \in H^1(\Omega)$ imply $\varphi \in X(\Omega)$ (the distributional curl vanishes, since φ is a gradient, and the distributional divergence is η). Concerning the boundary condition we calculate, for test functions $\xi \in H^2(\Omega)$,

$$\int_\Omega \operatorname{curl} \varphi \cdot \xi = \int_\Omega 0 \cdot \xi = 0\,,$$

and, exploiting that Φ has vanishing boundary values,

$$\int_\Omega \varphi \cdot \operatorname{curl} \xi = \int_\Omega \nabla\Phi \cdot \operatorname{curl} \xi = \int_\Omega \Phi\ \nabla \cdot \operatorname{curl} \xi = 0\,.$$

By density, the equality of the two expressions remains valid for all test-functions $\xi \in H^1(\Omega)$. By definition of $X_N(\Omega)$, this provides $\varphi = \nabla\Phi \in X_N(\Omega)$. From now on, we may therefore use φ as a test function in (22).

Relation (22) allows to calculate

$$\begin{aligned} 0 &= \int_\Omega \tilde{f} \cdot \operatorname{curl} \varphi = b(\psi, \varphi) \\ &= \int_\Omega \{(\nabla \cdot \psi)\,(\nabla \cdot \varphi) + (\operatorname{curl} \psi) \cdot (\operatorname{curl} \varphi)\} = \int_\Omega (\nabla \cdot \psi)\,\eta\,. \end{aligned}$$

Since η was arbitrary, we obtain $\nabla \cdot \psi = 0$.

Step 4. Properties of the remainder. We introduce the remainder $R := \tilde{f} - \operatorname{curl} \psi$ and claim that R vanishes.

We start with the observation that the property $\nabla \cdot \psi = 0$ simplifies relation (22), which now reads

$$\int_\Omega R \cdot \operatorname{curl} \varphi = \int_\Omega (\tilde{f} - \operatorname{curl} \psi) \cdot \operatorname{curl} \varphi = 0 \qquad \forall \varphi \in X_N(\Omega)\,. \tag{23}$$

This shows $\operatorname{curl} R = 0$ in the sense of distributions.

Furthermore, R is a solenoidal field: The divergence of $\tilde{f}$ vanishes by the construction in Step 1, and the divergence of curl ψ also vanishes.

We finally want to check the normal boundary condition for R. For every $\varphi \in H^1(\Omega)$ holds, using $\psi \in X_N(\Omega)$ in the last step,

$$\int_\Omega R \cdot \nabla\varphi = \int_\Omega \tilde{f} \cdot \nabla\varphi - \int_\Omega \operatorname{curl} \psi \cdot \nabla\varphi \overset{(20)}{=} - \int_\Omega \operatorname{curl} \psi \cdot \nabla\varphi = 0\,.$$

This shows $R \in X_T(\Omega)$.

The Friedrichs estimate (8) on the space $X_T(\Omega)$ allows to conclude from curl $R = 0$ and div $R = 0$ the equality $R = 0$. This shows the decomposition result $f = \nabla\phi + \operatorname{curl}\psi$.

Proof of Item (2). The proof of Item (2) follows along the same lines. In Step 1, the scalar potential ϕ is constructed as the solution $\phi \in H_0^1(\Omega)$ of the Dirichlet problem $\Delta\phi = \nabla \cdot f$. We consider the function $\tilde{f} = f - \nabla\phi$, which has vanishing divergence (but, in general, not vanishing normal boundary data). In Step 2 we consider once more the bilinear form

$$b(u, v) := \int_\Omega \{(\nabla \cdot u)(\nabla \cdot v) + (\operatorname{curl} u) \cdot (\operatorname{curl} v)\} \,, \tag{24}$$

but now on the space $X_T(\Omega)$; the bilinear form is now $b : X_T(\Omega) \times X_T(\Omega) \to \mathbb{R}$. The vector potential ψ is once more constructed with the Lax-Milgram theorem; now $\psi \in X_T(\Omega)$ satisfies the identity of (22) for every test-function $\varphi \in X_T(\Omega)$. Step 3 can be performed as above and we obtain $\nabla \cdot \psi = 0$; the test function $\varphi = \nabla\Phi$ must now be constructed by solving a Neumann problem for Φ in order to have $\varphi \in X_T(\Omega)$.

We provide some more details concerning Step 4: As in the proof of Item (1), we define the remainder $R := \tilde{f} - \operatorname{curl} \psi$ and show that R vanishes. By construction of $\tilde{f}$, there holds $\nabla \cdot R = 0$. The fact $\nabla \cdot \psi = 0$ simplifies the identity in (22) and we find, in analogy to relation (23),

$$\int_\Omega R \cdot \operatorname{curl} \varphi = 0 \qquad \forall \varphi \in X_T(\Omega)\,. \tag{25}$$

This equality shows curl $R = 0$ in the sense of distributions.

Because of curl $R = 0$, the equality of (25) is also satisfied for every function $\varphi \in X_N(\Omega)$. Indeed, the formal calculation for this fact is

$$\int_\Omega R \cdot \operatorname{curl} \varphi = \int_\Omega \operatorname{curl} R \cdot \varphi = 0\,.$$

The integration by parts is justified by definition of $X_N(\Omega)$. A rigorous proof is obtained by first regularizing R and then considering the limit.

An arbitrary function $\varphi \in H^1(\Omega)$ can be written as the sum $\varphi = \varphi_N + \varphi_T$ with $\varphi_N \in X_N(\Omega)$ and $\varphi_T \in X_T(\Omega)$ (the proof of this fact can easily been performed using charts under the regularity assumption $\partial\Omega \in \mathscr{C}^{1,1}$). By linearity of the expression in φ we obtain that the equality of (25) is satisfied for every function $\varphi \in H^1(\Omega)$. Since curl R vanishes, this shows $R \in X_N(\Omega)$. The Friedrichs estimate (8) on the space $X_N(\Omega)$ allows to conclude $R = 0$ and hence the decomposition result.

5 Construction of Vector Potentials

We next present a consequence on the existence of vector potentials: Given f with $\nabla \cdot f = 0$, we look for a vector potential ψ such that curl $\psi = f$.

Classically, the construction of ψ is performed with Fourier transformation methods, see [4]. In this approach, little regularity on $\partial\Omega$ is needed (Lipschitz is sufficient). On the other hand, without further arguments, one cannot prescribe boundary conditions for the potential ψ; see also [7], Lemma 4.4. The results of [4] are stated and proved in [1].

The latter reference includes many extensions. In particular, very general domains can be considered. The notion of pseudo-Lipschitz domains is introduced and the existence of vector potentials is shown on pseudo-Lipschitz domains (domains with cuts that are not Lipschitz domains can still be pseudo-Lipschitz domains). The results of [1] include boundary conditions, see Theorems 3.12 and 3.17 of that reference.

The following result makes a strong statement on the existence of vector potentials. We note that we have to assume a high regularity of the domain. Our emphasis is on the fact that, essentially, the result can be obtained from Friedrichs inequality (8). We use the boundary regularity only in Item (2), compare Remark 2.

Corollary 2 *Let $\Omega \subset \mathbb{R}^3$ be a simply connected bounded Lipschitz domain with connected boundary $\partial\Omega \in \mathscr{C}^{1,1}$. Then there exists a constant $C_V > 0$ such that, for every $f \in L^2(\Omega, \mathbb{R}^3)$, we have:*

1. f **with boundary condition.** *If f has vanishing divergence and vanishing normal boundary data, i.e. $f \in W_0$ of (10), then there exists a vector potential $\psi \in X_N(\Omega)$ with*

$$f = \operatorname{curl} \psi\,, \qquad \|\psi\|_{H^1(\Omega)} \le C_V \|f\|_{L^2(\Omega)}\,. \tag{26}$$

2. f **without boundary condition.** *If f has vanishing divergence, i.e. $f \in W$ of (12), then there exists a vector potential $\psi \in X_T(\Omega)$ with*

$$f = \operatorname{curl} \psi\,, \qquad \|\psi\|_{H^1(\Omega)} \le C_V \|f\|_{L^2(\Omega)}\,. \tag{27}$$

Proof Item (1). We use the Helmholtz decomposition according to (16), $f = \nabla\phi + \operatorname{curl}\psi$ with $\phi \in H^1(\Omega, \mathbb{R})$ and $\psi \in X_N(\Omega)$. Upon multiplication with the gradient $\nabla\varphi$ of a test function $\varphi \in H^1(\Omega)$, we obtain

$$0 \overset{f \in W_0}{=} \int_\Omega f \cdot \nabla\varphi = \int_\Omega (\nabla\phi + \operatorname{curl}\psi) \cdot \nabla\varphi \overset{\psi \in X_N}{=} \int_\Omega \nabla\phi \cdot \nabla\varphi\,.$$

This shows that ϕ solves the homogeneous Neumann problem $\Delta\phi = 0$ and is therefore constant. This shows $\nabla\phi = 0$ and hence $f = \operatorname{curl}\psi$.

Item (2). The proof is analogous to that of Item (1). We now use the Helmholtz decomposition according to (17), $f = \nabla\phi + \operatorname{curl}\psi$ with $\phi \in H_0^1(\Omega, \mathbb{R})$ and $\psi \in X_T(\Omega)$. Testing $f = \nabla\phi + \operatorname{curl}\psi$ with the gradient $\nabla\varphi$ of a test function $\varphi \in H_0^1(\Omega)$, we obtain that ϕ solves the homogeneous Dirichlet problem $\Delta\phi = 0$. This shows $\phi = 0$ and hence $\nabla\phi = 0$. We obtain $f = \operatorname{curl}\psi$ and have therefore found the vector potential.

6 Global div-curl Lemma

One of our motivations to study the above classical decomposition results is the div-curl lemma. Most often, this lemma is formulated in a local version, with the claim that the product $f_k \cdot p_k$ of two weakly convergent sequences converges *in the sense of distributions.* We are interested here in global results, i.e. in results that provide the convergence of the integrals $\int_\Omega f_k \cdot p_k$.

Lemma 1 (Global div-curl Lemma) *Let $\Omega \subset \mathbb{R}^3$ be a simply connected bounded Lipschitz domain with connected boundary $\partial\Omega \in \mathscr{C}^{1,1}$. Let $f_k \rightharpoonup f$ in $L^2(\Omega, \mathbb{R}^3)$ and $p_k \rightharpoonup p$ in $L^2(\Omega, \mathbb{R}^3)$ be two weakly convergent sequences. We assume that the distributional derivatives satisfy, for some $C > 0$,*

$$\|\nabla \cdot f_k\|_{L^2(\Omega)} \le C\,, \qquad \|\operatorname{curl} p_k\|_{L^2(\Omega)} \le C\,, \tag{28}$$

for every $k \in \mathbb{N}$. Let furthermore be one of the two boundary conditions (i) or (ii) be satisfied for every $k \in \mathbb{N}$:

(i) $f_k \cdot \nu|_{\partial\Omega} = 0$
(ii) $p_k \times \nu|_{\partial\Omega} = 0$

Then there holds, as $k \to \infty$,

$$\int_\Omega f_k \cdot p_k \to \int_\Omega f \cdot p\,. \tag{29}$$

Let us include two remarks concerning the proof of the above lemma. Concerning boundary condition (i), we could rely the proof also on the convexity of the domain Ω and work without the assumption $\partial\Omega \in \mathscr{C}^{1,1}$. In the proof of boundary condition (ii), we do not exploit the boundary condition $\psi \in X_T(\Omega)$ on ψ. This means that case (ii) can be proved also with a weaker version of Theorem 2.

Proof Proof for boundary condition (i). We write p_k as $p_k = \nabla\phi_k + \operatorname{curl}\psi_k$ with potentials as in Theorem 2, Item (1), i.e. with $\psi_k \in X_N(\Omega)$ and $w_k := \operatorname{curl}\psi_k$. We recall that $w_k \in W_0$ and hence $w_k \in X_T(\Omega)$ is satisfied, compare (19).

We claim that $w_k := \operatorname{curl}\psi_k$ converges strongly in $L^2(\Omega)$. Indeed, we have boundedness of w_k in $L^2(\Omega)$ by boundedness of p_k in $L^2(\Omega)$, furthermore the obvious boundedness of $\nabla \cdot w_k = 0$. Finally, the boundedness of curl

$w_k = \text{curl } p_k$ in $L^2(\Omega)$ holds by (28). The coercivity inequality (6) in $X_T(\Omega)$ provides boundedness of w_k in $H^1(\Omega)$ and hence the compactness in $L^2(\Omega)$.

With this compactness property for w_k and the strong convergence $\phi_k \to \phi$ in $L^2(\Omega)$ (which follows from the compact Rellich embedding $H^1(\Omega) \subset L^2(\Omega)$) we can calculate

$$\begin{aligned} \int_\Omega f_k \cdot p_k &= \int_\Omega f_k \cdot (\nabla \phi_k + \text{curl } \psi_k) = \int_\Omega (-\nabla \cdot f_k)\, \phi_k + f_k \cdot \text{curl } \psi_k \\ &\to \int_\Omega (-\nabla \cdot f)\, \phi + f \cdot \text{curl } \psi = \int_\Omega f \cdot (\nabla \phi + \text{curl } \psi) = \int_\Omega f \cdot p \,. \end{aligned}$$

We used in the last step that the limits ϕ and ψ are indeed the Helmholtz decomposition functions for the limit p. This provides the claim for boundary condition (i).

Proof for boundary condition (ii). We now decompose f_k as $f_k = \nabla \phi_k + \text{curl } \psi_k$ using Theorem 2, Item (2) (but we will not exploit $\psi_k \in X_T(\Omega)$). The a priori bound $\|\psi_k\|_{H^1(\Omega)} \le C_0$ from (18) allows to select a subsequence $k \to \infty$ with the strong convergence $\psi_k \to \psi$ in $L^2(\Omega)$.

The functions $\phi_k \in H^1_0(\Omega)$ solve a Dirichlet problem: For every $\varphi \in H^1_0(\Omega)$ there holds

$$\int_\Omega \nabla \phi_k \cdot \nabla \varphi = \int_\Omega (f_k - \text{curl } \psi_k) \cdot \nabla \varphi = -\int_\Omega \nabla \cdot f_k \, \varphi \,.$$

This is the weak form of the Dirichlet problem $\Delta \phi_k = \nabla \cdot f_k$. Since the solution map $H^{-1}(\Omega) \to H^1_0(\Omega)$ of this Dirichlet problem is linear and continuous, the strong convergence of $\nabla \cdot f_k$ in $H^{-1}(\Omega)$ implies the strong convergence $\nabla \phi_k \to \nabla \phi$ in $L^2(\Omega)$.

After this preparation we can calculate, using boundary condition (ii) for p_k and for p, in the limit $k \to \infty$,

$$\begin{aligned} \int_\Omega f_k \cdot p_k &= \int_\Omega (\nabla \phi_k + \text{curl } \psi_k) \cdot p_k = \int_\Omega \nabla \phi_k \cdot p_k + \psi_k \cdot \text{curl } p_k \\ &\to \int_\Omega \nabla \phi \cdot p + \psi \cdot \text{curl } p = \int_\Omega \nabla \phi \cdot p + \text{curl } \psi \cdot p = \int_\Omega f \cdot p \,. \end{aligned}$$

This was the claim in (29).

Corollary 3 (The Usual div-curl Lemma on Arbitrary Domains) *Let $\Omega \subset \mathbb{R}^3$ be an open set and let p_k and f_k be sequences with $f_k \rightharpoonup f$ and $p_k \rightharpoonup p$ in $L^2(\Omega, \mathbb{R}^3)$. We assume the* div- *and* curl-*control of* (28)*. Then $f_k \cdot p_k \to f \cdot p$ in the sense of distributions on Ω.*

Proof Upon subtracting f and p from the sequences, we can assume $p = 0$ and $f = 0$. The distributional convergence is a local property, it suffices to show that, for an arbitrary open ball $B \subset \bar{B} \subset \Omega$ and an arbitrary smooth function $\varphi \in C_c^\infty(B, \mathbb{R})$ there holds

$$\int_B f_k \cdot p_k\, \varphi \to 0\,. \tag{30}$$

The relation (30) is a direct consequence of the global div-curl lemma 1 (ii), applied to the sequences f_k and $p_k\, \varphi$. The sequence $p_k\, \varphi$ converges weakly to 0, has $L^2(B)$-bounded curl and satisfies the homogeneous tangential boundary condition. The ball B has a $\mathscr{C}^{1,1}$-boundary. Lemma 1 provides (30) and thus the claim.

7 Comments and Generalizations

We emphasize that there is another route to prove the above div-curl result. One can start the analysis from the simple Helmholtz decomposition of Proposition 1. This requires no properties of Ω. When needed, one can use Theorem 3.12 or 3.17 of [1] to write a solenoidal field w as a curl, $w = \operatorname{curl}\, \psi$. This requires less regularity on Ω than our Corollary 2 (essentially, Lipschitz domains with cuts are allowed). Furthermore, on less regular domains (or for mixed boundary conditions), one must avoid the space $H^1(\Omega)$ and work with the Maxwell compactness property to find strongly convergent sequences. We make this observation more precise with the following remark.

Remark 3 (Global div-curl Lemma on Lipschitz Domains) The statement of Lemma 1 remains valid on bounded Lipschitz domains Ω that are simply connected and have a connected boundary.

Proof Let us start with the boundary condition (i). We proceed as in the proof of Lemma 1 (i) and decompose p_k. We use the simple Helmholtz decomposition of Proposition 1, Item (1), and write $p_k = \nabla \phi_k + w_k$ with $\phi_k \in H^1(\Omega)$ and $w_k \in W_0$. We use the existence result for vector potentials from Theorem 3.17 of [1]: there exists a potential $\psi_k \in X_N(\Omega)$ with $w_k = \operatorname{curl}\, \psi_k$. The boundary condition for ψ_k allows to conclude strong convergence of w_k from

$$\begin{aligned}\int_\Omega |w_k|^2 = \int_\Omega w_k \cdot \operatorname{curl}\, \psi_k = \int_\Omega \operatorname{curl}\, w_k \cdot \psi_k = \int_\Omega \operatorname{curl}\, p_k \cdot \psi_k \\ \to \int_\Omega \operatorname{curl}\, p \cdot \psi = \int_\Omega w \cdot \operatorname{curl}\, \psi = \int_\Omega |w|^2\,.\end{aligned}$$

In this calculation, the strong convergence of ψ_k cannot be concluded from the compactness of the Rellich embedding $H^1(\Omega) \subset L^2(\Omega)$, since no $H^1(\Omega)$ property for ψ_k is available. Instead, one has to use the Maxwell compactness property, see e.g. [2].

The proof for boundary condition (ii) follows closely the one of Lemma 1 (ii) and uses no boundary conditions for the potentials ψ_k. We decompose $f_k = \nabla\phi_k + w_k$ with $\phi_k \in H_0^1(\Omega)$ and exploit the strong convergence of $\nabla\phi_k$. The functions $w_k \in W$ are written as $w_k = \operatorname{curl}\psi_k$. In the calculation of the integral, we can integrate by parts in the term (curl $\psi_k) \cdot p_k$, due to the boundary condition for p_k.

We conclude this contribution with a simple remark. As soon as higher integrability properties of the functions are known, the case of general domains and the case without boundary conditions can be treated easily:

Remark 4 (Global div-curl Lemma on Arbitrary Domains) The statement of Lemma 1 without boundary conditions on either p_k or f_k remains valid on general bounded domains Ω if the sequence f_k (or the sequence p_k) is bounded in $L^q(\Omega)$ for some $q > 2$.

Proof The sequence $f_k \cdot p_k$ is bounded in the reflexive space $L^{1+\delta}(\Omega)$ for some $\delta > 0$. It therefore converges weakly in $L^{1+\delta}(\Omega)$ to its distributional limit, which is $f \cdot p$ by Corollary 3. The weak convergence implies the convergence of integrals (29). This proves the claim.

References

1. Amrouche, C., Bernardi, C., Dauge, M., Girault, V.: Vector potentials in three-dimensional non-smooth domains. Math. Methods Appl. Sci. **21**(9), 823–864 (1998)
2. Bauer, S.,Pauly, D., Schomburg, M.: The Maxwell compactness property in bounded weak Lipschitz domains with mixed boundary conditions. SIAM J. Math. Anal. **48**(4), 2912–2943 (2016)
3. Costabel, M.: A coercive bilinear form for Maxwell's equations. J. Math. Anal. Appl. **157**(2), 527–541 (1991)
4. Girault, V., Raviart, P.-A.: Finite element approximation of the Navier-Stokes equations. Lecture Notes in Mathematics, vol. 749. Springer, Berlin-New York (1979)
5. Heida, M., Schweizer, B.: Non-periodic homogenization of infinitesimal strain plasticity equations. ZAMM Z. Angew. Math. Mech. **96**(1), 5–23 (2016)
6. Heida, M., Schweizer, B.: Stochastic homogenization of plasticity equations. ESAIM: COCV **24**(1), 153–176 (2018). https://doi.org/10.1051/cocv/2017015
7. Jikov, V., Kozlov, S., Oleinik, O.: Homogenization of Differential Operators and Integral Functionals. Springer, New York (1994)
8. Křížek, M., Neittaanmäki, P.: On the validity of Friedrichs' inequalities. Math. Scand. **54**(1), 17–26 (1984)
9. Leis, R.: Zur Theorie elektromagnetischer Schwingungen in anisotropen inhomogenen Medien. Math. Z. **106**, 213–224 (1968)
10. Röger, M., Schweizer, B.: Strain gradient visco-plasticity with dislocation densities contributing to the energy. M3AS: Math. Models Methods Appl. Sci. **27**(14), 2595 (2017)
11. Saranen, J.: On an inequality of Friedrichs. Math. Scand. **51**(2), 310–322 (1982/1983)
12. Schweizer, B., Veneroni, M.: The needle problem approach to non-periodic homogenization. Netw. Heterog. Media **6**(4), 755–781 (2011)
13. von Wahl, W.: Estimating ∇u by div u and curl u. Math. Methods Appl. Sci. **15**(2), 123–143 (1992)

Variational Analysis of Nematic Shells

Giacomo Canevari and Antonio Segatti

Abstract In this note we present some recent results on the Mathematical Analysis of *Nematic Shells*. The type of results we present deal with the analysis of defectless configurations as well as the analysis of defected configurations. The mathematical tools include Topology, Analysis of Partial Differential Equations as well as Variational Techniques like Γ convergence.

1 Introduction: The Model and the Role of the Topology

The occasion of writing this note came because the second author of this paper was invited to lecture at the

INdAM-ISIMM Workshop on Trends on Applications of Mathematics to Mechanics

in Rome. These note contains the results presented in the seminar. More precisely, these results are the outcome of a research line started in 2012 and culminated in the papers [9, 35, 36] and [8]. In this note we try to convey the main ideas behind the results and leave the detailed proofs to the above mentioned papers.

A *Nematic Shell* is a rigid colloidal particle with a typical dimension in the micrometer scale coated with a thin film of nematic liquid crystal whose molecular orientation is subjected to a tangential anchoring. The study of these structures has received a good deal of interest, especially in the physics community (see, e.g., [6, 23, 26, 27, 30, 37, 39, 41, 42] and [28]).

From a mathematical point of view, a *Nematic Shell* is usually identified with a two dimensional compact (oriented by the choice of the unit normal field

G. Canevari
Basque Center for Applied Mathematics, Bilbao, Spain
e-mail: gcanevari@bcamath.org

A. Segatti (✉)
Dipartimento di Matematica "F. Casorati", Università di Pavia, Pavia, Italy
e-mail: antonio.segatti@unipv.it

E. Rocca et al. (eds.), *Trends in Applications of Mathematics to Mechanics*,
Springer INdAM Series 27, https://doi.org/10.1007/978-3-319-75940-1_5

$\boldsymbol{\gamma} : M \to \mathbb{R}^3$) surface M without boundary with the local orientation of the molecules described via a unit norm tangent vector field, named director in analogy with the "flat" case. More precisely, the local orientation of the molecules is described via a unit-norm tangent vector field $\mathbf{n} : M \to \mathbb{R}^3$ with $\mathbf{n}(x) \in \mathrm{T}_x M$ for any $x \in M$, $\mathrm{T}_x M$ being the tangent plane at the point x.

The study of these structures is particularly interesting and challenging due to its interdisciplinary character as it combines in a non trivial way physics, geometry, topology and variational techniques. In particular, the interplay between the geometry and the topology of the fixed substrate and the tangential anchoring constraint is a source of difficulties that will accompany us for the whole analysis. Indeed, as observed in [41] and [6], the liquid crystal equilibrium (and all its stable configurations, in general) is the result of the competition between two driving principles: on the one hand the minimization of the "curvature of the texture" penalized by the elastic energy, and on the other the frustration due to constraints of geometrical and topological nature, imposed by anchoring the nematic to the surface of the underlying particle. Different theoretical approaches for the treatment of *Nematic Shells* are available. Differences arise in the choice of the form of the elastic part of the free energy which could be of *intrinsic* or *extrinsic* nature. More precisely, theories which employ only covariant derivatives will be named *intrinsic* (see [26, 38, 39, 41]) while theories that comprise also how the shell sits in the three dimensional space will be named *extrinsic* (see [27] and [28]). When restricting to the simpler *one-constant approximation*, the *extrinsic energy* has the form

$$W(\mathbf{n}) := \frac{\kappa}{2} \int_M |\mathrm{D}\mathbf{n}|^2 + |\mathrm{d}\boldsymbol{\gamma}(\mathbf{n})\mathbf{n}|^2 \, \mathrm{d}S, \tag{1}$$

while the *intrinsic energy* has the form

$$W_{\mathrm{intr}}(\mathbf{n}) := \frac{\kappa}{2} \int_M |\mathrm{D}\mathbf{n}|^2 \, \mathrm{d}S. \tag{2}$$

In the definitions above $\mathbf{n}$ is a tangent vector field with unit norm, κ is a positive constant (from now on κ will be taken equal to one), the symbol D denotes the covariant derivative on M, and $\mathrm{d}\boldsymbol{\gamma}$, the differential of the Gauss map, is the so called shape operator. We refer to the quantity $\int_M |\mathrm{D}\mathbf{n}|^2$ as the *Dirichlet (or elastic) energy* of $\mathbf{n}$.

The extrinsic energy (1) has been derived by Napoli & Vergori (see [27] and [28]) by using a formal dimension reduction. More precisely, starting from the Oseen-Frank energy W^{OF} (see [40]) on a tubular neighborhood M_h of thickness h (satisfying a suitable constraint related to the curvature of M), Napoli and Vergori obtain that the limit

$$\lim_{h \searrow 0} \frac{1}{h} W^{\mathrm{OF}}(\mathbf{n}, M_h) = W_{\mathrm{extr}}(\mathbf{n}).$$

is well defined for any fixed and sufficiently smooth field $\mathbf{n}$ with the property of being independent of the thickness direction and tangent to leaf of the foliation M_h. The form of the limit energy is as follows

$$W_{\text{extr}}(\mathbf{n}) := \frac{1}{2}\int_M K_1(\operatorname{div}_{\mathrm{s}}\mathbf{n})^2 + K_2(\mathbf{n}\cdot\operatorname{curl}_{\mathrm{s}}\mathbf{n})^2 + K_3|\mathbf{n}\times\operatorname{curl}_{\mathrm{s}}\mathbf{n}|^2\,\mathrm{d}S, \tag{3}$$

where the differential operators $\operatorname{div}_{\mathrm{s}}$ and $\operatorname{curl}_{\mathrm{s}}$ in the display above are proper surface counterparts of the divergence and the curl operators (see [28]). The positive constants K_1, K_2, K_3 are the analogous of the Frank's constants in the euclidean case (see [40]). Finally, the energy (1) corresponds to the one-constant approximation, namely the energy W_{extr} with $K_1 = K_2 = K_3 = \kappa$.

It is worthwhile noting that the above formal argument can be made rigorous using the theory of Γ-convergence in the spirit of [24] (see [14] for the derivation of the surface Q tensor energy).

An important problem in the modern Materials Science is the analysis and the control of the complex microstructures that the material may develop. As observed in [19], the appearance of microstructures is usually related to the occurrence of the so-called defects, which are localized regions where the material behavior appears to be drastically different from the prototypical one. This is the case of Nematic Liquid Crystals for which defects can be easily seen in experiments. Defects are regions where the director field changes abruptly, due to the topological behavior of the field surrounding them. A prominent example of the appearance of defects is that of *Nematic Shells* which may develop topological defects due to the interplay between the topology of the substrate, the boundary conditions and the constraints on the director field (see [9]).

More precisely, when dealing with *Nematic Shells*, the topology of the shell and, possibly, of the boundary conditions is responsible for the emergence of defects which manifest in points in the shell where the director field is not well defined and consequently its energy ((1) or (2)) is infinite. The link between the topology of the shell M and the number of singularities that a unit norm vector field must have is given by the Poincaré-Hopf Index Theorem: If a unit norm has singularities of degree d_i located at the points $x_1, \dots, x_k$ then

$$\sum_{i=1}^{k} d_i = \chi(M),$$

where $\chi(M)$ is the Euler Characteristic of M. For example, a spherical shell has $\chi(M) = 2$, thus implying the necessity of having defects with total degree equal to 2. A crucial step in the analysis of a variational problem is the understanding of the correct functional framework where to set, for example, the minimization of the given energy. In the context of *Nematic Shells*, a closer inspection of the energy (3) reveals that there exist constants such that (see Proposition 1)

$$\frac{K_*}{2}\int_M \big(|\mathrm{D}\mathbf{n}|^2 + |\mathrm{d}\boldsymbol{\gamma}(\mathbf{n})|^2\big)\mathrm{d}S \le W_{\mathrm{extr}}(\mathbf{n}) \le \frac{K^*}{2}\int_M \big(|\mathrm{D}\mathbf{n}|^2 + |\mathrm{d}\boldsymbol{\gamma}(\mathbf{n})|^2\big)\mathrm{d}S,$$

Consequently, the natural choice for the functional framework would be to set the analysis in the space of tangent vector fields such that $|\mathbf{n}|$ and $|\mathrm{D}\mathbf{n}|$ belong to $L^2(M)$, which means that we have to consider the Sobolev set

$$W^{1,2}_{\tan}(M, \mathbb{S}^2) = \Big\{\mathbf{v} : M \to \mathbb{R}^3,\ \ |\mathbf{v}(x)| = 1,\ \ \mathbf{v}(x) \in \mathrm{T}_x M \ \text{ for a.a. } x \in M, |\mathrm{D}\mathbf{v}| \in L^2(M)\Big\}.$$

As it happens for smooth vector fields, the topology of the shell may introduce possible obstructions to this program. This is again related to the Poincaré-Hopf index Theorem. In particular, the following theorem clarifies the situation for vector fields with $W^{1,2}$ regularity

Theorem 1.1 *Let M be a compact smooth surface without boundary, embedded in $\mathbb{R}^3$. Let $\chi(M)$ be the Euler characteristic of M. Then*

$$W^{1,2}_{\tan}(M, \mathbb{S}^2) \neq \emptyset \ \Leftrightarrow \chi(M) = 0.$$

The proof of this theorem is given in [36] and it is based on a purely PDE argument. Interestingly, Theorem 1.1 is a consequence of the more general results contained in [9] regarding the extension of the Poincaré-Hopf Theorem to vector fields with VMO regularity defined on compact manifolds with, possibly, boundary. Theorem 1.1 is in a certain sense a borderline case for the existence of unit norm vector fields with Sobolev regularity. In fact, defining for $p \ge 1$, the Sobolev set of tangent vector fields

$$W^{1,p}_{\tan}(M; \mathbb{S}^2) := \Big\{\mathbf{v} : M \to \mathbb{R}^3,\ \ |\mathbf{v}(x)| = 1,\ \ \mathbf{v}(x) \in \mathrm{T}_x M \ \text{ for a.a. } x \in M, |\mathrm{D}\mathbf{v}| \in L^p(M)\Big\},$$

we have that

- For $p \ge 2$, $W^{1,p}_{\tan}(M; \mathbb{S}^2) \neq \emptyset$ if and only if $\chi(M) = 0$
- For $1 \le p < 2$, $W^{1,p}_{\tan}(M; \mathbb{S}^2) \neq \emptyset$.

The first item when $p > 2$ is a consequence of the classical Poincaré-Hopf Theorem and of the embedding $W^{1,p} \subset C^0$ for $p > 2$ in two dimensions. The case $p = 2$ follows from Theorem 1.1. The second item follows from the fact that a vector field that behaves like $\frac{x}{|x|}$ around the singularities belong to $W^{1,p}_{\tan}(M; \mathbb{S}^2)$ for $1 \le p < 2$ as a direct computation shows.

Coming back to the analysis of the energy (1), Theorem 1.1 gives that for shells M with $\chi(M) \neq 0$ the energy (1) is infinite and thus clearly not adequate to describe this situation.

The rest of the paper is divided according to Theorem 1.1. More precisely, we will first discuss in Sect. 2 the results for shells with Euler Characteristic equal to zero and then in Sect. 3 we will concentrate on shells with non zero Euler Characteristic.

When $\chi(M) = 0$, we obtain results regarding the existence of minimizers, the existence of the gradient flow and also some quantitative results on the structure of the minimizers for axisymmetric toroidal shells. The proofs of the results use ideas borrowed from the theory of harmonic maps.

Moreover, starting from a variant of the well known XY spin model, we perform the rigorous derivation via Γ-convergence of the energy (1) in terms of a discrete to continuum limit. More precisely, we consider a family of triangulations $\mathcal{T}_\varepsilon$ of M with the vertices $i \in \mathcal{T}_\varepsilon^0$ lying on M and with mesh size ε, i.e. $\varepsilon = \max_{T\in\mathcal{T}_\varepsilon} \mathrm{diam}(T)$. At any point $i \in \mathcal{T}_\varepsilon^0$ sits a unit-norm tangent vector $\mathbf{v}_\varepsilon(i) \in \mathrm{T}_i M$ named spin. We consider the following discrete energy

$$XY_\varepsilon(\mathbf{v}_\varepsilon) := \frac{1}{2} \sum_{i\neq j\in\mathcal{T}_\varepsilon^0} \kappa_\varepsilon^{ij} \left|\mathbf{v}_\varepsilon(i) - \mathbf{v}_\varepsilon(j)\right|^2 , \tag{4}$$

where the coefficients κ_ε^{ij} are the entries of the stiffness matrix of the Laplace-Beltrami operator of M. We show that, as $\varepsilon \to 0$, the discrete energy XY_ε converges to the continuum energy (1), in the sense of Γ-convergence.

The XY spin model has been widely used in the physics community due to its simple use and effectiveness (among the others, we refer to the works of Berezinskii [4] and of Kosterlitz and Thouless [22] who were awarded the 2016 Nobel Prize for Physics, together with Haldane) but has also attracted the attention of the mathematics community, see for instance [1, 2, 7].

For shells M with $\chi(M) \neq 0$, the energy (1) is clearly not well defined due to Theorem 1.1 and we have to face the emergence of configurations with defects. In Sect. 3 we will discuss the location of defects and their energetics. A possible strategy would be to relax one the above constraints, for instance the unit-norm constraint as in the Ginzburg-Landau theory (see, for instance, [5, 20, 21, 31, 32] and the recent papers [17] and [18] for the analysis on a Riemannian manifold). In this note, we present the approach of [8] and instead of a continuous model we rather consider the discrete XY spin model (4). Defects emerge when we let $\varepsilon \to 0$ in (4). In particular, we will address the Γ-convergence of the energy

$$XY_\varepsilon(\cdot) - \pi\mathscr{K}|\log\varepsilon|,$$

where $\mathscr{K}$ is an even, positive integer, such that $|\chi(M)| \leq \mathscr{K}$. What appears in the limit is the so called Renormalized Energy (introduced and studied first in [5] and then in many other contributions, see [3, 32] and references therein) that describes the energetics and the interaction between defects. The Renormalized Energy we obtain is given by the sum of a purely intrinsic part and of an extrinsic part related to the shape operator of M and thus the location of the defects also depends on how the shell "sits" in the three dimensional space. At the level of minimizers, we have the following expansion

$$\min XY_\varepsilon = \pi|\chi(M)||\log\varepsilon| + \mathbb{W}(\mathbf{v}) + \sum_{i=1}^{|\chi(M)|} \gamma(x_i) + o_{\varepsilon\to 0}(1),$$

where $\mathbf{v} \in W^{1,2}_{\text{tan,loc}}(M \setminus \{x_1, \dots, x_{|\chi(M)|}\}; \mathbb{S}^2)$ is the "continuum limit" of the sequence of discrete minimizers and $\gamma(x_i)$ is a positive quantity that takes into account the energy located in the core of the defects x_i of $\mathbf{v}$. An interesting feature that is not shared in the planar case, both continuous and discrete (see [5] and [2]), nor in the curved continuous case (see [17] and [18]), is that the core energy $\gamma(x_i)$ depends on the singularity x_i.

The interest in analyzing configurations with defects goes beyond the aesthetic appeal of the question. In fact, the defect's points could serve as anchoring bonds between colloidal particles, as precognized by Nelson [29] and recently realized in [43]. Thus, the understanding of the defects formation and of their energetics and location could be of impact for this new chemistry for meta materials.

We conclude this long introduction with some differential geometry notation that we use. We refer to the book [13] for all the material regarding differential geometry.

Given a compact two dimensional surface M with metric g, embedded in $\mathbb{R}^3$ and oriented with the normal $\boldsymbol{\gamma}$, we denote the area element induced by the choice of the orientation with dS. We denote with ∇ the connection with respect to the standard metric of $\mathbb{R}^3$, and we let $\mathrm{D}_\mathbf{v}\mathbf{u}$ be the covariant derivative of $\mathbf{u}$ in the direction $\mathbf{v}$ ($\mathbf{u}$ and $\mathbf{v}$ are smooth tangent vector fields in M), with respect to the Levi Civita (or Riemannian) connection D of the metric g on M.

Now, if $\mathbf{u}$ and $\mathbf{v}$ are extended arbitrarily to smooth vector fields on $\mathbb{R}^3$, we have the Gauss Formula :

$$\nabla_\mathbf{v}\mathbf{u} = \mathrm{D}_\mathbf{v}\mathbf{u} + \langle \mathrm{d}\boldsymbol{\gamma}(\mathbf{u}), \mathbf{v}\rangle\boldsymbol{\gamma}. \tag{5}$$

This decomposition is orthogonal, thus there holds

$$|\nabla\mathbf{u}|^2 = |\mathrm{D}\mathbf{u}|^2 + |\mathrm{d}\boldsymbol{\gamma}(\mathbf{u})|^2. \tag{6}$$

Beside the covariant derivative, we introduce another differential operator for vector fields on M, which takes into account also the way that M embeds in $\mathbb{R}^3$. Let $\mathbf{u}$ be a smooth vector field on M. We extend it smoothly to a vector field $\tilde{\mathbf{u}}$ on $\mathbb{R}^3$ and we denote its standard gradient by $\nabla\tilde{\mathbf{u}}$ on $\mathbb{R}^3$. For $x \in M$, we define

$$\nabla_\mathrm{s}\mathbf{u}(x) := \nabla\tilde{\mathbf{u}}(x)P_M(x),$$

where $P_M(x) := (\mathrm{Id} - \boldsymbol{\gamma} \otimes \boldsymbol{\gamma})(x)$ is the orthogonal projection on T_xM. In other words, ∇_s is the restriction of the standard derivative in $\mathbb{R}^3$ to directions that are tangent to M. This differential operator is well-defined, as it does not depend on the particular extension $\tilde{\mathbf{u}}$. In general, $\nabla_\mathrm{s}\mathbf{u} \neq \mathrm{D}\mathbf{u} = P_M(\nabla\mathbf{u})$ since the matrix product is non commutative. Moreover, thanks to (5) and (6) there holds

$$|\nabla_{\mathrm{s}}\mathbf{u}|^2 = |\mathrm{D}\mathbf{u}|^2 + |\mathrm{d}\boldsymbol{\gamma}(\mathbf{u})|^2.$$

Note that, by identifying $\mathbf{u}$ with a map $\mathbf{u} = (\mathbf{u}^1, \mathbf{u}^2, \mathbf{u}^3) \colon M \to \mathbb{R}^3$, the k-th row of the matrix representing $\nabla_{\mathrm{s}}\mathbf{u}$ coincides with the Riemannian gradient (that we still denote with ∇_{s}) of $\mathbf{u}^k$.

2 Shells of Zero Euler Characteristic

According to Theorem 1.1, unless otherwise stated, throughout this section we will consider M to be a compact and smooth two-dimensional surface without boundary such that

$$\boxed{M \texttt{ has Euler characteristic equal to zero, that is } \chi(M) = 0} \tag{7}$$

and we will leave to the next Sect. 3 the case of a shell M with $\chi(M) \neq 0$. This section is organized as follows. First of all, in Sect. 2.1 we will discuss the minimization of the full energy (3) while in Sect. 2.2 we will study the gradient flow of the energy (1) with respect to the scalar product of L^2. Finally, in Sect. 2.3 we discuss the rigorous derivation (in terms of Γ-convergence) of the energy (1) from the discrete energy (4). It is an open problem to justify in terms of a microscopic derivation the full energy (3), even in the euclidean case.

2.1 Existence of Minimizers

We let M satisfy (7), in such a way that $W^{1,2}_{\tan}(M, \mathbb{S}^2) \neq \emptyset$, we have the following (see [15] for the flat case)

Proposition 1 *Let M be a smooth, compact surface in $\mathbb{R}^3$, without boundary, satisfying (7) and let $W : W^{1,2}_{\tan}(M, \mathbb{S}^2) \to \mathbb{R}$ be the energy functional (1). Set $K_* := \min\{K_1, K_2, K_3\}$ and $K^* := 3(K_1 + K_2 + K_3)$. We have that*

$$\frac{K_*}{2}\int_M \left(|\mathrm{D}\mathbf{n}|^2 + |\mathrm{d}\boldsymbol{\gamma}(\mathbf{n})|^2\right)\mathrm{d}S \le W_{\mathrm{extr}}(\mathbf{n}) \le \frac{K^*}{2}\int_M \left(|\mathrm{D}\mathbf{n}|^2 + |\mathrm{d}\boldsymbol{\gamma}(\mathbf{n})|^2\right)\mathrm{d}S.$$

Moreover, the energy W is sequentially lower semicontinuous with respect to the weak convergence of $W^{1,2}(M; \mathbb{R}^3)$.

Thus, the existence of a minimizer of the energy W follows from the direct method of calculus of variations.

Proposition 2 *There exists* $\mathbf{n} \in W^{1,2}_{\tan}(M,\,\mathbb{S}^2)$ *such that* $W_{\mathrm{extr}}(\mathbf{n}) = \inf_{\mathbf{u}\in W^{1,2}_{\tan}(M,\,\mathbb{S}^2)} W_{\mathrm{extr}}(\mathbf{u})$.

The proof of the existence of minimizers is simple being the energy quadratic. It is interesting to discuss how the energy selects the minimizers. We leave to [36] the discussion on the relation between the different tunings of K_1, K_2, K_3 and the energy landscape for constant deviation angle (namely the angle that $\mathbf{n}$ forms with one of vectors generating the tangent plane to M, see formula (8) below) and we rather concentrate on the one-constant approximation. We observe that the energy (1) has the form of a "phase transition energy" since it is the sum of a Dirichlet part and of a (vectorial) double well potential part. In fact the purely extrinsic part $|\mathrm{d}\boldsymbol{\gamma}(\mathbf{n})|^2$ is minimized when $\mathbf{n}$ is oriented along the direction of minimal principal curvature (i.e. minimal normal curvature). Thus, the energy (1) favors a parallel configuration (i.e. a vector field such that $\mathrm{D}\mathbf{n} = 0$) in the direction of minimal principal curvature. Already considering only the Dirichlet part (i.e. the intrinsic energy) the minimization experiences an interesting frustration of geometric nature due to the fact that the existence of globally defined unit norm parallel vector fields requires the Gaussian curvature to vanish. The effect of the competition between the two terms of the energy is particularly interesting on the axisymmetric torus. Thus, we fix M to be the axisymmetric torus, namely the surface parametrized by $X : [0, 2\pi] \times [0, 2\pi] \to \mathbb{R}^3$ where

$$X(\theta,\phi) = \begin{pmatrix} (R + r\cos\theta)\cos\phi \\ (R + r\cos\theta)\sin\phi \\ r\sin\theta \end{pmatrix}.$$

R and r are usually known as major and minor radius, respectively. We let $\mathbf{e}_1$ and $\mathbf{e}_2$ be the unit tangent vectors given by

$$\mathbf{e}_1 = \frac{X_\theta}{|X_\theta|}, \quad \mathbf{e}_2 = \frac{X_\phi}{|X_\phi|},$$

and we let c_1 and c_2 be the principal curvatures

$$c_1 = \frac{1}{r}, \quad c_2(\theta,\phi) = \frac{\cos\theta}{R + r\cos\theta}.$$

Then, we proceed as in [36] and we represent the director field $\mathbf{n}$ as

$$\mathbf{n} = \cos\alpha\,\mathbf{e}_1 + \sin\alpha\,\mathbf{e}_2. \tag{8}$$

The angle α is named deviation angle. We restrict to vector fields $\mathbf{n} \in W^{1,2}_{\tan}(M,\,\mathbb{S}^2)$ with zero winding number (the general case is discussed in [36]). Thus the deviation angle turns out to be periodic, namely $\alpha \in H^1_{\mathrm{per}}(Q)$. The energy expressed in terms of α is particularly appealing for the analysis. Setting $W(\alpha) = W(\mathbf{n})$, with $\mathbf{n}$ given by (8), we have

$$W(\alpha) = \frac{1}{2}\int_Q \left\{\kappa|\nabla_s \alpha|^2 + \eta\cos(2\alpha)\right\} \mathrm{d}S + \kappa\pi^2\left(\frac{2-\mu^2}{\sqrt{\mu^2-1}} + 2\mu\right), \tag{9}$$

where $\eta(\theta,\phi) := \kappa\frac{c_1^2 - c_2^2(\theta,\phi)}{2} = \kappa\frac{R^2+2Rr\cos\theta}{2r^2(R+r\cos\theta)^2}$, and $\mu := \frac{R}{r}$. The number μ is called *aspect ratio* and plays a prominent role in the minimization. In the next Proposition we discuss the dependence of minimizers on the aspect ratio μ. In particular, we discuss the stability of the minimizers.

Proposition 3 *Let $\mu := R/r$. There exists $\mu^* \in (2/\sqrt{3}, 2]$ such that the constant values $\alpha = \pi/2 + m\pi$, $m \in \mathbb{Z}$, are local minimizers for W in $H^1_{\mathrm{per}}(Q)$ if and only if $\mu \geq \mu^*$. Moreover, if $\mu \geq 2$, there exists no non-constant solution w to the Euler Lagrange equation*

$$-\Delta_s\alpha = \frac{\kappa}{2}(c_1^2 - c_2^2)\sin(2\alpha) \quad \textit{in } Q$$

such that

$$\frac{\pi}{2} + m\pi \leq w \leq \frac{\pi}{2} + (m+1)\pi.$$

The proof of the proposition is in [36]. It is worthwhile noting that it is an interesting open problem to analytically determine the exact value of the critical threshold μ^*. Numerics indicates that $\mu^* \approx 1.52$.

Proposition 3 is important since it describes how the Napoli-Vergori energy (1) acts. In particular, it shows the differences—for a toroidal shell—with the classical intrinsic energy (2). It turns out that the presence of the extrinsic term related to the shape operator acts as a selection principle for equilibrium configurations. More precisely, when $\mu := R/r$ is sufficiently large then (see Proposition 3) the only constant solution is $\alpha = \pi/2 + m\pi$ ($m \in \mathbb{Z}$). Moreover, when $R/r < \mu^*$ a new class of non constant solutions appears (see Fig. 1, obtained discretizing the gradient flow equation). We make the following observation: This new solution tries to minimize the effect of the curvature by orienting the director field along the meridian lines ($\alpha = 0$), which are geodesics on the torus, near the hole of the torus, while near the external equator the director is oriented along the parallel lines $\alpha = \pi/2$, which are lines of curvature. The fact that the solution $\alpha = \pi/2$ is no longer stable for sufficiently small μ is due to the high bending energy associated to $\alpha = \pi/2$ in the internal hole of the torus. In fact, in a small strip close to the internal equator of the torus, we can approximate (see [36])

$$c_1^2 - c_2^2 \approx \frac{1}{r^2} - \frac{1}{(R-r)^2}, \quad \mathrm{d}S \approx r(R-r)\mathrm{d}\theta\,\mathrm{d}\phi,$$

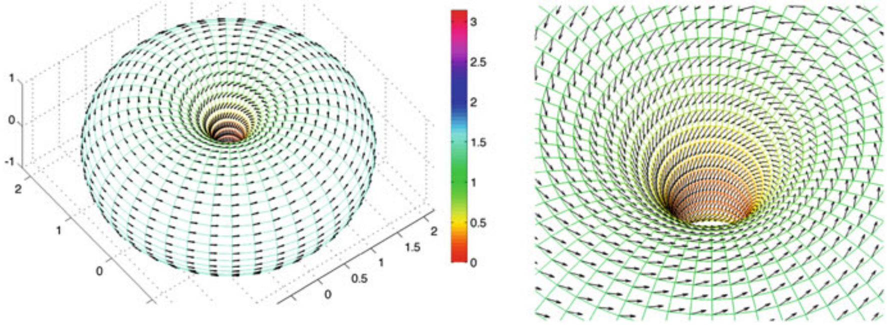

Fig. 1 Configuration of the scalar field α and of the vector field $\mathbf{n}$ of a numerical solution to the gradient flow of (9) in the case $R/r = 1.2$ (left). Zoom-in of the central region of the same fields (right). The colour represents the angle $\alpha \in [0, \pi]$, the arrows represent the corresponding vector field $\mathbf{n}$

and therefore

$$(c_1^2 - c_2^2)\cos(\pi)\mathrm{d}S \approx \mu\frac{2-\mu}{\mu - 1}\mathrm{d}\theta\,\mathrm{d}\phi,$$

which tends to $+\infty$ as $\mu \to 1$.

Due to its “double well”-like structure, the energy (9) favors a smooth transition between $\alpha = \pi/2$ and $\alpha = 0$. In this sense, the new solution can be understood as an interpolation between $\alpha = \pi/2$ and $\alpha = 0$, which are the two constant stationary solutions of the system.

2.2 Existence of Solutions of the Gradient Flow of (1)

We then focus on the L^2-gradient flow of the one-constant approximation energy (1). The study of the gradient flow for the energy (1) could be seen as a starting point for the analysis of an Ericksen-Leslie type model for nematic shells. This problem has already been addressed in [38] where various well-posedness and long-time behavior results have been obtained for an Ericksen-Leslie type model on Riemannian manifolds. However, it should be pointed out that the model in [38] is purely intrinsic and does not take into account the way the substrate on which the nematic is deposited sits in the three-dimensional space. Moreover, in the equation describing the evolution of $\mathbf{n}$ (called $\mathbf{d}$ therein) the constraint $|\mathbf{n}| = 1$ is not considered.

We prove (see Theorem 2.1) the well-posedness of the L^2-gradient flow of (1), i.e.

$$\begin{cases} \partial_t \mathbf{n} - \Delta_g \mathbf{n} + \mathfrak{B}^2\mathbf{n} = |\mathrm{D}\mathbf{n}|^2\mathbf{n} + |\mathrm{d}\boldsymbol{\gamma}(\mathbf{n})|^2\mathbf{n} \quad \text{in } M \times (0, +\infty), \\ \mathbf{n}(x, 0) = \mathbf{n}_0 \quad \text{a.e. in } M. \end{cases} \tag{10}$$

Here Δ_g is the *rough Laplacian* and $\mathfrak{B}^2$ is the linear operator $(\mathfrak{B}^2\mathbf{u}, \mathbf{v})_{\mathbb{R}^3} := (\mathrm{d}\boldsymbol{\gamma}(\mathbf{u}), \mathrm{d}\boldsymbol{\gamma}(\mathbf{v}))_{\mathbb{R}^3}$ for any $\mathbf{u}, \mathbf{v}$ tangent vector fields. The right-hand side of (10) is a result of the unit-norm constraint on the director $\mathbf{n}$. The initial datum $\mathbf{n}_0$ is taken in $W^{1,2}_{\tan}(M, \mathbb{S}^2)$ and we look for weak solutions with bounded energy. A proof of the existence relying on *(i) discretization, (ii) a priori estimates, (iii) convergence of discrete solutions*, would encounter a difficulty here, as the nonlinear term $|\mathrm{D}\mathbf{n}|^2$ in the right-hand side of (10) is not continuous with respect to the weak-$W^{1,2}$ convergence expected from the a priori estimates. We overcome this problem with techniques employed in the study of the heat flow for harmonic maps (see [10, 11]): we first relax the unit-norm constraint with a Ginzburg-Landau approximation, i.e., we allow for vectors $\mathbf{n}$ with $|\mathbf{n}| \neq 1$, but we penalize deviations from unitary length at the order $1/\varepsilon^2$, for a small parameter $\varepsilon > 0$. More precisely, we construct (via a time discretization argument) a sequence of fields $\mathbf{n}^\varepsilon$ which solve

$$\begin{cases} \partial_t \mathbf{n}^\varepsilon - \Delta_g \mathbf{n}^\varepsilon + \mathfrak{B}^2\mathbf{n}^\varepsilon + \frac{1}{\varepsilon^2}(|\mathbf{n}^\varepsilon|^2 - 1)\mathbf{n}^\varepsilon = 0, \ \text{a.e. in } M \times (0, +\infty), \\ \mathbf{n}^\varepsilon(x, 0) = \mathbf{n}_0 \quad \text{a.e. in } M. \end{cases}$$

The above equation has a gradient flow structure. Thus, we have

$$\|\partial_t \mathbf{n}^\varepsilon(t)\|^2 + \frac{\mathrm{d}}{\mathrm{d}t} W(\mathbf{n}^\varepsilon(t)) + \frac{1}{4\varepsilon^2}\frac{\mathrm{d}}{\mathrm{d}t}\int_M (|\mathbf{n}^\varepsilon(t)|^2 - 1)^2 \mathrm{d}S = 0.$$

which produces the following energy estimate when the initial condition $\mathbf{n}_0$ has finite energy

$$\|\partial_t \mathbf{n}^\varepsilon\|^2_{L^2(0,T;L^2_{\tan}(M))} + \|\mathrm{D}\mathbf{n}^\varepsilon\|^2_{L^\infty(0,T;L^2_{\tan}(M))} + \|\mathrm{d}\boldsymbol{\gamma}(\mathbf{n}^\varepsilon)\|^2_{L^\infty(0,T;L^2_{\tan}(M))} \\ + \sup_{t\in(0,T)} \frac{1}{4\varepsilon^2}\int_M (|\mathbf{n}^\varepsilon(t)|^2 - 1)^2 \mathrm{d}S \leq 3W(\mathbf{n}_0).$$

Via standard compactness arguments, one obtains the existence of limit vector field $\mathbf{n}$ with the energy regularity specified by the above estimate. The difficult part is clearly to pass to the limit in the approximate equation and to show that the field $\mathbf{n}$ is indeed a weak solution of (10) The crucial observation (borrowed from [10, 11]) is that for a smooth unit-norm field $\mathbf{n}$, (10) is equivalent to

$$(\partial_t \mathbf{n} - \Delta_g \mathbf{n} + \mathfrak{B}^2\mathbf{n}) \times \mathbf{n} = 0. \tag{11}$$

To highlight the importance of the reformulation (11), let us consider the case of an harmonic map $u : \Omega \to \mathbb{S}^2$ with $\Omega \subset \mathbb{R}^n$ an open set. Being an harmonic map, u solves the nonlinear elliptic equation

$$\Delta u + u|\nabla u|^2 = 0 \quad \text{in } \Omega. \tag{12}$$

Now, taking the vector product of the equation with u, one obtains that u solves (12) if and only if it solves

$$\Delta u \times u = 0 \quad \text{in } \Omega,$$

which is equivalent to

$$\sum_{i=1}^{n} \frac{\partial}{\partial x_i}(u \times \frac{\partial u}{\partial x_i}) = 0 \quad \text{in } \Omega. \tag{13}$$

Note that, differently from (12), the equation (13) is in divergence form and thus is it more treatable in weak regularity contexts. The above strategy can be implemented in our case and gives that (see [25, Lemma 7.5.4] for a similar argument)

Lemma 1 *A vector field* $\mathbf{n} \in W^{1,2}(0,T; L^2_{\tan}(M)) \cap L^\infty(0,T; W^{1,2}_{\tan}(M, \mathbb{S}^2))$ *is a weak solution of* (10) *if and only if it solves*

$$-\int_M (\partial_t \mathbf{n} \times \mathbf{n}, \boldsymbol{\gamma})_{\mathbb{R}^3}\, \psi \,\mathrm{d}S + \int_M g^{ij}(\mathrm{D}_i \mathbf{n}, \boldsymbol{\gamma} \times \mathbf{n})_{\mathbb{R}^3}\, \partial_j \psi \,\mathrm{d}S - \int_M (\mathfrak{B}^2 \mathbf{n} \times \mathbf{n}, \boldsymbol{\gamma})_{\mathbb{R}^3}\, \psi \,\mathrm{d}S = 0 \tag{14}$$

for any smooth function $\psi : M \to \mathbb{R}$.

Thus the strategy is as follows. First of all, we test the weak formulation of (10) with the vector field $\phi = \psi \boldsymbol{\gamma} \times \mathbf{n}^\varepsilon$ where $\psi : M \to \mathbb{R}$ is smooth. We obtain

$$\begin{aligned} -\int_M (\partial_t \mathbf{n}^\varepsilon \times \mathbf{n}^\varepsilon, \boldsymbol{\gamma})_{\mathbb{R}^3}\, \psi \,\mathrm{d}S + \int_M g^{ij}(\mathrm{D}_i \mathbf{n}^\varepsilon, \boldsymbol{\gamma} \times \mathbf{n}^\varepsilon)_{\mathbb{R}^3}\, \partial_j \psi \,\mathrm{d}S \\ - \int_M (\mathfrak{B}^2 \mathbf{n}^\varepsilon \times \mathbf{n}^\varepsilon, \boldsymbol{\gamma})_{\mathbb{R}^3} \psi \,\mathrm{d}S = 0, \end{aligned} \tag{15}$$

where the penalization term has disappeared thanks to $(a, b \times a)_{\mathbb{R}^3} = (b, a \times a)_{\mathbb{R}^3} = 0$, for $a, b \in \mathbb{R}^3$. Now, (15) has a “divergence” structure and thus is adequate for the limit procedure with respect to the convergences given by the energy estimate. Consequently, we pass to the limit in (15) and we obtain that $\mathbf{n}$ solves (14) that is equivalent to (10) thanks to the above lemma. Thus, we have (see [36, Theorem 5.1] for the details)

Theorem 2.1 *Let M be a two-dimensional compact surface satisfying* (7)*. Given $\mathbf{n}_0 \in W^{1,2}_{\tan}(M, \mathbb{S}^2)$ there exists a global weak solution to* (10) *with $\mathbf{n}(\cdot, 0) = \mathbf{n}_0(\cdot)$ in M.*

2.3 Justification of the Energy (1): A Discrete to Continuum Approach

In this subsection we show how the energy (1) emerges as the discrete to continuum limit of a discrete energy of XY type. We recall that we will use the very same discrete energy to understand the generation of defects for shells with non zero Euler Characteristic in the next Sect. 3. The main tool of our analysis will be the concept of Γ-convergence for which we refer to the book of G. Dal Maso [12].

The discrete energy we consider is defined on a triangulation of the surface M. Thus, before introducing the discrete energy, we have to (briefly) introduce the discrete formalism. We refer to the paper [8] for the details of the construction.

For any $\varepsilon \in (0, \varepsilon_0]$, we let $\mathcal{T}_\varepsilon$ be a triangulation of M, that is, a finite collection of non-degenerate affine triangles $T \subseteq \mathbb{R}^3$ with the following property: the intersection of any two triangles $T, T' \in \mathcal{T}_\varepsilon$ is either empty or a common subsimplex of T, T'. The parameter ε is the mesh size, namely we assume $\varepsilon = \max_{T\in\mathcal{T}_\varepsilon} \operatorname{diam}(T)$. The set of vertices of $\mathcal{T}_\varepsilon$ will be denoted by $\mathcal{T}^0_\varepsilon$. We will always assume that $\mathcal{T}^0_\varepsilon \subseteq M$. We set $\widehat{M}_\varepsilon := \cup_{T\in\mathcal{T}_\varepsilon} T$, so $\widehat{M}_\varepsilon$ is the piecewise-affine approximation of M induced by $\mathcal{T}_\varepsilon$. Given a piecewise-smooth function $\mathbf{u}\colon \widehat{M}_\varepsilon \to \mathbb{R}^k$, we denote by $\nabla_\varepsilon \mathbf{u}$ the restriction of the derivative $\nabla \mathbf{u}$ to directions that lie in the triangles of $\widehat{M}_\varepsilon$.

We will only consider family of triangulations $(\mathcal{T}_\varepsilon)$ that satisfy the following conditions.

(H$_1$) There exists a constant $\Lambda > 0$ such that, for any $\varepsilon \in (0, \varepsilon_0]$ and any $T \in \mathcal{T}_\varepsilon$, the (unique) affine bijection $\phi\colon T_{\mathrm{ref}} \to T$ satisfies

$$\operatorname{Lip}(\phi) \le \Lambda\varepsilon, \qquad \operatorname{Lip}(\phi^{-1}) \le \Lambda\varepsilon^{-1},$$

where $T_{\mathrm{ref}} \subseteq \mathbb{R}^2$ be a reference triangle of vertices $(0, 0)$, $(1, 0)$ and $(0, 1)$. Here $\operatorname{Lip}(\phi)$ denotes the Lipschitz constant of ϕ, $\operatorname{Lip}(\phi) := \sup_{x\neq y} |x - y|^{-1}|\phi(x) - \phi(y)|$.

(H$_2$) For any $\varepsilon \in (0, \varepsilon_0]$ and any $i, j \in \mathcal{T}^0_\varepsilon$ with $i \neq j$, the stiffness matrix κ^{ij}_ε of the Laplace Beltrami operator on M satisfies

$$\kappa^{ij}_\varepsilon := -\int_{\widehat{M}_\varepsilon} \nabla_\varepsilon \widehat{\varphi}_{\varepsilon,i} \cdot \nabla_\varepsilon \widehat{\varphi}_{\varepsilon,j} \,\mathrm{d}S \ge 0,$$

where the hat function $\widehat{\varphi}_{\varepsilon,i}$ is the unique piecewise-affine, continuous function $\widehat{M}_\varepsilon \to \mathbb{R}$ such that $\widehat{\varphi}_{\varepsilon,i}(j) = \delta_{ij}$ for any $j \in \mathcal{T}^0_\varepsilon$.

(H$_3$) For any $\varepsilon \in (0, \varepsilon_0]$, $\widehat{M}_\varepsilon \subseteq U$ and the restriction of the nearest-point projection $\widehat{P}_\varepsilon := P_{|\widehat{M}_\varepsilon} \colon \widehat{M}_\varepsilon \to M$ has a Lipschitz inverse. Moreover, we have $\mathrm{Lip}(\widehat{P}_\varepsilon) + \mathrm{Lip}(\widehat{P}_\varepsilon^{-1}) \leq \Lambda$ for some ε-independent constant Λ.

An important consequence of the assumption (H$_3$) is that the restriction of the nearest-point projection $\widehat{P}_\varepsilon \colon \widehat{M}_\varepsilon \to M$ has a Lipschitz inverse $\widehat{P}_\varepsilon^{-1} \colon M \to \widehat{M}_\varepsilon$. Following [16], we use $\widehat{P}_\varepsilon$ and $\widehat{P}_\varepsilon^{-1}$ to construct the so called *metric distorsion tensor*. This object will be important in our analysis since it will permit to rewrite our discrete energy as an energy for a proper vector field interpolating the discrete spins. To introduce the *metric distorsion tensor*, we proceed as follow. For any $x \in M$ such that $\widehat{P}_\varepsilon^{-1}(x)$ falls in the interior of a triangle of $\widehat{M}_\varepsilon$ (so that $\widehat{P}_\varepsilon^{-1}$ is smooth in a neighbourhood of x), we let the *metric distorsion tensor* $\mathbf{A}_\varepsilon(x)$ to be the unique linear operator $\mathrm{T}_x M \to \mathrm{T}_x M$ that satisfies

$$(\mathbf{A}_\varepsilon(x)\mathbf{X},\, \mathbf{Y}) = \left(\mathrm{d}\widehat{P}_\varepsilon^{-1}(x)[\mathbf{X}],\, \mathrm{d}\widehat{P}_\varepsilon^{-1}(x)[\mathbf{Y}]\right) \tag{16}$$

for any $\mathbf{X}, \mathbf{Y} \in \mathrm{T}_x M$. The metric distorsion tensor is symmetric and positive definite, since the right-hand side of (16) is. Consequently, we introduce a norm $\|\cdot\|_{L^\infty(M)}$ on $L^\infty(M;\, \mathrm{T}M \otimes \mathrm{T}^*M)$ by

$$\|\mathbf{A}\|_{L^\infty(M)} := \operatorname*{ess\,sup}_{x\in M} \|\mathbf{A}(x)\|_{\mathrm{T}M\otimes\mathrm{T}^*M},$$

where $\|\cdot\|_{\mathrm{T}M\otimes\mathrm{T}^*M}$ is the operator norm. The following lemma (see [8, Lemma 2]) is important.

Lemma 2 *Suppose that* $(\mathcal{T}_\varepsilon)$ *satisfies* (H1) *and* (H3). *Then, there holds*

$$\|\mathbf{A}_\varepsilon - \mathrm{Id}\,\|_{L^\infty(M)} + \|\mathbf{A}_\varepsilon^{-1} - \mathrm{Id}\,\|_{L^\infty(M)} \leq C\varepsilon.$$

Let $g_\varepsilon \in L^\infty(M;\, \mathrm{T}^*M^{\otimes 2})$ be the metric on M defined by $g_\varepsilon(\mathbf{X}, \mathbf{Y}) := (\mathbf{A}_\varepsilon\mathbf{X},\, \mathbf{Y})$, for any smooth fields $\mathbf{X}$ and $\mathbf{Y}$ on M. Given a function $u \in W^{1,2}(M)$, one can define the Sobolev $W^{1,2}$-seminorm of u with respect to g_ε, i.e.

$$|u|^2_{W^{1,2}_\varepsilon(M)} := \int_M \left(\mathbf{A}_\varepsilon^{-1}\nabla_{\mathrm{s}} u,\, \nabla_{\mathrm{s}} u\right) (\det \mathbf{A}_\varepsilon)^{1/2}\, \mathrm{d}S, \tag{17}$$

where ∇_{s} denotes the Riemaniann gradient and dS the volume form on M (with respect to the metric induced by $\mathbb{R}^3$). By construction (16), the map $\widehat{P}_\varepsilon^{-1}$ is an isometry between M, equipped with the metric g_ε, and $\widehat{M}_\varepsilon$, with the metric induced by $\mathbb{R}^3$. Therefore, given $v \in W^{1,2}(\widehat{M}_\varepsilon;\, \mathbb{R})$ and a Borel set $U \subseteq M$, there holds

$$|v \circ \widehat{P}_\varepsilon^{-1}|^2_{W^{1,2}_\varepsilon(U)} = \int_{\widehat{P}_\varepsilon^{-1}(U)} |\nabla_\varepsilon v|^2\, \mathrm{d}S.$$

Arguing component-wise, we see that the same equality holds for a (not necessarily tangent) vector field $\mathbf{v}\colon \widehat{M}_\varepsilon \to \mathbb{R}^3$ in place of v.

Using assumption (H3), to any discrete vector field $\mathbf{v}_\varepsilon \in \mathrm{T}(\mathcal{T}_\varepsilon;\, \mathbb{S}^2)$ we can associate a continuous field $\mathbf{w}_\varepsilon\colon M \to \mathbb{R}^3$ by setting

$$\mathbf{w}_\varepsilon := \widehat{\mathbf{v}}_\varepsilon \circ \widehat{P}_\varepsilon^{-1}, \tag{18}$$

where $\widehat{\mathbf{v}}_\varepsilon\colon \widehat{M}_\varepsilon \to \mathbb{R}^3$ is the affine interpolant of $\mathbf{v}_\varepsilon$. The field $\mathbf{w}_\varepsilon$ is Lipschitz-continuous and satisfies $\mathbf{w}_\varepsilon = \mathbf{v}_\varepsilon$ on $\mathcal{T}_\varepsilon^0$, but it is not tangent to M nor unit-valued, in general. However, one can still prove some useful properties that we collect in a single lemma (see [8, Lemma 3, Lemma 4, Lemma 5] for the proofs).

Lemma 3 *Suppose that* (H1)*,* (H2)*,* (H3) *are satisfied. Then, for any* $\varepsilon \in (0, \varepsilon_0]$ *and any discrete field* $\mathbf{v}_\varepsilon \in \mathrm{T}(\mathcal{T}_\varepsilon;\, \mathbb{S}^2)$*,* $\mathbf{w}_\varepsilon$ *is Lipschitz-continuous with Lipschitz constant*

$$\mathrm{Lip}(\mathbf{w}_\varepsilon) \le C\varepsilon^{-1}.$$

Moreover, $\mathbf{w}_\varepsilon$ *satisfies the following*

- *For any subset* $\widehat{U} \subseteq \widehat{M}_\varepsilon$ *that can be written as union of triangles of* $\mathcal{T}_\varepsilon$*, there holds*

$$XY_\varepsilon(\mathbf{v}_\varepsilon,\, \widehat{U}) := \frac{1}{2} \sum_{i,j \in \mathcal{T}_\varepsilon^0 \cap \widehat{U}} \kappa_\varepsilon^{ij}\, |\mathbf{v}_\varepsilon(i) - \mathbf{v}_\varepsilon(j)|^2 = \frac{1}{2} |\mathbf{w}_\varepsilon|^2_{W_\varepsilon^{1,2}(P(\widehat{U}))}. \tag{19}$$

- *There exists a positive constant C such that*

$$\|(\mathbf{w}_\varepsilon,\, \boldsymbol{\gamma})\|_{L^\infty(M)} \le C\varepsilon, \;\; \textit{and}\; \frac{1}{\varepsilon^2} \int_{\widehat{M}_\varepsilon} \left(1 - |\mathbf{w}_\varepsilon|^2\right)^2 \le C\, XY_\varepsilon(\mathbf{v}_\varepsilon). \tag{20}$$

Another immediate but important consequence of the lemma above is a compactness result for discrete sequences $\mathbf{v}_\varepsilon$ with equi-bounded energy with respect to ε.

Lemma 4 *Let* $\mathbf{v}_\varepsilon \in \mathrm{T}(\mathcal{T}_\varepsilon;\, \mathbb{S}^2)$ *be a sequence such that*

$$XY_\varepsilon(\mathbf{v}_\varepsilon) \le C \qquad \textit{for any } \varepsilon > 0,$$

then there exists $\mathbf{v} \in W^{1,2}_{\tan}(M,\, \mathbb{S}^2)$ *and a subsequence of* ε *such that, defining* $\mathbf{w}_\varepsilon$ *as in* (18)*, there holds*

$$\mathbf{w}_\varepsilon \xrightarrow{\varepsilon \to 0} \mathbf{v} \quad \textit{strongly in } L^2(M; \mathbb{R}^3). \tag{21}$$

Then, we have the following.

Theorem 2.2 *Suppose that the assumptions* (H1), (H2) *and* (H3) *are satisfied. Then, XY_ε Γ-converges with respect to weak convergence of $L^2(M;\mathbb{R}^3)$ to the functional*

$$W(\mathbf{v}) := \begin{cases} \frac{1}{2}\int_M |\mathrm{D}\mathbf{v}|^2 + |\mathrm{d}\boldsymbol{\gamma}[\mathbf{v}]|^2 \mathrm{d}S, & \textit{if } \mathbf{v} \in W^{1,2}_{\tan}(M;\mathbb{S}^2) \\ +\infty, & \textit{otherwise in } L^2(M;\mathbb{R}^3). \end{cases}$$

The proof follows standard argument in the analysis of discrete to continuum limits via Γ-convergence (see, e.g., [1, 7] and the Lecture Notes [34] for a slightly different model). We highlight the main points for future reference since, to the best of our knowledge, the proof of this result is not contained in any contribution.

Proof (Proof—Γ-liminf Inequality) We are given a sequence of discrete vector fields $\mathbf{v}_\varepsilon$ and we aim to prove that there exists a unit norm tangent vector field $\mathbf{v}$ such that $\mathbf{w}_\varepsilon \to \mathbf{v}$ weakly in $L^2(M;\mathbb{R}^3)$ (actually much more is true) and

$$\liminf_{\varepsilon\to 0} XY_\varepsilon(\mathbf{v}_\varepsilon) \geq W(\mathbf{v}). \tag{22}$$

Without loss of generality, we may assume that there exists a constant C such that $XY_\varepsilon(\mathbf{v}_\varepsilon) \leq C$ for any ε (if not (22) is trivially satisfied). Thus, we have that the sequence $\mathbf{w}_\varepsilon$ defined in (18) is bounded, uniformly with respect to ε, in $W^{1,2}_\varepsilon(M)$. Then, the compactness result in Lemma 4 gives that there exists a subsequence, still denoted with $\mathbf{w}_\varepsilon$, and a vector field $\mathbf{v} \in W^{1,2}(M;\mathbb{R}^3)$ for which

$$\mathbf{w}_\varepsilon \xrightarrow{\varepsilon\to 0} \mathbf{v} \quad \text{strongly in } L^2(M;\mathbb{R}^3). \tag{23}$$

This convergence, combined with (20), give that $\mathbf{v}$ is tangent and $|\mathbf{v}| = 1$, namely $\mathbf{v} \in W^{1,2}_{\tan}(M,\ \mathbb{S}^2)$. Finally, since there holds (see (17))

$$XY_\varepsilon(\mathbf{v}_\varepsilon) = \frac{1}{2}|\mathbf{w}_\varepsilon|^2_{W^{1,2}_\varepsilon(M)} = \frac{1}{2}\sum_{i=1}^{3}|\mathbf{w}^i_\varepsilon|^2_{W^{1,2}_\varepsilon(M)} = \frac{1}{2}\sum_{i=1}^{3}\int_M \left(\mathbf{A}_\varepsilon^{-1}\nabla_{\mathrm{s}}\mathbf{w}^i_\varepsilon,\ \nabla_{\mathrm{s}}\mathbf{w}^i_\varepsilon\right)(\det\mathbf{A}_\varepsilon)^{1/2}\,\mathrm{d}S,$$

Lemma 2 and the semicontinuity of norms with respect to weak convergence gives

$$\liminf_{\varepsilon\to 0}\frac{1}{2}|\mathbf{w}_\varepsilon|^2_{W^{1,2}_\varepsilon(M)} \geq \frac{1}{2}\int_M |\nabla_{\mathrm{s}}\mathbf{v}|^2 \mathrm{d}S = W(\mathbf{v}),$$

that is (22).

Proof (Proof—Γ-limsup Inequality: Existence of a Recovery Sequence) Given $\mathbf{v} \in W^{1,2}_{\tan}(M,\ \mathbb{S}^2)$, we have to construct a sequence of discrete vector fields $\mathbf{v}_\varepsilon \in \mathrm{T}(\mathcal{T}_\varepsilon;\ \mathbb{S}^2)$ such that $\mathbf{w}_\varepsilon \to \mathbf{v}$ weakly in $L^2(M;\mathbb{R}^3)$ and

$$\limsup_{\varepsilon\to 0} XY_\varepsilon(\mathbf{v}_\varepsilon) \leq W(\mathbf{v}).$$

The construction is as follows. First of all, we can assume that $\mathbf{v}$ is smooth, otherwise we can approximate it with a density argument (see [33] and [9]). Now, we let $\mathbf{v}_\varepsilon$ be the discrete vector field given by the restriction of $\mathbf{v}$ to the nodes of the triangulation, namely $\mathbf{v}_\varepsilon(i) := \mathbf{v}(i)$ for $i \in \mathcal{T}_0^\varepsilon$. Then, constructing $\mathbf{w}_\varepsilon$ as in (18), it is not difficult to realize that $\mathbf{w}_\varepsilon \to \mathbf{v}$ strongly in $L^2(M; \mathbb{R}^3)$ and that

$$\limsup_{\varepsilon \to 0} XY_\varepsilon(\mathbf{v}_\varepsilon) \leq W(\mathbf{v}),$$

hence the thesis follows.

3 Shells of Non-Zero Euler Characteristic: Emergence of Defects

In this last section we are interested in understanding the energetics of defected configurations and, consequently, locate the defects on the surface M. The results we present are taken from [8] to which we refer for all the details and proofs. First of all, we introduce the notion of vorticity and its discrete counterpart which, as it happens for the discrete flat case and for the Ginzburg Landau case, encodes the topological informations of the discrete sequence $\mathbf{v}_\varepsilon$. Moreover, the concentration of the discrete vorticity in the $\varepsilon \to 0$ limit will be the indication of the emergence of defects. We leave the precise introduction of this measure to the paper [8]. However, for the sake of clarity we briefly sketch it here.

We first consider the continuum setting. Given a map $\mathbf{u} \in (W^{1,1} \cap L^\infty)(M;\ \mathbb{R}^3)$, we define the vorticity of $\mathbf{u}$ as the 1-form

$$\jmath(\mathbf{u}) := (\boldsymbol{\gamma},\ \mathbf{u} \wedge \mathrm{d}\mathbf{u}),$$

whose action on a smooth, *tangent* field $\mathbf{w}$ on M is given by

$$\langle \jmath(\mathbf{u}),\ \mathbf{w} \rangle = (\boldsymbol{\gamma},\ \mathbf{u} \times \nabla_{\mathbf{w}} \mathbf{u}).$$

The role of the vorticity (actually of its differential) is expressed in the following lemma (see [8, Lemma 6]).

Lemma 5 *Let $\mathbf{u} \in W^{1,1}_{\tan}(M;\ \mathbb{S}^2)$ be a unit, tangent field. Suppose that there exist a finite number of points $x_1, \ldots, x_p$ such that*

$$\mathbf{u} \in W^{1,2}_{\mathrm{loc}}(M \setminus \{x_1, \ldots, x_p\};\ \mathbb{R}^3).$$

Then

$$\star \mathrm{d}\jmath(\mathbf{u}) = 2\pi \sum_{i=1}^{p} \mathrm{ind}(\mathbf{u},\ x_i)\delta_{x_i} - G \qquad \textit{in } \mathscr{D}'(M).$$

In the lemma, $\star$ is the Hodge dual operator and $\mathrm{ind}(\mathbf{u},\ x_i)$ the local degree of $\mathbf{u}$ at the point x_i, that is, the winding number of $\mathbf{u}$ around the boundary of a small disk centred at x_i (see e.g. [9] for more details).

Now, given a discrete field $\mathbf{v}_\varepsilon \in \mathrm{T}(\mathcal{T}_\varepsilon;\ \mathbb{S}^2)$, we define the *discrete vorticity measure* $\widehat{\mu}_\varepsilon(\mathbf{v}_\varepsilon)$ as follows. For any given triangle $T \in \mathcal{T}_\varepsilon$ we let $(i_0,\ i_1,\ i_2)$ be its vertices, sorted in counter-clockwise order with respect to the orientation induced by $\boldsymbol{\gamma}$ and we let $i_3 := i_0$. The measure $\widehat{\mu}_\varepsilon(\mathbf{v}_\varepsilon)$ is defined as a linear combination of Dirac delta measures supported on the baricenters of triangles $T \in \mathcal{T}_\varepsilon$, and the weights are given in such a way that

$$\widehat{\mu}_\varepsilon(\mathbf{v}_\varepsilon)[T] = \sum_{k=0}^{2} \left(\frac{\boldsymbol{\gamma}(i_k) + \boldsymbol{\gamma}(i_{k+1})}{2},\ \mathbf{v}_\varepsilon(i_k) \times \mathbf{v}_\varepsilon(i_{k+1}) \right).$$

It turns out that the right-hand side approximates the integral $\int_T \mathrm{d}\jmath(\widehat{\mathbf{v}}_\varepsilon)$, where $\widehat{\mathbf{v}}_\varepsilon \colon \widehat{M}_\varepsilon \to \mathbb{R}^3$ is the affine interpolant of $\mathbf{v}_\varepsilon$, hence $\widehat{\mu}_\varepsilon(\mathbf{v}_\varepsilon)$ is a discretization of $\mathrm{d}\jmath(\widehat{\mathbf{v}}_\varepsilon)$. In the limit $\varepsilon \to 0$, the appearance of defects is related to the convergence $\widehat{\mu}_\varepsilon(\mathbf{v}_\varepsilon) \to 2\pi\mu - G\,\mathrm{d}S$, where μ is a measure concentrated on a finite number of points $\{x_1, \ldots, x_k\}$ in M. This convergence is to be intended in the sense of the *flat topology*, that is, the dual-norm topology on $W^{1,\infty}(\mathbb{R}^3)'$.

The location of the defects is achieved by the analysis of the so called Renormalized Energy $\mathbb{W}$ introduced by Brezis, Bethuel and Hélein for the Ginzburg Landau equation in [5]. In [8], we obtain the Renormalized Energy as the (first order) Γ-limit of the discrete energy XY_ε as in [2, 3, 32] for the euclidean case.

Following [2], we introduce the following class of vector fields in M: for any k, $\mathcal{V}_k$ is the set of fields $\mathbf{v} \in L^2(M;\ \mathbb{S}^2)$ such that there exist $(x_i)_{i=1}^k \in M^k$, $(d_i)_{i=1}^k \in \{-1,\ 1\}^k$ such that

$$\mathbf{v} \in W^{1,2}_{\mathrm{tan,loc}}\left(M \setminus \bigcup_{i=1}^{k} x_i;\ \mathbb{S}^2 \right), \qquad \star\mathrm{d}\jmath(\mathbf{v}) = 2\pi \sum_{i=1}^{k} d_i \delta_{x_i} - G.$$

Given an even number $\mathscr{K} \in \mathbb{N}$ such that $\mathscr{K} \geq |\chi(M)|$, we define the intrinsic Renormalized Energy as (see [2, Eq. (4.22)]):

$$\mathbb{W}_{\mathrm{intr}}(\mathbf{v}) := \begin{cases} \lim_{\delta\to 0} \left(\frac{1}{2} \int_{M_\delta} |\mathrm{D}\mathbf{v}|^2 \mathrm{d}S - \mathscr{K}\pi |\log\delta| \right) & \text{for } \mathbf{v} \in \mathcal{V}_{\mathscr{K}} \\ -\infty & \text{for } \mathbf{v} \in \mathcal{V}_k,\ k < \mathscr{K}, \\ +\infty & \text{otherwise in } L^2(M;\ \mathbb{R}^3), \end{cases}$$

where, given $\mathbf{v} \in \mathcal{V}_{\mathscr{K}}$ and $\delta > 0$ so small that the balls $B_\delta(x_i)$ are pairwise disjoint, we have set $M_\delta := M \setminus \bigcup_{i=1}^{\mathscr{K}} B_\delta(x_i)$. The definition above is shown to be well posed (see [8]). It is important to note that for $\mathbf{v} \in \mathcal{V}_{\mathscr{K}}$ there holds

$$|\mathrm{d}\boldsymbol{\gamma}[\mathbf{v}]| \leq C \quad \text{a.e. in } M, \tag{24}$$

where the constant C depends only on M. Thus, the following quantity exists in $[-\infty, +\infty]$:

$$\mathbb{W}(\mathbf{v}) := \mathbb{W}_{\mathrm{intr}}(\mathbf{v}) + \frac{1}{2}\int_M |\mathrm{d}\boldsymbol{\gamma}[\mathbf{v}]|^2 \mathrm{d}S.$$

$\mathbb{W}$ will be called the Renormalized Energy. Note that $\mathbb{W}$ contains both an intrinsic and an extrinsic term but, due to (24), the latter is always finite. This shows, as expected, that the concentration of the energy is due to the Dirichlet part of $\mathbb{W}$ in (1).

A source of difficulties that emerges in the analysis of this discrete energy is related to the fact that for a curved shell the vertices of the triangulation do not necessarily sit on a structured lattice. In particular, this problem reflects on the study of the so called core energy, namely the energy concentrated in each defect, for which the typical scaling arguments used in the planar case (see [5] and [2]) are not available. As already anticipated, as a result of our analysis we will obtain a core energy that depends of the singularity and moreover it will depend on the (limit) triangulation around each defect x_i. To obtain such a result, we have to enforce our assumptions on the triangulation $\mathcal{T}_\varepsilon$ around the singularities in the limit $\varepsilon \to 0$. At base, we require that our triangulation $\mathcal{T}_\varepsilon$ is somehow scale invariant. We express this requirement as follows.

(H$_4$) For any $x \in M$ there exists a triangulation $\mathcal{S} = \mathcal{S}(x)$ on $\mathbb{R}^2$ such that, for any $\delta > 0$ smaller than the injectivity radius of M, there holds

$$\lim_{\varepsilon \searrow 0} d(\mathcal{S}_\varepsilon, \mathcal{S}_{|B_{\delta/\varepsilon}}) \, |\log \varepsilon| = 0,$$

where $d(\cdot, \cdot)$ is a properly defined distance between triangulations (see [8] for the details) and $\mathcal{S}_{|B_{\delta/\varepsilon}}$ denotes the restriction of $\mathcal{S}$ to the ball $B_{\delta/\varepsilon}$.

In [8, Theorem B] the following theorem is proved

Theorem 3.1 *Suppose that the assumptions* (H1)*,* (H2)*,* (H3) *and* (H4) *are satisfied. Then the following Γ-convergence result holds.*

(i) Compactness. *Let $\mathscr{K} \in \mathbb{N}$ and let $\mathbf{v}_\varepsilon$ be a sequence in* $\mathrm{T}(\mathcal{T}_\varepsilon; \mathbb{S}^2)$ *for which there exists a positive constant $C_{\mathscr{K}}$ such that*

$$XY_\varepsilon(\mathbf{v}_\varepsilon) - \mathscr{K}\pi |\log \varepsilon| \le C_{\mathscr{K}}. \tag{25}$$

Then, up to a subsequence, there holds

$$\hat{\mu}_\varepsilon(\mathbf{v}_\varepsilon) \overset{\text{flat}}{\longrightarrow} 2\pi\mu - G\mathrm{d}S \tag{26}$$

for some $\mu = \sum_{i=1}^k d_i \delta_{x_i}$ with $\sum_{i=1}^k |d_i| \le \mathscr{K}$. If $|\mu| = \mathscr{K}$, then $k = \mathscr{K} \equiv \chi(M) \mod 2$, $|d_i| = 1$ for any i. Moreover, there exists $\mathbf{v} \in \mathcal{V}_{\mathscr{K}}$ and a *subsequence such that*

$$\mathbf{w}_\varepsilon \to \mathbf{v} \textit{ strongly in } L^2(M;\mathbb{R}^3) \textit{ and weakly in } W^{1,2}_{\mathrm{loc}}(M\setminus \bigcup_{i=1}^{\mathscr{K}} x_i;\mathbb{R}^3), \tag{27}$$

where $\mathbf{w}_\varepsilon$ *is the interpolant of* $\mathbf{v}_\varepsilon$ *defined by* (18).

(ii) Γ-lim inf inequality. *Let* $\mathbf{v}_\varepsilon \in \mathrm{T}(\mathcal{T}_\varepsilon;\,\mathbb{S}^2)$ *be a sequence satisfying* (25) *with* $\mathscr{K} \equiv \chi(M) \mod 2$ *and converging to some* $\mathbf{v}\in\mathcal{V}_{\mathscr{K}}$ *as in* (26)–(27). *Then, there holds*

$$\liminf_{\varepsilon\to 0}\,(XY_\varepsilon(\mathbf{v}_\varepsilon) - \mathscr{K}\pi|\log\varepsilon|) \geq \mathbb{W}(\mathbf{v}) + \sum_{i=1}^{\mathscr{K}} \gamma(x_i),$$

where $\gamma(x_i)$ *is the core energy around each defect* x_i.

(iii) Γ-lim sup inequality. *Given* $\mathbf{v}\in\mathcal{V}_{\mathscr{K}}$, *there exists* $\mathbf{v}_\varepsilon \in \mathrm{T}(\mathcal{T}_\varepsilon;\,\mathbb{S}^2)$ *such that* $\hat{\mu}_\varepsilon(\mathbf{v}_\varepsilon) \xrightarrow{\text{flat}} \star\mathrm{d}j(\mathbf{v})$, $\mathbf{w}_\varepsilon \to \mathbf{v}$ *as in* (27) *and*

$$\lim_{\varepsilon\to 0}\,(XY_\varepsilon(\mathbf{v}_\varepsilon) - \mathscr{K}\pi|\log\varepsilon|) = \mathbb{W}(\mathbf{v}) + \sum_{i=1}^{\mathscr{K}} \gamma(x_i).$$

Acknowledgements G.C.'s research has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013) / ERC grant agreement n° 291053. A.S. gratefully acknowledges the financial support of the FP7-IDEAS-ERC-StG #256872 (EntroPhase). G.C. and A.S. acknowledge the partial support of the GNAMPA (Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni) group of INdAM (Istituto Nazionale di Alta Matematica).

References

1. Alicandro, R., Cicalese, M.: Variational analysis of the asymptotics of the XY model. Arch. Ration. Mech. Anal. **192**(3), 501–536 (2009)
2. Alicandro, R., De Luca, L., Garroni, A., Ponsiglione, M.: Metastability and dynamics of discrete topological singularities in two dimensions: a Γ-convergence approach. Arch. Ration. Mech. Anal. **214**(1), 269–330 (2014)
3. Alicandro, R., Ponsiglione, M.: Ginzburg-Landau functionals and renormalized energy: a revised Γ-convergence approach. J. Funct. Anal. **266**(8), 4890–4907 (2014)
4. Berezinskii, V.L.: Destruction of long-range order in one-dimensional and two-dimensional systems possessing a continuous symmetry group. i. Classical systems. J. Exp. Theor. Phys. **61**(3), 1144 (1972)
5. Bethuel, F., Brezis, H., Hélein, F.: Ginzburg-Landau vortices. Progress in Nonlinear Differential Equations and their Applications, vol. 13. Birkhäuser Boston, Inc., Boston, MA (1994)
6. Bowick, M.J., Giomi, L.: Two-dimensional matter: order, curvature and defects. Adv. Phys. **58**(5), 449–563 (2009)

7. Braides, A., Cicalese, M., Solombrino, F.: Q-Tensor continuum energies as limits of head-to-tail symmetric spin systems. SIAM J. Math. Anal. **47**(4), 2832–2867 (2015)
8. Canevari, G., Segatti, A.: Defects in nematic shells: a Γ-convergence discrete to continuum approach. Arch. Ration. Mech. Anal. (2018, to appear)
9. Canevari, G., Segatti, A., Veneroni, M.: Morse's index formula in VMO for compact manifolds with boundary. J. Funct. Anal. **269**(10), 3043–3082 (2015)
10. Chen, Y.M.: The weak solutions to the evolution problems of harmonic maps. Math. Z. **201**(1), 69–74 (1989)
11. Chen, Y.M., Struwe, M.: Existence and partial regularity results for the heat flow for harmonic maps. Math. Z. **201**(1), 83–103 (1989)
12. Dal Maso, G.: An introduction to Γ-convergence. Progress in Nonlinear Differential Equations and their Applications, vol. 8. Birkhäuser Boston, Inc., Boston, MA (1993)
13. do Carmo, M.P.: Riemannian geometry. Mathematics: Theory & Applications. Birkhäuser Boston Inc., Boston, MA (1992). Translated from the second Portuguese edition by Francis Flaherty.
14. Golovaty, D., Montero, A., Sternberg, P.: Dimension reduction for the landau-de gennes model on curved nematic thin films. arXiv, arXiv:1611.03011v1 (2016)
15. Hardt, R., Kinderlehrer, D., Lin, F.-H.: Existence and partial regularity of static liquid crystal configurations. Comm. Math. Phys. **105**(4), 547–570 (1986)
16. Hildebrandt, K., Polthier, K., Wardetzky, M.: On the convergence of metric and geometric properties of polyhedral surfaces. Geom. Dedicata **123**, 89–112 (2006)
17. Ignat, R., Jerrard, R.: Interaction energy between vortices of vector fields on Riemannian surfaces. ArXiv: 1701.06546 (2017)
18. Ignat, R., Jerrard, R.: Renormalized energy between vortices in some Ginzburg-Landau models on Riemannian surfaces. Preprint (2017)
19. Ignat, R., Nguyen, L., Slastikov, V., Zarnescu, A.: Stability of the melting hedgehog in the Landau–de Gennes theory of nematic liquid crystals. Arch. Ration. Mech. Anal. **215**(2), 633–673 (2015)
20. Jerrard, R.L.: Lower bounds for generalized Ginzburg-Landau functionals. SIAM J. Math. Anal. **30**(4), 721–746 (1999)
21. Jerrard, R.L., Soner, H.M.: The Jacobian and the Ginzburg-Landau energy. Calc. Var. Partial Differ. Equ. **14**(2), 151–191 (2002)
22. Kosterlitz, J.M., Thouless, D.J.: Ordering, metastability and phase transitions in two-dimensional systems. J. Phys. C Solid State Phys. **6**(7), 1181 (1973)
23. Kralj, S., Rosso, R., Virga, E.G.: Curvature control of valence on nematic shells. Soft Matter **7**, 670–683 (2011)
24. Le Dret, H., Raoult, A.: The membrane shell model in nonlinear elasticity: a variational asymptotic derivation. J. Nonlinear Sci. **6**(1), 59–84 (1996)
25. Lin, F., Wang, C.: The Analysis of Harmonic Maps and Their Heat Flows. World Scientific Publishing, Hackensack, NJ (2008)
26. Lubensky, T.C., Prost, J.: Orientational order and vesicle shape. J. Phys. II France **2**(3), 371–382 (1992)
27. Napoli, G., Vergori, L.: Extrinsic curvature effects on nematic shells. Phys. Rev. Lett. **108**(20), 207803 (2012)
28. Napoli, G., Vergori, L.: Surface free energies for nematic shells. Phys. Rev. E **85**(6), 061701 (2012)
29. Nelson, D.R.: Toward a tetravalent chemistry of colloids. Nano Lett. **2**(10), 1125–1129 (2002)
30. Rosso, R., Virga, E.G., Kralj, S.: Parallel transport and defects on nematic shells. Continum Mech. Thermodyn. **24**(4–6), 643–664 (2012)
31. Sandier, É.: Lower bounds for the energy of unit vector fields and applications. J. Funct. Anal. **152**(2), 379–403 (1998); See Erratum, ibidem **171**(1), 233 (2000)
32. Sandier, É., Serfaty, S.: Vortices in the magnetic Ginzburg-Landau model. Progress in Nonlinear Differential Equations and their Applications, vol. 70. Birkhäuser Boston, Inc., Boston, MA (2007)

33. Schoen, R., Uhlenbeck, K.: A regularity theory for harmonic maps. J. Differ. Geom. **17**(2), 307–335 (1982)
34. Segatti, A.: Variational models for nematic shells. Lecture Notes for a PhD course at Universidad Autonoma, Madrid (October 2015)
35. Segatti, A., Snarski, M., Veneroni, M.: Equilibrium configurations of nematic liquid crystals on a torus. Phys. Rev. E **90**(1), 012501 (2014)
36. Segatti, A., Snarski, M., Veneroni, M.: Analysis of a variational model for nematic shells. Math. Models Methods Appl. Sci. **26**(10), 1865–1918 (2016)
37. Selinger, R.L., Konya, A., Travesset, A., Selinger, J.V.: Monte Carlo studies of the XY model on two-dimensional curved surfaces. J. Phys. Chem B **48**, 12989–13993 (2011)
38. Shkoller, S.: Well-posedness and global attractors for liquid crystals on Riemannian manifolds. Comm. Partial Differ. Equ. **27**(5–6), 1103–1137 (2002)
39. Straley, J.P.: Liquid crystals in two dimensions. Phys. Rev. A **4**(2), 675–681 (1971)
40. Virga, E.G.: Variational theories for liquid crystals. Applied Mathematics and Mathematical Computation, vol. 8. Chapman & Hall, London, 1994.
41. Vitelli, V., Nelson, D.: Nematic textures in spherical shells. Phys. Rev. E **74**(2), 021711 (2006)
42. Vitelli, V., Nelson, D.R.: Defect generation and deconfinement on corrugated topographies. Phys. Rev. E **70**, 051105 (2004)
43. Wang, X., Miller, D.S., Bukusoglu, E., de Pablo, J.J., Abbott, N.L.: Topological defects in liquid crystals as templates for molecular self-assembly. Nat. Mater. **15**(1), 106–112 (2016)

Modeling of Microstructures in a Cosserat Continuum Using Relaxed Energies

Muhammad Sabeel Khan and Klaus Hackl

Abstract Granular materials tend to exhibit distinct patterns under deformation consisting of layers of counter-rotating particles. In this article, we are going to model this phenomenon on a continuum level by employing the calculus of variations, specifically the concept of energy relaxation. In the framework of Cosserat continuum theory the free energy of the material is enriched with an interaction energy potential taking into account the counter rotations of the particles. The total energy thus becomes non-quasiconvex, giving rise to the development of microstructures. Relaxation theory is then applied to compute its exact quasiconvex envelope. It is worth mentioning that there are no further assumptions necessary here. The computed relaxed energy yields all possible displacement and micro-rotation field fluctuations as minimizers. Based on a two-field variational principle the constitutive response of the material is derived. Results from numerical computations demonstrating the properties of relaxed potential are shown.

1 Introduction

This paper focuses on the treatment of a non-quasiconvex, and therefore ill-posed variational model for granular materials that arises as a consequence of the particle counter rotations at the microscale. In continuum mechanics non-quasiconvex potentials may arise due to various reasons, e.g., in the case of strain-softening plasticity [34, 40] they can be caused by non-monotone constitutive behavior, in the case of single slip plasticity they can be due to single slip constraints on the deformation of crystal in association with cross-hardening [23, 24], for twinning induced plasticity they stem from multi-phase energy potentials corresponding to different martensitic variants [8, 14, 32, 35, 36].

M. Sabeel Khan · K. Hackl (✉)
Lehrstuhl für Mechanik-Materialtheorie, Ruhr-Universität, Bochum, Germany
e-mail: klaus.hackl@rub.de

E. Rocca et al. (eds.), *Trends in Applications of Mathematics to Mechanics*,
Springer INdAM Series 27, https://doi.org/10.1007/978-3-319-75940-1_6

So far, different approaches have been discussed in the literature to treat non-quasiconvex variational problems. One possibility is to use regularization techniques which are based on a gradient-type enhancement of the original non-quasiconvex energy function in (5). But the regularization method has its own limitations as far as the physical properties of the unrelaxed problems are concerned.

Contrary to this is the method of relaxation is a more effective and natural way to deal with non-quasiconvex energies. There are two ways to relax the original non-quasiconvex energy minimization problem (5). Either to enlarge the space of admissible deformations $\left(W^{1,p\in(1,\infty)}(\Omega,\mathbb{R}^n)\right)$ to the space of parametrized measures [8, 47, 64], or, to replace the original non-quasiconvex energy with its relaxed energy envelope. The methodology of constructing a relaxed minimization problem by using parametrized measures is discussed by Carstensen and Roubíček [15, 16], Nicolaides and Walkington [42, 43], Pedregal [47–49] and Roubíček [51–53]. The references which suggests to replace the non-quasiconvex energy with its corresponding relaxed energy function are found in Carstensen et al. [13], Conti and Ortiz [23], Conti and Theil [24], Hackl and Heinen [35], Govindjee et al. [32], Miehe and Gürses [34]. Numerical schemes for calculating relaxed envelopes have been worked out by Aranda and Pedregal [4], Bartels [10], Carstensen, Conti and Orlando [12], Carstensen and Plechac [14], Carstensen and Roubíček [15], Chipot [19], Chipot and Collins [20], Collins, Kinderlehrer and Luskin [21], Dolzmann and Walkington [29], Pedregal [49] and Roubíček [53]. For a detailed discussion on the methods of relaxation the reader is referred to the work by Dacorogna [25], Ball [7] and references therein.

Exact analytical results for the relaxed energy are known only for few variational problems in the literature so far. For example the work of DeSimone and Dolzmann [28] where they give an exact envelope of the relaxed energy potential for the free energy of the nematic elastomers undergoing a transition from isotropic to nematic-phase. Dret and Raoult [30] compute an exact quasiconvex envelope for the Saint Venant-Kirchhoff stored energy function expressed in terms of singular values. Some analytical examples of quasiconvex envelopes are also mentioned by Raoult in [50] for different models in nonlinear elasticity. Kohn and Strang [37, 38] gave an exact formula (see Theorem 1.1 in [37]) for the relaxed energy for a variational problem which has its emergence from the shape optimization problems for electrical conduction. Another exact relaxed result is given by Conti and Theil in [24] for the incremental variational problem for rate-independent single slip elastoplasticity. Conti and Ortiz [23] determine an exact analytical expression for the relaxed energy in single crystal plasticity with a non-convex constraint on the deformation of the crystal requiring all material points must deform via single slip. They extended their analytical expression in [22] to the case of crystal plasticity with arbitrary hardening features. Kohn and Vogelius studied the inverse problem of applied potential tomography and come up with an analytical formula [39] for the relaxed energy by using results from homogenization. In a similar manner but this time with the use of Fourier analysis Kohn presents in Theorem 3.1 of [36] an exact analytical expression for a two well energy function with application to solid-solid phase transitions.

In this paper, we provide an exact relaxation for the non-quasiconvex energy which arises during our study on the rotational microstructures in granular materials. Due to a large number of industrial applications and their use in everyday life granular materials have been studied extensively throughout the past years. Numerous investigations have been performed in order to model the mechanical behavior of these materials [2, 3, 18, 31, 44–46, 54–57, 59, 60]. In this work, the focus is to consider the counter-rotations of granular particles at the microscale and to develop a mechanical model that can predict the formation of distinct deformation patterns that are related to the microstructures in these materials. For an overview on the experimental observations of such patterns the reader is referred to the book by Aranson and Tsimring [5]. For this purpose the continuum description of granular materials is used, specifically the theory of Cosserat continuum.

The present work is organized as follows. In Sect. 2 the intergranular kinematics is discussed and an interaction energy potential contributing to the strain energy of the material is proposed. In Sect. 3 a relaxed variational model for granular materials is presented where we state and prove a theorem on the explicit computation of the relaxed envelope. Employing this result, the exact relaxed energy is derived where all the material regimes are explicitly characterized. In Sect. 4 numerical results demonstrating on the properties of computed relaxed potential are presented. Finally in Sect. 5 conclusions are drawn.

2 Intergranular Interactions and Counter Rotations

Intergranular interactions and particle counter rotations in a granular medium subjected to deformation are intriguing and experimentally well recognized [44, 54] phenomenon that contribute in the development of material microstructures [9, 55, 58]. Because of intricate nature of particle rotations and complex behavior of granular materials under deformation it is therefore difficult to understand the intergranular cohesive interactions completely. In literature almost no comprehensive study appeals which discuss the intergranular interactions and the arising phenomenon in detail that can truly justify the naturally observed microstructural patterns in deforming granular materials. Although the particle rotations at the microscale has been considered by a number of authors, see e.g. [1, 17, 18, 45, 55, 58], the essence of their counter rotations especially their interactions in observing the formation of distinct deformation patterns is not well understood. It is therefore our aim to reconsider the intergranular kinematics of counter-rotating particles at microscale and to develop an interaction energy potential for a granular medium that arises as a consequence of these particle counter rotations.

Here, we develop an interaction energy potential that takes into account the intergranular kinematics at the continuum scale and define two new material parameters as a suitable measure for the observation of microstructural phases of granular materials. In this spirit, consider the granular material where two neighboring particles are in contact with each other as shown in Fig. 1. These particle interactions

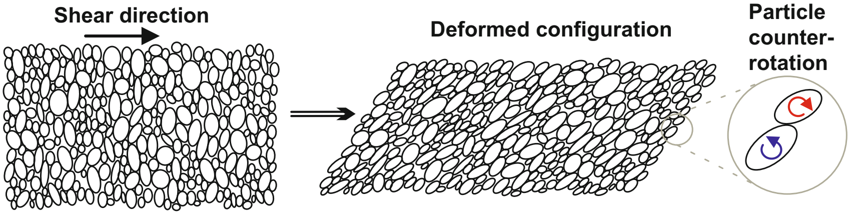

Fig. 1 Schematic of a granular medium subjected to shear with phenomenon of particle counter rotations

leads to two important modes of deformations called translational and micro-rotational motions of the particles which can play a crucial role in the dissipation of the material energy [1, 45] at the continuum scale and therefore contribute to the material strain energy. These independent translational and rotational motions of the granules at the microscale are interlinked with a suitable deformation measure analogous to the concept used in the theory of generalized continuum. Consider now that at the continuum scale the translational motion of the two interacting particles is represented by the vector field $\{u_i\, \mathbf{e}_i\} : \mathbb{R}^d \mapsto \mathbb{R}^d$ and the rotational motion is represented by a field vector analogous to the micro-rotational vector $\{\varphi_i\, \mathbf{e}_i\} : \mathbb{R}^d \mapsto \mathbb{R}^d$ of the Cosserat continuum. Associated with these deformation field vectors are the strain measures. Corresponding to translational and microrotational vector field these measure are the deformation tensor $\left[u_{j,i}\, \mathbf{e}_i \otimes \mathbf{e}_j\right] : \mathbb{R}^d \mapsto \mathbb{R}^{d\times d}$ and $\left[\varphi_{j,i}\, \mathbf{e}_i \otimes \mathbf{e}_j\right] : \mathbb{R}^d \mapsto \mathbb{R}^{d\times d}$ respectively. The symmetric part of $u_{j,i}\, \mathbf{e}_i \otimes \mathbf{e}_j$ is the classical strain tensor $\varepsilon_{ij}\, \mathbf{e}_i \otimes \mathbf{e}_j$. An investigation of the rotating phenomenon of the interacting particles reveals that the macroscopic shear $\left(\varepsilon_{ij} - \frac{1}{d}\varepsilon_{kk}\, \delta_{ij}\right) \mathbf{e}_i \otimes \mathbf{e}_j$ influence the microrotational deformation $\varphi_{j,i}\, \mathbf{e}_i \otimes \mathbf{e}_j$ of the granular particles. This leads us to suggest a proportionality relation between the gradient of the microrotational vector field and the macroscopic shear strain which in mathematical terms is given by

$$\sqrt{\sum_{i,j=1}^{d} \left(\varphi_{j,i}\right)^2} \propto \sqrt{\sum_{i,j=1}^{d} \left(\varepsilon_{ij} - \frac{1}{d}\varepsilon_{kk}\, \delta_{ij}\right)^2}, \tag{1}$$

where d is the dimension of the problem under consideration. This proportionality relation is solved with the introduction of the length scale parameter β with the dimension of the inverse of a length. Thus we can write

$$\sqrt{\sum_{i,j=1}^{d} \left(\varphi_{j,i}\right)^2} = \beta \sqrt{\sum_{i,j=1}^{d} \left(\varepsilon_{ij} - \frac{1}{d}\varepsilon_{kk}\, \delta_{ij}\right)^2}. \tag{2}$$

Equation (2) is indeed the simplest possible assumption taking into account such an intergranular relationship. More complex forms can be envisioned, but we will demonstrate in the sequel that the present one already leads to a very intricate kinetics.

This brief but comprehensive discussion on itergranular kinematics enables us to propose an interaction energy potential that will contribute to the material strain energy function. This interaction energy potential is stated as

$$\mathscr{I} = \alpha \left(\sum_{i,j=1}^{d} \left(\varphi_{j,i}\right)^2 - \beta^2 \sum_{i,j=1}^{d} \left(\varepsilon_{ij} - \frac{1}{d} \varepsilon_{kk} \, \delta_{ij} \right)^2 \right)^2 , \tag{3}$$

where Einstines summation convention is assumed. In tensorial notation it takes the following form

$$\mathscr{I} \; = \; \alpha \, \left(\|\nabla \boldsymbol{\varphi}\|^2 - \beta^2 \, \|\mathrm{sym}\,\mathrm{dev}\,\nabla \mathbf{u}\|^2 \right)^2 , \tag{4}$$

where α and β are non-negative material constants, α is the interaction modulus having information regarding frictional effect in the interacting particles and β is related to the particle size having information regarding intrinsic length scale in Cosserat continuum. The proposed interaction energy potential not only bridges the gap between microstructural properties and the macroscopic behavior of the material but also enables us to characterize different microstructural regimes in granular materials.

3 A Relaxed Variational Model for Granular Materials

3.1 Variational Model

The mechanical response of granular materials can be computed from variational models defined within the context of Cosserat continuum theory. Let Ω be a bounded domain with Lipschitz boundary $\partial\Omega$ and $\mathbf{u} : \Omega \subset \mathbb{R}^d \mapsto \mathbb{R}^d$ be the displacement vector field where d being the dimension of the problem under consideration, $\boldsymbol{\Phi} : \Omega \subset \mathbb{R}^d \mapsto \mathfrak{so}(d) := \left\{ R \in \mathbb{M}^{d\times d} \mid R^T = -R \right\}$ be the microrotations such that the micromotions of the particles are collected in the vector field $\boldsymbol{\varphi} = axl(\boldsymbol{\Phi}) : \Omega \subset \mathbb{R}^d \mapsto \mathbb{R}^d$, then the deformed configuration of these materials can be completely determined from the following minimization problem

$$\inf_{\mathbf{u},\boldsymbol{\Phi},\boldsymbol{\varphi}} \left\{ I\left(\mathbf{u}, \boldsymbol{\Phi}, \boldsymbol{\varphi}\right) \; ; \; \left(\mathbf{u}, \boldsymbol{\Phi}, \boldsymbol{\varphi}\right) \in W^{1,p}\left(\Omega, \mathbb{R}^d\right) \times W^{1,p}\big(\Omega, \mathfrak{so}\,(d)\big) \times W^{1,p}\left(\Omega, \mathbb{R}^d\right) \right\}, \tag{5}$$

along with the prescribed boundary conditions $\mathbf{u}|_{\partial\Omega_u} = \mathbf{u}_\circ$ and $\boldsymbol{\varphi}|_{\partial\Omega_\varphi} = \boldsymbol{\varphi}_\circ$. Here $W^{1,p}$ is the space of admissible deformations (also known as Sobolev space) with

$p \in (1, \infty)$ related to the growth of the energy function W. The integral functional I is defined as

$$I(\mathbf{u}, \boldsymbol{\Phi}, \boldsymbol{\varphi}) = \int_{\Omega} W(\nabla\mathbf{u}, \boldsymbol{\Phi}, \nabla\boldsymbol{\varphi})\ dV - \ell(\mathbf{u}, \boldsymbol{\varphi}), \tag{6}$$

where the potential ℓ takes the contribution of external forces $\mathbf{b}$, external couples $\mathbf{m}$, traction forces $\mathbf{t}_u$ and traction moments $\mathbf{t}_\varphi$ such that

$$\ell(\mathbf{u}, \boldsymbol{\varphi}) = \int_{\Omega} (\mathbf{b}\cdot\mathbf{u} + \mathbf{m}\cdot\boldsymbol{\varphi})\ dV + \int_{\partial\Omega_u} \mathbf{t}_u \cdot \mathbf{u}\, dS + \int_{\partial\Omega_\varphi} \mathbf{t}_\varphi \cdot \boldsymbol{\varphi}\, dS. \tag{7}$$

In reality, the deformation of granular media is a dissipative process which should not be discussed in terms of energies and displacements. In this sense, our model only covers the initiation of material microstructures. For a full description of extended time-intervals, the variables $\mathbf{u}, \boldsymbol{\Phi}, \boldsymbol{\varphi}$ would have to be replaced by their corresponding velocities and the energy W by a dissipation function. An exposition of this procedure in the case of rigid elasticity can be found in [61–63].

Within the framework of generalized elasticity the mechanical response of granular materials can be determined with the specification of an energy potential that depends, in an independent way, on the particle displacement and microrotations. It is therefore possible to replace the energy potential W in the integral functional (6) by the following Cosserat energy function

$$W^{csrt}(\nabla\mathbf{u}, \boldsymbol{\Phi}, \nabla\boldsymbol{\varphi}) = \frac{1}{2}\boldsymbol{e}(\mathbf{u}, \boldsymbol{\varphi}) : \mathbb{C} : \boldsymbol{e}(\mathbf{u}, \boldsymbol{\varphi}) + \frac{1}{2}\boldsymbol{\kappa}(\boldsymbol{\varphi}) : \overline{\mathbb{C}} : \boldsymbol{\kappa}(\boldsymbol{\varphi}), \tag{8}$$

which do not only depends on the gradients of the macro and micro-motions of the particles but also on a relative macro-rotational deformation tensor $\boldsymbol{\Phi}$ that associates the macro-deformation with the micro-deformation of the particles. Here, $\boldsymbol{e} = \nabla\mathbf{u} - \boldsymbol{\Phi}$ is the Cosserat deformation strain tensor, $\boldsymbol{\kappa} = \nabla\boldsymbol{\varphi}$ is the rotational deformation strain tensor, $\mathbb{C}$ and $\overline{\mathbb{C}}$ are the fourth order constitutive tensors of elastic constants.

The earlier discussion in Sect. 2 on the intergranular interactions and counter rotations of the particles leads us to introduce an enhanced energy potential for the granular materials. In this spirit, the interaction energy potential (4) is integrated with the Cosserat energy function (8) to model the microstructures of the granular materials. This enables us to define a new enhanced energy potential for the granular materials in a Cosserat medium which is given by

$$W(\nabla\mathbf{u}, \boldsymbol{\Phi}, \nabla\boldsymbol{\varphi}) = \underbrace{W^{csrt}(\nabla\mathbf{u}, \boldsymbol{\Phi}, \nabla\boldsymbol{\varphi})}_{\text{Cosserat energy function}} + \underbrace{\alpha\left(\|\nabla\boldsymbol{\varphi}\|^2 - \beta^2\,\|\mathrm{dev\,sym}\,\nabla\mathbf{u}\|^2\right)^2}_{\text{Interaction energy potential}}. \tag{9}$$

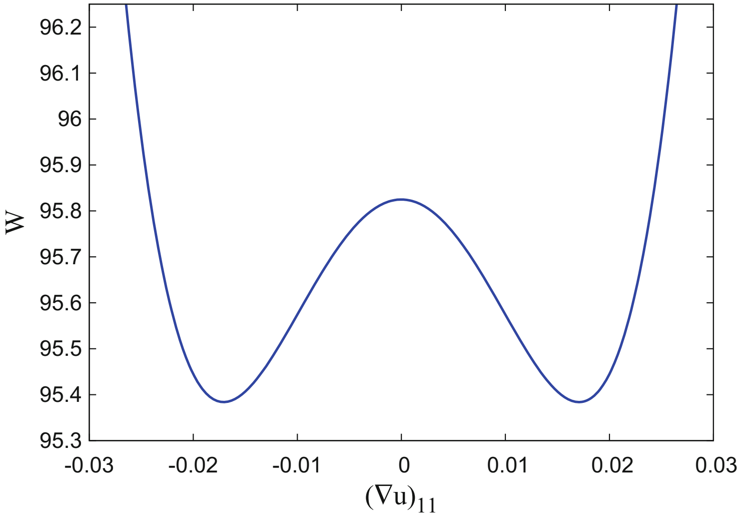

Fig. 2 Unrelaxed energy (10) curve for E $= 2.0 \times 10^2$ (MPa), $\nu = 0.3$, $\mu_c = 1.0 \times 10^{-2}$ (MPa), $\overline{\lambda} = 1.15 \times 10^2$ (N), $\overline{\mu} = 7.69 \times 10^1$ (N), $\overline{\mu}_c = 1.00 \times 10^1$ (N), $\alpha = 1.0 \times 10^1$ (N.mm^2) and $\beta = 1.20 \times 10^2$ (mm^{-1})

In an isotropic elastic Cosserat medium the enhanced energy potential (9) takes the form

$$W\left(\nabla \mathbf{u}, \boldsymbol{\Phi}, \nabla \boldsymbol{\varphi}\right) = \left(\frac{\lambda}{2} + \frac{\mu}{d}\right)\left(\operatorname{tr} \boldsymbol{\varepsilon}\right)^2 + \mu \left\|\operatorname{dev} \boldsymbol{\varepsilon}\right\|^2 + \mu_c \left\|\operatorname{asy} \nabla \mathbf{u} - \boldsymbol{\Phi}\right\|^2 + \frac{\overline{\lambda}}{2}\left(\operatorname{tr} \boldsymbol{\kappa}\right)^2 + \overline{\mu}\left\|\operatorname{sym} \boldsymbol{\kappa}\right\|^2 + \overline{\mu}_c \left\|\operatorname{asy} \boldsymbol{\kappa}\right\|^2 + \alpha\left(\left\|\operatorname{sym} \boldsymbol{\kappa}\right\|^2 + \left\|\operatorname{asy} \boldsymbol{\kappa}\right\|^2 - \beta^2 \left\|\operatorname{dev} \boldsymbol{\varepsilon}\right\|^2\right)^2 \tag{10}$$

Here, λ, μ, μ_c, $\bar{\lambda}$, $\bar{\mu}$, $\bar{\mu}_c$ are the Cosserat material constants.

The nonconvexity and hence the non-quasiconvexity of the energy potential (10) along some chosen strain paths can be seen from Fig. 2. Such non-quasiconvex energy potential when enters in (6) will lead to work with non-quasiconvex energy minimization problem whose general analytical solutions are always of interest. But, the solutions to such non-quasiconvex energy minimization problems do not exist in general, which is highly due to fine scale oscillations of the gradients of infimizing deformations. Here, in this case, the non-existence of these solutions is due to the possible displacement and microrotation field fluctuations at fine scales. The fine scale oscillations of the minimizing displacement and microrotation field variables will lead to the development of internal structures in the material. Formation of such microstructures can be extended microstructures [6, 33] which is distributed through the material domain or the localized microstructures [11, 27] which appear in the form of narrow shearing bands. Moreover, the existence of the unique minimizing translational and microrotational deformations are not guaranteed in this situation.

Thus to avoid these problems and to resolve the internal structures of the materials in consideration it is therefore necessary to compute a quasiconvex (relaxed) energy potential W^{rel}. The relaxed potential when enters in the minimization problem (5) now assures the ellipticity of the resulting boundary value problem, since it satisfy the Legendre-Hadamard condition (see definition by Ball and Dacorogna [7, 25]). The study by Morrey [41], Dacorogna [25, 26] gives sufficient justification for the relation of Legendre-Hadamard (ellipticity) condition with the constitutive description of a related mechanical problem.

If possible to compute the exact relaxed envelope of the corresponding non-quasiconvex energy in the energy minimization problem (5) one do not only guarantee general solutions of the associated energy minimization problem but also can predict on the formation of both the extended and localized microstructures in the materials. It is worth mentioning that, in this case, we are enable to compute an exact relaxed (quasi-convex) energy envelope corresponding to the non-quasiconvex energy potential in (10).

Since quasiconvex envelopes possess only degenerate ellipticity, only existence of minimizers can be guaranteed, no uniqueness. For numerical purposes it is therefore advantageous to add a very small strongly elliptic regularization term. This does not alter the character of the calculated solutions.

3.2 *Computation of Relaxed Energy Envelope*

In this section, we present our main result concerning the solutions of non-quasiconvex energy minimization problem in (5). In this respect, we compute an exact quasiconvex envelope of the energy function in (10). For other cases where it was possible to construct exact relaxed envelopes corresponding to energy minimization problems addressing different mechanical aspects the reader is referred to the work by Conti and Theil [24], Conti and Ortiz [23], Conti et al. [22], DeSimone and Dolzmann [28], Dret and Raoult [30], Kohn [36], Kohn and Strang [37, 38], Kohn and Vogelius [39], Raoult [50]. The quasiconvex envelope which here termed as the relaxed energy W^{rel} is thus stated as

Theorem 1 *Assume $d = 3$, λ, μ, μ_c, $\bar{\lambda}$, $\bar{\mu}$, $\bar{\mu}_c$, α, $\beta \geq 0$, $\mu_\circ = \min\{\bar{\mu}, \bar{\mu}_c\}$. Let*

$$f = \mu_\circ\, s + \mu\, c + \alpha \left(s - \beta^2\, c\right)^2, \qquad h = \begin{cases} (\bar{\mu} - \bar{\mu}_c)\ \|\mathrm{sym}\ \boldsymbol{\kappa}\|^2 & \text{if } \bar{\mu} \geq \bar{\mu}_c \\ (\bar{\mu}_c - \bar{\mu})\ \|\mathrm{asy}\ \boldsymbol{\kappa}\|^2 & \text{otherwise} \end{cases}$$

and define g by

$$g = \min_{\substack{s,c;\ c \geq \|\mathrm{dev}\, \boldsymbol{\varepsilon}\|^2, \\ s \geq \left(\|\mathrm{sym}\ \boldsymbol{\kappa}\|^2 + \|\mathrm{asy}\ \boldsymbol{\kappa}\|^2\right)}} f\,(s, c)\,. \tag{11}$$

Then, the quasicovnex envelope of the Cosserat strain energy defined in (10) is given by

$$W^{rel} = \left(\frac{\lambda}{2} + \frac{\mu}{d}\right)\left(\mathrm{tr}\ \boldsymbol{\varepsilon}\right)^2 \ + \ \mu_c \left\|\mathrm{asy}\ \nabla\mathbf{u} - \boldsymbol{\Phi}\right\|^2 \ + \ \frac{\overline{\lambda}}{2}\left(\mathrm{tr}\ \boldsymbol{\kappa}\right)^2 \ + \ h \\ + \ g\left(\|\mathrm{sym}\ \boldsymbol{\kappa}\|^2, \|\mathrm{asy}\ \boldsymbol{\kappa}\|^2, \|\mathrm{dev}\,\boldsymbol{\varepsilon}\|^2\right). \tag{12}$$

Proof Consider the rank-one line $\boldsymbol{\kappa}_t = \boldsymbol{\kappa} + t\,\mathbf{a}\otimes\mathbf{b}; \quad \mathbf{a},\mathbf{b}\in\mathbb{R}^d, \quad t\in\mathbb{R}$, then

$$W\left(\boldsymbol{e},\boldsymbol{\kappa}_t\right) = \left(\frac{\lambda}{2} + \frac{\mu}{d}\right)\left(\mathrm{tr}\ \boldsymbol{\varepsilon}\right)^2 \ + \ \mu\,\|\mathrm{dev}\,\boldsymbol{\varepsilon}\|^2 \ + \ \mu_c\,\|\mathrm{asy}\,\nabla\mathbf{u} - \boldsymbol{\Phi}\|^2 \ + \ \frac{\overline{\lambda}}{2}\left(\mathrm{tr}\ \boldsymbol{\kappa}\right)^2 \\ + \overline{\mu}\,\|\mathrm{sym}\ \boldsymbol{\kappa}_t\|^2 \ + \ \overline{\mu}_c\,\|\mathrm{asy}\ \boldsymbol{\kappa}_t\|^2 \ + \ \alpha\left(\|\mathrm{sym}\ \boldsymbol{\kappa}_t\|^2 \ + \ \|\mathrm{asy}\ \boldsymbol{\kappa}_t\|^2 \ - \ \beta^2\,\|\mathrm{dev}\,\boldsymbol{\varepsilon}\|^2\right)^2 \tag{13}$$

Now, for any $s \ge \|\boldsymbol{\kappa}\|^2$ we can select $t_- < t \le 0$ such that $\|\boldsymbol{\kappa}_t\|^2 = s$. A lamination in this direction gives

$$W^{rc} \le \left(\frac{\lambda}{2} + \frac{\mu}{d}\right)\left(\mathrm{tr}\ \boldsymbol{\varepsilon}\right)^2 \ + \ \mu_c\,\|\mathrm{asy}\,\nabla\mathbf{u} - \boldsymbol{\Phi}\|^2 \ + \ \frac{\overline{\lambda}}{2}\left(\mathrm{tr}\ \boldsymbol{\kappa}\right)^2 \ + \ h \\ + \min_{s\ge\|\mathrm{sym}\ \boldsymbol{\kappa}\|^2+\|\mathrm{asy}\ \boldsymbol{\kappa}\|^2}\left\{\mu_\circ\, s \ + \ \mu\,\|\mathrm{dev}\,\boldsymbol{\varepsilon}\|^2 \ + \ \alpha\left(s \ - \ \beta^2\,\|\mathrm{dev}\,\boldsymbol{\varepsilon}\|^2\right)^2\right\}. \tag{14}$$

Here, rc in the superscript stands for rank-one convex envelope. Working along the rank-one line $\boldsymbol{e}_t = \boldsymbol{e} + t\,\mathbf{c}\otimes\mathbf{d}; \quad \mathbf{c},\mathbf{d}\in\mathbb{R}^d$ and following the arguments above, we obtain

$$W^{rc} \le \left(\frac{\lambda}{2} + \frac{\mu}{d}\right)\left(\mathrm{tr}\ \boldsymbol{\varepsilon}\right)^2 \ + \ \mu_c\,\|\mathrm{asy}\,\nabla\mathbf{u} - \boldsymbol{\Phi}\|^2 \ + \ \frac{\overline{\lambda}}{2}\left(\mathrm{tr}\ \boldsymbol{\kappa}\right)^2 \ + \ h \\ + \min_{c\ge\|\mathrm{dev}\,\boldsymbol{\varepsilon}\|^2}\left\{\mu_\circ\left(\|\mathrm{sym}\ \boldsymbol{\kappa}\|^2 \ + \ \|\mathrm{asy}\ \boldsymbol{\kappa}\|^2\right) \ + \ \mu\, c \right. \\ \left. + \,\alpha\left(\|\mathrm{sym}\ \boldsymbol{\kappa}\|^2 \ + \ \|\mathrm{asy}\ \boldsymbol{\kappa}\|^2 \ - \ \beta^2 c\right)^2\right\}. \tag{15}$$

Hence the upper bound is proved. The lower bound is based on Lemma 1 below and on the fact that, for $h_1 : [0,\infty)^d \mapsto \mathbb{R}^d$ convex and non-decreasing in each variable and $h_2 : \mathbb{R}^{d\times d} \mapsto \mathbb{R}^d$ component-wise convex, the function $h_1 \circ h_2$ is convex. This completes the proof.

Lemma 1 *Let $f : [0, \infty)^2 \mapsto [0, \infty)$ be convex. Then the function g defined by*

$$g(x) = \inf_{s_1 \geq x_1, s_2 \geq x_2} f(s) \tag{16}$$

is convex and non-decreasing in each variable.

Proof Fix $x', x'', \lambda \in (0, 1)$. For any $\epsilon > 0$ there are s', s'' such that $x' \leq s', x'' \leq s''$, and

$$f(s') \leq g(x') + \epsilon, \quad f(s'') \leq g(x'') + \epsilon. \tag{17}$$

Then $\lambda s' + (1 - \lambda)s'' \geq \lambda x' + (1 - \lambda)x''$, and since f is convex we obtain

$$\begin{aligned} g(\lambda x' + (1 - \lambda)x'') \leq f(\lambda s' + (1 - \lambda)s'') &\leq \lambda f(s') + (1 - \lambda) f(s'') \\ &\leq \lambda g(x') + (1 - \lambda)g(x'') + \epsilon. \end{aligned} \tag{18}$$

Therefore g is convex. Monotonicity is clear from the definition.

To compute the exact relaxed envelope in (12) one needs to solve the minimization problem (11). The stationarity conditions to this minimization problem are as follows

3.2.1 Stationarity Conditions

$$(1). \quad \text{for} \quad s = \|\text{sym}\, \boldsymbol{\kappa}\|^2 + \|\text{asy}\, \boldsymbol{\kappa}\|^2 \ \text{and} \quad c \geq \|\text{dev}\, \boldsymbol{\varepsilon}\|^2 \ : \quad \frac{\partial g}{\partial c} = 0, \quad \frac{\partial g}{\partial s} \geq 0, \tag{19a}$$

$$(2). \quad \text{for} \quad s = \|\text{sym}\, \boldsymbol{\kappa}\|^2 + \|\text{asy}\, \boldsymbol{\kappa}\|^2 \ \text{and} \quad c = \|\text{dev}\, \boldsymbol{\varepsilon}\|^2 \ : \quad \frac{\partial g}{\partial c} \geq 0, \quad \frac{\partial g}{\partial s} \geq 0, \tag{19b}$$

$$(3). \quad \text{for} \quad c = \|\text{dev}\, \boldsymbol{\varepsilon}\|^2 \ \text{and} \quad s \geq \|\text{sym}\, \boldsymbol{\kappa}\|^2 + \|\text{asy}\, \boldsymbol{\kappa}\|^2 \ : \quad \frac{\partial g}{\partial s} = 0, \quad \frac{\partial g}{\partial c} \geq 0. \tag{19c}$$

On the basis of these three stationarity conditions the material energy can be characterized into the following three phases

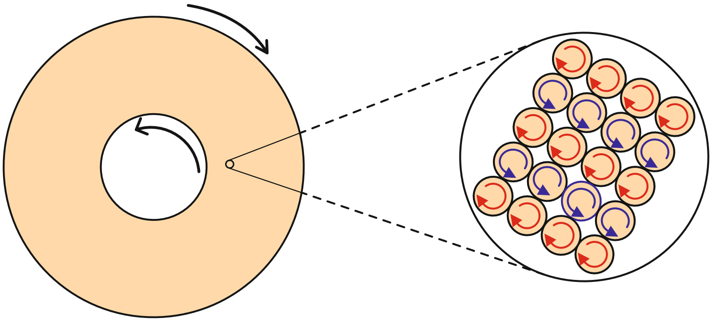

Fig. 3 A Couette shear cell where the two arrows indicates the shearing direction of the inner and outer boundaries of the annular domain. In inset the microstructure patterns due to microrotational motions of the particles is shown

3.2.2 Material Phase with Microstructure in Microrotational Motions (Micromotions) (Phase 1)

This phase is corresponding to the material regime where there are microstructures due to the micromotions (which are in fact the rotational degrees of freedom assembled in the microrotational vector field $\boldsymbol{\varphi}$) of the continuum particles. A schematic representation of such microstructure is given in Fig. 3. The enhanced energy potential (10) is nonconvex in this microstructural phase. It is observed that whenever the norm of the curvature strain tensor is dominating over the norm of the macroscopic shear strain tensor for some specific choice of the material parameters μ, α and β, the material experiences a microstructure in micromotions. This microstructural material phase is characterized by the following inequality relation

$$\|\boldsymbol{\kappa}\|^2 \geq \beta^2 \, \|\operatorname{dev} \boldsymbol{\varepsilon}\|^2 + \frac{\mu}{2\alpha\beta^2}. \tag{20}$$

It is important to note the effect of shear modulus μ, internal length scale (e.g., the diameter of particles) β and the coherency interaction modulus or frictional modulus α in conjunction with the curvature and macroscopic shear strains which plays very crucial role in the observation of this internal structural phase of the material. Using the first stationarity condition (19a) the minimizers of the problem in (11) are obtained as

$$s = \|\operatorname{sym} \boldsymbol{\kappa}\|^2 + \|\operatorname{asy} \boldsymbol{\kappa}\|^2 , \qquad c = \frac{1}{\beta^2}\left(\|\operatorname{sym} \boldsymbol{\kappa}\|^2 + \|\operatorname{asy} \boldsymbol{\kappa}\|^2\right) - \frac{\mu}{2\alpha\beta^4}. \tag{21}$$

Thus, the scalar convex function g is given by

$$g = \begin{cases} \left(\overline{\mu} - \overline{\mu}_c + \mu_\circ + \dfrac{\mu}{\beta^2}\right) \|\text{sym } \boldsymbol{\kappa}\|^2 + \left(\mu_\circ + \dfrac{\mu}{\beta^2}\right) \|\text{asy } \boldsymbol{\kappa}\|^2 - \dfrac{\mu^2}{4\alpha\beta^4} & \text{if} \quad \bar{\mu} \geq \bar{\mu}_c \\[2ex] \left(\mu_\circ + \dfrac{\mu}{\beta^2}\right) \|\text{sym } \boldsymbol{\kappa}\|^2 + \left(\overline{\mu}_c - \overline{\mu} + \mu_\circ + \dfrac{\mu}{\beta^2}\right) \|\text{asy } \boldsymbol{\kappa}\|^2 - \dfrac{\mu^2}{4\alpha\beta^4} & \text{if} \quad \bar{\mu} < \bar{\mu}_c \end{cases} \tag{22}$$

The relaxed energy of the material in this phase is obtained as

$$W_1^{rel} = \begin{cases} \left\{ \begin{array}{l} \left(\dfrac{\lambda}{2} + \dfrac{\mu}{d}\right) \left(\text{tr } \boldsymbol{\varepsilon}\right)^2 + \mu_c \|\text{asy } \nabla \mathbf{u} - \mathscr{E} \cdot \boldsymbol{\varphi}\|^2 - \dfrac{\mu^2}{4\alpha\beta^4} \\[2ex] + \dfrac{\bar{\lambda}}{2} (\text{tr } \boldsymbol{\kappa})^2 + (\bar{\mu} - \bar{\mu}_c) \|\text{sym } \boldsymbol{\kappa}\|^2 + \left(\mu_\circ + \dfrac{\mu}{\beta^2}\right) \|\boldsymbol{\kappa}\|^2 \end{array} \right. & \text{if} \quad \bar{\mu} \geq \bar{\mu}_c, \\[6ex] \left\{ \begin{array}{l} \left(\dfrac{\lambda}{2} + \dfrac{\mu}{d}\right) \left(\text{tr } \boldsymbol{\varepsilon}\right)^2 + \mu_c \|\text{asy } \nabla \mathbf{u} - \mathscr{E} \cdot \boldsymbol{\varphi}\|^2 - \dfrac{\mu^2}{4\alpha\beta^4} \\[2ex] + \dfrac{\bar{\lambda}}{2} (\text{tr } \boldsymbol{\kappa})^2 - (\bar{\mu} - \bar{\mu}_c) \|\text{asy } \boldsymbol{\kappa}\|^2 + \left(\mu_\circ + \dfrac{\mu}{\beta^2}\right) \|\boldsymbol{\kappa}\|^2 \end{array} \right. & \text{if} \quad \bar{\mu} < \bar{\mu}_c \end{cases} \tag{23}$$

3.2.3 Material Phase with No Microstructure (Phase 2)

This phase is connected with the material regime where there is no internal structure in the material. The second stationarity condition (19b) clearly shows that the minimizers of the functional in (11) are itself $\left(\|\text{sym } \boldsymbol{\kappa}\|^2 + \|\text{asy } \boldsymbol{\kappa}\|^2\right)$ and $\|\text{dev}\, \boldsymbol{\varepsilon}\|^2$ respectively. This indicates that the original energy potential in (10) is convex in this material phase. The criteria for the recognition of this material phase is given by the following inequality relation

$$\beta^2 \|\text{dev}\, \boldsymbol{\varepsilon}\|^2 - \frac{\mu_\circ}{2\alpha} \leq \|\boldsymbol{\kappa}\|^2 \leq \beta^2 \|\text{dev}\, \boldsymbol{\varepsilon}\|^2 + \frac{\mu}{2\alpha\beta^2}. \tag{24}$$

The function g in this phase is given by

$$\begin{aligned} g \; = \; & \overline{\mu}\, \|\text{sym } \boldsymbol{\kappa}\|^2 \; + \; \overline{\mu}_c\, \|\text{asy } \boldsymbol{\kappa}\|^2 \; + \; \mu\, \|\text{dev}\, \boldsymbol{\varepsilon}\|^2 \\ & + \; \alpha \left(\|\text{sym } \boldsymbol{\kappa}\|^2 \; + \; \|\text{asy } \boldsymbol{\kappa}\|^2 \; - \; \beta^2\, \|\text{dev}\, \boldsymbol{\varepsilon}\|^2\right)^2. \end{aligned} \tag{25}$$

The relaxed energy potential in this phase is thus the original energy potential (10) itself and we write

$$
\begin{aligned}
W_2^{rel} = \left(\frac{\lambda}{2} + \frac{\mu}{d}\right)\left(\operatorname{tr} \boldsymbol{\varepsilon}\right)^2 + \mu \, \|\operatorname{dev} \boldsymbol{\varepsilon}\|^2 + \mu_c \, \left\|\operatorname{asy} \nabla \mathbf{u} - \mathscr{E} \cdot \boldsymbol{\varphi}\right\|^2 + \frac{\overline{\lambda}}{2}\left(\operatorname{tr} \boldsymbol{\kappa}\right)^2 \\
+ \overline{\mu} \, \|\operatorname{sym} \boldsymbol{\kappa}\|^2 + \overline{\mu}_c \, \|\operatorname{asy} \boldsymbol{\kappa}\|^2 + \alpha \left(\|\operatorname{sym} \boldsymbol{\kappa}\|^2 + \|\operatorname{asy} \boldsymbol{\kappa}\|^2 - \beta^2 \, \|\operatorname{dev} \boldsymbol{\varepsilon}\|^2\right)^2
\end{aligned}
\tag{26}
$$

3.2.4 Material Phase with Microstructure in Translational Motions (Phase 3)

This phase constitutes an unexpected outcome of the theory presented. It consists of laminates formed by alternating displacements as for example formed by phase-transforming materials. It would be interesting to see whether such structures can be observed experimentally.

This phase is related to the material regime where there is a microstructure in translational motions (which are in fact the displacement degrees of freedom of the continuum particles and are assembled in the displacement vector field **u**) of the continuum particles. A schematic representation of such microstructure formation is shown in Fig. 4. The enhanced energy potential (10) thus becomes nonconvex in this phase. Using the third stationarity condition (19c) it is observed that the norm of the macroscopic shear strain tensor is dominating over the norm of the rotational strain tensor. The material is said to be in this phase whenever the following criteria is satisfied

$$
\beta^2 \, \|\operatorname{dev} \boldsymbol{\varepsilon}\|^2 - \frac{\mu_\circ}{2\alpha} \geq \|\boldsymbol{\kappa}\|^2 \, . \tag{27}
$$

It is important to note the effect the coherency modulus α and the Cosserat material modulus $\mu_\circ$ in the characterization of this microstructural phase. The minimizers of the functional in (11) are obtained after solving the third stationarity condition (19c)

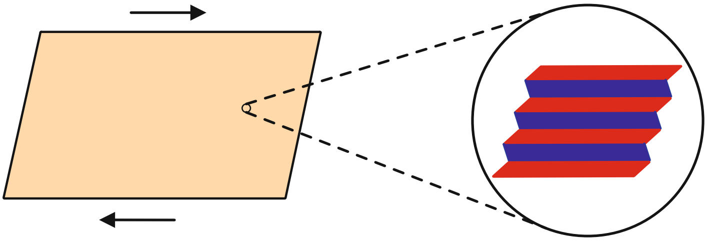

Fig. 4 A rectangular specimen under shear with two arrow head pointing towards the shearing direction. In inset the microstructure patterns formed due to the translational motions of the continuum particles is shown

which are given as

$$c \;=\; \|\text{dev}\,\boldsymbol{\varepsilon}\|^2 \qquad \text{and} \qquad s \;=\; \beta^2\; \|\text{dev}\,\boldsymbol{\varepsilon}\|^2 - \frac{\mu_\circ}{2\alpha}. \tag{28}$$

Thus minimum potential g in (11) takes the following form

$$g \;=\; \begin{cases} \left(\overline{\mu} - \overline{\mu}_c\right) \|\text{sym}\;\boldsymbol{\kappa}\|^2 + \left(\mu_\circ \beta^2 + \mu\right)\; \|\text{dev}\,\boldsymbol{\varepsilon}\|^2 - \dfrac{\mu_\circ^2}{4\alpha} & \text{if} \quad \bar{\mu} \geq \bar{\mu}_c \\[2ex] \left(\overline{\mu}_c - \overline{\mu}\right) \|\text{asy}\;\boldsymbol{\kappa}\|^2 + \left(\mu_\circ \beta^2 + \mu\right)\; \|\text{dev}\,\boldsymbol{\varepsilon}\|^2 - \dfrac{\mu_\circ^2}{4\alpha} & \text{if} \quad \bar{\mu} < \bar{\mu}_c \end{cases} \tag{29}$$

Hence the relaxed energy potential in this phase is obtained as

$$W_3^{rel} \;=\; \begin{cases} \begin{cases} \left(\dfrac{\lambda}{2} + \dfrac{\mu}{d}\right)\left(\text{tr}\;\boldsymbol{\varepsilon}\right)^2 + \mu_c\; \|\text{asy}\;\nabla\mathbf{u} - \mathscr{E}\cdot\boldsymbol{\varphi}\|^2 + \dfrac{\bar{\lambda}}{2}(\text{tr}\;\boldsymbol{\kappa})^2 \\[2ex] +(\bar{\mu} - \bar{\mu}_c)\; \|\text{sym}\;\boldsymbol{\kappa}\|^2 + \left(\mu_\circ {\beta}^2 + \mu\right) \|\text{dev}\,\boldsymbol{\varepsilon}\|^2 - \dfrac{{\mu_\circ}^2}{4\alpha} \end{cases} & \text{if} \quad \bar{\mu} \geq \bar{\mu}_c \\[6ex] \begin{cases} \left(\dfrac{\lambda}{2} + \dfrac{\mu}{d}\right)\left(\text{tr}\;\boldsymbol{\varepsilon}\right)^2 + \mu_c\; \|\text{asy}\;\nabla\mathbf{u} - \mathscr{E}\cdot\boldsymbol{\varphi}\|^2 + \dfrac{\bar{\lambda}}{2}\;(\text{tr}\;\boldsymbol{\kappa})^2 \\[2ex] -\,(\bar{\mu} - \bar{\mu}_c)\; \|\text{asy}\;\boldsymbol{\kappa}\|^2 + \left(\mu_\circ \beta^2 + \mu\right) \|\text{dev}\,\boldsymbol{\varepsilon}\|^2 - \dfrac{{\mu_\circ}^2}{4\alpha} \end{cases} & \text{if} \quad \bar{\mu} < \bar{\mu}_c \end{cases} \tag{30}$$

3.2.5 Relaxed Energy

The total relaxed energy thus comprises all the three energies in each of the phase and it acquires finally the following form

$$W^{rel} \;=\; \begin{cases} W_1^{rel} & \text{if} \quad \|\boldsymbol{\kappa}\|^2 \geq {\beta}^2\; \|\text{dev}\,\boldsymbol{\varepsilon}\|^2 + \dfrac{\mu}{2\alpha\beta^2} \\[3ex] W_2^{rel} & \text{if} \quad -\dfrac{\mu_\circ}{2\alpha} \leq \|\boldsymbol{\kappa}\|^2 - {\beta}^2\; \|\text{dev}\,\boldsymbol{\varepsilon}\|^2 \leq \dfrac{\mu}{2\alpha\beta^2} \\[3ex] W_3^{rel} & \text{if} \quad \|\boldsymbol{\kappa}\|^2 \leq \beta^2\; \|\text{dev}\,\boldsymbol{\varepsilon}\|^2 - \dfrac{\mu_\circ}{2\alpha} \end{cases} \tag{31}$$

where W_1^{rel}, W_2^{rel} and W_3^{rel} are explicitly given as in (23), (26) and (30), respectively. The computation of this analytical expression for the relaxed energy corresponding to non-quasiconvex energy function in (10) thus enable us to predict all microstructural features of the material which are carried safely from the microscopic to macroscopic computational scale. Hence we have extracted all possible information regarding the development of microstructural regimes in the granular materials pertinent to observing its macro-mechanical behavior. For practical applications it is now more efficient and effective to reformulate the original non-quasiconvex problem in (5) to a relaxed energy minimization problem using this relaxed potential.

3.2.6 Nonlinear Constitutive Relations

The proposed granular material model is completed with the formulation of constitutive relations between stress and strain tensors in a Cosserat medium. The constitutive structure of the proposed theory thus comprises of three phases (as discussed in Sect. 3.2) where in each phase the force-stress are explicitly related to the Cosserat strain tensors according to the following formulas:

$$
\boldsymbol{\sigma} = \begin{cases}
2\left(\dfrac{\lambda}{2}+\dfrac{\mu}{d}\right)\left(\operatorname{tr}\boldsymbol{\varepsilon}\right)\mathbf{I}+2\,\mu_c\left(\operatorname{asy}\,\nabla\mathbf{u}-\boldsymbol{\Phi}\right), & \text{(Phase 1)} \\[2ex]
\begin{cases} \lambda\left(\operatorname{tr}\boldsymbol{\varepsilon}\right)\mathbf{I} + 2\,\mu\,\boldsymbol{\varepsilon} + 2\,\mu_c\left(\operatorname{asy}\,\nabla\mathbf{u}-\boldsymbol{\Phi}\right) \\ -\,4\,\alpha\,\beta^2\left(\,\|\boldsymbol{\kappa}\|^2-\beta^2\,\|\operatorname{dev}\boldsymbol{\varepsilon}\|^2\right)\left(\operatorname{dev}\boldsymbol{\varepsilon}\right), \end{cases} & \text{(Phase 2)} \\[4ex]
\lambda\left(\operatorname{tr}\boldsymbol{\varepsilon}\right)\mathbf{I} + 2\,\mu\,\boldsymbol{\varepsilon} + 2\,\mu_{\circ}\beta^2\left(\operatorname{dev}\boldsymbol{\varepsilon}\right) + 2\,\mu_c\left(\operatorname{asy}\,\nabla\mathbf{u}-\boldsymbol{\Phi}\right). & \text{(Phase 3)}
\end{cases} \tag{32}
$$

The couple-stress tensor is related to the curvature strain tensors by the following formulas:

$$
\boldsymbol{\mu} = \begin{cases}
\begin{cases} \bar{\lambda}\left(\operatorname{tr}\boldsymbol{\kappa}\right)\mathbf{I}+2\left(\bar{\mu}-\bar{\mu}_c\right)\left(\operatorname{sym}\boldsymbol{\kappa}\right)+2\left(\mu_{\circ}+\dfrac{\mu}{\beta^2}\right)\boldsymbol{\kappa} & \text{if}\quad \bar{\mu}\geq\bar{\mu}_c, \\ \bar{\lambda}\left(\operatorname{tr}\boldsymbol{\kappa}\right)\mathbf{I}-2\left(\bar{\mu}-\bar{\mu}_c\right)\left(\operatorname{asy}\boldsymbol{\kappa}\right)+2\left(\mu_{\circ}+\dfrac{\mu}{\beta^2}\right)\boldsymbol{\kappa} & \text{if}\quad \bar{\mu}<\bar{\mu}_c. \end{cases} & \text{(Phase 1)} \\[4ex]
\begin{cases} \bar{\lambda}\left(\operatorname{tr}\boldsymbol{\kappa}\right)\mathbf{I} + 2\,\bar{\mu}\left(\operatorname{sym}\boldsymbol{\kappa}\right) + 2\,\bar{\mu_c}\left(\operatorname{asy}\boldsymbol{\kappa}\right) \\ +\,4\,\alpha\left(\,\|\operatorname{sym}\boldsymbol{\kappa}\|^2 + \|\operatorname{asy}\boldsymbol{\kappa}\|^2 - \beta^2\,\|\operatorname{dev}\boldsymbol{\varepsilon}\|^2\right)\boldsymbol{\kappa} \end{cases} & \text{(Phase 2)} \\[4ex]
\begin{cases} \bar{\lambda}\left(\operatorname{tr}\boldsymbol{\kappa}\right)\mathbf{I} + 2\left(\bar{\mu}-\bar{\mu}_c\right)\left(\operatorname{sym}\boldsymbol{\kappa}\right) & \text{if}\quad \bar{\mu}\geq\bar{\mu}_c, \\ \bar{\lambda}\left(\operatorname{tr}\boldsymbol{\kappa}\right)\mathbf{I} - 2\left(\bar{\mu}-\bar{\mu}_c\right)\left(\operatorname{asy}\boldsymbol{\kappa}\right) & \text{if}\quad \bar{\mu}<\bar{\mu}_c. \end{cases} & \text{(Phase 3)}
\end{cases} \tag{33}
$$

4 Numerical Results

Based on one-dimensional numerical computations the mechanical response of the material is analyzed along some chosen macroscopic strain paths. A simple shear and a tension-compression tests are briefly presented to observe the development of microstructures which is characterized by the activation of different material regimes as discussed in the Sect. 3.2.

4.1 A Simple Shear Test

Consider a two dimensional domain $\Omega = (0, X_1) \times (0, X_2)$ where $(X_1, X_2) \in \mathbb{R}^2$. We choose the macroscopic strain paths as follows

$$\begin{aligned}
\boldsymbol{\varepsilon} &= \frac{\gamma}{2}\left(\mathbf{e}_1 \otimes \mathbf{e}_2 + \mathbf{e}_2 \otimes \mathbf{e}_1\right), \\
\boldsymbol{e} &= \gamma\, \mathbf{e}_2 \otimes \mathbf{e}_1 + \varphi_3 \left(\mathbf{e}_2 \otimes \mathbf{e}_1 - \mathbf{e}_1 \otimes \mathbf{e}_2\right), \\
\boldsymbol{\omega}_e &= \left(\frac{\gamma}{2} + \varphi_3\right)\left(\mathbf{e}_2 \otimes \mathbf{e}_1 - \mathbf{e}_1 \otimes \mathbf{e}_2\right), \\
\boldsymbol{\kappa} &= b\left(\mathbf{e}_1 \otimes \mathbf{e}_3 + \mathbf{e}_2 \otimes \mathbf{e}_3\right).
\end{aligned} \tag{34}$$

Here, γ is the macroscopic shear, φ_3 is the material microrotational degree of freedom and b is some fixed curvature. We assume that φ_3 linearly depends on both of the material coordinates X_1 and X_2 such that $\varphi_3 = b(X_1 + X_2)$. In this analysis we take $b = \frac{\pi}{6}$ and calculate φ_3 for all those material points which lies on the line $X_1 + X_2 = 1$. Other than Lame's constants $\lambda = \frac{\nu \mathrm{E}}{(1+\nu)(1-2\nu)}$ and $\mu = \frac{\mathrm{E}}{2(1+\nu)}$ there are eight additional material parameters that are pertinent to the material microstructures and are described in Table 1. Initially the material experiences a microstructure in micromotions of the particles. Upon further loading it transforms its structure and enter into a regime where there is no microstructure in the material. Further, upon increasing the load it changes its state to a material regime where it experiences a microstructure in translational motions of the particles. It is observed that all three phases of the material structure with two microstructural regimes and one non-microstructural regime coexists. In Fig. 5a the constitutive response of the material is shown, where it is observed that the non-monotone stress-strain curve is replaced by its energetically equivalent Maxwell line corresponding to a uniform vanishing stress. This vanishing stress regime is corresponding to the regime of the material where it experiences a microstructure in micromotions of the particles. In the material regime where there is no internal structure a nonlinear constitutive response is seen. Whereas, in the material regime where there is a microstructure in translational motions of the particles we observe a linear constitutive response in this

Table 1 Material parameters for the analytical computations in a simple shear test

Parameter	Numerical value	Units	Parameter	Numerical value	Units
E	2.0×10^2	(MPa)	$\overline{\lambda}$	λ	(N)
μ_c	1.0×10^{-1}	(MPa)	$\overline{\mu}$	μ	(N)
ν	0.3	(—)	$\overline{\mu}_c$	μ_c	(N)
α	5.0×10^{-1}	(N.mm^2)	β	1.0×10^1	(mm^{-1})

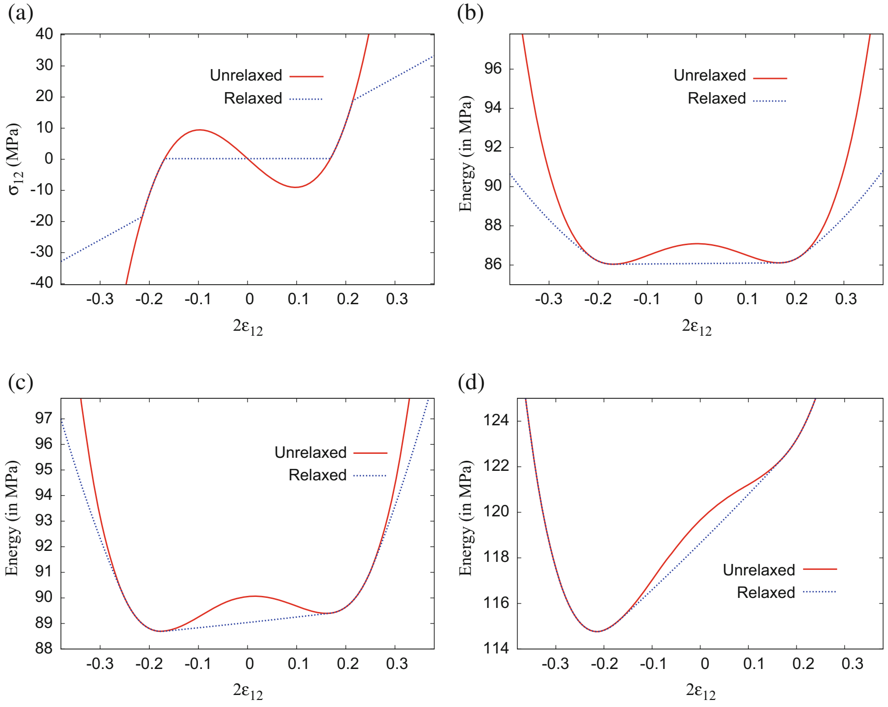

Fig. 5 (**a**) Relaxed and unrelaxed stress-strain curve in different material regimes; (**b**) Relaxed and unrelaxed curve for the Cosserat coupled modulus $\mu_c = 0.1$; (**c**) Relaxed and unrelaxed curve for the Cosserat coupled modulus $\mu_c = 1.0$; and, (**d**) Relaxed and unrelaxed curve for the Cosserat coupled modulus $\mu_c = 10.0$

one dimensional analysis. The corresponding nonconvex and relaxed energy plots are shown in Fig. 5b. In Fig. 5c and d the relaxed and unrelaxed energy is plotted for two different values of the Cosserat coupled modulus $\mu_c = 1.0$ and $\mu_c = 10.0$ respectively. These figures demonstrate that not only the particle size in granular material effects the development of microstructures but also the Cosserat coupled shear modulus do have influence in the development of material microstructures in granular materials.

Table 2 Material parameters for the analytical computations in a tension-compression test

Parameter	Numerical value	Units	Parameter	Numerical value	Units
E	2.0×10^2	(MPa)	$\overline{\lambda}$	1.15×10^2	(N)
μ_c	1.0×10^{-2}	(MPa)	$\overline{\mu}$	7.69×10^1	(N)
ν	0.3	(—)	$\overline{\mu}_c$	1.00×10^1	(N)
α	1.0×10^{-1}	(N.mm^2)	β	1.20×10^2	(mm^{-1})

4.2 A Tension-Compression Test

In this example the material behavior in a plain strain tension-compression test is investigated. The macroscopic strain tensors for this analysis takes the following form

$$\begin{aligned}
\boldsymbol{\varepsilon} &= \delta\, \mathbf{e}_1 \otimes \mathbf{e}_1 , \\
\boldsymbol{e} &= \delta\, \mathbf{e}_1 \otimes \mathbf{e}_1 + \varphi_3 \left(\mathbf{e}_2 \otimes \mathbf{e}_1 - \mathbf{e}_1 \otimes \mathbf{e}_2 \right) , \\
\boldsymbol{\omega}_e &= \varphi_3 \left(\mathbf{e}_2 \otimes \mathbf{e}_1 - \mathbf{e}_1 \otimes \mathbf{e}_2 \right) .
\end{aligned} \tag{35}$$

Here δ is the macroscopic stretch. The Cosserat rotational strain tensor $\boldsymbol{\kappa}$ is taken to be the same as mentioned in the previous test. Moreover, the micro-rotational degree of freedom, φ_3 at each material point is calculated according to similar assumption as in the case of simple shear test. The material parameters are chosen as described in Table 2.

It is observed that all the three phases of material structure coexists in this case. The constitutive behavior in the material microstructural and non-microstructural regimes is shown in Fig. 6a where contrary to the case of shear test it is observed that the stress do not vanish in the regime where material experiences a microstructure in micromotions. Here the non-monotone stress-strain curve is replaced by its energetically equivalent monotone curve. This is due to the non-constant slope of the relaxed energy envelope in the globally non-convex range of the unrelaxed energy potential, as seen in magnified picture in Fig. 6b. Moreover, the properties of unrelaxed and relaxed energy envelope are studied for different values of the interaction modulus α and the material parameter β related to the particle size. A two-well energy structure is seen in Figure for three different values of the interaction modulus. Both the wells have same local minima. In Fig. 6c it is observed that by varying the interaction modulus the local minima of the energy envelope do not change. This is because the globally nonconvex range of these energy curves do not vary. However it is important to note that the locally non-convex range of these unrelaxed energy curves decreases with the increase in the interaction modulus. The computed relaxed energy is plotted in Fig. 6d where it is seen that by varying the interaction modulus the global minima of all the three energy curves do not change. The influence of the particle size on the material strain energy is observed in Figs. 6e and f. It is seen that the particle size do not only influence the range of local non-convexity of the energy potential but also its global non-convexity range.

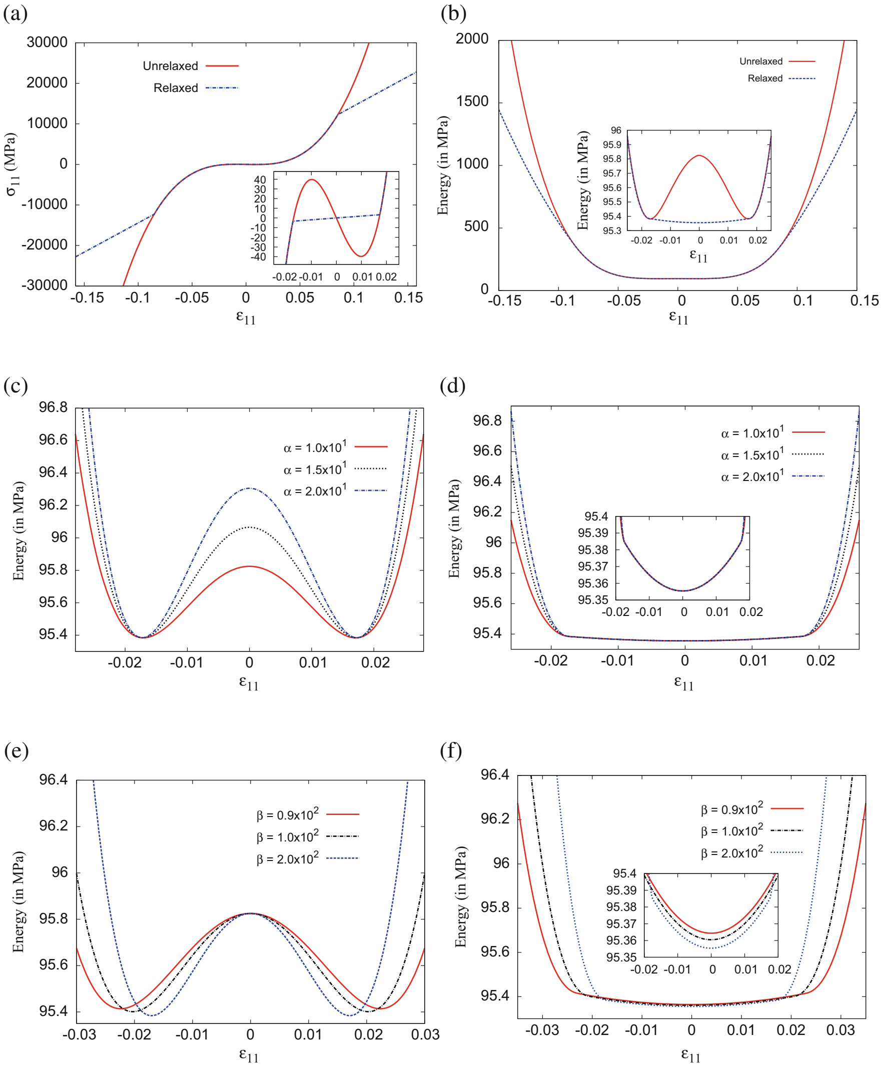

Fig. 6 (**a**) Relaxed and unrelaxed stress-strain curve in different regimes of the material; (**b**) Relaxed and unrelaxed energy curve in different material regimes; (**c**) Unrelaxed energy curves for varying values of the material parameter α; (**d**) Relaxed energy curves for varying values of the material parameter α; (**e**) Unrelaxed energy curves for varying values of the material parameter β; and, (**f**) Relaxed energy curves for varying values of the material parameter β

It is important to note that the local maxima of the energy potential do not change with the varying particle size. This is contrary to the case seen in Fig. 6c. Moreover, the local and global minima of the potential are shifted and get a lower values with the increased value of the material parameter β as seen in Figs. 6e and f.

5 Conclusion

In nature granular materials exhibit distinct patterns under deformation. The formation of these patterns is strongly influenced by counter-rotations of the interacting particle at the microscale. In this article, we study the counter-rotations of the particles and the formation of rotational microstructures in granular materials.

By employing the direct methods in the calculus of variations it turns out to be possible to derive an exact quasiconvex envelope of the energy potential. It is worth mentioning that there are no further assumptions necessary to derive this quasiconvex envelope. The computed relaxed potential yields all the possible displacement and micro-rotation field fluctuations as minimizers. Hence, by doing so we do not only resolve the issues concerning related non-quasiconvex variational problem but also guarantee the existence and uniqueness of energy minimizers. Moreover, the independence of these minimizers on the discretization of the spatial domain is ensured. We conclude with the result that the granular material behavior can be divided into three different regimes. Two of the material regimes are exhibiting microstructures in rotational and translational motions of the particles, respectively, and the third one is corresponding to the case where there is no internal structure of the deformation field.

The proposed model is analyzed numerically in one-dimensional case where the numerical computations performed are based on some chosen strain paths. We demonstrate on different properties of the computed relaxed potential in a simple shear and a tension-compression test. Moreover, It has been shown that all the material phases can co-exist.

Acknowledgements The first author gratefully acknowledge the funding by Higher Education Commission (HEC) of Pakistan and highly appreciate the support by Deutscher Akademischer Austausch Dienst (DAAD) for this research work.

References

1. Alonso-Marroquín, F., Vardoulakis, I., Herrmann, H.J., Weatherley, D., Mora, P.: Effect of rolling on dissipation in fault gouges. Phys. Rev. E **74**, 301–306 (2006). http://dx.doi.org/doi:10.1103/PhysRevE.74.031306
2. Alsaleh, M.I., Voyiadjis, G.Z., Alshibli, K.A.: Modeling strain localization in granular materials using micropolar theory: Mathematical formulations. Int. J. Numer. Anal. Meth. Goemech. **30**, 1501–1524 (2006). http://dx.doi.org/doi:10.1002/nag.533
3. Alshibli, K.A., Alsaleh, M.I., Voyiadjis, G.Z.: Modelling strain localization in granular materials using micropolar theory: Numerical implementation and verification. Int. J. Numer. Anal. Meth. Goemech. **30**, 1525–1544 (2006). http://dx.doi.org/doi:10.1002/nag.534
4. Aranda, E., Pedregal, P.: Numerical approximation of non-homogeneous, non-convex vector variational problems. Numer. Math. **89**, 425–444 (2001). http://dx.doi.org/doi:10.1007/s002110100294
5. Aranson, I., Tsimring, L.: Granular Patterns. Oxford University Press, Oxford (2009)

6. Bagnold, R.A.: The Physics of Blown Sand and Desert Dunes. Methuen and Co. Ltd., London (1941)
7. Ball, J.M.: Convexity conditions and existence theorems in nonlinear elasticity. Arch. Ration. Mech. Anal. **63**, 337–403 (1976). http://dx.doi.org/doi:10.1007/BF00279992
8. Ball, J.M., James, R.D.: Fine phase mixtures as minimizers of energy. Arch. Ration. Mech. Anal. **100**, 13–52 (1987). http://dx.doi.org/doi:10.1007/bf00281246
9. Bardet, J.P.: Observation on the effects of particle rotations on the failure of idealized granular materials. Mech. Mater. **8**, 159–182 (1994). http://dx.doi.org/doi:10.1016/0167-6636(94)00006-9
10. Bartels, S.: Numerical Analysis of Some Non-Convex Variational Problems. PhD thesis. Christian-Alberechts-Universität, Kiel (2001)
11. Bauer, E., Huang, W.: Numerical investigation of strain localization in a hypoplastic cosserat material under shearing. In: Desai (ed.) Proceedings of the 10th International Conference on Computer Methods and Advances in Geomechanics, pp. 525–528. Taylor & Francis (2001)
12. Carstensen, C., Conti, S., Orlando, A.: Mixed analytical-numerical relaxation in finite single-slip crystal plastictiy. Continuum Mech. Thermodyn. **20**, 275–301 (2008). http://dx.doi.org/doi:10.1007/s00161-008-0082-0
13. Carstensen, C., Hackl, K., Mielke, A.: Nonconvex potentials and microstructures in finite-strain plasticity. Proc. R. Soc. Lond. A **458**, 299–317 (2002). http://dx.doi.org/doi:10.1098/rspa.2001.0864
14. Carstensen, C., Plech$\acute{a}\check{c}$, P.: Numerical solution of the scalar double-well problem allowing microstructure. Math. Comp. **66**, 997–1026 (1997). http://dx.doi.org/doi:10.1090/S0025-5718-97-00849-1
15. Carstensen, C., Roub$\acute{i}\check{c}$ek, T.: Numerical approximation of young measures in non-convex variational problems. Numer. Math. **84**, 395–415 (2000). http://dx.doi.org/doi:10.1007/s002119900122
16. Carstensen, C., Roub$\acute{i}\check{c}$ek, T.: Numerical approximation of young measures in non-convex variational problems. Tech. Rep., 97–18 (1997). Universität Kiel
17. Chang, C.S., Hicher, P.Y.: An elasto-plastic model for grnaular materials with microstructural consideration. Int. J. Solids Struct. **42**, 4258–4277 (2005). http://dx.doi.org/doi:10.1016/j.ijsolstr.2004.09.021
18. Chang, C.S., Ma, L.: Elastic material constants for isotropic granular solids with particle rotation. Int. J. Solids Struct. **29**, 1001–1018 (1992). http://dx.doi.org/doi:10.1016/0020-7683(92)90071-Z
19. Chipot, M.: The appearance of microstructures in problems with incompatible wells and their numerical approach. Numer. Math. **83**, 325–352 (1999). http://dx.doi.org/doi:10.1007/s002110050452
20. Chipot, M., Collins, C.: Numerical approximation in variational problems with potential wells. SIAM J. Numer. Anal. **29**, 1002–1019 (1992). http://dx.doi.org/doi:10.1137/0729061
21. Collins, C., Kinderlehrer, D., Luskin, M.: Numerical approximation of the solution of a variational problem with a double well potential. SIAM J. Numer. Anal. **28**, 321–332 (1991). http://dx.doi.org/doi:10.1137/0728018
22. Conti, S., Hauret, P., Ortiz, M.: Conurrent multiscale computing of deformation microstructure by relaxation and local enrichment with application to single-crystal plasticity. Multiscale Model. Simul. **6**, 135–157 (2007). http://dx.doi.org/doi:10.1137/060662332
23. Conti, S., Ortiz, M.: Dislocation microstructures and the effective behavior of single crystals. Arch. Rational Mech. Anal. **176**, 103–147 (2005). http://dx.doi.org/doi:10.1007/s00205-004-0353-2
24. Conti, S., Theil, F.: Single-slip elastoplastic microstructures. Arch. Ration. Mech. Anal. **178**, 125–148 (2005). http://dx.doi.org/doi:10.1007/s00205-005-0371-8
25. Dacorogna, B.: Direct Methods in the Calculus of Variations. Springer, Berlin-Heidelberg-New York (1989)

26. Dacorogna, B.: Necessary and sufficient conditions for strong ellipticity of isotropic functions in any dimension. Discrete Contin. Dyn. Syst. B. **1**, 257–263 (2001). http://dx.doi.org/doi:10.3934/dcdsb.2001.1.257
27. de Borst, R.: Simulation of strain localization: a reappraisal of the Cosserat-continuum. Eng. Comp. **8**, 317–332 (1991). http://dx.doi.org/doi:10.1108/eb023842
28. DeSimone, A., Dolzmann, G.: Macroscopic response of nematic elastomers via relaxation of a class of SO(3)-invariant energies. Arch. Ration. Mech. Anal. **161**, 181–204 (2002). http://dx.doi.org/doi:10.107/s002050100174
29. Dolzmann, G., Walkington, N.J.: Estimates for numerical approximations of rank one convex envelopes. Numer. Math. **85**, 647–663 (2000). http://dx.doi.org/doi:10.1007/PL00005395
30. Le Dret, H., Raoult, A.: The quasiconvex envelope of the Saint Venant-Kirchhoff stored energy function. Proc. Roy. Soc. Edinburgh **125A**, 1179–1192 (1995). http://dx.doi.org/doi:10.1017/S0308210500030456
31. Ehlers, W., Volk, W.: On shear band localization phenomena of liquid-saturated granular elastoplastic porous solid materials accounting for fluid viscosity and micropolar solid rotations. Mech. Cohes.-Frict. Mat. **2**, 301–320 (1997). http://dx.doi.org/doi:10.1002/(SICI)1099-1484(199710)2:4<301::AID-CFM34>3.0.CO;2-D(10.1002/(SICI)1099-1484(199710)2:4<301::AID-CFM34>3.0.CO;2-D)
32. Govindjee, S., Hackl, K., Heinen, R.: An upper bound to the free energy of mixing by twin-compatible lamination for n-variant martensitic phase transformations. Continuum Mech. Thermodynam. **18**, 443–453 (2007). http://dx.doi.org/doi:10.1007/s00161-006-0038-1(10.1007/s00161-006-0038-1)
33. Gudehus, G., Nübel, K.: Evolution of shear bands in sand. Géotechnique **54**, 187–201 (2004). http://dx.doi.org/doi:10.1680/geot.2004.54.3.187(10.1680/geot.2004.54.3.187)
34. Gürses, E., Miehe, C.: On evolving deformation microstructures in non-convex partially damaged solids. J. Mech. Phys. Solids **59**, 1268–1290 (2011). http://dx.doi.org/doi:10.1016/j.jmps.2011.01.002(10.1016/j.jmps.2011.01.002)
35. Hackl, K., Heinen, R.: An upper bound to the free energy of n-variant polycrystalline shape memory alloys. J. Mech. Phys. Solids. **56**, 2832–2843 (2008). http://dx.doi.org/doi:10.1016/j.jmps.2008.04.005(10.1016/j.jmps.2008.04.005)
36. Kohn, R.V.: The relaxation of a double-well energy. Continum Mech. Thermodynam. **3**, 193–236 (1991). http://dx.doi.org/doi:10.1007/BF01135336
37. Kohn, R.V., Strang, G.: Optimal design and relaxation of variational problems I. Comm. Pure Appl. Math. **39**, 113–137 (1986). http://dx.doi.org/doi:10.1002/cpa.3160390107
38. Kohn, R.V., Strang, G.: Optimal design and relaxation of variational problems II. Comm. Pure Appl. Math. **39**, 139–182 (1986). http://dx.doi.org/doi:10.1002/cpa.3160390202
39. Kohn, R.V., Vogelius, M.: Relaxation of a variational method for impedance computed tomography. Comm. Pure Appl. Math. **40**, 745–777 (1987). http://dx.doi.org/doi:10.1002/cpa.3160400605
40. Lambrecht, M., Miehe, C., Dettmar, J.: Energy relaxation of non-convex incremental stress potentials in a strain-softening elastic-plastic bar. Int. J. Soids Struct. **40**, 1369–1391 (2003). http://dx.doi.org/doi:10.1016/S0020-7683(02)00658-3
41. Morrey, C.B.: Quasi-convextiy and the lower seimicontinuity of multiple integrals. Pac. J. Math. **2**, 25–53 (1952). See http://projecteuclid.org/euclid.pjm/1103051941http://projecteuclid.org/euclid.pjm/1103051941
42. Nicolaides, R.A., Walkington, N.J.: Computation of microsturcture utilizing Young measures representations. In: Rogers, C.A., Rogers, R.A. (eds.) Recent Advances in Adaptive and Sensory Materials and their Applications, pp. 131–141.Technomic Publ., Lancaster (1992)
43. Nicolaides, R.A., Walkington, N.J.: Strong convergence of numerical solutions to degenrate variational problems. Math. Comp. **64**, 117–127 (1992). See http://www.jstor.org/stable/2153325http://www.jstor.org/stable/2153325
44. Oda, M., Kazama, H.: Microstructure of shear bands and its relation to the mechanisms of dilatancy and failure of dense granular soils. Géotechnique **48**, 465–481 (1998). http://dx.doi.org/doi:10.1680/geot.1998.48.4.465

45. Papanicolopulos, S.A., Veveakis, E.: Sliding and rolling dissipation in Cosserat plasticity. Granular Matter **13**, 197–204 (2011). http://dx.doi.org/doi:10.1007/s10035-011-0253-8
46. Pasternak, E., Mühlhaus, H.B.: Cosserat continuum modelling of granulate materials. In: Valliappan, S., Khalili, N. (eds.) Computational Mechanics - New Frontiers for New Millennium, pp. 1189–1194. Elsevier Science (2001)
47. Pedregal, P.: Parametrized Measures and Variational Principles. Birkhäuser (1997)
48. Pedregal, P.: Numerical approximation of parametrized measures. Numer. Funct. Anal. Optim. **16**, 1049–1066 (1995). http://dx.doi.org/doi:10.1080/01630569508816659
49. Pedregal, P.: On numerical analysis of non-convex variational problems. Numer. Math. **74**, 325–336 (1996). http://dx.doi.org/doi:10.1007/s002110050219
50. Raoult, A.: Quasiconvex envelopes in nonlinear elasticity. In: Schröder, J., Neff, P. (ed.) Poly-, Quasi- and Rank-One Convexity in Applied Mechanics, pp. 17–51. Springer, Vienna (2010). http://dx.doi.org/doi:10.1007/978-3-7091-0174-2
51. Roubíček, T.: Relaxation in Optimization Theory and Variational Calculus. Valter de Gruyter, Berlin, New York (1997)
52. Roubíček, T.: Finite element approximation of a microstructure evolution. Math. Methods Appl. Sci. **17**, 377–393 (1994). http://dx.doi.org/doi:10.1002/mma.1670170505
53. Roubíček, T.: Numerical approximation of relaxed variational problems. J. Convex Anal. **3**, 329–347 (1996). See http://eudml.org/doc/i33027http://eudml.org/doc/233027
54. Sawada, K., Zhang, F., Yashima, A.: Rotation of granular material in laboratory tests and its numerical simulation using TIJ-Cosserat continuum theory. Comput. Methods, 1701–1706 (2006). http://dx.doi.org/doi:10.1007/978-1-4020-3953-9_104
55. Suiker, A.S.J., de Borst, R., Chang, C.S.: Micro-mechanical modelling of granular material. Part 1: Derivation of a second-gradient micro-polar constitutive theory. Acta Mechanica **149**, 161–180 (2001). http://dx.doi.org/doi:10.1007/BF01261670
56. Suiker, A.S.J., de Borst, R., Chang, C.S.: Micro-mechanical modelling of granular material. Part 2: Plane wave propagation in infinite media. Acta Mechanica **149**, 181–200 (2001). http://dx.doi.org/doi:10.1007/BF01261671
57. Tejchman, J., Niemunis, A.: FE-studies on shear localization in an anisotropic micro-polar hypoplastic granular material. Granular Matter. **8**, 205–220 (2006). http://dx.doi.org/doi:10.1007/s10035-006-0009-z
58. Tordesillas, A., Peters, J.F., Muthuswamy, M.: Role of particle rotations and rolling resistance in a semi-infinite particulate solid indented by a rigid flat punch. ANZIAM J. **46**, C260–C275 (2005)
59. Tordesillas, A., Walsh, S.D.C.: Incorporating rolling resistance and contact anisotropy in micromechanical models of granular media. Powder Technol. **124**, 106–111 (2002). http://dx.doi.org/doi:10.1016/S0032-5910(01)00490-9
60. Tordesillas, A., Walsh, S.D.C., Gardiner, B.: Bridging the length scales: Micromechanics of granular media. BIT Numer. Maths. **44**, 539–556 (2004). http://dx.doi.org/doi:10.1023/B:BITN.0000046817.60322.ed
61. Trinh, B.T., Hackl, K.: Performance of mixed and enhanced finite elements for strain localization in hypoplasticity. Int. J. Numer. Anal. Methods Geomech. **35**, 1125–1150 (2012). https://doi.org/10.1002/nag.1042
62. Trinh, B.T., Hackl, K.: Modelling of shear localization in solids by means of energy relaxation. Asia Pac. J. Comput. Eng. **1**, 1–21 (2014)
63. Trinh, B.T., Hackl, K.: A model for high temperature creep of single crystal superalloys based on nonlocal damage and viscoplastic material behavior. Contin. Mech. Thermodyn. **26**, 551–562 (2014)
64. Young, L.C.: Generalized Curves and the Existence of an Attained Absolute Minimum in the Calculus of Variations, pp. 212–234 (1937)

From Nonlinear to Linear Elasticity in a Coupled Rate-Dependent/Independent System for Brittle Delamination

Riccarda Rossi and Marita Thomas

Abstract We revisit the weak, energetic-type existence results obtained in (Rossi and Thomas, ESAIM Control Optim. Calc. Var. **21**, 1–59, (2015)) for a system for rate-independent, brittle delamination between two visco-elastic, *physically nonlinear* bulk materials and explain how to rigorously extend such results to the case of visco-elastic, *linearly* elastic bulk materials. Our approximation result is essentially based on deducing the Mosco-convergence of the functionals involved in the energetic formulation of the system. We apply this approximation result in two different situations at small strains: Firstly, to pass from a nonlinearly elastic to a linearly elastic, brittle model on the time-continuous level, and secondly, to pass from a time-discrete to a time-continuous model using an adhesive contact approximation of the brittle model, in combination with a vanishing, super-quadratic regularization of the bulk energy. The latter approach is beneficial if the model also accounts for the evolution of temperature.

1 Introduction

In the spirit of generalized standard materials, cf. e.g. [12], delamination processes along a prescribed interface Γ_{C} between two elastic materials $\Omega_+, \Omega_- \subset \mathbb{R}^d$ can be modeled with the aid of an internal delamination variable $z : [0, T] \times \Gamma_{\mathrm{C}} \to [0, 1]$, which describes the state of the glue located in Γ_{C} during a time interval $[0, T]$. In particular, in our notation $z(t, x) = 1$, resp. $z(t, x) = 0$, shall indicate that the

R. Rossi (✉)
Università degli studi di Brescia, DIMI, Brescia, Italy
e-mail: riccarda.rossi@unibs.it

M. Thomas
Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany
e-mail: marita.thomas@wias-berlin.de

E. Rocca et al. (eds.), *Trends in Applications of Mathematics to Mechanics*,
Springer INdAM Series 27, https://doi.org/10.1007/978-3-319-75940-1_7

glue is fully intact, resp. broken, at $(t, x) \in [0, T] \times \Gamma_C$. Such a type of modeling approach in the framework of delamination dates back to e.g. [10, 13]. In the case of a *rate-independent* evolution law for z, analytical results for delamination models have been obtained e.g. in [14, 22] in the case of *adhesive contact* and *brittle delamination* in the framework of the energetic formulation of rate-independent processes. Instead, [26], also in the fully rate-independent setting, constructed for the brittle system *local* (or *semistable energetic*) solutions, i.e. fulfilling a minimality property for the displacements and a semistability inequality for the internal variable, combined with an energy-dissipation inequality, cf. also [20]. The approach in [26] was based on time discretization using an alternate minimization scheme. Semistable energetic solutions to the adhesive contact system were also obtained in [27] by a vanishing-viscosity approach. In [21] existence of semistable energetic solutions for an adhesive contact model with rate-independent evolution of the delamination variable was discussed for the first time in combination with other rate-dependent effects: Therein, the displacements are subjected to viscosity and acceleration, and in addition also the evolution of temperature is taken into account. Based on this, [23] addressed the existence of (weak, energetic-type) solutions for a *brittle delamination* system, extending the isothermal, fully rate-independent model addressed in [22] to the coupled rate-independent/rate-dependent setting of [21]. The aim of this work is to further extend the analytical results that were developed in [23] for rate-independent delamination in visco-elastic *physically nonlinear* materials at small strains, to the case of *physically linear* materials at small strains.

More precisely, the existence of solutions to the coupled rate-dependent/independent system for brittle delamination was shown in [23] by passing to the limit in an approximate system for *adhesive contact*, under the condition that the elastic energy density $W = W(e)$ fulfilled

$$c|e|^p \leq W(e) \leq C(|e|^p + 1) \quad \text{with } p > d. \tag{1.1}$$

This kind of nonlinear growth is used in the engineering literature to model strain hardening or softening of so-called power-law materials, see e.g. [11, 15]. In particular, the exponent $p > d$ is applied at small strains in [4] to describe strain hardening. Yet, for our analytical results in [23], the condition $p > d$ also had a very specific, technical motivation. In fact, our analysis relied on the validity of a Hardy inequality, applied to the displacement variable u, which at that time was only available for functions in $W^{1,p}(\Omega; \mathbb{R}^d)$ with $p > d$. In the meantime, an improved version of this Hardy inequality, also valid for $p = 2$, was obtained in [8], thus making the restriction $p > d$ unnecessary. This was already reflected in [26], where the existence of semistable energetic solutions was shown by a constructive approach combining the adhesive-to-brittle limit and the discrete-to-continuous limit passages in a time discretization scheme. A quadratic growth for the elastic energy density was also allowed in [25], where the existence of solutions to the brittle delamination system in *visco-elastodynamics* (i.e., encompassing inertial effects) was still obtained by passing to the limit in the adhesive contact approximate system.

The aim of this note is to close the gap between the results in [23] and those in [25, 26]. Namely, we will perform

(1) the limit passage from nonlinear to linear small-strain elasticity in the mechanical force balance for the brittle delamination system;
(2) the joint adhesive-to-brittle, discrete-to-continuous, nonlinearly elastic-to-linearly elastic limit passage in a delamination system at small strains, also encompassing thermal effects.

We do not consider the case of geometrically nonlinear materials, which would be treated in a different way in the framework of finite-strain elasticity, e.g. using tools like polyconvexity.

In Sect. 2.1, we are going to describe the brittle delamination and adhesive contact systems, confining the discussion to the *quasistatic* (without inertia in the mechanical force balance for the displacements) and isothermal case. Yet, as we discuss in more detail in Sect. 4.2, it is possible to encompass thermal effects in our analysis, still remaining quasistatic for the displacements. But here, unhampered by the technical problems related to the handling of inertia and temperature, we will focus on the analytical difficulties attached to the adhesive-to-brittle limit. We will then explain the technique for taking the adhesive-to-brittle limit passage in the equation for the displacements first developed in [23]. This will help us put into context the main result of this paper, Theorem 3, stating the MOSCO-convergence of the energy functionals underlying the brittle (small-strain) mechanical force balance from the nonlinearly to linearly elastic case. While Theorem 3 will be stated in Sect. 2.2 and proved throughout Sect. 3, its applications to the limit passages **(1)** & **(2)** will be carried out in Sect. 4.

Let us finally fix some notation that will be used throughout the paper: We will denote by $\|\cdot\|_X$ both the norm of a Banach space X and, often, the norm in any power of it, and by $\langle\cdot,\cdot\rangle_X$ the duality pairing between X^* and X. Moreover, we shall often denote by the symbols c, c', C, C' various positive constants, whose meaning may vary from line to line, depending only on known quantities.

2 Our Main Result: Motivation and Statement

2.1 The Brittle Delamination System, Its Adhesive Contact Approximation, and the Adhesive-to-Brittle Limit

Let us now gain insight into the PDE system for brittle delamination between two bodies Ω_+ and $\Omega_- \subset \mathbb{R}^d$, $d \geq 2$. We enforce the

$$\text{brittle constraint:} \qquad [\![u(t)]\!] = 0 \quad \text{a.e. on } (0,T) \times \operatorname{supp} z(t)\,, \tag{2.1}$$

where $[\![u]\!] = u^+|_{\Gamma_C} - u^-|_{\Gamma_C}$ is the jump of u across the interface $\Gamma_C = \overline{\Omega_-} \cap \overline{\Omega_+}$, $u^\pm|_{\Gamma_C}$ denoting the traces on Γ_C of the restrictions $u^\pm$ of u to $\Omega_\pm$, and $\operatorname{supp} z$ the support of the delamination variable $z \in L^\infty(\Gamma_C)$, cf. (2.19) ahead. Hence, (2.1) ensures the continuity of the displacements, i.e. $[\![u(t,x)]\!] = 0$, in the (closure of the) set of points where (a portion of) the bonding is still active, i.e. $z(t,x) > 0$, and it allows for displacement jumps only in points $x \in \Gamma_C$ where the bonding is completely broken, where $z(t,x) = 0$. Therefore, (2.1) distinguishes between the crack set $\Gamma_C \backslash \operatorname{supp} z(t)$, where the displacements may jump, and the complementary set with active bonding, where it imposes a transmission condition on the displacements. We also enforce the

$$\text{non-penetration condition:} \qquad [\![u(t)]\!] \cdot \mathbf{n} \geq 0 \quad \text{a.e. on } (0,T) \times \operatorname{supp} z(t)\,, \tag{2.2}$$

with $\mathbf{n}$ the unit normal to Γ_C, oriented from Ω_+ to Ω_-.

The PDE system for brittle delamination between two visco-elastic bodies addressed in this paper consists of the *quasistatic* mechanical force balance for the displacements

$$-\operatorname{div}(\sigma(e,\dot{e})) = F \qquad \text{in } (0,T) \times (\Omega_+ \cup \Omega_-), \tag{2.3a}$$

where $e = e(u) := \frac{1}{2}(\nabla u + \nabla u^\top)$ is the linearized strain tensor and $\dot{e} = e(\dot{u})$, while F is a time-dependent applied volume force. The stress tensor σ, encompassing the visco-elastic response of the body, is given by the following constitutive law

$$\sigma(e,\dot{e}) = \mathbb{D}\dot{e} + \mathrm{D}W(e),$$

where $\mathbb{D} \in \mathbb{R}^{d\times d\times d\times d}$ is the symmetric and positive definite viscosity tensor and the elastic energy density $W : \mathbb{R}^{d\times d} \to [0,\infty)$, with Gâteaux derivative $\mathrm{D}W$, is specified by (2.18) below. Equation (2.3a) is supplemented with homogeneous Dirichlet boundary conditions on the Dirichlet part Γ_D of the boundary $\partial\Omega$, where $\Omega := \Omega_+ \cup \Gamma_C \cup \Omega_-$, and subject to an applied traction f on the Neumann part $\Gamma_N = \partial\Omega \setminus \Gamma_D$, i.e.

$$u = 0 \quad \text{on } (0,T) \times \Gamma_D, \qquad \qquad \sigma(e,\dot{e})|_{\Gamma_N}\nu = f \quad \text{on } (0,T) \times \Gamma_N, \tag{2.3b}$$

with ν the outward unit normal to $\partial\Omega$. For technical reasons, we will require Γ_D to have positive distance from Γ_C, cf. Assumption 1 ahead. The evolution of u and of the delamination parameter z are coupled through the following (formally written) boundary condition on the contact surface Γ_C

$$\sigma(e,\dot{e})|_{\Gamma_C}\mathbf{n} + \partial_u \widetilde{J}_\infty([\![u]\!], z) + \partial I_{C(x)}([\![u]\!]) \ni 0 \quad \text{on } (0,T) \times \Gamma_C, \tag{2.4}$$

where the subdifferential terms render the brittle and non-penetration constraints, respectively. Indeed, $\partial_u \widetilde{J}_\infty : \mathbb{R}^d \times \mathbb{R} \rightrightarrows \mathbb{R}^d$ is the subdifferential (in the sense of convex analysis) of the functional $\widetilde{J}_\infty : \mathbb{R}^d \times \mathbb{R} \to [0,\infty]$ defined by the indicator function of the set individuated by (a slightly weaker version of) the brittle constraint, namely

$$\widetilde{J}_\infty(v,z) := I_{\{vz=0\}}(v,z) = \begin{cases} 0 & \text{if } vz = 0, \\ \infty & \text{otherwise.} \end{cases} \tag{2.5}$$

The non-penetration constraint is imposed through the multivalued mapping $C : \Gamma_{\mathrm{C}} \rightrightarrows \mathbb{R}^d$ defined by

$$C(x) := \{v \in \mathbb{R}^d \ : \ v \cdot \mathbf{n}(x) \geq 0\} \qquad \text{for a.a. } x \in \Gamma_{\mathrm{C}}. \tag{2.6}$$

Further coupling is provided by the flow rule for the delamination parameter

$$\partial \mathrm{R}(\dot{z}) + \partial \mathcal{G}(z) + \partial_z \widetilde{J}_\infty([\![u]\!], z) \ni 0 \quad \text{on } (0,T) \times \Gamma_{\mathrm{C}}, \tag{2.7}$$

featuring the dissipation potential density

$$\mathrm{R}(\dot{z}) := \begin{cases} a_1|\dot{z}| & \text{if } \dot{z} \leq 0, \\ \infty & \text{otherwise,} \end{cases}$$

(with $a_1 > 0$ the phenomenological specific energy per area dissipated by disintegrating the adhesive) and $\partial\mathcal{G}$ the (still formally written) subdifferential of a functional $\mathcal{G}$ encompassing a suitable gradient regularization, given in (2.16) below.

The brittle and non-penetration constraints are reflected in the variational formulation of the mechanical force balance for the displacements. To properly give it, we introduce the time-dependent spaces

$$\begin{aligned} \mathbf{V}_z^q(t) := \{v \in W_{\mathrm{D}}^{1,q}(\Omega \backslash \Gamma_{\mathrm{C}}; \mathbb{R}^d) \ : \ [\![v]\!] = 0 \text{ a.e. on } \operatorname{supp} z(t) \subset \Gamma_{\mathrm{C}} \text{ and} \\ [\![v(x)]\!] \in C(x) \text{ for a.a. } x \in \Gamma_{\mathrm{C}}\}\,, \end{aligned}$$

where the exponent $q > 1$ depends on the growth properties of the density W and we use the notation $W_{\mathrm{D}}^{1,q}(A;\mathbb{R}^d)$ for the space of $W^{1,q}$-functions on a domain A with null trace on Γ_{D}. In this work, we will in particular deal with the cases $q = p > d$ and $q = 2$. Thus, the weak formulation of (2.3) reads

$$\begin{aligned} & u(t) \in \mathbf{V}_z^q(t) \quad \text{for a.a. } t \in (0,T), \\ & \int_{\Omega\backslash\Gamma_{\mathrm{C}}} \big(\mathbb{D}e(\dot{u}(t)) + \mathrm{D}W(e(u(t)))\big) : e(v - u(t))\,\mathrm{d}x \geq \langle L(t), v - u(t)\rangle \\ & \qquad \text{for all } v \in \mathbf{V}_z^q(t), \ \text{for a.a. } t \in (0,T), \end{aligned} \tag{2.8}$$

with $L : (0, T) \to W^{1,q}_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}}; \mathbb{R}^d)^*$ a functional subsuming the external forces F and f, i.e.

$$\langle L(t), v\rangle := \int_{\Omega} F(t)\cdot v\,\mathrm{d}x + \int_{\Gamma_{\mathrm{N}}} f(t)\cdot v\,\mathrm{d}\mathcal{H}^{d-1}(x)\,; \tag{2.9}$$

more details on the above duality pairing and the conditions on the forces F and f will be given in Sect. 4.1. In this paper, along the footsteps of [19, 24], we will weakly formulate the coupled rate-dependent/independent system (2.3, 2.4, 2.7) by means of an extension of the concept of semistable energetic solution from [20]. As we will see in Definition 3 ahead, the *semistable energetic* solutions of system (2.3, 2.4, 2.7) are defined by fulfilling the weak mechanical force balance for the displacements (2.8) combined with a suitable energy-dissipation inequality and a semistability condition, weakly rendering the flow rule (2.7).

In [23] we showed the existence of semistable energetic solutions of the brittle system, by passing to the limit in an approximate system where the brittle constraint (2.1) is penalized by the

adhesive contact term:

$$\int_{\Gamma_{\mathrm{C}}} J_k([\![u]\!], z)\,\mathrm{d}\mathcal{H}^{d-1}(x) \text{ with } J_k([\![u]\!], z) := \frac{k}{2} z|[\![u]\!]|^2 \quad \text{for } k > 0, \tag{2.10}$$

featured in the energy functional underlying the mechanical force balance for the displacements. Above, $\mathcal{H}^{d-1}$ denotes the $(d-1)$-dimensional Hausdorff measure. In fact, the existence of *energetic solutions* to the purely rate-independent brittle system was proved in [22] by passing to the limit in this adhesive contact approximation, as the parameter $k \to \infty$. For our coupled rate-dependent/independent brittle system, the adhesive contact approximation consists of the mechanical force balance (2.3) for the displacements coupled with the following contact surface condition and flow rule for the delamination parameter

$$\sigma(e, \dot{e})|_{\Gamma_{\mathrm{C}}}\mathbf{n} + \partial_u J_k([\![u]\!], z) + \partial I_{C(x)}([\![u]\!]) \ni 0 \quad \text{on } (0, T)\times\Gamma_{\mathrm{C}}, \tag{2.11}$$

$$\partial\mathrm{R}(\dot{z}) + \partial\mathcal{G}(z) + \partial_z J_k([\![u]\!], z) \ni 0 \quad \text{on } (0, T)\times\Gamma_{\mathrm{C}}, \tag{2.12}$$

which replace (2.4) and (2.7), respectively. Accordingly, the weak formulation of the mechanical force balance for the adhesive contact system (2.3, 2.11, 2.12) reads

$$\begin{aligned}&\int_{\Omega\backslash\Gamma_{\mathrm{C}}}\!\big(\mathbb{D}e(\dot{u}(t)) + \mathrm{D}W(e(u(t)))\big) : e(v-u(t))\,\mathrm{d}x + \int_{\Gamma_{\mathrm{C}}} kz(t)[\![u(t)]\!]\cdot[\![v-u(t)]\!]\,\mathrm{d}\mathcal{H}^{d-1}(x)\\ &\geq \langle L(t), v-u(t)\rangle \quad \text{for all } v\in\mathbf{V}^q, \text{ for a.a. } t\in(0,T),\end{aligned} \tag{2.13}$$

where the (no longer time-dependent) space for the test functions now only encompasses the non-penetration condition (2.2), i.e.

$$\mathbf{V}^q := \{v \in W^{1,q}_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d)\ :\ [\![v(x)]\!] \in C(x) \text{ for a.a. } x \in \Gamma_{\mathrm{C}}\}\,.$$

The limit-passage argument for the adhesive-to-brittle limit developed in [22] was based on the Evolutionary Gamma-convergence theory for (purely) rate-independent systems from [17]: Basically, it only necessitated the Gamma-convergence of the underlying energy and dissipation functionals, combined with a mutual recovery sequence condition that ensured the limit passage in the global stability condition. For coupled rate-dependent/independent systems, it is not sufficient to solely rely on the abstract toolbox of [17]: In particular, in our specific context, the Gamma-convergence of the energies no longer guarantees the limit passage, as $k \to \infty$, from the weak mechanical force balance for the displacements (2.13) to its brittle analogue (2.8). For that, given a sequence of semistable energetic solutions $(u_k, z_k)_k$ converging to a pair (u, z), which is a candidate semistable energetic solution of the brittle system, it is indeed necessary to construct, for every admissible test function $v \in \mathbf{V}^q_z(t)$ for the brittle mechanical force balance (2.8), with $t \in (0, T)$ fixed, a sequence $(v_k)_k$ of test functions for (2.13) such that

1. $(v_k)_k$ converge to v in a suitable sense, ensuring the limit passage in the bulk terms of (2.13);
2. the functions v_k also satisfy the non-penetration condition (2.2);
3. there holds

$$\limsup_{k\to\infty} \int_{\Gamma_{\mathrm{C}}} k z_k(t) [\![u_k(t)]\!] \cdot [\![v_k - u_k(t)]\!] \,\mathrm{d}\mathcal{H}^{d-1}(x) \le 0\,.$$

Since $\liminf_{k\to\infty} \int_{\Gamma_{\mathrm{C}}} k z_k(t) |[\![u_k(t)]\!]|^2 \,\mathrm{d}\mathcal{H}^{d-1}(x) \ge 0$ for almost all $t \in (0, T)$, it is immediate to check that the above property is ensured as soon as

$$\limsup_{k\to\infty} \int_{\Gamma_{\mathrm{C}}} k z_k(t) [\![u_k(t)]\!] \cdot [\![v_k]\!] \,\mathrm{d}\mathcal{H}^{d-1}(x) \le 0\,. \tag{2.14}$$

In [23] we were able to construct a sequence $(v_k)_k$ complying with (2.14), starting from a test function v such that $[\![v]\!] = 0$ a.e. on $\operatorname{supp} z(t)$, by modifying v in such a way that the support of the obtained $[\![v_k]\!]$ fitted to the null set of z_k, approximating z. This construction hinged on two crucial ingredients:

1. First, we preliminarily obtained refined convergence properties of the delamination variables $(z_k)_k$. In particular, we proved the *support convergence*

$$\operatorname{supp} z_k(t) \subset \operatorname{supp} z(t) + B_{\rho_k}(0) \quad \text{and} \quad \rho_k \to 0 \text{ as } k \to \infty, \tag{2.15}$$

at every $t \in (0, T)$ via arguments from geometric measure theory. In fact, our proof of (2.15) heavily relied on the following, specific choice for the gradient regularizing term for the delamination flow rule

$$\mathcal{G}(z) := \begin{cases} \mathrm{b}|\mathrm{D}z|(\Gamma_{\mathrm{C}}) & \text{if } z \in \mathrm{SBV}(\Gamma_{\mathrm{C}}; \{0, 1\}), \\ \infty & \text{otherwise,} \end{cases} \tag{2.16}$$

with $\mathrm{b} > 0$, $\mathrm{SBV}(\Gamma_{\mathrm{C}}; \{0, 1\})$ the set of the special bounded variation functions on Γ_{C}, taking values in $\{0, 1\}$, and $|\mathrm{D}z|(\Gamma_{\mathrm{C}})$ the variation on Γ_{C} of the Radon measure $\mathrm{D}z$. The set $\mathrm{SBV}(\Gamma_{\mathrm{C}}; \{0, 1\})$ thus only consists of characteristic functions of sets with finite perimeter in Γ_{C}, and the total variation $|\mathrm{D}z|(\Gamma_{\mathrm{C}})$ of $z = \chi_Z \in \mathrm{SBV}(\Gamma_{\mathrm{C}}; \{0, 1\})$ is given by the perimeter of Z in Γ_{C}. With (2.16) we thus imposed that z only takes the values 0 and 1, i.e. we encompassed in the model only two states of the bonding between Ω_+ and Ω_-, the fully effective and the completely ineffective ones. Relying on the information $z_k \in \{0, 1\}$ and on the support convergence (2.15), we in fact constructed a sequence $(v_k)_k$ such that

$$z_k(t)|[\![v_k(t)]\!]|^2 = 0 \quad \text{for all } k \in \mathbb{N} \text{ and all } t \in [0, T]. \tag{2.17}$$

2. Second, for establishing the convergence properties of the recovery sequence of test functions for the displacements, we resorted to a Hardy inequality given in [16] for closed sets of arbitrarily low regularity, but applicable only to functions in $W^{1,p}(\Omega; \mathbb{R}^d)$, with $p > d$. To enforce this integrability property for the gradients of the displacements, we thus had to impose the growth condition (1.1) on the elastic energy density and, accordingly, to consider the variational formulation of the adhesive contact and of the brittle equations for the displacements in the spaces $\mathbf{V}^p$ and $\mathbf{V}^p_z(t)$, respectively.

   However, this condition can be weakened to *quadratic* growth in view of the improved Hardy's inequality recently proved in [8].

As a matter of fact, our construction of recovery test functions did guarantee the Mosco-convergence of the energy functionals underlying the adhesive contact mechanical force balance (2.13) to that of the brittle mechanical force balance (2.8).

Indeed, in Sect. 2.2, we are going to state the main result of this paper in terms of Mosco-convergence of functionals. This result will ensure the passage from elastic energy densities with $(p{>}d)$-growth to quadratic densities in the following two situations:

1. in the brittle delamination system: for this, we will resort to the convergence of the functionals $(\Phi_k)_k$ to Φ_∞, cf. (2.21) & (2.23);
2. jointly with the adhesive-to-brittle and discrete-to-continuous limit passage in thermo-visco-elastic delamination systems: for this, we will resort to the convergence of the functionals $(\Phi_k^{\mathrm{adh}})_k$ to Φ_∞, cf. (2.22) & (2.23).

2.2 Our Main Result

Definition of MOSCO-Convergence

We recall the definition from, e.g. [3, Sec. 3.3, p. 295]): Given a Banach space X and proper functionals $\Phi_k,\ \Phi_\infty : \mathbb{R} \to (-\infty, \infty]$, $k \in \mathbb{N}$, we say that the sequence $(\Phi_k)_k$ MOSCO-converges to Φ as $k \to \infty$ if the following conditions hold:

- lim inf-inequality: for every $u \in X$ and all $(u_k)_k \subset X$ there holds

$$u_k \rightharpoonup u \text{ weakly in } X \ \Rightarrow\ \liminf_{k\to\infty} \Phi_k(u_k) \geq \Phi_\infty(u);$$

- lim sup-inequality: for every $v \in X$ there exists a sequence $(v_k)_k \subset X$ such that

$$v_k \to v \text{ strongly in } X \ \text{ and } \ \limsup_{k\to\infty} \Phi_k(v_k) \leq \Phi_\infty(v).$$

The Functionals

Throughout the paper, we will consider elastic energy densities of the type

$$\begin{aligned} &W_q : \mathbb{R}^{d\times d} \to [0,\infty) \text{ convex, differentiable, and such that} \\ &\exists\, c_q,\ C_q > 0\ \forall\, e \in \mathbb{R}^{d\times d}\ :\ \ c_q|e|^q \leq W_q(e) \leq C_q(|e|^q{+}1) \end{aligned} \tag{2.18}$$

for some $q \in (1,\infty)$ and the associated integral functionals on $\Omega\backslash\Gamma_{\mathrm{C}}$. We will also consider the integral functional induced by J_k from (2.10), i.e.

$$\mathcal{J}_k(v,z) := \int_{\Gamma_{\mathrm{C}}} J_k(v(x), z(x))\, \mathrm{d}\mathcal{H}^{d-1}(x)\,,$$

whose domain of definition depends on the choice of q from (2.18), cf. Remark 1 for more details. While $\mathcal{J}_k$ will contribute to Φ_k^{adh}, the functionals Φ_k and Φ_∞ will feature a term $\mathcal{J}_\infty$ accounting for the brittle constraint (2.1), which in turn involves the closed set $\operatorname{supp} z$. We will consider $\mathcal{J}_\infty$ to be defined for $z \in \mathrm{SBV}(\Gamma_{\mathrm{C}}; \{0,1\})$, which can be thus identified with the characteristic function of a finite perimeter set Z. In a measure-theoretic sense, $\operatorname{supp} z$ is given by

$$\operatorname{supp} z := \bigcap\{A \subset \Gamma_{\mathrm{C}} \subset \mathbb{R}^{d-1};\ A \text{ closed },\ \mathcal{H}^{d-1}(Z\backslash A) = 0\}. \tag{2.19}$$

We now define

$$\begin{aligned} &\mathcal{J}_\infty : L^1(\Gamma_{\mathrm{C}}; \mathbb{R}^d) \times \mathrm{SBV}(\Gamma_{\mathrm{C}}; \{0,1\}) \to [0,\infty]\,, \\ &\mathcal{J}_\infty(v,z) := \begin{cases} 0 & \text{if } v = 0\ \mathcal{H}^{d-1}\text{-a.e. on } \operatorname{supp} z, \\ \infty & \text{otherwise.} \end{cases} \end{aligned} \tag{2.20}$$

Finally, we introduce the integral functional induced by the indicator functions of the sets $C(x)$ from (2.6), i.e.

$$\mathcal{I}_C : L^1(\Gamma_C;\mathbb{R}^d) \to [0,\infty], \qquad \mathcal{I}_C(v) := \int_{\Gamma_C} I_{C(x)}(v(x))\,\mathrm{d}\mathcal{H}^{d-1}(x)\,.$$

Then, we define the functionals

$$\Phi_k : H^1_D(\Omega\backslash\Gamma_C;\mathbb{R}^d) \times \mathrm{SBV}(\Gamma_C;\{0,1\}) \to [0,\infty] \quad \text{given by}$$

$$\Phi_k(u,z) := \begin{cases} \int_{\Omega\backslash\Gamma_C} \Big(W_2(e(u)) + \frac{1}{k^p} W_p(e(u))\Big)\,\mathrm{d}x + \mathcal{J}_\infty([\![u]\!],z) & \text{if } u \in W^{1,p}_D(\Omega\backslash\Gamma_C;\mathbb{R}^d), \\ \infty & \text{otherwise,} \end{cases} \tag{2.21}$$

with $p > d$,

$$\Phi_k^{\mathrm{adh}} : H^1_D(\Omega\backslash\Gamma_C;\mathbb{R}^d) \times L^1(\Gamma_C) \to [0,\infty] \quad \text{given by}$$

$$\Phi_k^{\mathrm{adh}}(u,z) := \begin{cases} \int_{\Omega\backslash\Gamma_C} \Big(W_2(e(u)) + \frac{1}{k^p} W_p(e(u))\Big)\,\mathrm{d}x + \mathcal{J}_k([\![u]\!],z) & \text{if } u \in W^{1,p}_D(\Omega\backslash\Gamma_C;\mathbb{R}^d), \\ \infty & \text{otherwise,} \end{cases} \tag{2.22}$$

with $p > d$. We will show that, given a sequence $(z_k)_k \subset \mathrm{SBV}(\Gamma_C;\{0,1\})$ and suitably converging to some $z \in \mathrm{SBV}(\Gamma_C;\{0,1\})$ (cf. Theorem 3 below), both functionals $\Phi_k(\cdot,z_k)$ and $\Phi_k^{\mathrm{adh}}(\cdot,z_k)$ Mosco-converge in the $H^1_D(\Omega\backslash\Gamma_C;\mathbb{R}^d)$-topology, as $k\to\infty$, to the functional $\Phi_\infty(\cdot,z)$ defined by

$$\begin{aligned} &\Phi_\infty : H^1_D(\Omega\backslash\Gamma_C;\mathbb{R}^d) \times \mathrm{SBV}(\Gamma_C;\{0,1\}) \to [0,\infty], \\ &\Phi_\infty(u,z) := \int_{\Omega\backslash\Gamma_C} W_2(e(u))\,\mathrm{d}x + \mathcal{J}_\infty([\![u]\!],z)\,. \end{aligned} \tag{2.23}$$

Remark 1

1. Due to the condition $p > d$ and to trace theorems, for every $u \in W^{1,p}_D(\Omega\backslash\Gamma_C;\mathbb{R}^d)$ there holds

$$[\![u]\!] \in W^{1-1/p,p}(\Gamma_C;\mathbb{R}^d) \subset \mathrm{C}^0(\Gamma_C)\,. \tag{2.24}$$

   Therefore, for the term $\mathcal{J}_k([\![u]\!],z)$ to be well defined, it is in principle sufficient to have $z \in L^1(\Gamma_C)$.
2. As already mentioned, in [23] we performed the adhesive-to-brittle limit passage in the mechanical force balance staying in the context of nonlinear (small-strain) elasticity, with an elastic energy having p-growth, with $p > d$. In fact, we proved the Mosco-convergence of the functionals (w.r.t. the variable u, with the

second entry given by a sequence $(z_k)_k$ in $\mathrm{SBV}(\varGamma_{\mathrm{C}}; \{0, 1\})$ suitably converging to some z)

$$\begin{aligned} &\varPhi_k^{\mathrm{adh},p} : W_{\mathrm{D}}^{1,p}(\varOmega\backslash\varGamma_{\mathrm{C}}; \mathbb{R}^d) \times L^1(\varGamma_{\mathrm{C}}) \to [0, \infty), \\ &\varPhi_k^{\mathrm{adh},p}(u, z) := \int_{\varOmega\backslash\varGamma_{\mathrm{C}}} W_p(e(u))\,\mathrm{d}x + \mathfrak{J}_k([\![u]\!], z), \end{aligned}$$

to the functional

$$\widetilde{\varPhi}_\infty^p : W_{\mathrm{D}}^{1,p}(\varOmega\backslash\varGamma_{\mathrm{C}}; \mathbb{R}^d) \times L^1(\varGamma_{\mathrm{C}}) \to [0, \infty], \quad \widetilde{\varPhi}_\infty^p(u, z) := \int_{\varOmega\backslash\varGamma_{\mathrm{C}}} W_p(e(u))\,\mathrm{d}x + \widetilde{\mathfrak{J}}_\infty([\![u]\!], z),$$

with $\widetilde{\mathfrak{J}}_\infty$ the integral functional induced by the indicator function $\widetilde{J}_\infty$ from (2.5). Observe that, in view of (2.24), for $u \in W_{\mathrm{D}}^{1,p}(\varOmega\backslash\varGamma_{\mathrm{C}}; \mathbb{R}^d)$ there holds

$$\begin{gathered} z[\![u]\!] = 0\ \mathcal{H}^{d-1}\text{-a.e. on } \varGamma_{\mathrm{C}} \iff [\![u]\!] = 0\ \mathcal{H}^{d-1}\text{-a.e. on } \operatorname{supp} z, \\ \text{hence } \widetilde{\mathfrak{J}}_\infty([\![u]\!], z) = \mathfrak{J}_\infty([\![u]\!], z)\,. \end{gathered}$$

Instead, for the functional $\varPhi_\infty$, defined with $u \in H_{\mathrm{D}}^1(\varOmega\backslash\varGamma_{\mathrm{C}}; \mathbb{R}^d)$ it is essential to have the contribution with $\mathfrak{J}_\infty$, which enforces constraint (2.1) in terms of $\operatorname{supp} z$, stronger than $z[\![u]\!] = 0$ a.e. on $\varGamma_{\mathrm{C}}$. In fact, our argument for Mosco-convergence relies on the support convergence (2.15).

Assumptions

Let us now specify our geometric assumptions on the domain $\varOmega$, as well as the properties required of a sequence $(z_k)_k \subset \mathrm{SBV}(\varGamma_{\mathrm{C}}; \{0, 1\})$, converging to some $z \in \mathrm{SBV}(\varGamma_{\mathrm{C}}; \{0, 1\})$, to ensure that the functionals $\varPhi_k(\cdot, z_k)$ and $\varPhi_k^{\mathrm{adh}}(\cdot, z_k)$ Mosco-converge to $\varPhi_\infty(\cdot, z)$. In order to obtain a result as independent as possible from the problem of passing to the limit in the *coupled* system for brittle delamination, we will directly impose here certain additional regularity properties on $(z_k)_k$ and z, which are in fact induced by semistability, see Sect. 4.1.

We will suppose that the Dirichlet boundary $\varGamma_{\mathrm{D}}$ and the finite perimeter sets Z_k and Z associated with z_k and z enjoy a regularity property, which prevents outward cusps, introduced by Campanato as the Property $\mathfrak{a}$, cf. e.g. [6, 7], and also known as *lower density estimate* in e.g. [2, 9]. We recall it in the following definition.

Definition 1 (Property $\mathfrak{a}$) A set $M \subset \mathbb{R}^n$ has the Property $\mathfrak{a}$ if there exists a constant C such that

$$\forall\, y \in M \ \ \forall\, \rho_\star > 0 : \quad \mathcal{L}^n(M \cap B_{\rho_\star}(y)) \geq C\rho_\star^n\,. \tag{2.25}$$

Here, $B_{\rho_\star}(y)$ denotes the open ball of radius $\rho_\star$ with center in y.

We now fix our conditions on the domain Ω.

Assumption 1 *We suppose that*

$$\Omega \subset \mathbb{R}^d, d \geq 2, \text{ is bounded } \Omega_-,\ \Omega_+,\ \Omega \text{ are Lipschitz domains, } \Omega_+ \cap \Omega_- = \emptyset\,, \tag{2.26a}$$

$$\partial\Omega = \Gamma_{\mathrm{D}} \cup \Gamma_{\mathrm{N}},\ \text{s.th. } \Gamma_{\mathrm{N}} = \partial\Omega\backslash\Gamma_{\mathrm{D}},\ \Gamma_{\mathrm{D}} \subset \partial\Omega \text{ is closed with Property } \mathfrak{a}, \text{ and} \tag{2.26b}$$

$$\Gamma_{\mathrm{D}} \cap \overline{\Gamma_{\mathrm{C}}} = \emptyset,\ \mathcal{H}^{d-1}(\Gamma_{\mathrm{D}} \cap \overline{\Omega_-}) > 0\,,\ \mathcal{H}^{d-1}(\Gamma_{\mathrm{D}} \cap \overline{\Omega_+}) > 0\,, \tag{2.26c}$$

$$\mathrm{dist}(\Gamma_{\mathrm{D}}, \Gamma_{\mathrm{C}}) = \gamma > 0\,, \tag{2.26d}$$

$$\begin{aligned}\Gamma_{\mathrm{C}} = \overline{\Omega}_- \cap \overline{\Omega}_+ \subset \mathbb{R}^{d-1} \text{ is a “flat” surface, i.e. contained in a hyperplane of } \mathbb{R}^d, \\ \text{such that, in particular, } \mathcal{H}^{d-1}(\Gamma_{\mathrm{C}}) = \mathcal{L}^{d-1}(\Gamma_{\mathrm{C}}) > 0\,,\end{aligned} \tag{2.26e}$$

where $\mathcal{H}^{d-1}$, resp. $\mathcal{L}^{d-1}$, denotes the $(d-1)$-dimensional Hausdorff measure, resp. Lebesgue measure.

Here, the condition that Γ_{C} is contained in a hyperplane has no substantial role in our analysis, but to simplify arguments and notation.

As for the functions $(z_k)_k,\ z \subset \mathrm{SBV}(\Gamma_{\mathrm{C}}; \{0,1\})$, in addition to weak* convergence in $\mathrm{SBV}(\Gamma_{\mathrm{C}})$ we will suppose that they fulfill a *lower density estimate*, holding *uniformly* w.r.t. the parameter $k \in \mathbb{N} \cup \{\infty\}$.

Assumption 2 *There are constants R, $\mathfrak{a}(\Gamma_{\mathrm{C}}) > 0$ such that for every $k \in \mathbb{N} \cup \{\infty\}$ there holds*

$$\forall\, y \in \mathrm{supp}\, z_k\ \forall\, \rho_\star > 0: \quad \mathcal{L}^{d-1}(Z_k \cap B_{\rho_\star}(y)) \geq \begin{cases} \mathfrak{a}(\Gamma_{\mathrm{C}})\rho_\star^{d-1} & \text{if } \rho_\star < R, \\ \mathfrak{a}(\Gamma_{\mathrm{C}})R^{d-1} & \text{if } \rho_\star \geq R, \end{cases} \tag{2.27}$$

where Z_k is the finite perimeter set such that $z_k = \chi_{Z_k}$.

As we will see in Sect. 3.1, this condition, combined with the weak* convergence in $\mathrm{SBV}(\Gamma_{\mathrm{C}}; \{0,1\})$, ensures the support convergence (2.15) for the functions z_k.

We are now in a position to state the main result of this paper.

Theorem 3 *Under Assumption 1, let $(z_k)_k, z \in \mathrm{SBV}(\Gamma_{\mathrm{C}}; \{0,1\})$ fulfill as $k \to \infty$*

$$z_k \overset{*}{\rightharpoonup} z \text{ in } \mathrm{SBV}(\Gamma_{\mathrm{C}}; \{0,1\}) \tag{2.28}$$

and Assumption 2. Then, the functionals $\Phi_k(\cdot, z_k)$ and $\Phi_k^{\mathrm{adh}}(\cdot, z_k)$ Mosco-converge as $k \to \infty$ to $\Phi_\infty(\cdot, z)$, with respect to the topology of $H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}}; \mathbb{R}^d)$.

Its proof, carried out in Sect. 3, is based on a nontrivial adaptation of the arguments for the aforementioned Mosco-convergence result from [23].

3 Proof of Theorem 3

Let $(z_k)_k,\ z \in \mathrm{SBV}(\Gamma_{\mathrm{C}}; \{0,1\})$ fulfill the conditions of Theorem 3. In order to prove MOSCO-convergence of the functionals $\Phi_k(\cdot, z_k)$ and $\Phi_k^{\mathrm{adh}}(\cdot, z_k)$ to $\Phi_\infty(\cdot, z)$, we have to check the lim inf- and the lim sup-estimates. While the proof of the latter is more involved and will be carried out throughout Sects. 3.1 and 3.2, the argument for the former will be developed in the following lines. It relies on this key result.

Lemma 1 ([25], Lemma 4.5) *Let $z \in \mathrm{SBV}(\Gamma_{\mathrm{C}}; \{0,1\})$ and let $Z \subset \Gamma_{\mathrm{C}}$ be the associated finite perimeter set such that $z = \chi_Z$. Suppose that z fulfills the lower density estimate (2.25). Then,*

$$\mathcal{H}^{d-1}(\operatorname{supp} z \backslash Z) = 0. \tag{3.1}$$

The lim inf-Estimate

Let $(u_k),\ u \in H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}}; \mathbb{R}^d)$ fulfill $u_k \rightharpoonup u$. Since W_2 is convex and continuous on $H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}}; \mathbb{R}^d)$ and since $W_p \geq 0$ by (2.18), we have

$$\liminf_{k\to\infty} \int_{\Omega\backslash\Gamma_{\mathrm{C}}} \left(W_2(e(u_k)) + \frac{1}{k^p} W_p(e(u_k)) \right) \mathrm{d}x \geq \int_{\Omega\backslash\Gamma_{\mathrm{C}}} W_2(e(u))\, \mathrm{d}x\,.$$

We now distinguish the analysis for $\Phi_k(\cdot, z_k)$ from that for $\Phi_k^{\mathrm{adh}}(\cdot, z_k)$, cf. (2.21) & (2.22).

(i) We may of course suppose that $\sup_{k\in\mathbb{N}} \Phi_k(u_k, z_k) \leq C < \infty$. Therefore, we have

$$\sup_{k\in\mathbb{N}} \mathcal{J}_\infty([\![u_k]\!], z_k) \leq C, \quad \text{hence} \quad [\![u_k]\!] = 0\ \mathcal{H}^{d-1}\text{-a.e. on } \operatorname{supp} z_k\,.$$

Since $z_k \to z$ in $L^q(\Gamma_{\mathrm{C}})$ for every $1 \leq q < \infty$ by (2.28), and since $[\![u_k]\!] \to [\![u]\!]$ in $L^2(\Gamma_{\mathrm{C}}; \mathbb{R}^d)$ by the compact embedding $H^1(\Omega; \mathbb{R}^d) \subset L^2(\Gamma_{\mathrm{C}}; \mathbb{R}^d)$, we find a subsequence $(z_k, [\![u_k]\!])_k$ converging pointwise a.e. in Γ_{C} to $(z, [\![u]\!])$. More precisely, along this subsequence it holds $0 = z_k[\![u_k]\!] \to z[\![u]\!]$ a.e. in Γ_{C} and hence we conclude

$$z[\![u]\!] = 0\ \mathcal{H}^{d-1}\text{-a.e. on } \Gamma_{\mathrm{C}}, \text{ which implies } [\![u]\!] = 0\ \mathcal{H}^{d-1}\text{-a.e. on } \operatorname{supp} z \tag{3.2}$$

thanks to (3.1). Therefore,

$$\liminf_{k\to\infty} \mathcal{J}_\infty([\![u_k]\!], z_k) \geq 0 = \mathcal{J}_\infty([\![u]\!], z)\,,$$

which concludes the proof of the lower semicontinuity estimate.

(ii) From $\sup_k \Phi_k^{\mathrm{adh}}(u_k, z_k) \le C < \infty$ we now infer that $\sup_{k\in\mathbb{N}} \mathfrak{J}_k([\![u_k]\!], z_k) \le C$, which again yields (3.2), because of $0 \le \int_{\Gamma_\mathrm{C}} z_k |[\![u_k]\!]|^2 \,\mathrm{d}\mathcal{H}^{d-1}(x) \le C/k \to 0$. Then, also

$$\liminf_{k\to\infty} \mathfrak{J}_k([\![u_k]\!], z_k) \ge 0 = \mathfrak{J}_\infty([\![u]\!], z)\,.$$

Outline of the Proof of the lim sup-*Estimate*

Let $v \in H^1_\mathrm{D}(\Omega\backslash\Gamma_\mathrm{C};\mathbb{R}^d)$ fulfill $\Phi_\infty(v,z) < \infty$: in particular, z and v satisfy the brittle constraint (2.1). It is our task to construct a sequence $(v_k)_k$ with the following properties:

$$\begin{aligned} &v_k \in W^{1,p}_\mathrm{D}(\Omega\backslash\Gamma_\mathrm{C};\mathbb{R}^d) \text{ for all } k\in\mathbb{N},\ \sup_{k\in\mathbb{N}} \Phi_k(v_k,z_k) < \infty, \quad \text{and} \\ &v_k \to v \text{ in } H^1_\mathrm{D}(\Omega\backslash\Gamma_\mathrm{C};\mathbb{R}^d)\ \ \&\ \ \Phi_k(v_k,z_k) \to \Phi_\infty(v,z) \qquad \text{as } k\to\infty. \end{aligned} \tag{3.3}$$

Obviously, in order to improve the regularity of $v \in H^1_\mathrm{D}(\Omega\backslash\Gamma_\mathrm{C};\mathbb{R}^d)$ to $W^{1,p}_\mathrm{D}(\Omega\backslash\Gamma;\mathbb{R}^d)$ with $p > d$, v has to be mollified. For this, we will introduce a mollification operator $M^\pm_{\varepsilon_k}$, with a vanishing sequence $(\varepsilon_k)_k$, which involves the H^1-extension of $v|_{\Omega_\pm}$ from $\Omega_\pm$ to $\mathbb{R}^d$ and the convolution with a mollifier $\eta_{\varepsilon_k} \in C_0^\infty(\mathbb{R}^d)$. However, in general, the convolution of $v|_{\Omega_\pm}$ with a mollifier $\eta_k \in C_0^\infty(\mathbb{R}^d)$ will spoil its zero-trace on the Dirichlet boundary $\Gamma_\mathrm{D} \cap \overline{\Omega_\pm}$. In order to construct an element of $W^{1,p}_\mathrm{D}(\Omega\backslash\Gamma;\mathbb{R}^d)$ one has to set $v|_{\Omega_\pm}$ to zero in a sufficiently large, k-dependent neighborhood $\Gamma_\mathrm{D} + B_{r_k}(0)$ of Γ_D, before convolving with η_k. For this modification of a function $v \in H^1_\mathrm{D}(\Omega\backslash\Gamma;\mathbb{R}^d)$, leading to a function with zero values in a neighborhood of radius ρ of a closed set $M \subset \overline{\Omega}$, we will apply a suitably defined recovery operator that is a function of the radius ρ, of the points in M, and of the elements in $H^1_\mathrm{D}(\Omega\backslash\Gamma;\mathbb{R}^d)$. Namely,

$$\begin{aligned} \mathfrak{Rec} : \{\rho \in [0,\infty)\} \times M \times H^1_\mathrm{D}(\Omega\backslash\Gamma;\mathbb{R}^d) \\ \to \{\tilde v \in H^1_\mathrm{D}(\Omega\backslash\Gamma;\mathbb{R}^d) \ :\ \operatorname{supp} \tilde v \subset \Omega\backslash(M + B_\rho(0))\}; \end{aligned}$$

its definition is given in Definition 2 below. The now suitably mollified function $\tilde v_k$ given by $\tilde v_k|_{\Omega_\pm} = \tilde v_k^\pm := \eta_k * \mathfrak{Rec}(r_k, \Gamma_\mathrm{D} \cap \overline{\Omega_\pm}, v|_{\Omega_\pm}) \in W^{1,p}_\mathrm{D}(\Omega_\pm;\mathbb{R}^d)$, with a vanishing sequence $(r_k)_k$, will have to be further modified in such a way that the brittle constraint (2.1) is satisfied with the given sequence $(z_k)_k$. For this, the recovery operator $\mathfrak{Rec}$ will be once more applied to the triple $(\rho_k, \operatorname{supp} z, \tilde v_k^{\mathrm{anti}})$, where $\tilde v_k^{\mathrm{anti}}$ is the antisymmetric part of $\tilde v_k$, cf. (3.19), and

$$\rho_k := \inf\{\rho \in [0,\infty),\ \operatorname{supp} z_k \subset \operatorname{supp} z + B_{\rho_k}(0)\}\,. \tag{3.4}$$

In other words, the construction of the recovery sequence $(v_k)_k$ complying with (3.3) consists of the following three steps:

Step 1: Set $v|_{\Omega_\pm}$ to zero in $\Gamma_{\mathrm{D}} + B_{r_k}(0)$ using $\mathfrak{Rec}$, with a vanishing sequence $(r_k)_k$: this yields

$$\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^{\pm}, v|_{\Omega_\pm}), \quad \text{where } \Gamma_{\mathrm{D}}^{\pm} := \Gamma_{\mathrm{D}} \cap \overline{\Omega_\pm}. \tag{3.5a}$$

Here, the vanishing sequence $(r_k)_k$ has to be chosen in such a way that $(\Gamma_{\mathrm{D}} + B_{r_k}(0)) \cap \Gamma_{\mathrm{C}} = \emptyset$. This is possible thanks to Assumption (2.26d), which provides that $\mathrm{dist}(\Gamma_{\mathrm{D}}, \Gamma_{\mathrm{C}}) = \gamma > 0$.

Step 2: Mollify $\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^{\pm}, v|_{\Omega_\pm})$ using a suitably defined mollification operator $M_{\varepsilon_k}^{\pm} \in C_0^\infty(\mathbb{R}^d)$ for a vanishing sequence $(\varepsilon_k)_k$: this results in

$$\tilde{v}_k \in W^{1,p}(\Omega \backslash \Gamma_{\mathrm{C}}; \mathbb{R}^d) \;\text{ with }\; \tilde{v}_k^{\pm} := M_{\varepsilon_k}^{\pm}(\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^{\pm}, v|_{\Omega_\pm})). \tag{3.5b}$$

Step 3: Adapt $\tilde{v}_k$ to z_k in such a way as to obtain a sequence $(v_k)_k$ satisfying

$$z_k [\![v_k]\!] = 0 \qquad \mathcal{H}^{d-1}\text{-a.e. on } \Gamma_{\mathrm{C}} \;\text{ for each } k \in \mathbb{N}. \tag{3.6}$$

The technical tools for this construction will be provided in Sect. 3.1, whereas in Sect. 3.2 we will carry out the proof that the sequence $(v_k)_k$ indeed converges to v as stated in (3.3), cf. Theorem 4.

3.1 Preliminary Definitions and Results

We start by introducing the mollification operators. Since $\Omega_\pm \subset \mathbb{R}^d$ are Lipschitz domains, by [1, p. 91, Thm. 4.32], they are extension domains (for Sobolev functions); we introduce the linear extension operator

$E_\pm : H^1(\Omega_\pm; \mathbb{R}^d) \to H^1(\mathbb{R}^d; \mathbb{R}^d)$ with the properties:

- $\forall\, v \in H^1(\Omega_\pm; \mathbb{R}^d) : \quad E_\pm(v)(x) = v(x)$ a.e. in $\Omega_\pm$,
- $\exists C_\pm > 0\; \forall\, v \in H^1(\Omega_\pm; \mathbb{R}^d) : \quad \|E_\pm(v)\|_{H^1(\mathbb{R}^d;\mathbb{R}^d)} \leq C_\pm \|v\|_{H^1(\Omega_\pm;\mathbb{R}^d)}\,.$ (3.7)

In order to define a suitable mollification operator, we make use of the standard mollifier $\eta_1 \in C_0^\infty(\mathbb{R}^d)$, cf. e.g. [1, p. 29, 2.17],

$$\eta_1(x) := \begin{cases} \zeta \exp\big(-1/(1-|x|^2)\big) & \text{if } |x| < 1, \\ 0 & \text{if } |x| \geq 1, \end{cases} \tag{3.8a}$$

with a constant $\zeta > 0$ such that $\int_{\mathbb{R}^d} \eta_1(x)\,\mathrm{d}x = 1$, and for $\varepsilon > 0$ we set

$$\eta_\varepsilon(x) := \varepsilon^{-d}\eta_1(x/\varepsilon)\,. \tag{3.8b}$$

The Mollification Operator $M_\varepsilon^\pm$

Now, for $\varepsilon > 0$ we define the mollification operator

$$\begin{aligned} &M_\varepsilon^\pm : H^1(\Omega_\pm;\mathbb{R}^d) \to C^\infty(\Omega_\pm;\mathbb{R}^d),\\ &M_\varepsilon^\pm(v) := \eta_\varepsilon * E_\pm(v)|_{\Omega_\pm} = \int_{\mathbb{R}^d} \eta_\varepsilon(x-y)E_\pm(v)(y)\,\mathrm{d}y\bigg|_{\Omega_\pm} \end{aligned} \tag{3.9}$$

and collect its properties in the following result.

Proposition 1 (Properties of $M_\varepsilon^\pm$) *Let $p \in (1,\infty)$ fixed.*

1. *For every $\varepsilon > 0$ the linear operator $M_\varepsilon^\pm : H^1(\Omega_\pm;\mathbb{R}^d) \to H^1(\Omega_\pm;\mathbb{R}^d)$ satisfies*

$$\exists\,\overline{C} > 0\,\forall\, v \in H^1(\Omega_\pm;\mathbb{R}^d)\,: \qquad \|M_\varepsilon^\pm(v)\|_{H^1(\Omega_\pm;\mathbb{R}^d)} \le \overline{C}\|v\|_{H^1(\Omega_\pm;\mathbb{R}^d)}\,. \tag{3.10}$$

2. *Consider a sequence $\varepsilon \to 0$ and let $v \in H^1(\Omega_\pm;\mathbb{R}^d)$. Then, $M_\varepsilon^\pm v \to v$ in $H^1(\Omega_\pm;\mathbb{R}^d)$.*
3. *Let $p > d$ fixed. There is a constant $C_{d,p} > 0$, only depending on Ω, on d, and p, such that for all $v \in H^1(\Omega_\pm;\mathbb{R}^d)$*

$$\|\nabla M_\varepsilon^\pm(v)\|_{L^p(\Omega_\pm;\mathbb{R}^d)} \le \varepsilon^{-d/2}C_p\|v\|_{H^1(\Omega_\pm;\mathbb{R}^d)} \tag{3.11}$$

Proof The proof of Items 1 & 2 is a direct consequence of classical results on mollifiers for $W^{1,p}(\mathbb{R}^d)$-functions, see e.g. [5, p. 39, Lemma 1], combined with the continuity of the extension operator. Indeed, we have

$$\begin{aligned} \|M_\varepsilon^\pm(v)\|_{H^1(\Omega_\pm;\mathbb{R}^d)} &\le \|\eta_\varepsilon * E_\pm(v)\|_{H^1(\mathbb{R}^d;\mathbb{R}^d)}\\ &\le \|\eta_1\|_{L^1(\mathbb{R}^d)}\|E_\pm(v)\|_{H^1(\mathbb{R}^d;\mathbb{R}^d)} \le C_\pm\|\eta_1\|_{L^1(\mathbb{R}^d)}\|v\|_{H^1(\Omega_\pm;\mathbb{R}^d)}\,, \end{aligned}$$

whence (3.10) with $\overline{C} := \max\{C_+,\, C_-\}$, and Item 2.

Ad 3.: For the mollifiers defined in (3.8), observe that

$$\begin{aligned} &\nabla_z\eta_1(z) = \zeta\exp(-(1-|z|^2)^{-1})(-(1-|z|^2)^{-2}2z) && \text{for all } z \text{ with } |z|<1,\\ &\nabla_x\eta_\varepsilon(x) = \varepsilon^{-d}\nabla_x\left(\eta_1\left(\frac{x}{\varepsilon}\right)\right) = \varepsilon^{-(d+1)}\nabla_z\eta_1\left(\frac{x}{\varepsilon}\right) && \text{for all } x \text{ with } |x|<\varepsilon\,. \end{aligned} \tag{3.12}$$

Let $q' \geq 1$; using the transformation $(y-x)/\varepsilon = z$, $\mathrm{d}z_i = \varepsilon^{-1}\mathrm{d}y_i$ for $i \in \{1,\dots,d\}$, the $L^{q'}$-norm of $\nabla\eta_\varepsilon$ reads as follows

$$\begin{aligned}
&\|\nabla_x\eta_\varepsilon(x-\bullet)\|_{L^{q'}(\mathbb{R}^d)}\\
&=\Big(\int_{\mathbb{R}^d}\big|\nabla_x\eta_\varepsilon(x-y)\big|^{q'}\,\mathrm{d}y\Big)^{1/q'}\\
&=\Big(\int_{\mathbb{R}^d}\varepsilon^{(d-q'(d+1))}\big|\nabla_z\eta_1(z)\big|^{q'}\,\mathrm{d}z\Big)^{1/q'} = \varepsilon^{(d-q'(d+1))/q'}\|\nabla_z\eta_1\|_{L^{q'}(\mathbb{R}^d)}\,.
\end{aligned} \tag{3.13}$$

For $v \in H^1(\Omega_\pm;\mathbb{R}^d)$ the above considerations are now used to estimate $\|\nabla M_\varepsilon^\pm(v)\|_{L^p(\mathbb{R}^d;\mathbb{R}^d)}$. For this, we will in particular apply Hölder's inequality with the Sobolev exponent $q = 2d/(d-2)$, for which $\|v\|_{L^q(\Omega_\pm;\mathbb{R}^d)}$ is well-defined due to the continuous embedding $H^1(\Omega_\pm;\mathbb{R}^d) \subset L^q(\Omega_\pm;\mathbb{R}^d)$, i.e. there is $C_\mathrm{S} > 0$ such that

$$\|v\|_{L^q(\Omega_\pm;\mathbb{R}^d)} \leq C_\mathrm{S}\|v\|_{H^1(\Omega_\pm;\mathbb{R}^d)}. \tag{3.14}$$

Furthermore, note that, for $q = 2d/(d-2)$, it is $q' = q/(q-1) = 2d/(d+2)$ and hence, $\varepsilon^{(d-q'(d+1))/q'} = \varepsilon^{-d/2}$ in (3.13) above. Thus, we obtain

$$\begin{aligned}
\|\nabla_x M_\varepsilon(v)\|^p_{L^p(\Omega_\pm;\mathbb{R}^d)} &\leq \int_{\Omega_\pm}\Big(\sum_{i=1}^d\Big(\int_{\mathbb{R}^d}\big|\nabla_x\eta_\varepsilon(x-y)E_\pm(v_i)(y)\big|\,\mathrm{d}y\Big)^2\Big)^{p/2}\mathrm{d}x\\
&\leq \int_{\Omega_\pm}\Big(\sum_{i=1}^d\|\nabla_x\eta_\varepsilon(x-\bullet)\|^2_{L^{q'}(\mathbb{R}^d)}\|E_\pm(v_i)\|^2_{L^q(\mathbb{R}^d)}\Big)^{p/2}\mathrm{d}x\\
&\leq C_{d,p}\sum_{i=1}^d\int_{\Omega_\pm}\|\nabla_x\eta_\varepsilon(x-\bullet)\|^p_{L^{q'}(\mathbb{R}^d)}\|E_\pm(v_i)\|^p_{L^q(\mathbb{R}^d)}\,\mathrm{d}x\\
&\leq \varepsilon^{-dp/2}C_{d,p}\|\nabla_z\eta_1\|^p_{L^{q'}(\mathbb{R}^d)}\|v\|^p_{H^1(\Omega_\pm;\mathbb{R}^d)}\,.
\end{aligned}$$

where the positive constant $C_{d,p}$, varying from the third to the fourth line, only depends on d and p, and Ω, and for the fourth estimate we have used relation (3.13), as well as the continuity of the extension and the embedding operators, cf. (3.7) and (3.14).

The Recovery Operator $\mathfrak{Rec}$

We now introduce the recovery operator $\mathfrak{Rec}$.

Definition 2 (Recovery Operator $\mathfrak{Rec}$) Suppose that M is a closed subset of $\partial\Omega_\pm$ fulfilling property $\mathfrak{a}$ from Definition 1. Set

$$W^{1,r}_M(\Omega_\pm;\mathbb{R}^d) := \{v \in W^{1,r}(\Omega_\pm;\mathbb{R}^d),\ v = 0 \text{ on } M\},$$
$$d_M(x) := \min_{\tilde{x}\in M} |x - \tilde{x}| \quad \text{for all } x \in \overline{\Omega_\pm}.$$

Let $\rho \geq 0$. Then, for all $v \in W^{1,r}_M(\Omega_\pm;\mathbb{R}^d)$ and every $x \in \overline{\Omega_\pm}$, we define

$$\mathfrak{Rec}(\rho, M, v)(x) := v(x)\xi_\rho(x) \quad \text{with} \quad \xi_\rho(x) := \min\left\{\frac{1}{\rho}(d_M(x) - \rho)^+, 1\right\}, \tag{3.15}$$

where $(\cdot)^+$ denotes the positive part, i.e. $(z)^+ := \max\{0, z\}$.

The proof that $\mathfrak{Rec}(\rho, M, v) \to v$ in $H^1(\Omega_\pm;\mathbb{R}^d)$ is based on a Hardy-type inequality recently deduced in [8, Thm. 3.4]:

Proposition 2 (Hardy's Inequality for $r \in (1,\infty)$) *Let $\Omega_\pm$ satisfy (2.26a). Suppose that the closed set $M \subset \partial\Omega_\pm$ has Property $\mathfrak{a}$. Then, for all $r \in (1,\infty)$ there exists a constant $C_M = C(M, r)$ such that the following Hardy's inequality is fulfilled in $W^{1,r}_M(\Omega_\pm, \mathbb{R}^d)$:*

$$\forall\, v \in W^{1,r}_M(\Omega_\pm, \mathbb{R}^d): \quad \left\| v/d_M \right\|_{L^r(\Omega_\pm,\mathbb{R}^d)} \leq C_M \left\| \nabla v \right\|_{L^r(\Omega_\pm,\mathbb{R}^{d\times d})}. \tag{3.16}$$

With this Hardy's inequality at hand it is possible to deduce the following properties of $\mathfrak{Rec}$. We refer to [18, Cor. 2] for the proof of Proposition 3 below.

Proposition 3 (Properties of $\mathfrak{Rec}$) *Let the assumptions of Proposition 2 hold true. Keep $r \in (1,\infty)$ fixed. Consider a countable family $\{\rho\}$ with $\rho \to 0$ and let $v \in W^{1,r}_M(\Omega_\pm, \mathbb{R}^d)$.*

1. *There is a constant $c_r = c_r(\Omega_\pm)$ such that for every $\rho > 0$ the following estimates hold:*

$$\begin{aligned} &\|\mathfrak{Rec}(\rho, M, v)\|^r_{L^r(\Omega_\pm)} \leq \|v\|^r_{L^r(\Omega_\pm)} \quad \text{and} \\ &\|\nabla\mathfrak{Rec}(\rho, M, v)\|^r_{L^r(\Omega_\pm)} \leq c_r \|\nabla v\|^r_{L^r(\Omega_\pm)}. \end{aligned} \tag{3.17}$$

2. *$\mathfrak{Rec}(\rho, M, v) \to v$ strongly in $W^{1,r}(\Omega_\pm)$ as $\rho \to 0$.*

The bounds (3.17) will later be applied for the exponent $r = p$, whereas the strong convergence result shall be exploited for $r = 2$. As already mentioned, the recovery operator will be applied with $M = \Gamma_{\mathrm{D}}$, which is indeed required to fulfill property $\mathfrak{a}$. It will also be applied with $M = \operatorname{supp} z$, with the sequence of radii defined by (3.4). That is why, we need to impose on z the lower density estimate from Assumption 2 in Theorem 3. Assumption 2 is also at the basis of the following result, proved in

[23, Prop. 6.7, 6.8], which ensures that the sequence $(\rho_k)_k$ from (3.4) tends to 0 as $k \to \infty$.

Proposition 4 *Assume* (2.26e) *on* Γ_C. *Let* $(z_k)_k$, $z \in \mathrm{SBV}(\Gamma_C; \{0,1\})$ *fulfill* (2.28) *and Assumption* 2. *Then, for the sequence* $(\rho_k)_k$ *of radii given by* (3.4) *we have*

$$\operatorname{supp} z_k \subset \operatorname{supp} z + B_{\rho_k}(0) \quad \textit{and} \quad \rho_k \to 0 \textit{ as } k \to \infty. \tag{3.18}$$

3.2 *Construction of the Recovery Sequence and Proof of the Γ-lim sup Inequality*

We are now in a position to carry out the construction of the recovery sequence outlined at the beginning of this Section. In order to simplify the subsequent arguments, in accordance with condition (2.26e) ensuring the "flatness" of Γ_C, we suppose without loss of generality that Ω is rotated in such a way that the normal n on Γ_C points in the x_1-direction and that the origin $0 \in \Gamma_C$. Moreover, for every $x \in \Omega$ we may use the notation $x = (x_1, y)$ with $y = (x_2, \dots, x_d) \in \mathbb{R}^{d-1}$. We then define the symmetric and antisymmetric parts of a function $v = (v_{\mathrm{sym}} + v_{\mathrm{anti}}) \in H^1_D(\Omega\backslash\Gamma_C; \mathbb{R}^d)$ via

$$v_{\mathrm{sym}}(x) := \tfrac{1}{2}\big(v(x_1, y) + v(-x_1, y)\big) \quad \text{and} \quad v_{\mathrm{anti}}(x) := \tfrac{1}{2}\big(v(x_1, y) - v(-x_1, y)\big)\,. \tag{3.19}$$

In particular, $v_{\mathrm{sym}} \in H^1(\Omega, \mathbb{R}^d)$. Moreover, for $v \in H^1_D(\Omega\backslash\Gamma_C; \mathbb{R}^d)$ with $\Phi_\infty(v, z) < \infty$, there holds $v_{\mathrm{anti}} = 0$ a.e. on $\operatorname{supp} z$.

With our next result we give the precise definition of the recovery sequence and prove the Γ-lim sup inequality for the functionals Φ_k and Φ_k^{adh}.

Theorem 4 *Let Assumptions* (2.26) *be satisfied. Let* $(z_k)_k, z \in \mathrm{SBV}(\Gamma_C; \{0,1\})$ *satisfy* (2.28) *and Assumption* 2. *Let* $(\rho_k)_k$ *be defined by* (3.4). *For every* $k \in \mathbb{N}$ *set* $r_k := \frac{\gamma}{4k}$, *with* $\gamma = \mathrm{dist}(\Gamma_D, \Gamma_C)$, *and consider* M_{ε_k} *from* (3.9) *with* $\varepsilon_k := k^{-\alpha}$ *for* $\alpha \in (0, 2/d)$. *Then, for* $v \in H^1_D(\Omega\backslash\Gamma_C; \mathbb{R}^d)$ *with* $\Phi_\infty(v, z) < \infty$, *set*

$$v_k := \tilde{v}_k^{\mathrm{sym}} + \mathfrak{Rec}(\rho_k, \operatorname{supp} z, \tilde{v}_k^{\mathrm{anti}})\,, \tag{3.20}$$

with $\tilde{v}_k$ *from* (3.5), $(\rho_k)_k$ *from* (3.4), *and the recovery operator* $\mathfrak{Rec}$ *from* (3.15). *Then, for the functionals from* (2.21)–(2.23) *there holds*

$$\lim_{k\to\infty} \Phi_k(v_k, z_k) = \Phi_\infty(v, z) \quad \textit{and} \quad \lim_{k\to\infty} \Phi_k^{\mathrm{adh}}(v_k, z_k) = \Phi_\infty(v, z)\,. \tag{3.21}$$

Proof First of all, recall that both $M_\varepsilon^\pm$ from (3.9) and $\mathfrak{Rec}(\rho, M, \cdot)$ from (3.15) are linear operators. Hence, in (3.20) we have

$$v_k^\pm = M_{\varepsilon_k}^\pm(\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^\pm, v_{\mathrm{sym}}|_{\Omega_\pm})) + \mathfrak{Rec}(\rho_k, \operatorname{supp} z, M_{\varepsilon_k}^\pm(\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^\pm, v_{\mathrm{anti}}|_{\Omega_\pm}))) \,. \tag{3.22}$$

With $\tilde{v} \in H^1(\Omega_\pm, \mathbb{R}^d)$ as a placeholder for $u_{\mathrm{sym}}|_{\Omega_\pm}$, resp. $u_{\mathrm{anti}}|_{\Omega_\pm}$, and using (3.10), we deduce

$$\begin{aligned}
&\|M_{\varepsilon_k}^\pm(\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^\pm, \tilde{v})) - \tilde{v}\|_{H^1(\Omega_\pm)} \\
&\le \|M_{\varepsilon_k}^\pm(\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^\pm, \tilde{v})) - M_{\varepsilon_k}^\pm(\tilde{v})\|_{H^1(\Omega_\pm)} + \|M_{\varepsilon_k}^\pm(\tilde{v}) - \tilde{v}\|_{H^1(\Omega_\pm)} \\
&\le \overline{C}\|\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^\pm, \tilde{v}) - \tilde{v}\|_{H^1(\Omega_\pm)} + \|M_{\varepsilon_k}^\pm(\tilde{v}) - \tilde{v}\|_{H^1(\Omega_\pm)} \to 0\,,
\end{aligned} \tag{3.23}$$

and both terms on the right-hand side tend to 0 according to Propositions 1 & 3, since both sequences $(\varepsilon_k)_k$ and $(r_k)_k$ are null and since $r_k = \gamma/(4k) < \operatorname{dist}(\Gamma_{\mathrm{D}}, \Gamma_{\mathrm{C}})$ by assumption. Furthermore, thanks to (3.11), the L^p-norm of the gradient can be estimated as follows

$$\|\nabla M_{\varepsilon_k}^\pm(\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^\pm, \tilde{v}))\|_{L^p(\Omega_\pm)} \le \varepsilon_k^{-d/2} C_{d,p} \|\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^\pm, \tilde{v})\|_{H^1(\Omega_\pm)} \le \varepsilon_k^{-d/2} C\,. \tag{3.24}$$

Estimate (3.23) implies that

$$M_{\varepsilon_k}^\pm(\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^\pm, v_{\mathrm{sym}}|_{\Omega_\pm})) \to v_{\mathrm{sym}}|_{\Omega_\pm} \quad \text{strongly in } H^1(\Omega_\pm, \mathbb{R}^d). \tag{3.25}$$

Moreover, by estimate (3.24) we conclude that

$$k^{-p}\|\nabla M_{\varepsilon_k}^\pm(\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^\pm, \tilde{v}))\|_{L^p(\Omega_\pm)} \le k^{-p} \varepsilon_k^{-dp/2} C^p \to 0 \quad \text{as } k \to \infty, \tag{3.26}$$

due to $\varepsilon_k = k^{-\alpha}$ with $\alpha \in (0, 2/d)$.

It remains to verify similar relations for the term involving $v_{\mathrm{anti}}|_{\Omega_\pm}$, again abbreviated with $\tilde{v}$. With the aid of (3.17) and the linearity of $\mathfrak{Rec}$, we obtain

$$\begin{aligned}
&\|\mathfrak{Rec}(\rho_k, \operatorname{supp} z, M_{\varepsilon_k}^\pm(\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^\pm, \tilde{v}))) - \tilde{v}\|_{H^1(\Omega_\pm)} \\
&\le \|\mathfrak{Rec}(\rho_k, \operatorname{supp} z, M_{\varepsilon_k}^\pm(\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^\pm, \tilde{v}))) - \mathfrak{Rec}(\rho_k, \operatorname{supp} z, \tilde{v})\|_{H^1(\Omega_\pm)} \\
&\quad + \|\mathfrak{Rec}(\rho_k, \operatorname{supp} z, \tilde{v}) - \tilde{v}\|_{H^1(\Omega_\pm)} \\
&\le C\|M_{\varepsilon_k}^\pm(\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^\pm, \tilde{v})) - \tilde{v}\|_{H^1(\Omega_\pm)} + \|\mathfrak{Rec}(\rho_k, \operatorname{supp} z, \tilde{v}) - \tilde{v}\|_{H^1(\Omega_\pm)} \to 0
\end{aligned} \tag{3.27}$$

by (3.23) and Proposition 3. In order to deduce an estimate for the L^p-norm of the gradient we rewrite $\mathfrak{Rec}(\rho_k, \operatorname{supp} z, M^\pm_{\varepsilon_k}(\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^\pm, \tilde{v}))) = \xi_{\rho_k}^{\operatorname{supp} z} M^\pm_{\varepsilon_k}(\xi_{r_k}^{\Gamma_{\mathrm{D}}^\pm} v)$ with the aid of (3.15), and hence find that $\nabla\mathfrak{Rec}(\rho_k, \operatorname{supp} z, M^\pm_{\varepsilon_k}(\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^\pm, \tilde{v})))$ $= \xi_{\rho_k}^{\operatorname{supp} z}\nabla M^\pm_{\varepsilon_k}(\xi_{r_k}^{\Gamma_{\mathrm{D}}^\pm}\tilde{v}) + M^\pm_{\varepsilon_k}(\xi_{r_k}^{\Gamma_{\mathrm{D}}^\pm}\tilde{v}) \otimes \nabla\xi_{\rho_k}^{\operatorname{supp} z}$. Thus, by (3.17) and (3.11) it is

$$\begin{aligned} &\|\nabla\mathfrak{Rec}(\rho_k, \operatorname{supp} z, M^\pm_{\varepsilon_k}(\mathfrak{Rec}(r_k, \Gamma_{\mathrm{D}}^\pm, \tilde{v})))\|_{L^p(\Omega_\pm)} \\ &\leq \|\nabla M^\pm_{\varepsilon_k}(\xi_{r_k}^{\Gamma_{\mathrm{D}}^\pm}\tilde{v})\|_{L^p(\Omega_\pm)} + \|M^\pm_{\varepsilon_k}(\xi_{r_k}^{\Gamma_{\mathrm{D}}^\pm}\tilde{v}) \otimes \nabla\xi_{\rho_k}^{\operatorname{supp} z}\|_{L^p(\Omega_\pm)} \\ &\leq \varepsilon_k^{-d/2}(C_{d,p} + C)\|\xi_{r_k}^{\Gamma_{\mathrm{D}}^\pm}\tilde{v}\|_{H^1(\Omega_\pm)} \leq \varepsilon_k^{-d/2}C'\,. \end{aligned} \tag{3.28}$$

Let us now conclude the proof of (3.21). It follows from (3.23) and (3.27) that $v_k \to v$ as $k \to \infty$ strongly in $H^1(\Omega\backslash\Gamma_{\mathrm{C}}, \mathbb{R}^d)$. Hence we can choose a (not relabeled) subsequence that converges pointwise a.e. in $\Omega\backslash\Gamma_{\mathrm{C}}$. Then, for the quadratic part W_2 of the elastic energy we easily conclude that

$$\int_{\Omega\backslash\Gamma_{\mathrm{C}}} W_2(e(v_k))\,\mathrm{d}x \to \int_{\Omega\backslash\Gamma_{\mathrm{C}}} W_2(e(v))\,\mathrm{d}x \tag{3.29}$$

via the the dominated convergence theorem. As for the term $k^{-p}W_p$, we have that

$$\int_{\Omega\backslash\Gamma_{\mathrm{C}}} k^{-p}W_p(e(v_k))\,\mathrm{d}x \to 0\,. \tag{3.30}$$

due to growth property of W_p in combination with estimates (3.24) & (3.28). Finally, there holds

$$z_k[\![v_k]\!] = z_k[\![\tilde{v}_k^{\mathrm{sym}}]\!] + z_k[\![\mathfrak{Rec}(\rho_k, \operatorname{supp} z, \tilde{v}_k^{\mathrm{anti}})]\!] = 0 \quad \mathcal{H}^{d-1}\text{-a.e. on } \Gamma_{\mathrm{C}}, \tag{3.31}$$

since for the symmetric part we have $[\![\tilde{v}_k^{\mathrm{sym}}]\!] = 0$ a.e. on Γ_{C}, while, by construction, $[\![\mathfrak{Rec}(\rho_k, \operatorname{supp} z, \tilde{v}_k^{\mathrm{anti}})]\!] = 0$ on $\operatorname{supp} z + B_{\rho_k}$ which contains $\operatorname{supp} z_k$, cf. (3.18). Since the functions z_k fulfill the lower density estimate from Assumption 2, Lemma 1 is applicable. Therefore, from (3.31) we infer that $[\![v_k]\!] = 0$ a.e. on $\operatorname{supp} z_k$, i.e. that

$$\text{both} \quad \mathcal{J}_\infty([\![v_k]\!], z_k) = 0 \quad \text{and}\, \mathcal{J}_k([\![v_k]\!], z_k) = 0 \quad \text{for every } k \in \mathbb{N}. \tag{3.32}$$

From (3.29), (3.30), and (3.32) we conclude (3.21) and thus complete the proof.

4 Applications

4.1 *From Nonlinear to Linear Elasticity in the Brittle Delamination System*

Let us now address the limit passage from nonlinear to linear (small-strain) elasticity in the coupled rate-dependent/independent system for brittle delamination consisting of

1. the mechanical force balance for the displacements (2.3), with the stored elastic energy density $W(e) = W_2(e) + \frac{1}{k^p} W_p(e)$, where we let $k \to \infty$;
2. the contact boundary condition (2.4);
3. the brittle delamination flow rule (2.7).

Due to the rate-independent character of the flow rule, which possibly leads to jump discontinuities of z as a function of time, system (2.3, 2.4, 2.7) has to be weakly formulated. As already mentioned in Sect. 2, for this we resort to the notion of *semistable energetic solution* for coupled rate-dependent/independent systems, first proposed in [19] for rate-independent processes in viscous solids, and recently extended and generalized in [24]. We now recall this definition in the context of

- the *nonlinearly elastic* brittle delamination system, i.e. (2.3, 2.4, 2.7) with $W(e) = W_2(e) + \frac{1}{k^p} W_p(e)$;
- the *linearly elastic* brittle delamination system, i.e. (2.3, 2.4, 2.7) with $W(e) = W_2(e)$,

where, of course, the terms 'nonlinearly elastic' and 'linearly elastic' have been used with slight abuse, only to refer to the nonlinear/linear character of the equation for the displacements (at small strains).

Prior to giving Definition 3, we need to fix our conditions on the forces F and f: we assume that $F \in W^{1,1}(0,T; H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d)^*)$ and $f \in W^{1,1}(0,T; L^{2(d-1)/d}(\Gamma_{\mathrm{N}};\mathbb{R}^d))$, so that the total loading L defined by (2.9) fulfills

$$L \in W^{1,1}(0,T; H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d)^*)\,. \tag{4.1}$$

We then introduce the energy functionals driving the nonlinearly and linearly elastic systems, respectively:

$$\begin{aligned}
&\mathcal{E}_k,\ \mathcal{E}_\infty : [0,T] \times H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d) \times \mathrm{SBV}(\Gamma_{\mathrm{C}};\{0,1\}), \to (-\infty,\infty],\\
&\quad \mathcal{E}_k(t,u,z) := \Phi_k(u,z) + \mathcal{G}(z) - \langle L(t), z\rangle_{H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d)},\\
&\quad \mathcal{E}_\infty(t,u,z) := \Phi_\infty(u,z) + \mathcal{G}(z) - \langle L(t), z\rangle_{H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d)},
\end{aligned} \tag{4.2}$$

with $\mathcal{G}$ defined by (2.16). Finally, we consider the dissipation potential

$$\mathcal{R} : L^1(\Gamma_\mathrm{C}) \to [0,\infty], \qquad \mathcal{R}(\dot z) := \int_{\Gamma_\mathrm{C}} \mathrm{R}(\dot z)\,\mathrm{d}x\,, \quad \text{with } \mathrm{R}(v) := \begin{cases} a_1|v| & \text{if } v \le 0, \\ \infty & \text{otherwise} \end{cases} \tag{4.3}$$

and $a_1 > 0$. The fact that $\mathrm{R}(v) = \infty$ if $v > 0$ ensures the unidirectionality of the delamination process, i.e. a crack can only increase or stagnate but its healing is excluded. With $\mathcal{R}$ we associate the total variation functional

$$\mathrm{Var}_\mathcal{R}(z;[s,t]) := \sup\left\{ \sum_{j=1}^N \mathcal{R}(z(r_j)-z(r_{j-1})) \ : \quad s = r_0 < r_1 < \ldots < r_{N-1} < r_N = t \right\}.$$

for all $[s,t] \subset [0,T]$. Observe that the unidirectionality encoded in $\mathcal{R}$ provides monotonicity with respect to time of functions z with $\mathrm{Var}_\mathcal{R}(z;[s,t]) < \infty$. Hence, $\mathrm{Var}_\mathcal{R}(z;[s,t]) = \mathcal{R}(z(t)-z(s))$ in this case.

We are now in a position to give the following

Definition 3 We say that a pair (u,z), with $u : [0,T] \to W^{1,p}_\mathrm{D}(\Omega\backslash\Gamma_\mathrm{C};\mathbb{R}^d)$ in the nonlinear case and $u : [0,T] \to H^1_\mathrm{D}(\Omega\backslash\Gamma_\mathrm{C};\mathbb{R}^d)$ for the linear case, and $z : [0,T] \to \mathrm{SBV}(\Gamma_\mathrm{C};\{0,1\})$, is a *semistable energetic* solution of the nonlinearly/linearly elastic brittle delamination system, if

$$u \in H^1(0,T;H^1_\mathrm{D}(\Omega\backslash\Gamma_\mathrm{C};\mathbb{R}^d)) \cap \begin{cases} L^\infty(0,T;W^{1,p}_\mathrm{D}(\Omega\backslash\Gamma_\mathrm{C};\mathbb{R}^d)) & \text{in the nonlinear case,} \\ L^\infty(0,T;H^1_\mathrm{D}(\Omega\backslash\Gamma_\mathrm{C};\mathbb{R}^d)) & \text{in the linear case,} \end{cases}$$

$$z \in L^\infty(0,T;\mathrm{SBV}(\Gamma_\mathrm{C};\{0,1\})) \cap \mathrm{BV}([0,T];L^1(\Gamma_\mathrm{C})),$$

the pair (u,z) fulfills

- the weak formulation (2.8) of the mechanical force balance, with $q = p > d$ for the nonlinear case and $q = 2$ for the linear one;
- the semistability condition

$$\mathcal{E}_k(t,u(t),z(t)) \le \mathcal{E}_k(t,u(t),\tilde z) + \mathcal{R}(\tilde z - z(t)) \text{ for all } \tilde z \in L^1(\Gamma_\mathrm{C}) \text{ and all } t \in [0,T], \tag{4.4}$$

- the energy-dissipation inequality for all $t \in [0,T]$

$$\begin{aligned} &\mathrm{Var}_\mathcal{R}(z;[0,t]) + \int_0^t \mathbb{D}e(\dot u) : e(\dot u)\,\mathrm{d}x + \mathcal{E}_k(t,u(t),z(t)) \\ &\le \mathcal{E}_k(0,u(0),z(0)) + \int_0^t \partial_t\mathcal{E}_k(r,u(r),z(r))\,\mathrm{d}r, \end{aligned} \tag{4.5}$$

with $k \in \mathbb{N}$ ($k = \infty$) for the nonlinearly (linearly, respectively) elastic system.

Note that the existence of semistable energetic solutions to the nonlinearly elastic brittle system was proved in [23].

The following result formalizes the limit passage from nonlinear to linear elasticity for semistable energetic solutions of the brittle delamination system. For technical reasons that will be expounded in the proof, we need to strengthen our Assumption 1 on the domain, by requiring in addition that $\varGamma_{\mathrm{C}}$ is *convex*.

Theorem 5 *Under Assumption 1 suppose, in addition, that* $\varGamma_{\mathrm{C}}$ *is convex. Let* $(u_0^k, z_0^k)_k \subset W_{\mathrm{D}}^{1,p}(\varOmega\backslash\varGamma_{\mathrm{C}};\mathbb{R}^d) \times \mathrm{SBV}(\varGamma_{\mathrm{C}};\mathbb{R}^d)$ *be a sequence of data for the nonlinearly elastic brittle systems, and suppose that*

$$\begin{aligned} &(u_0^k, z_0^k) \to (u_0, z_0) \quad \textit{in } H_{\mathrm{D}}^1(\varOmega\backslash\varGamma_{\mathrm{C}};\mathbb{R}^d) \times \mathrm{SBV}(\varGamma_{\mathrm{C}};\mathbb{R}^d) \textit{ with} \\ &\mathcal{E}_k(u_0^k, z_0^k) \to \mathcal{E}_\infty(u_0, z_0) \quad \textit{as } k \to \infty. \end{aligned} \tag{4.6a}$$

Also, suppose that (u_0, z_0) *fulfill the semistability condition at* $t = 0$*, vit.*

$$\mathcal{E}(0, u_0, z_0) \le \mathcal{E}(0, u_0, \tilde{z}) + \mathcal{R}(\tilde{z}-z_0) \quad \textit{for all } \tilde{z} \in L^1(\varGamma_{\mathrm{C}}). \tag{4.6b}$$

Let $(u_k, z_k)_k$ *be a sequence of semistable energetic solutions of the nonlinearly elastic brittle system emanating from the initial data* $(u_0^k, z_0^k)_k$*. Then, there exist a (not relabeled) subsequence and functions* $u \in H^1(0, T; H_{\mathrm{D}}^1(\varOmega\backslash\varGamma_{\mathrm{C}};\mathbb{R}^d))$ *and* $z \in L^\infty(0, T; \mathrm{SBV}(\varGamma_{\mathrm{C}};\{0, 1\})) \cap \mathrm{BV}([0, T]; L^1(\varGamma_{\mathrm{C}}))$ *such that, as* $k \to \infty$*,*

$$\begin{aligned} &u_k \rightharpoonup u && \textit{in } H^1(0, T; H_{\mathrm{D}}^1(\varOmega\backslash\varGamma_{\mathrm{C}};\mathbb{R}^d)), \\ &u_k(t) \rightharpoonup u(t) && \textit{in } H_{\mathrm{D}}^1(\varOmega\backslash\varGamma_{\mathrm{C}};\mathbb{R}^d) \quad \textit{for all } t \in [0, T], \\ &z_k \overset{*}{\rightharpoonup} z && \textit{in } L^\infty(0, T; \mathrm{SBV}(\varGamma_{\mathrm{C}};\{0, 1\})) \cap L^\infty((0, T) \times \varGamma_{\mathrm{C}}), \\ &z_k(t) \overset{*}{\rightharpoonup} z(t) && \textit{in } \mathrm{SBV}(\varGamma_{\mathrm{C}};\{0, 1\}) \cap L^\infty(\varGamma_{\mathrm{C}}) \quad \textit{for all } t \in [0, T], \end{aligned} \tag{4.7}$$

$u(0) = u_0$, $z(0) = z_0$*, and the pair* (u, z) *is a semistable energetic solution of the linearly elastic brittle system in the sense of Definition 3.*

Remark 2 (Alternative Scaling & Energy-Dissipation Balance) In [23, 25, 26] also an alternative scaling for certain energy contributions was investigated. More, precisely, we replaced the perimeter regularization $\mathcal{G}$ in (4.2) and dissipation potential $\mathcal{R}$ in (4.3) by their scaled versions

$$\mathcal{G}_k(z) := \tfrac{1}{k}\mathcal{G}(z) \quad \text{and} \quad \mathcal{R}_k(v) := \tfrac{1}{k}\mathcal{R}(v)\,. \tag{4.8}$$

In [26] this was shown to be beneficial for modeling the onset of rupture when performing the adhesive contact approximation of brittle delamination. Still, the associated semistability inequality yielded compactness for the perimeters and the dissipation terms of the approximate solutions, as can be verified by a multiplication with a factor k. The uniform bound on the perimeters independent of k thus entailed that $\mathcal{G}_k(z_k(t)) \to 0$ along semistable energetic solutions as $k \to \infty$. Thus, given that the initial data are well-prepared, it was possible in [26] to deduce an energy-dissipation balance for the limit system. A similar result is also expected if the scaling (4.8) is applied in the setup presented in Theorem 5.

Proof (Sketch of the Proof of Theorem 5) We will not develop the proof in its completeness but rather highlight its main ingredients, focusing in particular on the limit passage in the mechanical force balance for the displacements. We will often refer to [23] for all details. We now split the proof into five steps.

Step 0: A Priori Estimates and Compactness
Exploiting regularity assumption (4.1), which allows us to estimate the work of the external loadings, as well as the information that $\sup_{k\in\mathbb{N}} \mathcal{E}_k(u_0^k, z_0^k) \leq C < \infty$, from the energy-dissipation inequality for the nonlinearly elastic case (i.e. $k \in \mathbb{N}$), written on the interval $[0, T]$, we deduce that

$$\begin{aligned}&\exists\, C > 0\ \forall\, k \in \mathbb{N}\ :\\ &\mathrm{Var}_{\mathcal{R}}(z_k; [0, t]) + \int_0^t \mathbb{D}e(\dot{u}_k) : e(\dot{u}_k)\,\mathrm{d}x + \sup_{t\in[0,T]} |\mathcal{E}_k(t, u_k(t), z_k(t))| \leq C\,.\end{aligned} \tag{4.9}$$

This yields the uniform bounds

$$\sup_{k\in\mathbb{N}} \Big(\|u_k\|_{H^1(0,T;H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d))} + \|z_k\|_{L^\infty(0,T;\mathrm{SBV}(\Gamma_{\mathrm{C}};\{0,1\}))\cap\mathrm{BV}([0,T];L^1(\Gamma_{\mathrm{C}}))} \Big) \leq C,$$

also by exploiting Korn's inequality for the displacements. Then, standard compactness arguments imply convergences (4.7), cf. the proof of [23, Thm. 4.3], which in particular give $u(0) = u_0$, $z(0) = z_0$. It also follows from (4.7), via standard lower semicontinuity arguments, that

$$\liminf_{k\to\infty} \mathcal{E}_k(t, u_k(t), z_k(t)) \geq \mathcal{E}_\infty(t, u(t), z(t)) \quad \text{for every } t \in [0, T]. \tag{4.10}$$

Step 1: Fine Properties of the Semistable Sequence $(z_k)_k$
Exploiting the additional condition that Γ_{C} is convex, in [23, Thm. 6.6] it was proved that the semistability condition (4.4) guarantees the validity of the lower density estimate (2.27) for every $k \in \mathbb{N} \cup \{\infty\}$, with constants uniform w.r.t. $k \in \mathbb{N} \cup \{\infty\}$. Therefore, the sequence $(z_k)_k$ fulfills Assumption 2 of Theorem 3.

Step 2: Limit Passage in the Mechanical Force Balance for the Displacements
We apply Theorem 3 and conclude the Mosco-convergence of the functionals $\Phi_k(\cdot, z_k)$ to $\Phi(\cdot, z)$ w.r.t. the topology of $H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}}; \mathbb{R}^d))$. Then, in order to pass to the limit in the mechanical force balance (2.8) as $k \to \infty$, we easily adapt the arguments from the proof of [23, Prop. 5.6]. They are based on the fact

that, for $k \in \mathbb{N}$, the weak formulation (2.8) can be reformulated in terms of the subdifferential (in the sense of convex analysis) of Φ_k w.r.t. the variable u, namely $\partial_u \Phi_k : H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d)) \times \mathrm{SBV}(\Gamma_{\mathrm{C}};\{0,1\}) \rightrightarrows H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d))^*$ given by

$$\xi \in \partial_u \Phi_k(u,z) \text{ if and only if } u \in W^{1,p}_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d) \text{ and}$$

$$\langle \xi, v\rangle_{W^{1,p}_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d)} = \int_{\Omega\backslash\Gamma_{\mathrm{C}}} \big(\mathrm{D}W_2(e(u)) + k^{-p}\mathrm{D}W_p(e(u))\big) : e(v)\,\mathrm{d}x + \langle \lambda, v\rangle_{H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d)}$$

for all $v \in W^{1,p}_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d)$, with λ an element of the subdifferential $\partial_u(\mathcal{I}_C + \mathcal{J}_k(\cdot,z)) : H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d) \rightrightarrows H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d)^*$. Then, the nonlinearly elastic version of the mechanical force balance (2.8) is equivalent to

$$\begin{aligned} &\int_{\Omega\backslash\Gamma_{\mathrm{C}}} \big(\mathbb{D}\dot{e}(t) + \mathrm{D}W_2(e(u)) + k^{-p}\mathrm{D}W_p(e(u))\big) : e(v)\,\mathrm{d}x + \langle \lambda(t), v\rangle_{H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d)} \\ &= \langle L(t), v\rangle_{H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d)} \end{aligned} \tag{4.11}$$

for all $v \in W^{1,p}_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d)$, with $\lambda(t)$ a selection in $\partial_u(\mathcal{I}_C + \mathcal{J}_k(\cdot, z(t)))(u(t))$. Analogously, in the linearly elastic case (2.8) reformulates in terms of the subdifferential $\partial_u \Phi_\infty : H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d) \times \mathrm{SBV}(\Gamma_{\mathrm{C}};\{0,1\}) \rightrightarrows H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d)^*$. Now, the MOSCO-convergence of the functionals $\Phi_k(\cdot, z_k)$ to $\Phi_\infty(\cdot, z)$ guarantees the *convergence in the sense of graphs* of the corresponding subdifferentials $\partial_u \Phi_k(\cdot, z_k)$ to $\partial_u \Phi_\infty(\cdot, z)$, cf. [3]. This is the key observation for passing to the limit in (4.11), arguing in the very same way as for [23, Prop. 5.6]. These arguments also yield, as a by-product, that

$$u_k(t) \to u(t) \text{ in } H^1_{\mathrm{D}}(\Omega\backslash\Gamma_{\mathrm{C}};\mathbb{R}^d) \;\text{ and }\; k^{-p}\int_{\Omega\backslash\Gamma_{\mathrm{C}}} W_p(e(u_k(t))\,\mathrm{d}x \to 0 \text{ as } k \to \infty$$

for almost all $t \in (0,T)$, hence

$$\Phi_k(u_k(t), z_k(t)) \to \Phi_\infty(u(t), z(t)) \qquad \text{as } k \to \infty \quad \text{for a.a. } t \in (0,T). \tag{4.12}$$

Step 3*: Limit Passage in the Semistability Condition*

First of all, observe that, for $k \in \mathbb{N} \cup \{\infty\}$ condition (4.4) reduces to

$$\begin{aligned} \mathcal{J}_\infty([\![u_k(t)]\!], z_k(t)) + \mathcal{G}(z_k(t)) \le \mathcal{J}_\infty([\![u_k(t)]\!], \tilde{z}) + \mathcal{G}(\tilde{z}) + \mathcal{R}(\tilde{z} - z_k(t)) \\ \text{for all } \tilde{z} \in L^1(\Gamma_{\mathrm{C}}) \text{ and for all } t \in [0,T]. \end{aligned} \tag{4.13}$$

We now aim to pass to the limit as $k \to \infty$ in (4.13) for every $t \in (0, T]$ (the semistability condition holds at $t = 0$ thanks to (4.6b)) and show that the functions (u, z) fulfill it for $k = \infty$. Following a well-consolidated procedure for energetic solutions to purely rate-independent systems (cf. [17]), for $t \in (0, T]$ fixed and given $\tilde{z} \in L^1(\Gamma_{\mathrm{C}})$ such that $\mathcal{R}(\tilde{z} - z(t)) < \infty$ and $\mathcal{J}_\infty([\![u(t)]\!], \tilde{z}) + \mathcal{G}(\tilde{z}) < \infty$ (otherwise (4.13) trivially holds), we exhibit a recovery sequence $(\tilde{z}_k)_k$, suitably converging to $\tilde{z}$ and fulfilling

$$\begin{aligned}&\limsup_{k\to\infty} \big(\mathcal{J}_\infty([\![u_k(t)]\!], \tilde{z}_k) + \mathcal{G}(\tilde{z}_k) + \mathcal{R}(\tilde{z}_k - z_k(t)) - \mathcal{J}_\infty([\![u_k(t)]\!], z_k(t)) - \mathcal{G}(z_k(t))\big)\\ &\leq \mathcal{J}_\infty([\![u(t)]\!], \tilde{z}) + \mathcal{G}(\tilde{z}) + \mathcal{R}(\tilde{z} - z(t)) - \mathcal{J}_\infty([\![u_k(t)]\!], z) - \mathcal{G}(z(t)) .\end{aligned} \tag{4.14}$$

For this, we borrow the construction from the proof of [23, Prop. 5.9] and set

$$\tilde{z}_k := \tilde{z}\chi_{A_k} + z_k(1-\chi_{A_k}) \quad \text{with } A_k := \{x \in \Gamma_{\mathrm{C}} \,:\, 0 \leq \tilde{z}(x) \leq z_k(x)\}$$

and χ_{A_k} its characteristic function. Observe that $0 \leq \tilde{z}_k \leq z_k$ a.e. on Γ_{C} by construction, therefore from $\sup_{k\in\mathbb{N}} \sup_{t\in(0,T)} \mathcal{J}_\infty([\![u_k(t)]\!], z_k(t)) = 0$ due to (4.9) we gather that $\mathcal{J}_\infty([\![u_k(t)]\!], \tilde{z}_k) = 0$ for all $k \in \mathbb{N}$. Therefore,

$$\limsup_{k\to\infty} \big(\mathcal{J}_\infty([\![u_k(t)]\!], \tilde{z}_k) - \mathcal{J}_\infty([\![u_k(t)]\!], z_k(t))\big) = 0 = \mathcal{J}_\infty([\![u(t)]\!], \tilde{z}) - \mathcal{J}_\infty([\![u(t)]\!], z(t)) .$$

We refer to the proof of [23, Prop. 5.9] for the calculations on the remaining contributions to (4.14).

Step 4: Proof of the Energy-Dissipation Inequality (4.5)
It follows by taking the $\liminf_{k\to\infty}$ of (4.5) for the nonlinearly elastic brittle system. For the left-hand side, we rely on convergences (4.7), the lower semicontinuity properties of the dissipative contributions to (4.5), and (4.10). For the right-hand side, we resort to the energy convergence (4.6a) for the initial data and to the continuity properties of the power term $\partial_t \mathcal{E}$, in view of (4.1).

This concludes the proof of Theorem 5.

4.2 The Joint Discrete-to-Continuous and Adhesive-to-Brittle Limit in the Mechanical Force Balance of the Thermoviscoelastic System

In this final section we shortly discuss how the MOSCO-convergence statement of Theorem 3 concerning the functionals $(\Phi_k^{\mathrm{adh}})_k$ from (2.22) can be used to prove the existence of solutions for a model for brittle delamination, also encompassing thermal effects. More precisely, the evolution of the displacement u, of the delamination variable z, and of the absolute temperature ϑ is governed by the following PDE system:

$$-\operatorname{div}\sigma(e,\dot e,\vartheta)=F \qquad \text{in } (0,T)\times(\Omega_+\cup\Omega_-), \tag{4.15a}$$

$$\dot\vartheta-\operatorname{div}\big(\mathbb{K}(e,\vartheta)\nabla\vartheta\big)=\dot e{:}\mathbb{D}{:}\dot e-\vartheta\mathbb{B}{:}\dot e+G \qquad \text{in } (0,T)\times(\Omega_+\cup\Omega_-), \tag{4.15b}$$

$$u=0 \qquad \text{on } (0,T)\times\Gamma_{\mathrm{D}}, \tag{4.15c}$$

$$\sigma(e,\dot e,\vartheta)\big|_{\Gamma_{\mathrm{N}}}\mathbf{n}=f \qquad \text{on } (0,T)\times\Gamma_{\mathrm{N}}, \tag{4.15d}$$

$$(\mathbb{K}(e,\theta)\nabla\theta)\mathbf{n}=g \qquad \text{on } (0,T)\times\partial\Omega, \tag{4.15e}$$

$$\sigma(e,\dot e,\vartheta)|_{\Gamma_{\mathrm{C}}}\mathbf{n}+\partial_u\widetilde{J}_\infty([\![u]\!],z)+\partial I_{C(x)}([\![u]\!])\ni 0 \qquad \text{on } (0,T)\times\Gamma_{\mathrm{C}}, \tag{4.15f}$$

$$\partial\mathrm{R}(\dot z)+\partial\mathcal{G}(z)+\partial_z\widetilde{J}_\infty([\![u]\!],z)\ni 0 \qquad \text{on } (0,T)\times\Gamma_{\mathrm{C}}, \tag{4.15g}$$

$$\tfrac12\big(\mathbb{K}(e,\vartheta)\nabla\vartheta|^+_{\Gamma_{\mathrm{C}}}+\mathbb{K}(e,\vartheta)\nabla\vartheta|^-_{\Gamma_{\mathrm{C}}}\big)\cdot\mathbf{n}+\eta([\![u]\!],z)[\![\vartheta]\!]=0 \qquad \text{on } (0,T)\times\Gamma_{\mathrm{C}}, \tag{4.15h}$$

$$[\![\mathbb{K}(e,\vartheta)\nabla\vartheta]\!]\cdot\mathbf{n}=-a_1\dot z \qquad \text{on } (0,T)\times\Gamma_{\mathrm{C}}. \tag{4.15i}$$

Here, the stress tensor σ encompasses both Kelvin-Voigt rheology and thermal expansion in a *linearly elastic* way, i.e.

$$\sigma(e,\dot e,\vartheta)=\mathbb{D}\dot e+\mathrm{D}W_2(e)-\theta\mathbb{B}. \tag{4.16}$$

The heat equation (4.15b), featuring the positive definite matrix of heat conduction coefficients $\mathbb{K}(e,\vartheta)$ and the positive heat source G, is complemented by the two boundary conditions (4.15h) and (4.15i) (with $g\ge 0$ another external heat source on the boundary $\partial\Omega$), which balance the heat transfer across Γ_{C} with the ongoing crack growth. In particular, the function η is a heat-transfer coefficient, determining the heat convection through Γ_{C}, which depends on the state of the bonding and on the distance between the crack lips.

In [23] we proved the existence of semistable energetic solutions (with the heat equation formulated in a suitably weak way) for system (4.15) in the *nonlinearly elastic* (small-strain) case, i.e. with $\sigma(e,\dot e,\vartheta)=\mathbb{D}\dot e+\mathrm{D}W_p(e)-\theta\mathbb{B}$ and $p>d$. As explained in Sect. 2, the latter constraint can be now overcome. Nonetheless, in order to show the existence of solutions to system (4.15) with (4.16), it is necessary to resort to a *nonlinear* approximation of the mechanical force for the displacements.

In fact, mimicking [21, 23] one can construct approximate solutions for system (4.15) with (4.16) by a carefully devised time discretization scheme, illustrated below (however neglecting the boundary conditions). In this scheme the equation for the displacements is discretized in the following way

$$-\operatorname{div}\left(\mathbb{D}e\left(\frac{u_\tau^j - u_\tau^{j-1}}{\tau}\right) + \mathrm{D}W_2(e(u_\tau^j)) + \tau \mathrm{D}W_p(e(u_\tau^j)) - \vartheta_\tau^j \mathbb{B}\right) = F_\tau^j \text{ in } \Omega_+ \cup \Omega_-, \tag{4.17a}$$

where τ is the time-step associated with a (for simplicity equidistant) partition $\{0 = t_\tau^0 < t_\tau^1 < \ldots < t_\tau^j < \ldots < t_\tau^{J_\tau} = T\}$ of the interval $[0, T]$ and $F_\tau^j = \frac{1}{\tau}\int_{t_\tau^{j-1}}^{t_\tau^j} F(s)\,\mathrm{d}s$. The nonlinear regularizing term $\mathrm{D}W_p(e(u_\tau^j))$, with $p > 4$, is added to the discrete momentum balance in order to compensate the quadratic growth of the terms on the right-hand side of the (discretized) heat equation, namely

$$\begin{aligned} &\frac{\vartheta_\tau^j - \vartheta_\tau^{j-1}}{\tau} - \operatorname{div}\left(\mathbb{K}(e(u_\tau^j), \vartheta_\tau^j)\nabla\vartheta_\tau^j\right) \\ &= e\left(\frac{u_\tau^j - u_\tau^{j-1}}{\tau}\right):\mathbb{D}:e\left(\frac{u_\tau^j - u_\tau^{j-1}}{\tau}\right) - \vartheta_\tau^j \mathbb{B}:e\left(\frac{u_\tau^j - u_\tau^{j-1}}{\tau}\right) + G_\tau^j \end{aligned} \tag{4.17b}$$

in $\Omega_+ \cup \Omega_-$, with G_τ^j defined by local means like F_τ^j. In this way, the right-hand side of (4.17b) turns out to be in $L^2(\Omega)$, and classical Leray-Schauder fixed point arguments can be applied to prove the existence of solutions to (4.17a,4.17b). Finally, we mention that the flow rule for the delamination parameter is discretized and further approximated by penalizing the brittle constraint, i.e. replacing $\widetilde{J}_\infty$ in (4.15g) by J_k.

Semistable energetic solutions of the time-continuous system (4.15), with (4.16), then arise from taking the limit of its time-discrete version, as $\tau \downarrow 0$ and $k \to \infty$ *simultaneously*. Without entering into the analysis of the heat equation and of the delamination flow rule, let us only comment on the limit passage in the weak formulation of the (discrete) equation for the displacements. For that, a key role is played the MOSCO-convergence properties as $k \to \infty$ of the functionals

$$\Phi_k^{\mathrm{adh}}(u, z) := \begin{cases} \int_{\Omega\backslash\Gamma_{\mathrm{C}}} \left(W_2(e(u)) + \tau_k W_p(e(u))\right) \mathrm{d}x + \mathfrak{J}_k([\![u]\!], z) & \text{if } u \in W_{\mathrm{D}}^{1,p}(\Omega\backslash\Gamma_{\mathrm{C}}; \mathbb{R}^d), \\ \infty & \text{otherwise,} \end{cases}$$

with $(\tau_k = k^{-p})_k$ a null sequence as $k \to \infty$. We have denoted the above functionals with the same symbol used for the functionals (2.22), to highlight that Theorem 3 holds for them as well and guarantees the MOSCO-convergence of the functionals $(\Phi_k^{\mathrm{adh}})_k$ to Φ_∞ from (2.23), and thus the limit passage in the mechanical force balance for the displacements.

Acknowledgements This work has been partially supported by the DFG within the project *Finite element approximation of functions of bounded variation and application to models of damage, fracture, and plasticity* in the Priority Programme SPP 1748 *Reliable Simulation Techniques* *in Solid Mechanics. Development of Non-standard Discretisation Methods, Mechanical and Mathematical Analysis.*[1] During a research stay of M.T. at DICATAM, University of Brescia, and of R.R. at WIAS Berlin major steps were made for the progress of this work. The two authors acknowledge the kind hospitality at their host institutions.

References

1. Adams, R.A.: Sobolev Spaces. Academic Press, New York (1975)
2. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Oxford University Press, Oxford (2005)
3. Attouch, H.: Variational convergence for functions and operators. Applicable Mathematics Series. Pitman (Advanced Publishing Program), Boston, MA (1984)
4. Bell, J.F.: Mechanics of Solids, Vol. 1, The Experimental Foundations of Solid Mechanics. Springer, New York (1984)
5. Burenkov, V.I.: Sobolev Spaces on Domains. B. G. Teubner, Stuttgart (1998)
6. Campanato, S.: Proprietà di hölderianità di alcune classi di funzioni. Ann. Scuola Norm. Sup. Pisa (3) **17**, 175–188, (1963)
7. Campanato, S.: Proprietà di una famiglia di spazi funzionali. Ann. Scuola Norm. Sup. Pisa (3) **18**, 137–160, (1964)
8. Egert, M., Haller-Dintelmann, R., Rehberg, J.: Hardy's inequality for functions vanishing on a part of the boundary. Potential Anal. **43**, 49–78, (2015)
9. Fonseca, I., Francfort, G.: Relaxation in BV versus quasiconvexification in $W^{1,p}$; a model for the interaction between fracture and damage. Calc. Var. Partial Differ. Equ. **3**, 407–446, (1995)
10. Frémond, M.: Contact with adhesion, in Topics in Nonsmooth Mechanics, pp. 157–186. In: Moreau, J.J., Panagiotopoulos, P.D., Strang, G. (eds.) Birkhäuser, Basel, (1988)
11. Horganm, C.O., Knowles, J.K.: The effect of nonlinearity on a principle of Saint-Venant type. J. Elasticity **11**, 271–291, (1981)
12. Halphen, B., Nguyen, Q.S.: Sur les matériaux standards généralisés. J. Mécanique **14**, 39–63, (1975)
13. Kachanov, L.M.: Delamination Buckling of Composite Materials. Mechanics of Elastic Stability. Kluwer Academic Publishers, Dordrecht (1988)
14. Kočvara, M., Mielke, A., Roubíček, T.: A rate-independent approach to the delamination problem. Math. Mech. Solids **11**, 423–447, (2006)
15. Knowles, J.K.: The finite anti-plane shear near the tip of a crack for a class of incompressible elastic solids. Int. J. Fract.**13**, 611–639, (1977)
16. Lewis, J.L.: Uniformly fat sets. Trans. Am. Math. Soc. **308**, 177–196, (1988)
17. Mielke, A., Roubíček, T., Stefanelli, U.: Γ-limits and relaxations for rate-independent evolutionary problems. Calc. Var. Partial Differ. Equ. **31**, 387–416, (2008)
18. Mielke, A., Roubíček, T., Thomas, M.: From damage to delamination in nonlinearly elastic materials at small strains. J. Elasticity **109**, 235–273, (2012)
19. Roubíček, T.: Rate-independent processes in viscous solids at small strains. Math. Methods Appl. Sci. **32**, 825–862, (2009)
20. Roubíček, T.: Adhesive contact of visco-elastic bodies and defect measures arising by vanishing viscosity. SIAM J. Math. Anal. **45**, 101–126, (2013)
21. Rossi, R., Roubíček, T.: Thermodynamics and analysis of rate-independent adhesive contact at small strains. Nonlinear Anal. **74**, 3159–3190, (2011)

[1] R.R. also acknowledges support from GNAMPA (Indam).

22. Roubíček, T., Scardia, L., Zanini, C.: Quasistatic delamination problem. Continuum Mech. Thermodynam. **21**, 223–235, (2009)
23. Rossi, R., Thomas, M.: From an adhesive to a brittle delamination model in thermo-visco-elasticity. ESAIM Control Optim. Calc. Var. **21**, 1–59, (2015)
24. Rossi, R., Thomas, M.: Coupling rate-independent and rate-dependent processes: Existence results. SIAM J. Math. Anal. **49**, 1419–1494, (2017)
25. Rossi, R., Thomas, M.: From adhesive to brittle delamination in visco-elastodynamics. Math. Models Methods Appl. Sci. **27**, 1489–1546, (2017)
26. Roubíček, T., Thomas, M., Panagiotopoulos, C.J.: Stress-driven local-solution approach to quasistatic brittle delamination. Nonlinear Anal. Real World Appl. **22**, 645–663, (2015)
27. Scala, R.: Limit of viscous dynamic processes in delamination as the viscosity and inertia vanish. ESAIM Control Optim. Calc. Var. **23**, 593–625, (2017)

Three Examples Concerning the Interaction of Dry Friction and Oscillations

Alexander Mielke

Abstract We discuss recent work concerning the interaction of dry friction, which is a rate independent effect, and temporal oscillations. First, we consider the temporal averaging of highly oscillatory friction coefficients. Here the effective dry friction is obtained as an infimal convolution. Second, we show that simple models with state-dependent friction may induce a Hopf bifurcation, where constant shear rates give rise to periodic behavior where sticking phases alternate with sliding motion. The essential feature here is the dependence of the friction coefficient on the internal state, which has an internal relaxation time. Finally, we present a simple model for rocking toy animal where walking is made possible by a periodic motion of the body that unloads the legs to be moved.

1 Introduction

The phenomenon as well as the microscopic origins of dry friction are well studied (see e.g. [10, 16–18, 21]). Here we understand dry friction in a generalized sense, namely in the sense of rate-independent friction that includes an activation threshold (critical force) to enable motion but then the friction force does not increase with the velocity (or more generally the rate). New nontrivial phenomena arise in cases where the critical force depends periodically on time, either given by an external process or because of the dependence on another state variable of the system. The three examples emphasize different realizations of this dependence.

Partially supported by DFG via project B01 in SFB 1114.

A. Mielke (✉)
Weierstraß-Institut für Angewandte Analysis und Stochastik, Berlin, Germany

Institut für Mathematik, Humboldt-Universität zu Berlin, Berlin, Germany
e-mail: alexander.mielke@wias-berlin.de

E. Rocca et al. (eds.), *Trends in Applications of Mathematics to Mechanics*,
Springer INdAM Series 27, https://doi.org/10.1007/978-3-319-75940-1_8

We will study the effect that, in contrast to systems with viscous friction, systems with rate-independent friction tend to wait in a sticking mode until the relevant friction coefficient is small, and then they can make a very fast move (or even jump) to compensate for the past waiting time. To be more precise, we denote by (q, z) the state of a system, where z is the friction variable, and by $\mathscr{R}$ the dissipation potential for the dry friction. Then $\mathscr{R}(q, z, \dot{q}, \dot{z})$ is nonnegative, convex in $(\dot{q}, \dot{z})$ and positively homogeneous of degree 1 in $\dot{z}$, namely $\mathscr{R}(q, z, \dot{q}, \gamma\dot{z}) = \gamma\mathscr{R}(q, z, \dot{q}, \dot{z})$ for all $\gamma > 0$. For simplicity we will assume that $\mathscr{R}$ has an additive structure in the form

$$\mathscr{R}(q, z, \dot{q}, \dot{z}) = \mathscr{R}_{\mathrm{vi}}(q, z, \dot{q}) + \mathscr{R}_{\mathrm{r.i}}(q, \dot{z}),$$

where "vi" stands for the viscous friction in the variable q, while "r.i" stands for the rate-independent friction in the variable z. Note that we further simplified by assuming that $\mathscr{R}_{\mathrm{r.i}}$ does not depend on z itself (see [2, 11, 12] for more general cases).

The mathematical models we are interested in are given in the form

$$0 = M\ddot{q} + \partial_{\dot{q}}\mathscr{R}_{\mathrm{vi}}(q, z, \dot{q}) + \mathrm{D}_q\mathscr{E}(t, q, z), \quad 0 \in \partial_{\dot{z}}\mathscr{R}_{\mathrm{r.i}}(q, \dot{z}) + \mathrm{D}_z\mathscr{E}(t, q, z).$$

The simplest case of such a system occurs when $q(t)$ displays oscillatory behavior that is totally independent of the variable z, but $\mathscr{R}_{\mathrm{r.i}}$ depends on q. In that case we may reduce to the equation for z alone and study

$$0 \in \mathscr{R}_{\mathrm{r.i}}(t/\varepsilon, \dot{z}) + \mathrm{D}_z\mathscr{E}(t, z), \tag{1}$$

where $\varepsilon > 0$ is a small parameter indicating the ratio between the period of oscillations and the changes in the loading through $t \mapsto \mathscr{E}(t, z)$. A typical application is a plate compactor (see Fig. 1a), where an internal imbalance oscillates rapidly and thus changes the normal pressure in the contact friction. In Sect. 2 we summarize the results from [8], where an explicit formula for the effective homogenized friction for $\varepsilon \to 0$ was derived, see Theorem 2 below.

In Sect. 3 we consider a system of the form

$$0 \in \partial_{\dot{z}}\mathscr{R}_{\mathrm{r.i}}(\alpha, \dot{z}) + \nu\dot{z} + \mathrm{D}_z\mathscr{E}(t, \alpha, z), \quad \dot{\alpha} = F(\alpha, z).$$

Our system is stimulated by applications in geophysics that relate to earthquakes and fault evolution, see [14, 15, 20]. There so-called internal states α are needed to describe the relaxation effects after a sudden tectonic movement or change of shearing motions. We will show that a very simple system under constant shear loading can generate oscillatory behavior that is similar to the famous squeaking chalk on the blackboard or the vibrations arising when moving a rubber over a smooth surface.

Finally, Sect. 4 is devoted to the mechanism of walking of humans or animals. Clearly, an animal wants to reduce friction when moving the extremities on the

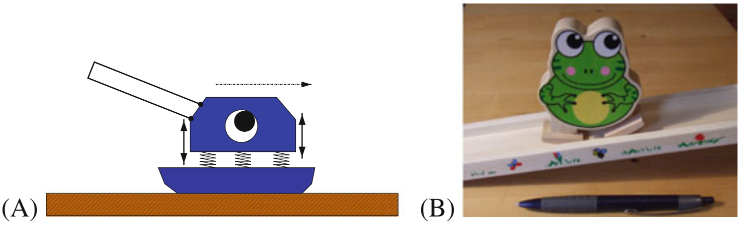

Fig. 1 (**a**) Because of the in-built unbalance, the *plate compactor* vibrates vertically leading to an oscillatory normal pressure. When pushing the plate compactor horizontally it will move only when the normal pressure is very low. (**b**) The *toy ramp walker* in form of a frog walks down only, when alternating the weight between the rigid downhill leg and the hinged uphill leg

ground. To do so, the weight on the leg to be moved has to be reduced. Thus, for making walking efficient it turns out that the body should oscillate in such a manner that without much extra energy the weight on the legs to be moved is minimal. Simple mechanical toys, where this interplay can easily be studied, are the so-called descending woodpecker (cf. [13]), the toy ramp walker, see Fig. 1b, and the rocking toy animal, see Fig. 6. We refer to [4–7] for models on locomotion for micro-machines or animals and to [19] for the slip-stick dynamics of polymers on inhomogeneous surfaces.

We suggest a simple ODE model for the walking of simple mechanical toys such as the *rocking toy animal*, where the essential point is that there is some internal oscillatory mechanism that moves the normal pressure from one leg to the other such that the leg with lowest friction can move. One non-trivial feature is that the natural damping of the rocking motion has to be compensated by some energy supply, where the walking motion feeds energy back into the rocking motion.

2 Prescribed Oscillatory Friction

In this section we summarize the results from [8] concerning the averaging of highly oscillatory rate-independent friction. As we will see there is a major difficulty intrinsic to rate-independent systems that we only obtain a priori bounds for the rate in $\mathrm{BV}([0,T];X)$, but not in a weakly closed Banach space like $\mathrm{W}^{1,p}([0,T];X)$ for $p \in \,]1,\infty[$. Thus, even in the case of classical evolutionary variational inequalities we will not be able to pass to the limit variational inequality but have to use the more flexible formulation in terms of *energetic solutions*.

2.1 Evolutionary Variational Inequalities

While [8] contains more general results, we restrict our discussion to the case of a Hilbert space Z and a quadratic energy $\mathcal{E}(t,z) = \frac{1}{2}\langle Az, z\rangle - \langle \ell(t), z\rangle$ with a loading $\ell \in \mathrm{W}^{1,\infty}([0,T], Z^*)$ and a bounded, symmetric and positive definite linear operator $A : Z \to Z^*$. The dissipation potential is given in the form $\mathscr{R}^{\varepsilon}_{\mathrm{r.i}}(t,\dot z) = \varPsi(t/\varepsilon, \dot z)$, where $\varPsi : \mathbb{S}^1 \times Z \to [0,\infty[$ is assumed to be continuous, and we assume $\varPsi(0,v) \le C\varPsi(s,v) \le C^2\varPsi(0,v)$ for some $C > 1$ and all $(s,v) \in \mathbb{S} \times Z$.

Clearly, the equation $0 \in \partial_{\dot z}\varPsi(t/\varepsilon, \dot z(t)) + Az(t) - \ell(t)$ is equivalent to the variational inequality

$$\forall_{\mathrm{a.a.}} t \in [0,T]\ \forall\, v \in Z : \quad \langle Az(t){-}\ell(t), v{-}\dot z(t)\rangle + \varPsi(t/\varepsilon, v) - \varPsi(t/\varepsilon, \dot z(t)) \ge 0. \tag{2}$$

The key to the analysis in [8] is that $z : [0,T] \to Z$ solves (2) if and only if it is an energetic solution, i.e.

$$\begin{aligned} &\text{(S)} \quad \forall\, t \in [0,T]\ \forall \widehat z \in Z : \quad \mathscr{E}(t, z(t)) \le \mathscr{E}(t,\widehat z) + \varPsi(t/\varepsilon, \widehat z{-}z(t));\\ &\text{(E)} \quad \mathscr{E}(T, z(T)) + \int_0^T \varPsi(s/\varepsilon, \dot z(s))\,\mathrm{d}s \le \mathscr{E}(0, z(0)) - \int_0^T \langle \dot\ell(s), z(s)\rangle\,\mathrm{d}s. \end{aligned} \tag{3}$$

2.2 A Scalar Hysteresis Operator

We now illustrate the difficulty in passing to the limit $\varepsilon \to 0$ in (2) by a very simple scalar hysteresis model by choosing $Z = \mathbb{R}$ and

$$\mathscr{E}(t,z) = \frac{1}{2}z^2 - \ell(t)z, \qquad \varPsi(s,\dot z) = \rho(s)|\dot y|, \quad \text{and} \quad z(0) = 0$$

with $\ell(t) = 5t - t^2$ and an arbitrary $\rho \in \mathrm{C}^1(\mathbb{S})$ (where $\mathbb{S} := \mathbb{R}/_{\mathbb{Z}}$) satisfying $\rho_{\min} := \min\{\,\rho(s) \mid s \in \mathbb{S}\,\} > 0$.

Starting from the initial condition $z(0) = 0$, we see that z cannot decrease but needs to lie in the stable interval $[\ell(t){-}\rho(t/\varepsilon), \ell(t){+}\rho(t/\varepsilon)]$, see (S) in (3). Thus, the solution $z_\varepsilon : [0,T] \to \mathbb{R}$ of $0 \in \rho(t/\varepsilon)\,\mathrm{Sign}(\dot z(t)) + z(t) - \ell(t)$ has, for sufficiently small $\varepsilon > 0$, the representation

$$z_\varepsilon(t) = \begin{cases} \max\{\,0, \ell(\tau) - \rho(\tau/\varepsilon) \mid \tau \in [0,t]\,\} & \text{for } t \in [0, \frac{5}{2}{+}\sqrt{\rho_{\min}}],\\ \min\{\,\frac{25}{4}{-}\rho_{\min}, \ell(\tau){+}\rho(\tau/\varepsilon) \mid \tau \in [\frac{5}{2}{+}\sqrt{\rho_{\min}}, t]\,\} & \text{for } t \ge \frac{5}{2}{+}\sqrt{\rho_{\min}}. \end{cases}$$

It can be checked by direct calculation that this is the unique solution. Moreover, we obtain uniform convergence to the limit solution given in the form

$$z_0(t) = \begin{cases} \max\{0, \ell(\tau) - \rho_{\min}\} & \text{for } t \in [0, \frac{5}{2}{+}\sqrt{\rho_{\min}}],\\ \min\{\frac{25}{4}{-}\rho_{\min}, \ell(\tau){+}\rho_{\min}\} & \text{for } t \in [\frac{5}{2}{+}\sqrt{\rho_{\min}}, T]. \end{cases}$$

In particular, we have $\|z_\varepsilon - z_0\|_\infty \le C\varepsilon$.

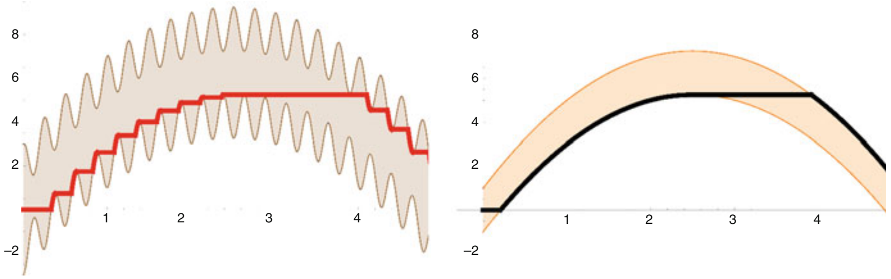

Fig. 2 Left: the energetic solution $z_\varepsilon : [0, 5] \to \mathbb{R}$ for $\varepsilon = 0.1$ lies in the ε-depending stable region (as shaded between wiggly boundaries). Right: the energetic solution $z_0 : [0, 5] \to \mathbb{R}$ for $\varepsilon = 0$. The limiting stable region (between parabolas) can be understood as the intersection of all stable regions for $\varepsilon > 0$

However, the situation for the rates $\dot z_\varepsilon : [0, T] \to \mathbb{R}$ is quite different. From the explicit formula we see that $\dot z_\varepsilon(t)$ either equals 0 (stiction) or $\dot z_\varepsilon(t) = \dot\ell(t) - \frac{1}{\varepsilon}\rho'(t/\varepsilon)$. Thus, within the intervals $[k\varepsilon, (k{+}1)\varepsilon]$ we typically have $\dot z_\varepsilon = 0$ for most of the time and $\dot z_\varepsilon \approx 1/\varepsilon$ for intervals of length $O(\varepsilon^2)$, see Fig. 2. As a consequence we conclude that $\dot z_\varepsilon$ does not converge weakly to $\dot z_0$ in $\mathrm{L}^p([0, T])$ for any $p \in [1, \infty[$. We only have $\dot z_\varepsilon \overset{*}{\rightharpoonup} \dot z_0$ in $M([0, T]) = \mathrm{C}^0([0, T])^*$, i.e. in the sense of measures when testing with continuous test functions.

2.3 The Averaging Result for Oscillatory Friction

We now provide the announced averaging result, which can be understood in terms of integral infimal convolutions as follows. We define

$$\Psi_{\mathrm{av}}(V) := \inf\Big\{ \int_{\mathbb{S}} \Psi(s, v(s))\,\mathrm{d}s \;\Big|\; v \in \mathrm{L}^1(\mathbb{S}),\ \int_{\mathbb{S}} v(s)\,\mathrm{d}s = V \Big\}. \tag{4}$$

This formulation justify the colloquial term that oscillatory rate-independent systems watch for the easiest opportunity to move: during the microscopic time $s = t/\varepsilon \in \mathbb{S}$ there is an instant such that moving in the direction $v(s) \in Z$ is optimal, hence the overall motion in direction $V \in Z$ will be decomposed into an oscillatory motion $s \mapsto v(s)$.

Example 1 For $Z = \mathbb{R}^2$ consider $\Psi(s, v) = (2{-}\cos(2\pi s))|v_1| + (2{+}\cos(2\pi s))\,|v_2|$. Then, $\Psi_{\mathrm{av}}(v) = |v_1| + |v_2|$, since moving in z_1-direction is optimal for $s \approx 0$ while motion in z_2-direction is optimal for $s \approx 1/2$.

The first observation is that Ψ_{av} can be characterized in terms of its conjugate $\Psi^*(s,\xi) = \sup\big\{ \langle \xi, v\rangle - \Psi(s,v) \,\big|\, v \in Z \big\}$ obtained by the Legendre-Fenchel transformation. From the 1-homogeneity of $\Psi(s,\cdot)$ we see that

$$\Psi^*(s,\cdot) = \chi_{K(s)}(\xi) = \begin{cases} 0 & \text{for } \xi \in K(s), \\ \infty & \text{otherwise,} \end{cases} \tag{5}$$

where $K(s) := \partial\Psi(s,0)$ is a closed convex set containing $\xi = 0 \in Z^*$. In [8, Prop. 3.6] it is shown that

$$\Psi_{\mathrm{av}}^*(\xi) = \chi_{K_{\mathrm{av}}}(\xi) \quad \text{with } K_{\mathrm{av}} = \bigcap_{s\in\mathbb{S}} K(s).$$

The averaging result now reads as follows.

Theorem 2 (See [8, Thm. 1.1]) *Consider a quadratic energetic system $(Z,\mathcal{E},\Psi)$ as in Sect. 2.1 and an initial condition $\widehat{z}_0 \in Z$ such that*

$$0 \in \partial_{\dot z}\Psi(s,0) + A\widehat{z}_0 - \ell(0) \text{ for all } s \in \mathbb{S}.$$

Under the unique solutions $z_\varepsilon : [0,T] \to Z$ of (2) with $z_\varepsilon(0) = \widehat{z}_0$ satisfy $z_\varepsilon(t) \rightharpoonup z_0(t)$ in Z, where z_0 is the unique solution of the averaged equation

$$0 \in \partial\Psi_{av}(\dot z(t)) + Az(t) - \ell(t), \quad z_0(0) = \widehat{z}_0.$$

The proof relies heavily on the following asymptotic equicontinuity result:

$$\begin{aligned} &\exists \text{ modulus of cont. } \omega \; \forall \varepsilon \in \,]0,1[\; \forall t_1,t_2 \in [0,T]: \\ &\qquad\qquad \|z_\varepsilon(t_2) - z_\varepsilon(t_1)\|_Z \le \omega(\varepsilon) + \omega\big(|t_2{-}t_1|\big). \end{aligned} \tag{6}$$

As is seen by the scalar example in Sect. 2.2 it is not possible to provide a better equicontinuity result. First it is then standard to extract a subsequence such that $z_{\varepsilon_n}(t)$ converges to some $z_0(t)$ weakly for all $t \in [0,T]$. The limit passage is then done in the energetic formulation (3). Using the definition of Ψ_{av} in (4) we have $\Psi_{\mathrm{av}} \le \Psi(s,\cdot)$, and it is easy to obtain the upper energy estimate (E), namely $\mathcal{E}(T, z_0(T) + \int_0^T \Psi_{\mathrm{av}}(\dot z_0)\,\mathrm{d}t \le \mathcal{E}(0,\widehat{z}_0) - \int_0^T \langle \dot\ell, z_0\rangle\,\mathrm{d}s$.

For the stability condition (S) we use the equivalent formulation $0 \in \partial\Psi(t/\varepsilon,0) + Az_\varepsilon(t) - \ell(t)$. Exploiting the equicontinuity (6) we can also have $z_\varepsilon(\widehat{\tau}(t,s,\varepsilon)) \rightharpoonup z_0(t)$ whenever $\widehat{\tau}(t,s,\varepsilon) \to 0$. Thus, we may choose $\widehat{\tau}(t,s,\varepsilon)$ such that $\widehat{\tau}(t,s,\varepsilon)\varepsilon \mod 1 = s$ and obtain $0 \in \partial\Psi(s,0) + Az_0(t) - \ell(t)$ for all $s \in \mathbb{S}$. By (5) we conclude $0 \in \partial\Psi_{\mathrm{av}}(0) + Az_0(t) - \ell(t)$ which is (S) for the limit equation. By standard arguments we then conclude that z_0 is the desired unique solution.

3 Self-Induced Oscillations in State-Dependent Friction

The modeling of rate-and-state dependent friction is a classical area in geophysics as it describes basic mechanisms in the frictional movement of tectonic plates or faults in the earth crust, see [1, 15] and [15, 20] for more mathematical approaches. In [9] the following work will be presented in the wider context of continuum mechanics. Here we rather restrict to a simple ODE in the spirit of the *spring-block sliders* studied in [1].

Our simple scalar model of a block slider is described by the position $z(t)$ over the flat surface and a *state variable* α (that may be interpreted as a local temperature). The importance is that the friction coefficient μ for the rate-independent friction occurring through $\dot{z}$ depends nontrivially on α, namely $\mu = \widetilde{\mu}(\alpha)$ with $\mu'(\alpha) < 0$, while friction $|\dot{z}|$ increases α.

For simplicity we restrict to the following simple coupled system:

$$0 \in \widetilde{\mu}(\alpha)\,\mathrm{Sign}(\dot{z}) + \nu\dot{z} + k\big(z-\ell\big), \quad \dot{\alpha} = \alpha_0 - \alpha + \widetilde{\mu}(\alpha)|\dot{z}| + \nu\dot{z}^2. \tag{7}$$

Here $k > 0$ is the elastic constant of the spring connecting the time-dependent external loading $\ell(t)$ with the body, $\nu \geq 0$ is a small viscosity coefficient in the friction law, and $\alpha_0 > 0$ is the constant rest state. Thus the friction is rate-dependent through $\nu\dot{z}$ as well as state-dependent through α, namely for $\dot{z} > 0$ we have $\xi_{\text{frict}} = \widetilde{\mu}(\alpha) + \nu\dot{z}$. Note that the relaxation time for the state variable α was set to 1 without loss of generality.

For the later analysis it is advantageous to rewrite the first equation in (7) as an explicit ODE. Defining the functions

$$G(\xi,\alpha) := \begin{cases} \big(\xi-\widetilde{\mu}(\alpha)\big)/\nu & \text{for } \xi \geq \widetilde{\mu}(\alpha), \\ \quad 0 & \text{for } |\xi| \leq \widetilde{\mu}(\alpha), \\ \big(\xi+\widetilde{\mu}(\alpha)\big)/\nu & \text{for } \xi \leq -\widetilde{\mu}(\alpha), \end{cases}$$

we find the equivalent form

$$\dot{z} = G\big(k(\ell-z,\alpha)\big), \quad \dot{\alpha} = 1 - \alpha + k(\ell-z)\,G\big(k(\ell-z),\alpha\big). \tag{8}$$

The typical experiment is the model with a constant shear velocity V, i.e. $\ell(t) = Vt$. Indeed, the problem is translationally invariant if ℓ and z are changed together. Thus, it is useful to work with $V(t) = \dot{\ell}(t)$ and to consider the difference $U(t) = \ell(t) - z(t)$, which satisfies the ODE system

$$\dot{U}(t) = V(t) - G\big(kU(t),\alpha(t)\big), \quad \dot{\alpha}(t) = \alpha_0 - \alpha(t) + kU(t)G\big(kU(t),\alpha(t)\big). \tag{9}$$

In [9] the response of the system to varying shear rates $V(t)$ is studied in regimes where the system prefers to return into a steady state, whenever $V(t)$ has a plateau.

Here, we want to show that under suitable conditions on the function $\alpha \mapsto \mu(\alpha)$ the system displays self-induced oscillations for constant shear rates $V(t) \equiv V_*$. In that case (9) is a planar autonomous system which can be discussed in the phase plane for (U, α). Without loss of generality we assume $V_* > 0$ and choose $k = \alpha_0 = 1$ for notational simplicity. We first calculate the equilibria (U_*, α_*) and note that no equilibria with $U_* \leq \widetilde{\mu}(\alpha_*)$ can exist because then $G(U_*, \alpha_*) \leq 0$. Hence, the relations for equilibria reduce to

$$U_* = \widetilde{\mu}(\alpha_*) + \nu V_* \quad \text{and} \quad \alpha_* = 1 + V_* U_*.$$

Using our major assumption $\widetilde{\mu}'(\alpha) \leq 0$ we immediately see that there is a unique equilibrium determined by the relation $\alpha_* = 1 + V_*\widetilde{\mu}(\alpha_*) + \nu V_*^2$. Clearly, α_* as a function of V_* is monotonously increasing from $\alpha_* = 1$ at $V_* = 0$.

To study the stability of the solution we calculate the linearization of the vector field $\frac{\mathrm{d}}{\mathrm{d}}\binom{U}{\alpha} = F(U, \alpha)$ in $q_* = (U_*, \alpha_*)$ giving the Jacobi matrix

$$\mathrm{D}F(q_*) = \begin{pmatrix} -\partial_U G(q_*) & -\partial_\alpha G(q_*)/\nu \\ V_*+U_*\partial_U G(q_*) & -1+U_*\partial_\alpha G(q_*) \end{pmatrix} = \begin{pmatrix} -1/\nu & -\widetilde{\mu}'(\alpha_*) \\ V_*+U_*/\nu & -1-U_*\widetilde{\mu}'(\alpha_*)/\nu \end{pmatrix}.$$

As a result we find that the determinant $\det \mathrm{D}F(q_*) = (1-V_*\widetilde{\mu}'(\alpha_*))/\nu$ is always positive. For the trace we obtain

$$\mathrm{tr}\big(\mathrm{D}F(q_*)\big) = -1 - \big(1+U_*\widetilde{\mu}'(\alpha_*)\big)/\nu = -1 - \widetilde{\mu}'(\alpha_*)V_* - \big(1 + \widetilde{\mu}(\alpha_*)\widetilde{\mu}'(\alpha_*)\big)/\nu.$$

Clearly the equilibrium is stable if $\mathrm{trace}\big(\mathrm{D}F(q_*)\big) < 0$, undergoes a Hopf-bifurcation for $\mathrm{trace}\big(\mathrm{D}F(q_*)\big) = 0$, and is unstable for $\mathrm{trace}\big(\mathrm{D}F(q_*)\big) > 0$.

Theorem 3 (Periodic Oscillations) *Assume that $V_* > 0$ is chosen such that the unique equilibrium $q_* = (U_*, \alpha_*)$ satisfies $\mathrm{trace}\big(\mathrm{D}F(q_*)\big) > 0$, then there exists a stable periodic orbit.*

Proof The result follows from standard phase-plane arguments, since the equilibrium is unstable, and there exists a positively invariant region. Indeed, setting $U_{\max} = \widetilde{\mu}(0) + \nu V_*$ we find $\dot{U} = V_* - G(U, \alpha) \leq 0$ for whenever $U \geq U_{\max}$. Hence, for $U \in [0, U_{\max}]$ we have $G(U, \alpha) \leq G_{\max} = U_{\max}^2/\nu$ and conclude that $\dot{\alpha} = 1 - \alpha + UG(U, \alpha) \leq 0$ for $\alpha \geq \alpha_{\max} = 1 + G_{\max}$. Thus, the rectangle $[0, U_{\max}] \times [0, \alpha_{\max}]$ is positively invariant. By the Poincaré–Bendixson the existence of at least one limit cycle follows. Standard argument show that there must also be one stable periodic orbit.

We also want to understand the limit behavior $\nu \to 0$, which means that the friction part converges to its rate-independent limit while the variable α remains rate dependent. In that case, we expect that the oscillations become very fast with a period of order $O(\nu^\delta)$ for some $\delta > 0$. To analyze this case we consider a special scaling limit that shows a non-standard bifurcation. In particular, we assume that V

is positive but also small with ν, i.e. we unfold ν and V simultaneously. Moreover, to simplify the notations we assume that the bifurcation takes place at $\alpha = 1$ already.

In particular, we consider the scalings

$$V = \nu\widehat{v}, \quad U = \widetilde{\mu}(\alpha) + \nu^2\beta, \quad \alpha = 1 + \nu\gamma, \quad \widetilde{\mu}(\alpha) = \mu(\nu\gamma) - \sqrt{\nu}B,$$

where $B \in \mathbb{R}$ is an unfolding parameter, which is chosen with a particular scaling to generate periodic solutions with a phase of sticking and a phase of frictional sliding. The function μ is assumed to satisfy

$$1 + \mu(0)\mu'(0) = 0 \quad \text{with } \mu_0 := \mu(0) > 0 \text{ and } \mu'(0) = -1/\mu_0 < 0. \tag{10}$$

This gives the following equivalent system

$$\nu\dot{\beta} = \widehat{v} - \beta^+ - \mu'(\nu\gamma)\dot{\gamma}, \quad \dot{\gamma} = -\gamma + \big(\mu(\nu\gamma) - \sqrt{\nu}B + \nu^2\beta\big)\beta^+,$$

where $\beta^+ := \max\{\beta, 0\}$. The special assumption in (10) leads to a cancellation when we insert the equation for $\dot{\gamma}$ into the equation for $\dot{\beta}$, namely

$$\begin{aligned} \dot{\beta} &= \frac{\widehat{v}}{\nu} + A(\nu,\gamma)\beta^+ + \frac{\mu'(\nu\gamma)}{\nu}\gamma - \nu\mu'(\nu\gamma)\beta\beta^+, \\ \dot{\gamma} &= -\gamma + \big(\mu(\nu\gamma) - \sqrt{\nu}B\big)\beta^+ + \nu^2\beta\beta^+, \end{aligned} \tag{11}$$

where the coefficient $A(\nu,\gamma)$ stays is order $1/\sqrt{\nu}$ for $\nu \to 0$, namely

$$A(\nu,\gamma) := \frac{\mu'(\nu\gamma)\big(\sqrt{\nu}B - \mu(\nu\gamma)\big) - 1}{\nu} = \frac{B}{\mu_0\sqrt{\nu}} + O(1)_{\nu\to 0},$$

where we used the first relation in (10).

The solutions we will construct below will satisfy estimates of the form $\gamma(t) \in [0, C]$ and $\beta(t) \leq [-C\widehat{v}/\nu, C/\sqrt{\nu}]$, hence it will be justified to drop the higher order terms. Using $b = B/\mu_0$ we will consider the simplified system

$$\dot{\beta} = \widehat{v}\nu\big(1 - b\sqrt{\nu}\big) + \frac{b}{\sqrt{\nu}}\beta^+ - \frac{1}{\nu\mu_0}\gamma, \qquad \dot{\gamma} = \mu_0\beta^+ - \gamma, \tag{12}$$

which is a piecewise linear system and has the unique steady state $(\beta_*, \gamma_*) = (\widehat{v}, \mu_0\widehat{v})$. Since the system is positively homogeneous of degree 1, the solutions for general $\widehat{v}$ are obtained from the solution $(\beta_1(t), \gamma_1(t))$ for $\widehat{v} = 1$ by a simple multiplication, namely $(\widehat{v}\beta_1(t), \widehat{v}\gamma_1(t))$.

We are especially interested in the case $b \in\,]0, 2[$ where the fixed point is an unstable focus with eigenvalues

$$\lambda_{1,2} = \frac{b/2}{\sqrt{\nu}} \pm \mathrm{i}\frac{\omega_b}{\sqrt{\nu}} + O(1), \quad \text{where } \omega_b = \sqrt{1 - b^2/4}.$$

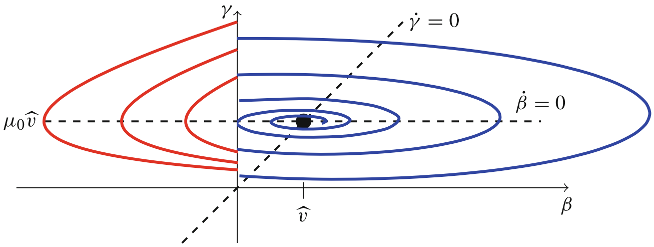

Fig. 3 The phase plane for the piecewise linear system (12) for $\widehat{v} = 1$ and $\mu_0 = 1$: For $\beta \geq 0$ we have an unstable focus, while for $\beta \leq 0$ we have the simple system $\dot\beta = 1/\nu - \gamma/(\nu\mu_0)$, $\dot\gamma = -\gamma$

In the phase plane for (β, γ) we can construct periodic solutions by piecing together the piecewise linear systems, see Fig. 3. For the explicit construction of a periodic orbits we decompose the axis $\{ (0, \gamma) \mid \gamma \geq 0 \}$ into the two parts $\{0\} \times A_j$ with

$$A_1 := [0, \mu_0\widehat{v}[\quad \text{and} \quad A_2 :=]\mu_0\widehat{v}, \infty[.$$

Then solutions starting in A_1 will move according to the unstable focus in $(\beta_*, \gamma_*) = (\widehat{v}, \mu_0\widehat{v})$: First they rapidly move to the right, then turn slowly upwards, and reach $\dot\beta = 0$ when β is of order $1/\sqrt{\nu}$. Then, the solutions move rapidly back to the axis $\beta = 0$. Let us denote this Poincaré mapping by $\Phi^+ : A_1 \to A_2$, see Fig. 4. Since the motion between $\beta = 0$ and $\beta = \widehat{v}$ only takes a time of order ν, it can be neglected compared to the travel time around the fixed point. Thus the travel time associated to Φ^+ is half the period, namely $\pi\omega_b/\sqrt{\nu}$. During that time the solutions are stretched, so that

$$\Phi^+(\nu, \cdot) : \begin{cases} A_1 \to & A_2, \\ \gamma \mapsto \mu_0\widehat{v} + \rho_b(\mu_0\widehat{v} - \gamma) + O(\sqrt{\nu}), \end{cases}$$

with a stretching factor $\rho_b := \mathrm{e}^{\pi b/(2\omega_b)} > 1$.

Similarly the linear flow for $\beta \leq 0$ provides a Poincaré map $\Phi^- : A_2 \to A_1$, see Fig. 4. As the solutions starting in A_2 are given by $\gamma(t) = \mathrm{e}^{-(t-t_0)}\gamma(t_0)$ and $\nu\beta(t) = \widehat{v}(t-t_0) + \frac{1}{\mu_0}(1-\mathrm{e}^{t_0-t})\gamma(t_0)$ we obtain $\Phi^-(\gamma(t_0)) = \gamma(t_1)$, where $t_1 = t_0 + T$ is defined via $\widehat{v}T = (1-\mathrm{e}^{-T})\gamma(t_0)/\mu_0$. Since the function $\mathscr{B} :]0, \infty[\to]0, 1[;\ T \mapsto (1-\mathrm{e}^{-T})/T$ is strictly decreasing it has a smooth inverse $\mathscr{C} :]0, 1[\to]1, \infty[$ which gives

$$\Phi^-(\nu, \cdot) : \begin{cases} A_2 \to & A_1, \\ \gamma \mapsto \mathrm{e}^{-\mathscr{C}(\mu_0\widehat{v}/\gamma)}\gamma, \end{cases}$$

which is even independent of ν, because this regime relates to the sticking phase $U < \mu(\alpha)$ where the viscosity ν is irrelevant. By construction it follows that Φ^- is convex and monotonously decreasing with slopes in $]-1, 0[$.

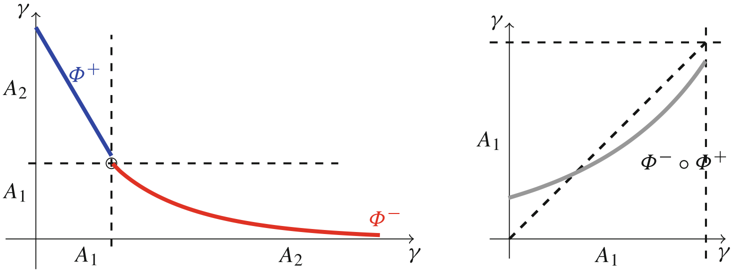

Fig. 4 On the left, the two Poincaré maps $\Phi^+ : A_1 \to A_2$ and $\Phi^- : A_2 \to A_1$ are displayed. The right shows $\Phi^- \circ \Phi^+ : A_1 \to A_1$, where the unique fixed point gives to the stable limit cycle

Periodic solutions are now obtained as fixed points of $\Psi := \Phi^- \circ \Phi^+ : A_1 \to A_1$. From the lowest order expansions of $\Phi^\pm$ we see that Ψ is convex and strictly increasing. Moreover $\Psi(\mu_0\widehat{v})$ is slightly below $\mu_0\widehat{v}$ and $\Psi'(\mu_0\widehat{v}) = \rho_b > 1$. Thus, there is a unique fixed point γ_b in the interior, while the fixed point at $\gamma = \mu_0\widehat{v}$ of the lowest-order expansion does not survive. As $\rho_b = \mathrm{e}^{\pi b/(2\omega_b)}$ is strictly increasing with $b \in]0, 2[$ from 1 to ∞, we see that $b \mapsto \gamma_b$ is strictly decreasing with limits $\gamma_0 = \mu_0\widehat{v}$ to $\gamma_2 = 0$. Since $0 < \Psi'(\gamma_b) < 1$, we also conclude that the associated periodic orbit is stable.

The important observation is that the travel times in the two Poincaré mappings Φ^+ and Φ^- are quite different. The time with $\beta > 0$ is of order $\sqrt{\nu}\pi/\omega_b + O(\nu)$ while the time with $\beta < 0$ is of order 1. Thus, looking at the temporal behavior we have a relatively long period of sticking, while there is a relatively short period of sliding. Transforming our solutions back into the original variables we obtain, in the case $\beta > 0$ the expansion

$$U(t) = \mu(\nu\gamma) - \sqrt{\nu}B + \nu^2\beta = \mu_0 - \sqrt{\nu}B - \nu\frac{\gamma(t)}{\mu_0} + O(\nu^{3/2}), \quad \alpha(t) = 1 + \nu\gamma(t),$$

whereas in the case $\beta(t) < 0$ we have $\beta = O(1/\nu)$ and thus

$$U(t) = \mu_0 - \sqrt{\nu}B + \nu\widehat{v}\,(t - t_k) - \nu\frac{\gamma(t_k)}{\mu_0} + O(\nu^{3/2}), \quad \alpha(t) = 1 + \nu\mathrm{e}^{-(t-t_k)}\gamma(t_k),$$

where t_k is the last time, where the solution switched from $\beta > 0$ to $\beta < 0$. The behavior is illustrated in Fig. 5.

We emphasize that all the solutions we have obtained in this scaling limit have a phase in the lower half plane, which means $U(t) \leq \widetilde{\mu}(\alpha(t))$ and hence $\dot{U} = V$. In the original variables this means $\dot{z} = 0$ which is the sticking phase. Physically this means that the system rest for a short time until the shear has build up to reach the critical threshold. However, then the state α (e.g. the temperature) is increased so that the friction coefficient drops. Thus $z(t) = Vt - U(t)$ moves forward a lot and reduces the shear stress significantly. But then α again decreases and thus the friction coefficient again raises, which leads to the next sticking phase.

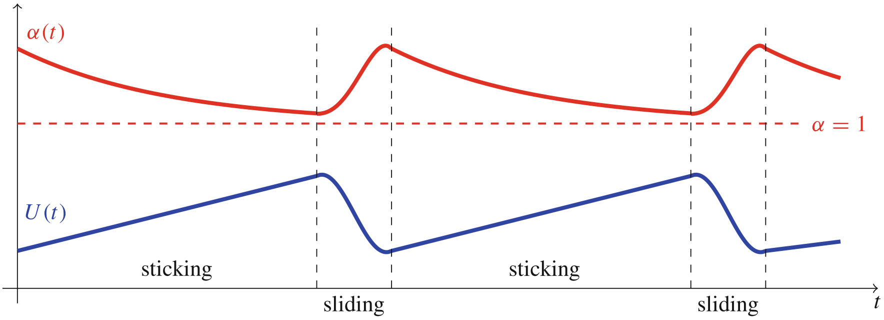

Fig. 5 The periodic functions $U(t), \alpha(t)$ displaying long phases of sticking and short phases of fast slip

4 A Model for the Rocking Toy Animal

Our third example concerning the interaction of Coulomb friction and oscillations relates to a very simplistic model for walking of so-called *rocking toy animals*. A similar model could be derived for the toy ramp walker shown in Fig. 1b.

4.1 Description of the Mechanical Toy

The toy animal has two right and two left legs that usually move together so we identify them and speak of the right and the left leg. The toy is pulled forward by a string that hangs over the edge of a table, where a suitable weight provides a constant pulling force. A related walking toy is the ramp walker, which oscillates in the direction of walking. It has only two legs, the forward and the backward one, which are alternately loaded and unloaded, see Fig. 6 for a pictures and two schematic views of a rocking toy cow.

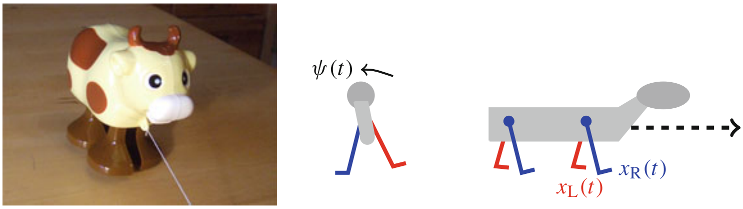

Fig. 6 Rocking toy animal. Left: A weight beyond the table edge pulls the toy animal forward, while the perpendicular rocking motions allows the lifted legs to swing forward because of the reduced normal pressure. Middle: changes in the perpendicular rocking angle $\psi(t)$ lifts either the right or the left leg. Right: the string pulls the animal forward and increases the potential energy slightly when the hinge of a leg is moved over the leg's contact point

This model has the following features:

(i) Walking is a periodic motion that is enabled by perpendicular oscillations, which change the weight on the left and right legs.

(ii) The force in the pulling string needs to be substantially less for the oscillating motion than for the sliding motion without oscillations. For very small pulling force no motion occurs.

(iii) To compensate for damping in the perpendicular oscillations, energy has to be transferred from the forward motion into the perpendicular oscillation.

4.2 A Model with Inertia

We model the system of the toy animal by three degrees of freedom, i.e. we assume that both legs on the right side and both legs on the left side move together respectively and can be described by the average position $x_{\mathrm{R}}(t) \in \mathbb{R}$ and $x_{\mathrm{L}}(t) \in \mathbb{R}$. To simplify notations we abbreviate $\boldsymbol{x} = (x_{\mathrm{R}}, x_{\mathrm{L}})$. The third degree of freedom is given by the angle ψ of the animals symmetry line against the vertical axis.

The total energy $\mathscr{E}(\boldsymbol{x}, \psi, \dot{\boldsymbol{x}}, \dot{\psi}) = \mathscr{E}_{\mathrm{ani}}(\boldsymbol{x}, \psi, \dot{\boldsymbol{x}}, \dot{\psi}) + \mathscr{E}_{\mathrm{weight}}(\boldsymbol{x}, \dot{\boldsymbol{x}})$ is given by

$$\mathscr{E}_{\mathrm{animal}}(\boldsymbol{x}, \psi, \dot{\boldsymbol{x}}, \dot{\psi}) = \Phi(x_{\mathrm{R}}-x_{\mathrm{L}}, \psi) + \frac{m_{\mathrm{b}}}{2}(\dot{x}_{\mathrm{R}}+\dot{x}_{\mathrm{L}})^2 + \frac{m_{\mathrm{leg}}}{2}\big((\dot{x}_{\mathrm{R}})^2+(\dot{x}_{\mathrm{L}})^2\big) + \frac{I_{\mathrm{b}}}{2}\dot{\psi}^2,$$

$$\mathscr{E}_{\mathrm{weight}}(\boldsymbol{x}, \dot{\boldsymbol{x}}) = -gm_{\mathrm{we}}\frac{1}{2}(x_{\mathrm{R}}+x_{\mathrm{L}}) + \frac{m_{\mathrm{we}}}{2}(\dot{x}_{\mathrm{R}}+\dot{x}_{\mathrm{L}})^2,$$

where I_{b} is the rotational inertia of the body, and m_{b}, m_{leg}, and m_{we} are the masses of the body, the legs, and the weight, respectively.

The main mechanism for walking originates from the dissipation, which we assume to have the form

$$\mathscr{R}(\boldsymbol{x}, \psi, \dot{\boldsymbol{x}}, \dot{\psi}) = \frac{\delta}{2}(\dot{\psi})^2+\big(\rho+H_{\mathrm{R}}(\psi)\big)|\dot{x}_{\mathrm{R}}|+\big(\rho+H_{\mathrm{L}}(\psi)\big)|\dot{x}_{\mathrm{L}}|+\frac{\nu}{2}\big(\dot{x}_{\mathrm{R}}\big)^2+\frac{\nu}{2}\big(\dot{x}_{\mathrm{L}}\big)^2,$$

where $\delta,\ \nu > 0$ induce simple viscous friction. The main feature of the model is the dependence of the rate-independent Coulomb friction of the two legs on the tilt angle ψ through the two functions H_{R} and H_{L}, which indicate the normal pressure times the friction coefficient on the right and the left leg, respectively, while $\rho > 0$ is the dry friction in the joints, which is independent of the normal pressure. We assume

$$H_{\mathrm{R}}(\psi) + H_{\mathrm{L}}(\psi) = H_* = \text{const.}, \qquad H_{\mathrm{R}}(\psi) = H_{\mathrm{L}}(-\psi),$$

$$H_{\mathrm{R}}(\psi) = H_{\mathrm{L}}(\psi) = \frac{1}{2}H_* \text{ for } |\psi| \le \psi_0, \qquad H_{\mathrm{R}}(\psi) = 0 \text{ for } \psi \ge \psi_1 > \psi_0.$$

An important point in the modeling is that $0 < \rho \ll H_*/2$, i.e. the friction in the joints is much smaller than the friction of moving the non-rocking animal.

Denoting by $q = \psi, x_{\mathrm{R}}, x_{\mathrm{L}})$ the state of the system, the equation to be studied is the damped Hamiltonian system

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\partial_{\dot q}\mathscr{E}(q,\dot q)\Big) + \partial_{\dot q}\mathscr{R}(q,\dot q) + \partial_q\mathscr{E}(q,\dot q) = 0.$$

Thus, the full model takes the form of a coupled three-degrees of freedom system:

$$I_{\mathrm{b}}\ddot\psi \qquad + \delta\dot\psi \qquad + \partial_\psi\Phi(x_{\mathrm{R}}-x_{\mathrm{L}},\psi) = 0, \tag{13a}$$

$$\begin{aligned}(m_{\mathrm{we}}+m_{\mathrm{b}})(\ddot x_{\mathrm{R}}+\ddot x_{\mathrm{L}}) + m_{\mathrm{leg}}\ddot x_{\mathrm{R}}&\\ +\nu\dot x_{\mathrm{R}} + \big(\rho+H_{\mathrm{R}}(\psi)\big)\,\mathrm{Sign}(\dot x_{\mathrm{R}}) + \partial_d\Phi(x_{\mathrm{R}}-x_{\mathrm{L}},\psi) &= gm_{\mathrm{we}}/2,\end{aligned} \tag{13b}$$

$$\begin{aligned}(m_{\mathrm{we}}+m_{\mathrm{b}})(\ddot x_{\mathrm{R}}+\ddot x_{\mathrm{L}}) + m_{\mathrm{leg}}\ddot x_{\mathrm{L}}&\\ +\nu\dot x_{\mathrm{L}} + \big(\rho+H_{\mathrm{L}}(\psi)\big)\,\mathrm{Sign}(\dot x_{\mathrm{L}}) - \partial_d\Phi(x_{\mathrm{R}}-x_{\mathrm{L}},\psi) &= gm_{\mathrm{we}}/2,\end{aligned} \tag{13c}$$

where $d = x_{\mathrm{R}} - x_{\mathrm{L}}$ is the (signed) distance between the right and the left leg.

The main mathematical task in studying this model is to show that there are time-periodic translating motions, i.e.

$$\psi(t) = \Psi_{\mathrm{per}}(t), \quad x_{\mathrm{R}}(t) = vt + R_{\mathrm{per}}(t), \quad y_{\mathrm{L}}(t) = vt + L_{\mathrm{per}}(t),$$

where v is the average walking speed while $(\Psi_{\mathrm{per}}, R_{\mathrm{per}}, L_{\mathrm{per}}) : \mathbb{R} \to \mathbb{R}^3$ is periodic. The trivial solution is the non-rocking solution $(\Psi_{\mathrm{per}}, R_{\mathrm{per}}, L_{\mathrm{per}}) \equiv 0$, where the velocity and the pulling force are related by

$$\nu v + \rho + \frac{1}{2}H_* = \frac{1}{2}gm_{\mathrm{we}}.$$

Thus, even for arbitrary small velocities $v > 0$, the pulling force must overcome the full Coulomb friction for the full weight of the toy. The point is that a symmetry breaking leading to an oscillatory behavior can lead to larger velocities v even for much lower pulling forces gm_{we}.

In principle, this model could be studied for the desired oscillatory behavior, but we will simplify the model further such that the existence of relevant periodic motions can be shown more easily.

4.3 A Simplified Model Without Translational Inertia

We consider a simplified model, where we neglect inertial effects in the translation direction but not in the transverse oscillations. This can be justified for the rocking toy animal, since the forward motions are relatively slow and masses are low; whereas the rocking motion in transverse direction has relatively fast giving rise to a transverse oscillations dominated by inertia. Numerically, it can be shown that the model with transverse inertia still displays the same solutions, but the associated mathematical analysis would be significantly more difficult and thus obscure the main mechanism of feeding energy from the forward motion into the rocking motion by a suitable coupling, see $c \neq 0$ below.

Thus, we neglect all terms in the energy arising through $(\dot{x}_\mathrm{R}, \dot{x}_\mathrm{L})$. Similarly, we may keep the

$$\text{pulling force } \; P := g m_\mathrm{we}$$

constant and then set $m_\mathrm{leg} = m_\mathrm{b} = m_\mathrm{we} = 0$. Moreover, we choose a simple quadratic energy potential, where it is important to couple the leg distance $d = x_\mathrm{R} - x_\mathrm{L}$ and the angle ψ, namely

$$\Phi(d, \psi) = \frac{a}{2} d^2 + \frac{b}{2} \psi^2 - c\, d\psi \quad \text{with } a,\ b,\ ab{-}c^2 \ > 0.$$

Hence, the trivial symmetric state $(x_\mathrm{R}{-}x_\mathrm{L}, \psi) = (0, 0)$ is stable. It is important to have $c \neq 0$ (we choose $c > 0$ without loss of generality), which reflects the fact of symmetry breaking for the walking toy: the tilt angle restoring force is $\partial_\psi \Phi(d, \psi) = b\psi - cd$, so if $d > 0$ (right leg before left one) then there is a stronger tendency to fall to the left than to fall to the right.

The simplified system now takes the form

$$I_\mathrm{b}\ddot{\psi} \qquad + \delta\dot{\psi} \qquad + b\psi - c(x_\mathrm{R}{-}x_\mathrm{L}) = 0, \tag{14a}$$

$$\big(\rho{+}H_\mathrm{R}(\psi)\big)\,\mathrm{Sign}(\dot{x}_\mathrm{R}) + a(x_\mathrm{R}{-}x_\mathrm{L}) - c\psi = P, \tag{14b}$$

$$\big(\rho{+}H_\mathrm{L}(\psi)\big)\,\mathrm{Sign}(\dot{x}_\mathrm{L}) - a(x_\mathrm{R}{-}x_\mathrm{L}) + c\psi = P. \tag{14c}$$

The equations (14b) and (14c) for x_R and x_L, respectively, are simple play operators (cf. [3, 22]), however the thresholds $\rho + H_{\mathrm{R,L}}(\psi(t))$ vary in time and are even influenced by $\boldsymbol{x}$ through (14a).

Nevertheless, we will be able to reduce this coupled system to an oscillator for ψ involving a hysteresis operator induced by the relations for x_R and x_L. For this we first observe that the relations (14b) and (14c) restrict the leg distance $d(t) := x_{\mathrm{R}}(t) - x_{\mathrm{L}}(t)$ because of $\mathrm{Sign}(\dot{x}_{\mathrm{R,L}}) \in [-1, 1]$ as follows:

$$
\begin{aligned}
g(t) \in [G_{\mathrm{R}}^{-}(\psi), G_{\mathrm{R}}^{+}(\psi)] &\cap [G_{\mathrm{L}}^{-}(\psi), G_{\mathrm{L}}^{+}(\psi)] \text{ with} \\
G_{\mathrm{R}}^{\pm}(\psi) &= \frac{1}{a}\Big(P + c\psi \pm \big(\rho + H_{\mathrm{R}}(\psi)\big)\Big), \\
G_{\mathrm{L}}^{\pm}(\psi) &= \frac{1}{a}\Big(-P + c\psi \pm \big(\rho + H_{\mathrm{L}}(\psi)\big)\Big).
\end{aligned}
$$

We now explain that for a given continuous function $t \mapsto \psi(t)$ there is a hysteresis operator $\mathscr{H}$ such that the output $d(t) = \mathscr{H}[\psi(\cdot)](t)$ is explicitly given through the boundary curves $G^{+} > G^{-}$ via the formulas

$$
G^{+}(\psi) := \min\{G_{\mathrm{L}}^{+}(\psi), G_{\mathrm{R}}^{+}(\psi)\} \quad \text{and } G^{-}(\psi) := \max\{G_{\mathrm{L}}^{-}(\psi), G_{\mathrm{R}}^{-}(\psi)\}.
$$

Of most interest are the local minimum of G^{+} at $\psi_1 > 0$ and the local maximum of G^{-} at $-\psi_1 < 0$ (see $\psi_1 = 1$ in Fig. 7). For simplicity, we choose constants $\psi_*, H_* > 0$ with $H_* > 2c\psi_*$ and restrict to the piecewise affine case

$$
H_{\mathrm{R}}(\psi) = \begin{cases} 0 & \text{for } \psi \le -\psi_*, \\ H_*(\psi + \psi_*)/(2\psi_*) & \text{for } |\psi| \le \psi_*, \\ H_* & \text{for } \psi \ge \psi. \end{cases}
$$

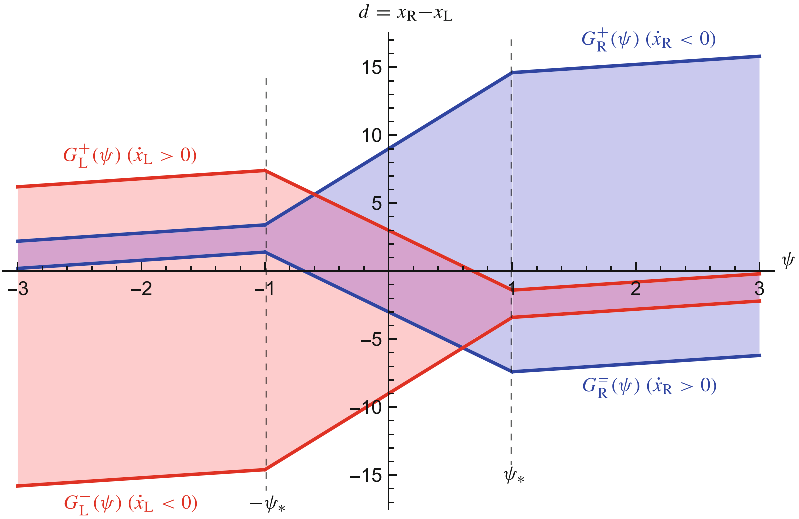

Fig. 7 Sketch of the sets $[G_{\mathrm{R}}^{-}(\psi), G_{\mathrm{R}}^{+}(\psi)]$ and $[G_{\mathrm{L}}^{-}(\psi), G_{\mathrm{L}}^{+}(\psi)]$. The solutions have to stay inside the intersection of the two shaded regions. We have $\dot{d} \le 0$ at the upper curve $G^{+} : \psi \mapsto \min\{G_{\mathrm{L}}^{+}(\psi), G_{\mathrm{R}}^{+}(\psi)\}$ and $\dot{d} \ge 0$ at the lower curve $G^{-} : \psi \mapsto \max\{G_{\mathrm{L}}^{-}(\psi), G_{\mathrm{R}}^{-}(\psi)\}$. Between these two curves we have $\dot{d} \equiv 0$

Thus, we can calculate the local minimum of G^+ and the local maximum of G^- explicitly, namely

$$(\psi_*, -\Xi) \text{ and } (-\psi_*, \Xi) \quad \text{with } \Xi := \frac{1}{a}\big(P - \rho - c\psi_*\big),$$

where we further assume $\Xi > 0$ (i.e. $P > \rho + c\psi_*$) and $H_* > 2P$.

4.4 Restriction to Simple Period Motions

We now restrict to a special period motion where the hysteresis operator can be replaced by an ordinary function, namely in the region

$$\psi \in \big[-\psi_2, \psi_2\big] \text{ with } \psi_2 := 2\rho/c + \psi_*.$$

where we set $d(t) = \mathscr{G}(\psi(t), \dot\psi(t))$ with

$$\mathscr{G}(\psi, \dot\psi) = \begin{cases} \Gamma(\psi) & \text{if } \dot\psi \geq 0, \\ -\Gamma(-\psi) & \text{if } \dot\psi < 0. \end{cases} \quad \text{with } \Gamma(\psi) := \begin{cases} \Xi & \text{for } \psi \in [-\psi_2, \psi_3], \\ G^+_{\mathrm{L}}(\psi) & \text{for } [\psi_3, \psi_*], \\ -\Xi & \text{for } \psi \in [\psi_*, \psi_2], \end{cases}$$

where ψ_3 is the unique solution of $\Xi = G^+_{\mathrm{L}}(\psi)$ in $[0, \psi_*]$. (Note that $G^+_{\mathrm{L}}(0) = (\rho + H_*/2 - D)/a > 0$ and $G^+_{\mathrm{L}}(\psi_*) = -\Xi < 0$.)

Thus, we have eliminated all dependence on the variables x_{R} and x_{L} and are left with a nonlinear oscillator equation for ψ, namely

$$I_{\mathrm{b}}\ddot\psi + \delta\dot\psi + b\psi - c\mathscr{G}(\psi, \dot\psi) = 0.$$

Note that this is a piecewise linear equation, where $\mathscr{G}$ switches between the two constant values $\pm\Xi$ with some linear transition region in between (Fig. 8). The point is that this switching feeds energy into the system which may compensate the damping through $\delta > 0$.

It is now possible to show that there are suitable parameters such that this equation has a periodic orbit. This can be done in a similar way using Poincaré sections as in the previous section. We refer to subsequent work for precise

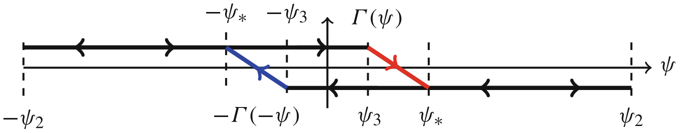

Fig. 8 Two branches of the function $\mathscr{G}(\psi, \dot\psi)$, namely $\Gamma(\psi)$ for $\dot\psi > 0$ and $-\Gamma(-\psi)$ for $\dot\psi < 0$

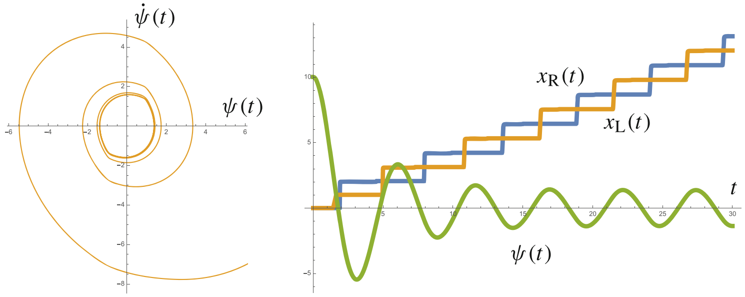

Fig. 9 Simulation for the simplified system (14). Left: $(\psi(t), \dot\psi(t)$ spirals towards a stable limit cycle. Right: The functions $\psi(t)$, $x_\mathrm{R}(t)$, and $x_\mathrm{L}(t)$ show periodic behavior up to a linear translational mode for $x_{\mathrm{R,L}}$

statements and proofs. We conclude with some numerical results, displayed in Fig. 9, that show the convergence into a stable periodic orbit for ψ and $\boldsymbol{x}(t)-v(t,t)$ with a suitable walking speed $v > 0$.

Acknowledgements The authors is grateful to Martin Heida and Elias Pipping for stimulating discussions. The research was partially supported by DFG via the project B01 *Fault networks and scaling properties of deformation accumulation* within the SFB 1114 *Scaling Cascades in Complex Systems*.

References

1. Abe, Y., Kato, N.: Complex earthquake cycle simulations using a two-degree-of-freedom spring-block model with a rate- and state-friction law. Pure Appl. Geophys. **170**(5), 745–765 (2013)
2. Brokate, M., Krejčí, P., Schnabel, H.: On uniqueness in evolution quasivariational inequalities. J. Convex Anal. **11**, 111–130 (2004)
3. Brokate, M., Sprekels, J.: Hysteresis and Phase Transitions. Springer, New York (1996)
4. DeSimone, A., Gidoni, P., Noselli, G.: Liquid crystal elastomer strips as soft crawlers. J. Mech. Phys. Solids **84**, 254–272 (2015)
5. Gidoni, P., DeSimone, A.: On the genesis of directional friction through bristle-like mediating elements crawler. arXiv:1602.05611 (2016)
6. Gidoni, P., DeSimone, A.: Stasis domains and slip surfaces in the locomotion of a bio-inspired two-segment crawler. Meccanica **52**(3), 587–601 (2017)
7. Gidoni, P., Noselli, G., DeSimone, A.: Crawling on directional surfaces. Int. J. Non-Linear Mech. **61**, 65–73 (2014)
8. Heida, M., Mielke, A.: Averaging of time-periodic dissipation potentials in rate-independent processes. Discr. Cont. Dynam. Syst. Ser. S **10**(6), 1303–1327 (2017)
9. Heida, M., Mielke, A., Pipping, E.: Rate-and-state friction from a thermodynamical viewpoint. In preparation (2017)
10. Mielke, A.: Emergence of rate-independent dissipation from viscous systems with wiggly energies. Contin. Mech. Thermodyn. **24**(4), 591–606 (2012)
11. Mielke, A., Rossi, R.: Existence and uniqueness results for a class of rate-independent hysteresis problems. Math. Models Meth. Appl. Sci. **17**(1), 81–123 (2007)

12. Mielke, A., Roubíček, T.: Rate-Independent Systems: Theory and Application. Applied Mathematical Sciences, vol. 193. Springer, New York (2015)
13. Pfeiffer, F.: Mechanische Systeme mit unstetigen Übergängen. Ingenieur-Archiv **54**, 232–240 (1984). (In German)
14. Pipping, E.: Existence of long-time solutions to dynamic problems of viscoelasticity with rate-and-state friction. arXiv:1703.04289v1 (2017)
15. Pipping, E., Kornhuber, R., Rosenau, M., Oncken, O.: On the efficient and reliable numerical solution of rate-and-state friction problems. Geophys. J. Int. **204**(3), 1858–1866 (2016)
16. Popov, V.L., Gray, J.A.T.: Prandtl-Tomlinson model: History and applications in friction, plasticity, and nanotechnologies. Z. Angew. Math. Mech. **92**(9), 692–708 (2012)
17. Popov, V.L.: Contact Mechanics and Friction. Springer, New York (2010)
18. Prandtl, L.: Gedankenmodel zur kinetischen Theorie der festen Körper. Z. Angew. Math. Mech. **8**, 85–106 (1928)
19. Radtke, M., Netz, R.R.: Shear-induced dynamics of polymeric globules at adsorbing homogeneous and inhomogeneous surfaces. Euro. Phys. J. E **37**(20), 11 (2014)
20. Roubíček, T.: A note about the rate-and-state-dependent friction model in a thermodynamical framework of the biot-type equation. Geophys. J. Int. **199**(1), 286–295 (2014)
21. Tomlinson, G.A.: A molecular theory of friction. Phil. Mag. **7**, 905–939 (1929)
22. Visintin, A.: Differential Models of Hysteresis. Springer, Berlin (1994)

Numerical Approach to a Model for Quasistatic Damage with Spatial BV-Regularization

Sören Bartels, Marijo Milicevic, and Marita Thomas

Abstract We address a model for rate-independent, partial, isotropic damage in quasistatic small strain linear elasticity, featuring a damage variable with spatial BV-regularization. Discrete solutions are obtained using an alternate time-discrete scheme and the *Variable*-ADMM algorithm to solve the constrained nonsmooth optimization problem that determines the damage variable at each time step. We prove stability of the method and show that a discrete version of a semistable energetic formulation of the rate-independent system holds. Moreover, we present our numerical results for two benchmark problems.

1 The Damage Model, Its Solution Concept, and Our Results

By damage evolution we understand the formation and growth of cracks and voids in the microstructure of a solid material. This process is monitored over a time interval $[0, \mathsf{T}]$ for a body with reference configuration $\Omega \subset \mathbb{R}^d$, $d > 1$. In the spirit of generalized standard materials [27] and continuum damage mechanics [32, 33] this degradation phenomenon is modeled by a volumetric internal damage variable $z : [0, \mathsf{T}] \times \Omega \to [0, 1]$ which is incorporated into the constitutive law in order to reflect the changes of the elastic behavior due to damage. It is assumed that the length scale of the specimen of the considered material is much larger than that of the respective *reference volume*. The reference volume of a material is a characteristic volume such that all relevant properties of the material are comprised in this amount of material and such that the material can be regarded as homogeneous if it is considered in a much larger length scale than the length

S. Bartels · M. Milicevic
Department of Applied Mathematics, Mathematical Institute, University of Freiburg, Freiburg i. Br., Germany
e-mail: bartels@mathematik.uni-freiburg.de; marijo.milicevic@mathematik.uni-freiburg.de

M. Thomas (✉)
Weierstrass-Institute for Applied Analysis and Stochastics, Berlin, Germany
e-mail: marita.thomas@wias-berlin.de

E. Rocca et al. (eds.), *Trends in Applications of Mathematics to Mechanics*,
Springer INdAM Series 27, https://doi.org/10.1007/978-3-319-75940-1_9

scale of the reference volume. The value $z(t,x)$ at $(t,x) \in [0,\mathsf{T}] \times \Omega$ can then be understood as the undamaged fraction of the reference volume at time t located in $x \in \Omega$.

The evolution of the damage variable is driven by time-dependent external loads, which cause the deformation of the body and increase its stresses. To relax, damage evolves and thus turns stored energy into dissipated energy. These two energy contributions can be described by an energy functional $\mathscr{E}$ and a dissipation potential $\mathscr{R}$. In literature many different assumptions have been made with regard to the growth properties of the two functionals, which directly affect the regularity properties of the damage variable with regard to time and space. In this way the contributions to damage processes in mathematical and engineering literature can be divided into two major classes: One class considers the evolution of damage as a rate-dependent phenomenon, mostly modeled by a viscous dissipation with quadratic growth, cf., e.g., [7, 8, 17, 18, 29, 46], and a further class understands damage as a rate-independent process described by a positively 1-homogeneous dissipation potential, cf., e.g., [11, 15, 28, 35, 42, 53–55]. While the first growth property leads comparably smooth evolution in time settled in $L^2(\Omega)$, the latter only provides bounded variations in time, so that the damage variable may jump in time. Indeed, the use of a rate-independent model, resp. the neglection of rate-effects, is also seen as a feasible approximation for certain damage processes observed in experiments, cf., e.g., [25]. We will follow the latter concept and consider the positively 1-homogeneous dissipation potential $\mathscr{R} : \mathbf{Z} \to \mathbb{R} \cup \{\infty\}$,

$$\mathscr{R}(v) := \int_\Omega \mathrm{R}(v)\,\mathrm{d}x, \quad \text{with } \mathrm{R}(v) := \begin{cases} \varrho|v|\,\mathrm{d}x & \text{if } v \in (-\infty, 0], \\ +\infty & \text{if } v > 0 \end{cases} \tag{1a}$$

$$\text{with } \mathbf{Z} := L^1(\Omega)\,, \tag{1b}$$

and with a constant dissipation rate $\varrho > 0$. Due to the convention $z = 1$ for the unbroken and $z = 0$ for the broken state of the material, the dissipation potential ensures the unidirectionality of the process and thus prevents healing of the material.

Also for the energy functional $\mathscr{E}$ different regularity assumptions have been made for the damage variable: By now, it has become a well-accepted approach to incorporate damage gradients into the energy, in order to account for nonlocal effects of damage from a physical point of view, and to benefit from its regularizing effect in the mathematical analysis and numerical simulations. The vast majority of contributions considers a damage gradient with growth of power $p = 2$ [2, 7, 8, 17, 18, 26, 34, 37–39, 52, 56]. For technical reasons, sometimes also $p > d$ is chosen, cf., e.g., [29, 41, 46]. It has to be remarked that this choice has direct influence on the effects of damage that can be observed with this model: For gradient regularizations of this type, mathematically, the damage variable is an element in a Sobolev space, and transitions between damaged and undamaged material phases have to be smooth and thus have to take place in zones of a certain positive width. The assumption $p > d$ enforces that the damage variable even has to be continuous in space. Yet, from own experience one can also observe situations where the transition between damaged and undamaged regions is very sharp. This effect cannot be described by a regularization in Sobolev spaces. Therefore it is the aim of this work to contribute to the toolbox for the investigation of damage processes with a model that allows for sharp transitions between damaged and undamaged material phases. To capture this effect, but still to benefit from regularizing effects of gradients, we propose to replace the Sobolev-gradient by a BV-gradient. More precisely, we shall consider the function spaces

$$\mathbf{U} := \{v \in H^1(\Omega, \mathbb{R}^d),\ v = 0 \text{ on } \Gamma_D \text{ in trace sense}\}\,, \tag{2a}$$

$$\mathbf{X} := BV(\Omega)\,, \tag{2b}$$

and an energy functional $\widehat{\mathscr{E}} : [0, \mathsf{T}] \times \mathbf{U} \times \mathbf{X} \to \mathbb{R} \cup \{\infty\}$ of the form

$$\begin{aligned}\widehat{\mathscr{E}}(t, u, z) := \frac{1}{2}\int_\Omega f(z)\big(\lambda\big|\operatorname{tr} e(u + g(t))\big|^2 + 2\mu|e(u + g(t))|^2\big)\,\mathrm{d}x \\ + \kappa|\mathrm{D}z|(\Omega) + \int_\Omega I_{[0,1]}(z)\,\mathrm{d}x - \int_{\Gamma_{\text{Neu}}} u_{\text{Neu}}(t)\cdot(u + g(t))\,\mathrm{d}s\end{aligned} \tag{3}$$

with the Lamé constants $\lambda, \mu > 0$, $e(u) := \frac{1}{2}(\nabla u + \nabla u^\top)$ the linear-strain tensor, $g : [0, \mathsf{T}] \times \Omega \to \mathbb{R}^d$ a suitable extension of a given Dirichlet datum into the domain Ω and $u_{\text{Neu}} : [0, \mathsf{T}] \times \Gamma_{\text{Neu}} \to \mathbb{R}^d$ a given surface loading acting along the Neumann-boundary Γ_{Neu}. Due to the mapping properties of the monotonously increasing function $f : [0, 1] \to [a, b]$ with constants $0 < a < b$ the model will capture partial damage only: It is $f(0) \geq a$ and hence, even in the state of maximal damage the solid has the ability to counteract external loadings with suitable stresses and displacements; for models allowing for complete damage, where this property is lost, we refer, e.g., to [9, 30, 43]. The compactness information needed to handle the product of $f(z)$ and quadratic terms in e is provided by the total variation $|\mathrm{D}z|(\Omega)$ of z in Ω, weighted with a constant $\kappa > 0$,. Finally, the indicator function $I_{[0,1]}$ confines the values of z to the interval $[0, 1]$, i.e., $I_{[0,1]}(z) = 0$ if $z \in [0, 1]$ and $I_{[0,1]}(z) = \infty$ otherwise. In view of (1b), we will work with the extended energy functional $\mathscr{E} : [0, \mathsf{T}] \times \mathbf{U} \times \mathbf{Z} \to \mathbb{R} \cup \{\infty\}$

$$\mathscr{E}(t, u, z) := \begin{cases} \widehat{\mathscr{E}}(t, u, z) & \text{if } (u, z) \in \mathbf{U} \times \mathbf{X}, \\ \infty & \text{otherwise.} \end{cases} \tag{4}$$

It is the aim of this paper to study the existence of solutions for the rate-independent system $(\mathbf{U} \times \mathbf{Z}, \mathscr{E}, \mathscr{R})$ given by (2), (4), (1a) by proving the convergence of a numerical method. For this, we will impose a partition $\Pi_N := \{t_N^k,\ k \in \{0, 1, \ldots, N\}, 0 = t_N^0 < \ldots < t_N^N = \mathsf{T}\}$ of the time-interval $[0, \mathsf{T}]$ and a space discretization in terms of $P1$ finite elements, yielding finite-element spaces

$\mathbf{U}_h, \mathbf{X}_h$. At each time-step $t_N^k \in \Pi_N$, we will determine approximate solutions in $\mathbf{U}_h, \mathbf{X}_h$ via an alternating minimization scheme, i.e., starting from an approximation $(u_{0h}, z_{0h}) \in \mathbf{U}_h \times \mathbf{X}_h$ of the initial datum (u_0, z_0) at t_N^0, we alternatingly compute for given $(u_{Nh}^0, z_{Nh}^0) = (u_{0h}, z_{0h})$

$$u_{Nh}^k = \operatorname{argmin}_{u \in \mathbf{U}_h} \mathscr{E}(t_k, u, z_{Nh}^{k-1}), \tag{5a}$$

$$z_{Nh}^k \in \operatorname{argmin}_{z \in \mathbf{X}_h} \mathscr{E}(t_k, u_{Nh}^k, z) + \mathscr{R}(z - z_{Nh}^{k-1})\,. \tag{5b}$$

While the computation of u_{Nh}^k reduces to the solution of a linear system of equations, the computation of z_{Nh}^k requires the solution of a constrained nonsmooth minimization problem. This problem is qualitatively of the form of the Rudin-Osher-Fatemi (ROF) problem [51] for which various numerical schemes have been proposed for its iterative solution, cf., e.g., [3, 6, 13, 14, 23, 24, 31, 36, 47, 57]. We approximate a minimizer z_{Nh}^k by converting the minimization problem into a saddle-point problem and use a variant of the alternate direction method of multipliers (ADMM) [16, 19–22] recently introduced in [5] as *Variable-ADMM* for the approximate solution of the saddle-point problem.

We show stability of the alternate minimization scheme and prove that suitable interpolants constructed from (5) satisfy a discrete version of a semistable energetic formulation of the system $(\mathbf{U} \times \mathbf{Z}, \mathscr{E}, \mathscr{R})$:

Definition 1.1 (Semistable Energetic Solution) A function $q = (u, z) : [0, \mathsf{T}] \to \mathbf{U} \times \mathbf{Z}$ is called semistable energetic solution for the system $(\mathbf{U} \times \mathbf{Z}, \mathscr{E}, \mathscr{R})$, if $t \to \partial_t \mathscr{E}(t, q) \in L^1((0, \mathsf{T}))$ and if for all $s, t \in [0, \mathsf{T}]$ we have $\mathscr{E}(t, q(t)) < \infty$, if for a.a. $t \in (0, \mathsf{T})$ minimality condition (6a) is satisfied and if for all $t \in [0, \mathsf{T}]$ semistability (6b) as well as the upper energy-dissipation estimate (6c) hold true, i.e.:

$$\text{for all } \tilde{u} \in \mathbf{U} : \quad \mathscr{E}(t, u(t), z(t)) \le \mathscr{E}(t, \tilde{u}, z(t))\,, \tag{6a}$$

$$\text{for all } \tilde{z} \in \mathbf{X} : \quad \mathscr{E}(t, u(t), z(t)) \le \mathscr{E}(t, u(t), \tilde{z}) + \mathscr{R}(\tilde{z} - z(t)), \tag{6b}$$

$$\mathscr{E}(t, q(t)) + \mathscr{R}(z(t) - z(0)) \le \mathscr{E}(0, q(0)) + \int_0^t \partial_\xi \mathscr{E}(\xi, q(\xi))\, \mathrm{d}\xi\,, \tag{6c}$$

where the dissipated energy up to time t is given by the total variation induced by the dissipation potential $\mathscr{R}$ with unidirectionality constraint and, by the induced monotonicity of $z : [0, \mathsf{T}] \to \mathbf{Z}$, takes the form $\mathscr{R}(z(t) - z(0))$.

Let us note here that the alternate minimization scheme (5) directly leads to the notion of semistable energetic solutions. In the quasistatic, rate-independent setting they form a much wider class than the well-known energetic solutions, cf., e.g., [40, 42], which replace conditions (6a) & (6b) by the joint global stability condition $\forall\, (\tilde{u}, \tilde{z}) \in \mathbf{U} \times \mathbf{Z} : \mathscr{E}(t, u(t), z(t)) \le \mathscr{E}(t, \tilde{u}, \tilde{z}) + \mathscr{R}(\tilde{z} - z(t))$ and the upper energy-dissipation estimate (6c) by an energy-dissipation *balance*. In fact, the existence of energetic solutions for the above system $(\mathbf{U} \times \mathbf{Z}, \mathscr{E}, \mathscr{R})$ was investigated in [53]. As a matter of concept, energetic solutions are obtained from a time-discrete scheme with a monolithic minimization in the pair (u, z) in each time step. In the case that $\mathscr{E}(t, \cdot, \cdot)$ is jointly convex in the pair (u, z) it can be shown that semistable energetic solutions are also energetic solutions. However, this is not true if the energy functional does not enjoy the property of joint convexity. In this case it can be observed that energetic solutions tend to evolve earlier than semistable energetic solutions, cf., e.g., [48]. Indeed, many energy functionals taken from engineering literature are separately convex in the variables u and z but not jointly convex, cf. [55, Sec. 5] for examples on convexity properties of damage models.

Our paper is organized as follows: In Sect. 2 we state the main assumptions needed for the analysis. Section 3 introduces the numerical algorithms used to calculate approximate solutions in the sense of (5). We present the Variable-ADMM adjusted to the present setting, address its stability and the monotonicity of the residual and prove that the residual controls the difference between the optimal energy and the energy of the iterates. Based on this, in Sect. 4 we prove the stability of the fully discretized problem. We also show that the solutions satisfy a discrete version of the semistable energetic formulation as well as uniform apriori estimates. This is the basis for the limit passage to the notion of solution given in Definition 1.1, which, however, we do not carry out in this work. Finally, in Sect. 5 we report our numerical results for an academic example and a benchmark problem from engineering.

2 Setup and Notation

Throughout this work, we consider the time interval $[0, \mathsf{T}]$ for some time horizon $\mathsf{T} > 0$ and an open bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, with Dirichlet boundary $\Gamma_D \subset \partial\Omega$ with $(d-1)$-dimensional Hausdorff-measure $\mathscr{H}^{d-1}(\Gamma_D) > 0$. We denote by $(\cdot, \cdot)$ the L^2-inner product, by $\|\cdot\|$ the L^2-norm, and by $|\cdot|$ the Euclidean norm on $\mathbb{R}^d$. Moreover, by $\mathrm{B}([0, \mathsf{T}], \bullet)$ we denote the space of functions f mapping time into a space $\bullet$, which are bounded and defined everywhere in $[0, \mathsf{T}]$.

Regarding the given data appearing in (3) we make the following assumptions:

Assumption 2.1 (Assumptions on the Given Data)

1. *The function* $f : \mathbb{R} \to \mathbb{R}$ *is continuously differentiable and convex and such that* $f|_{[0,1]} : [0, 1] \to [a, b]$ *is monotonically increasing.*
2. *The Lamé constants satisfy* $\lambda, \mu > 0$.
3. *The extension of the Dirichlet datum is of regularity* $g \in C^1([0, \mathsf{T}], H^1(\Omega; \mathbb{R}^d))$ *with* $C_g := \|g\|_{C^1([0,\mathsf{T}],H^1(\Omega;\mathbb{R}^d))}$.
4. *The Neumann datum* u_{Neu} *is of regularity* $u_{\mathrm{Neu}} \in C^1([0, \mathsf{T}], L^2(\Gamma_{\mathrm{Neu}}; \mathbb{R}^d))$ *with* $C_{u_{\mathrm{Neu}}} := \|u_{\mathrm{Neu}}\|_{C^1([0,\mathsf{T}],L^2(\Gamma_{\mathrm{Neu}};\mathbb{R}^d))}$

Moreover, for the space discretization we will use the following notation related to **finite element spaces**: Let $(\mathscr{T}_h)_{h>0}$ be a family of triangulations of Ω where the index h denotes the mesh size $h = \max_{T\in\mathscr{T}_h} h_T$ with h_T being the diameter of the simplex T. The minimal diameter is given by $h_{\min} = \min_{T\in\mathscr{T}_h} h_T$. The sets $\mathscr{N}_h$ and E_h contain all nodes and edges, respectively, of the triangulation $\mathscr{T}_h$. We will use the finite element space of continuous, piecewise affine functions ($r = 1$) or vector fields ($r = d$), denoted by $\mathscr{S}^1(\mathscr{T}_h)^r$ and of elementwise constant vector fields $\mathscr{L}^0(\mathscr{T}_h)^d$, i.e.,

$$\mathscr{S}^1(\mathscr{T}_h)^r := \{v_h \in C(\Omega;\mathbb{R}^r) : \; v_h|_T \text{ affine for all } T \in \mathscr{T}_h\}, \tag{7a}$$

$$\mathscr{L}^0(\mathscr{T}_h)^d := \{\tilde{p}_h \in L^\infty(\Omega;\mathbb{R}^d) : \; \tilde{p}_h|_T \text{ constant for all } T \in \mathscr{T}_h\}. \tag{7b}$$

Moreover, denoting by $\mathscr{I}_h : C^0(\overline{\Omega}) \to \mathscr{S}^1(\mathscr{T}_h)$ the standard nodal interpolation operator we will consider the discrete inner products

$$(v_h, w_h)_h := \int_\Omega \mathscr{I}_h[v_h w_h]\,\mathrm{d}x = \sum_{y\in\mathscr{N}_h} \beta_y v_h(y) w_h(y) \quad \text{on } \mathscr{S}^1(\mathscr{T}_h),$$

$$(p_h, \tilde{p}_h)_w := h^d_{\min}(p_h, \tilde{p}_h) \quad \text{on } \mathscr{L}^0(\mathscr{T}_h)^d,$$

where $\beta_y = \int_\Omega \varphi_y \,\mathrm{d}x$ with φ_y the nodal basis function associated to $y \in \mathscr{N}_h$. We have the relations

$$\|v_h\| \le \|v_h\|_h \le (d+2)^{1/2}\|v_h\|, \quad \text{and} \quad \|\tilde{p}_h\|_w \le c\|\tilde{p}_h\|_{L^1(\Omega)},$$

for all $v_h \in \mathscr{S}^1(\mathscr{T}_h)$ and $\tilde{p}_h \in \mathscr{L}^0(\mathscr{T}_h)^d$, see [4, Lemma 3.9] and [12, Thm. 4.5.11]. Finally, for a sequence of step sizes $(\tau_j)_{j\in\mathbb{N}}$ and functions $(a^j)_{j\in\mathbb{N}}$ we will denote the backward difference quotient by

$$d_t a^j = \frac{a^j - a^{j-1}}{\tau_j}.$$

3 Numerical Method

We now discuss the numerical algorithms used to solve the alternate minimization problem (5) on the discrete level. With $\mathscr{S}^1(\mathscr{T}_h)^d$ and $\mathscr{S}^1(\mathscr{T}_h)$ from (7) we set $\mathbf{U}_h := \mathscr{S}^1(\mathscr{T}_h)^d \cap \{v \in C(\overline{\Omega};\mathbb{R}^d),\, v = 0 \text{ on } \Gamma_D\} \subset H^1_D(\Omega;\mathbb{R}^d)$ in (5a) and $\mathbf{X}_h := \mathscr{S}^1(\mathscr{T}_h) \subset BV(\Omega)$ in (5b). While the minimization problem (5a) to determine u^k_{Nh} reduces to the solution of a linear system of equations, the minimization problem (5b) to find z^k_{Nh} is more difficult due to the non-differentiability of the BV-seminorm and the occurrence of non-smooth constraints in $\mathscr{E}$ and $\mathscr{R}$. We will deal with the minimization problem (5b) in Sect. 3.1 and subsequently explain the algorithm for the full alternate minimization problem in Sect. 3.2.

3.1 Minimization with Respect to z in (5b)

For the following discussion we consider a partition Π_N of $[0, \mathsf{T}]$ with $N \in \mathbb{N}$ fixed. We also keep $t_N^k \in \Pi_N$ and u_{Nh}^k the solution of (5a) fixed. For simpler notation we here write $t_k = t_N^k$, $u_h^k = u_{Nh}^k$, and $z_h^k = z_{Nh}^k$, i.e., we do not indicate the dependence of these quantities on $N \in \mathbb{N}$ fixed. We first of all note that a minimizer $z_h^k = z_{Nh}^k$ obtained in (5b) is required to satisfy $z_h^k - z_h^{k-1} \leq 0$ almost everywhere in Ω since otherwise $\mathscr{R}(z_h^k - z_h^{k-1})$ is infinite. Since $z_h^k, z_h^{k-1} \in \mathbf{X}_h = \mathscr{S}^1(\mathscr{T}_h)$ are globally continuous and piecewise affine this is equivalent to $z_h^k(x) \leq z_h^{k-1}(x)$ for all $x \in \mathscr{N}_h$. Particularly, $|z_h^k(x) - z_h^{k-1}(x)| = z_h^{k-1}(x) - z_h^k(x)$. Hence, letting for $k \geq 1$

$$K_k := \{v_h \in \mathscr{S}^1(\mathscr{T}_h) : \ 0 \leq v_h(x) \leq z_h^{k-1}(x) \ \forall \ x \in \mathscr{N}_h\} \tag{8}$$

we define the auxiliary functional $\widetilde{\mathscr{E}}(t_k, \cdot, \cdot) : \mathbf{U}_h \times \mathbf{X}_h \to \mathbb{R} \cup \{\infty\}$,

$$\begin{aligned}\widetilde{\mathscr{E}}(t_k, u_h, z_h) := \frac{1}{2}\int_\Omega f(z_h)\big(\lambda|\operatorname{tr} e(u_h + g(t_k))|^2 + 2\mu|e(u_h + g(t_k))|^2\big)\,\mathrm{d}x \\ - \int_{\Gamma_{\mathrm{Neu}}} u_{\mathrm{Neu}}(t_k) \cdot (u_h + g(t_k))\,\mathrm{d}s + \kappa\int_\Omega |\nabla z_h|\,\mathrm{d}x + I_{K_k}(z_h).\end{aligned}$$

We obtain that minimality property (5b) is equivalent to

$$z_h^k \in \operatorname{argmin}_{z_h \in \mathbf{X}_h} \widetilde{\mathscr{E}}(t_k, u_h^k, z_h) - \rho(z_h, 1).$$

In order to approximate a minimizer z_h^k we consider for $\tau_j > 0$ and $\mathbb{C}A = \lambda \operatorname{tr}(A)I + 2\mu A$ for $A \in \mathbb{R}^{d \times d}$ the augmented Lagrangian functional

$$\begin{aligned}L_h^k(z_h, p_h, s_h; \eta_h, \zeta_h) := \frac{1}{2}\int_\Omega f(z_h)e(u_h^k + g(t_k)) : \mathbb{C}e(u_h^k + g(t_k))\,\mathrm{d}x - \rho(z_h, 1) \\ + \kappa\int_\Omega |p_h|\,\mathrm{d}x + (\eta_h, \nabla z_h - p_h)_w + \frac{\tau_j}{2}\|\nabla z_h - p_h\|_w^2 \\ + I_{K_k}(s_h) + (\zeta_h, z_h - s_h)_h + \frac{\tau_j}{2}\|z_h - s_h\|_h^2.\end{aligned}$$

For the approximation of a minimizer z_h^k we use the following algorithm [5] which generalizes the alternating direction method of multipliers (ADMM) established and analyzed, e.g., in [16, 19–22] by using variable step sizes.

Algorithm 3.1 (Variable-ADMM) *Choose* $z_h^0 = z_h^{k-1}$, $\eta_h^0 = 0$ *and* $\zeta_h^0 = 0$. *Choose* $\underline{\tau}, \overline{\tau} > 0$ *with* $\underline{\tau} \leq \overline{\tau}$, $\delta \in (0, 1)$, $\underline{\gamma}, \overline{\gamma} \in (0, 1)$ *with* $\underline{\gamma} \leq \overline{\gamma}$, *and* $\overline{R} \gg 1$. *Set* $j = 1$.

(1) Set $\gamma_1 = \underline{\gamma}$, $\tau_1 = \overline{\tau}$ *and* $R_0 = \overline{R}$.

(2) Compute a minimizer $(p_h^j, s_h^j) \in \mathscr{L}^0(\mathscr{T}_h)^d \times \mathscr{S}^1(\mathscr{T}_h)$ *of*

$$(p_h, s_h) \mapsto L_h^k(z_h^{j-1}, p_h, s_h; \eta_h^{j-1}, \zeta_h^{j-1}).$$

(3) Compute a minimizer $z_h^j \in \mathscr{S}^1(\mathscr{T}_h)$ *of*

$$z_h \mapsto L_h^k(z_h, p_h^j, s_h^j; \eta_h^{j-1}, \zeta_h^{j-1}).$$

(4) Update

$$\begin{aligned}\eta_h^j &= \eta_h^{j-1} + \tau_j(\nabla z_h^j - p_h^j),\\ \zeta_h^j &= \zeta_h^{j-1} + \tau_j(z_h^j - s_h^j).\end{aligned}$$

(5) Define

$$R_j = \big(\|\eta_h^j - \eta_h^{j-1}\|_w^2 + \tau_j^2\|\nabla(z_h^j - z_h^{j-1})\|_w^2 + \|\zeta_h^j - \zeta_h^{j-1}\|_h^2 + \tau_j^2\|z_h^j - z_h^{j-1}\|_h^2\big)^{1/2}.$$

(6) Stop if R_j *is sufficiently small.*

(7) Define $(\tau_{j+1}, \gamma_{j+1})$ *as follows:*

- *If* $R_j \le \gamma_j R_{j-1}$ *or if* $\tau_j = \underline{\tau}$ *and* $\gamma_j = \overline{\gamma}$ *set*

$$\tau_{j+1} = \tau_j \quad \textit{and} \quad \gamma_{j+1} = \gamma_j.$$

- *If* $R_j > \gamma_j R_{j-1}$ *and* $\tau_j > \underline{\tau}$ *set*

$$\tau_{j+1} = \max\{\delta\tau_j, \underline{\tau}\} \quad \textit{and} \quad \gamma_{j+1} = \gamma_j.$$

- *If* $R_j > \gamma_j R_{j-1}$, $\tau_j = \underline{\tau}$ *and* $\gamma_j < \overline{\gamma}$ *set*

$$\tau_{j+1} = \overline{\tau},\ \gamma_{j+1} = \min\Big\{\frac{\gamma_j + 1}{2}, \overline{\gamma}\Big\},\ u^j = u^0 \textit{ and } \lambda^j = \lambda^0.$$

(8) Set $j = j + 1$ *and continue with (2).*

In the following proposition we prove that the iterates are bounded, that the algorithm terminates and that the residuals R_j are monotonically decreasing. To this extent we define the functionals

$$F(p_h) = \kappa \int_\Omega |p_h| \,\mathrm{d}x, \quad H(s_h) = I_{K_k}(s_h),$$

$$G(z_h) = \frac{1}{2}\int_\Omega f(z_h) e(u_h^k + g(t_k)) : \mathbb{C} e(u_h^k + g(t_k)) \,\mathrm{d}x - \rho(z_h, 1).$$

Proposition 3.1 (Termination of Algorithm 3.1 and Monotonicity of Residuals) *Let $(z_h, p_h, s_h; \eta_h, \zeta_h)$ be a saddle-point for L_h^k. For the iterates $(z_h^j, p_h^j, s_h^j; \eta_h^j, \zeta_h^j)$, $j \geq 0$, of Algorithm 3.1, the corresponding differences $\delta_\eta^j := \eta_h - \eta_h^j$, $\delta_\zeta^j := \zeta_h - \zeta_h^j$, $\delta_p^j := p_h - p_h^j$, $\delta_s^j := s_h - s_h^j$, and $\delta_z^j := z_h - z_h^j$, and the distance*

$$D_j^2 = \|\delta_\eta^j\|_w^2 + \|\delta_\zeta^j\|_h^2 + \tau_j^2 \|\nabla \delta_z^j\|_w^2 + \tau_j^2 \|\delta_z^j\|_h^2,$$

we have that for every $J \geq 1$ it holds

$$D_J^2 + \sum_{j=1}^{J} R_j^2 \leq D_0^2.$$

In particular, $R_j \to 0$ as $j \to \infty$ and Algorithm 3.1 terminates. Moreover, we have

$$R_{j+1}^2 \leq R_j^2,$$

i.e., the residual is non-increasing.

Proof The optimality conditions for a saddle-point of L_h^k are given by

$$\begin{aligned}
(\eta_h, \tilde{p}_h - p_h)_w + F(p_h) &\leq F(\tilde{p}_h) \quad \forall\, \tilde{p}_h \in \mathscr{L}^0(\mathscr{T}_h)^d,\\
(\zeta_h, r_h - s_h)_h + H(s_h) &\leq H(r_h) \quad \forall\, r_h \in \mathscr{S}^1(\mathscr{T}_h),\\
-(\eta_h, \nabla(w_h - z_h))_w - (\zeta_h, w_h - z_h)_h + G(z_h) &\leq G(w_h) \quad \forall\, w_h \in \mathscr{S}^1(\mathscr{T}_h),
\end{aligned} \tag{9}$$

and $p_h = \nabla z_h$ and $s_h = z_h$. On the other hand, with $\widetilde{\eta}_h^j = \eta_h^{j-1} + \tau_j(\nabla z_h^{j-1} - p_h^j)$ and $\widetilde{\zeta}_h^j = \zeta_h^{j-1} + \tau_j(z_h^{j-1} - s_h^j)$, the optimality conditions for the iterates of Algorithm 3.1 read

$$\begin{aligned}
(\widetilde{\eta}_h^j, \tilde{p}_h - p_h^j)_w + F(p_h^j) &\leq F(\tilde{p}_h) \quad \forall\, \tilde{p}_h \in \mathscr{L}^0(\mathscr{T}_h)^d,\\
(\widetilde{\zeta}_h^j, r_h - s_h^j)_h + H(s_h^j) &\leq H(r_h) \quad \forall\, r_h \in \mathscr{S}^1(\mathscr{T}_h),\\
-(\eta_h^j, \nabla(w_h - z_h^j))_w - (\zeta_h^j, w_h - z_h^j)_h + G(z_h^j) &\leq G(w_h) \quad \forall\, w_h \in \mathscr{S}^1(\mathscr{T}_h).
\end{aligned} \tag{10}$$

Testing (9) and (10) with $(\tilde{p}_h, r_h, w_h) = (p_h^j, s_h^j, z_h^j)$ and $(\tilde{p}_h, r_h, w_h) = (p_h, s_h, z_h)$, respectively, and adding corresponding inequalities gives

$$\begin{aligned}
(\widetilde{\eta}_h^j - \eta_h, p_h - p_h^j)_w &\leq 0,\\
(\widetilde{\zeta}_h^j - \zeta_h, s_h - s_h^j)_h &\leq 0,\\
(\eta_h - \eta_h^j, \nabla(z_h - z_h^j))_w + (\zeta_h - \zeta_h^j, z_h - z_h^j)_h &\leq 0.
\end{aligned}$$

The rest of the proof of the first estimate is analogous to the proof of [5, Thm. 3.7].

The proof of the monotonicity follows by testing (10) at iterations j and $j+1$ with $(\tilde{p}_h, r_h, w_h) = (p_h^{j+1}, s_h^{j+1}, z_h^{j+1})$ and $(\tilde{p}_h, r_h, w_h) = (p_h^j, s_h^j, z_h^j)$, respectively, and adding the inequalities, which gives

$$\begin{aligned} 0 \leq\ & -(\widetilde{\eta}_h^{j+1} - \widetilde{\eta}_h^j, p_h^j - p_h^{j+1})_w - (\eta_h^j - \eta_h^{j+1}, \nabla(z_h^j - z_h^{j+1}))_w \\ & - (\widetilde{\zeta}_h^{j+1} - \widetilde{\zeta}_h^j, s_h^j - s_h^{j+1})_h - (\zeta_h^j - \zeta_h^{j+1}, z_h^j - z_h^{j+1})_h. \end{aligned}$$

The monotonicity then follows as in the proof of [5, Prop. 3.11]. ■

In the next step, we show that the residual R_j controls the difference in the objective values.

Lemma 3.1 *Let $(z_h, p_h, s_h; \eta_h, \zeta_h)$ be a saddle-point of L_h^k. Then there exists a constant $C_0 > 0$ such that we have for any $j \geq 1$*

$$\widetilde{\mathscr{E}}(t_k, u_h^k, s_h^j) + \mathscr{R}(s_h^j - z_h^{k-1}) - \widetilde{\mathscr{E}}(t_k, u_h^k, z_h) - \mathscr{R}(z_h - z_h^{k-1}) \leq C_0 R_j. \tag{11}$$

Proof We use the short notation δ_η^j, δ_ζ^j, δ_p^j, δ_s^j and δ_z^j as in Proposition 3.1. Testing (10) with $(\tilde{p}_h, r_h, w_h) = (p_h, s_h, z_h)$, adding the inequalities, noting that $p_h = \nabla z_h$ and $s_h = z_h$ and using $\eta_h^j - \widetilde{\eta}_h^j = \tau_j \nabla(z_h^j - z_h^{j-1})$, $\zeta_h^j - \widetilde{\zeta}_h^j = \tau_j (z_h^j - z_h^{j-1})$ we obtain

$$\begin{aligned} & F(p_h^j) + G(z_h^j) + H(s_h^j) - F(p_h) - G(z_h) - H(s_h) \\ \leq\ & -(\widetilde{\eta}_h^j, \delta_p^j)_w + (\eta_h^j, \nabla \delta_z^j)_w - (\widetilde{\zeta}_h^j, \delta_s^j)_h + (\zeta_h^j, \delta_z^j)_h \\ =\ & -(\eta_h^j, d_t \eta_h^j)_w - \tau_j^2 (\nabla d_t \delta_z^j, \delta_p^j)_w - (\zeta_h^j, d_t \zeta_h^j)_h - \tau_j^2 (d_t \delta_z^j, \delta_s^j)_h. \end{aligned} \tag{12}$$

Testing the optimality conditions of z_h^j and z_h^{j-1} with $w_h = z_h^{j-1}$ and $w_h = z_h^j$, respectively, and adding the corresponding inequalities gives

$$0 \leq -\tau_j^2 (d_t \eta_h^j, \nabla d_t z_h^j)_w - \tau_j^2 (d_t \zeta_h^j, d_t z_h^j)_h.$$

Using $d_t \eta_h^j = \nabla z_h^j - p_h^j$ and $d_t \zeta_h^j = z_h^j - s_h^j$ and inserting $p_h = \nabla z_h$ and $s_h = z_h$ on the right-hand side gives

$$0 \leq -\tau_j^2 (\nabla \delta_z^j, \nabla d_t \delta_z^j)_w + \tau_j^2 (\delta_p^j, \nabla d_t \delta_z^j)_w - \tau_j^2 (\delta_z^j, d_t \delta_z^j)_h + \tau_j^2 (\delta_s^j, d_t \delta_z^j)_h. \tag{13}$$

Adding (12) and (13) we get

$$\begin{aligned} & F(p_h^j) + G(z_h^j) + H(s_h^j) - F(p_h) - G(z_h) - H(s_h) \\ \leq\ & -(\eta_h^j, d_t \eta_h^j)_w + \tau_j^2 (\nabla \delta_z^j, \nabla d_t z_h^j)_w - (\zeta_h^j, d_t \zeta_h^j)_h + \tau_j^2 (\delta_z^j, d_t z_h^j)_h \\ \leq\ & \|\eta_h^j\|_w \|d_t \eta_h^j\|_w + \tau_j^2 \|\nabla \delta_z^j\|_w \|\nabla d_t z_h^j\|_w + \|\zeta_h^j\|_h \|d_t \zeta_h^j\|_h + \tau_j^2 \|\delta_z^j\|_h \|d_t z_h^j\|_h \leq C_0 R_j, \end{aligned}$$

with C_0 being bounded due to Proposition 3.1.

Let us furthermore note that by Proposition 3.1 we have that s_h^j and z_h^j are bounded, particularly $0 \leq s_h^j \leq z_h^{k-1}$ for all $j \geq 0$. Since f is Lipschitz continuous on bounded intervals, the Hölder inequality, the Lipschitz continuity of f and the inverse estimate $\|w_h\|_{L^\infty(\Omega)} \leq h^{-d/2}\|w_h\|$ (cf. [12, Thm. 4.5.11]) yield

$$\frac{1}{2}\int_\Omega (f(s_h^j) - f(z_h^j))e(u_h^k + g(t_k)) : \mathbb{C}e(u_h^k + g(t_k))\,\mathrm{d}x \leq ch^{-d/2}\|s_h^j - z_h^j\|.$$

We finally observe that using $s_h^j \leq z_h^{k-1}$, $z_h \leq z_h^{k-1}$, the triangle inequality, the inverse estimate $\|\nabla w_h\|_{L^1(\Omega)} \leq ch^{-1}\|w_h\|_{L^1(\Omega)}$ and the equivalence of $\|\cdot\|$ and $\|\cdot\|_h$ we have

$$\begin{aligned}
&\widetilde{\mathscr{E}}(t_k, u_h^k, s_h^j) + \mathscr{R}(s_h^j - z_h^{k-1}) - \widetilde{\mathscr{E}}(t_k, u_h^k, z_h) - \mathscr{R}(z_h - z_h^{k-1})\\
&= F(p_h^j) + G(z_h^j) + H(s_h^j) - F(p_h) - G(z_h) - H(s_h) + \kappa\int_\Omega \left(|\nabla s_h^j| - |p_h^j|\right)\mathrm{d}x\\
&\quad + \frac{1}{2}\int_\Omega (f(s_h^j) - f(z_h^j))e(u_h^k + g(t_k)) : \mathbb{C}e(u_h^k + g(t_k))\,\mathrm{d}x + \rho\int_\Omega \left(z_h^j - s_h^j\right)\mathrm{d}x\\
&\leq C_0R_j + c\kappa h^{-d/2}\|\nabla z_h^j - p_h^j\|_w + c\kappa h^{-1}\|s_h^j - z_h^j\|_h + c(\rho + h^{-d/2})\|z_h^j - s_h^j\|_h\\
&\leq C_0R_j.
\end{aligned}$$

which proves the assertion. ■

Remark 3.1 In general, the iterates $(z_h^j)_{j\geq 0}$ of Algorithm 3.1 may penetrate the obstacles, i.e., $z_h^j \notin K_k$ for some $j \in \mathbb{N}$, cf. (8). Therefore, if $(z_h^{stop}, p_h^{stop}, s_h^{stop}; \eta_h^{stop}, \zeta_h^{stop})$ is the output of the algorithm, we set $z_h^k = s_h^{stop} \in K_k$ to ensure the coercivity of the bulk energy.

3.2 Alternate Minimization (5)

In order to solve the full problem (5) we apply the following scheme:

Algorithm 3.2 (Alternate Minimization) *Choose a stable initial pair* $(u_h^0, z_h^0) \in \mathscr{S}^1(\mathscr{T}_h)^d \times \mathscr{S}^1(\mathscr{T}_h)$ *and a partition* $0 = t = 0 < \ldots < t_N = \mathsf{T}$ *of the time interval and set* $k = 1$.

(1) Compute the unique minimizer u_h^k *of*

$$u_h \mapsto \widetilde{\mathscr{E}}(t_k, u_h, z_h^{k-1}).$$

(2) Compute an approximate minimizer z_h^k of

$$z_h \mapsto \widetilde{\mathscr{E}}(t_k, u_h, z_h) - \rho(z_h, 1)$$

by using Algorithm 3.1, i.e., set $z_h^k = s_h^{stop}$ with s_h^{stop} computed by Algorithm 3.1.

(3) Stop if $k = N$. Otherwise, increase $k \to k + 1$ and continue with (1).

The optimality condition for u_h^k in step (1) of the algorithm reads

$$\int_\Omega f(z_h^{k-1}) e(u_h^k) : \mathbb{C} e(v_h)\, \mathrm{d}x = -\int_\Omega e(g(t_k)) : \mathbb{C} e(v_h)\, \mathrm{d}x + \int_{\Gamma_{\mathrm{Neu}}} u_{\mathrm{Neu}}(t_k) \cdot v_h\, \mathrm{d}s$$

for all $v_h \in \mathbf{U}_h$. In our computation we replace g by $g_h = \mathscr{I}_h g$ on the right-hand side with $\mathscr{I}_h$ being the nodal interpolant and g sufficiently smooth. We further use the midpoint rule to compute for $T \in \mathscr{T}_h$ and $e \in E_h$ the integrals

$$\int_T f(z_h^{k-1})\, \mathrm{d}x, \quad \text{and} \quad \int_e u_{\mathrm{Neu}}(t_k) \cdot v_h\, \mathrm{d}s.$$

The computation of u_h^k then amounts to solving a linear system of equations with a weighted stiffness matrix.

4 Existence Result on a Discrete Level

In this section we show that suitable time-interpolants of the solutions $(u_{Nh}^k, z_{Nh}^k)_{Nh}$ obtained at each time step t_N^k via the alternate minimization problem (5) satisfy a discrete version of the semistable energetic formulation (6). To this end, with $\mathscr{S}^1(\mathscr{T}_h)^d$ and $\mathscr{S}^1(\mathscr{T}_h)$ from (7), we set in (5)

$$\mathbf{U}_h := \mathscr{S}^1(\mathscr{T}_h)^d \cap \{v \in C(\overline{\Omega}; \mathbb{R}^d),\, v = 0 \text{ on } \Gamma_D\} \text{ and } \mathbf{X}_h := \mathscr{S}^1(\mathscr{T}_h)\,. \tag{14}$$

We recall that $\mathbf{U}_h \subset H_D^1(\Omega; \mathbb{R}^d)$ and $\mathbf{X}_h \subset BV(\Omega)$ for all $h > 0$ and

$$\bigcup_h \mathbf{U}_h \subset H_D^1(\Omega; \mathbb{R}^d) \text{ densely and } \bigcup_h \mathbf{X}_h \subset BV(\Omega) \text{ densely }. \tag{15}$$

We now choose a sequence $(h(N))_{N\in\mathbb{N}}$ such that $h(N) \to 0$ as $N \to \infty$ and consider a sequence of partitions $(\Pi_N)_N$ of $[0, \mathsf{T}]$ such that the time-step size $\Delta_N \to 0$ as $N \to \infty$. With $\mathscr{E}$ from (4) we introduce the energy functionals $\mathscr{E}_N : [0, \mathsf{T}] \times \mathbf{U} \times \mathbf{Z} \to \mathbb{R} \cup \{\infty\}$,

$$\mathscr{E}_N(t, u, z) := \begin{cases} \mathscr{E}(t, u, z) \text{ if } (u, z) \in \mathbf{U}_{h(N)} \times \mathbf{X}_{h(N)}, \\ \quad \infty \quad \text{ otherwise,} \end{cases} \tag{16}$$

where the given data $g(t)$ and $u_{\mathrm{Neu}}(t)$ are replaced by suitably interpolated versions $g_N(t)$ and $u_{\mathrm{Neu}N}(t)$ in the discrete spaces, which are uniformly bounded and converge strongly to the original datum. We thus compute for every $N \in \mathbb{N}$ and $h(N) > 0$, for each $t_N^k \in \Pi_N$ a solution $(u_N^k, z_N^k) = (u_{Nh(N)}^k, z_{Nh(N)}^k)$ to (5) using Algorithm 3.2. In particular, according to Algorithm 3.1 the pair $(u_N^k, z_N^k) = (u_{Nh(N)}^k, z_{Nh(N)}^k)$ satisfies

$$\forall u \in \mathbf{U}: \ \mathscr{E}_N(t_N^k, u_N^k, z_N^{k-1}) \le \mathscr{E}_N(t_N^k, u, z_N^{k-1}), \tag{17a}$$

$$\forall z \in \mathbf{X}:$$

$$\mathscr{E}_N(t_N^k, u_N^k, z_N^k) + \mathscr{R}(z_N^k - z_N^{k-1}) \le \mathscr{E}_N(t_N^k, u_N^k, z) + \mathscr{R}(z - z_N^{k-1}) + \mathrm{TOL}(N) \tag{17b}$$

with some $h(N)$-dependent tolerance $\mathrm{TOL}(N)$, which bounds the residual R_j^h, cf. Algorithm 3.1, Step (5). In view of Lemma 3.1 a sequence $(\mathrm{TOL}(N))_N$ can be chosen such that

$$\mathrm{TOL}(N)N \to 0 \ \text{ as } N \to \infty. \tag{18}$$

We evaluate the given data in the partition $\{t_N^0, \dots, t_N^N\}$ which results in an $(N+1)$-tupel. Moreover, for any tupel $(v_N^0, \dots, v_N^N)$ we introduce the piecewise constant left-continuous (right-continuous) interpolant $\overline{v}_N$ $(\underline{v}_N)$:

$$\overline{v}_N(t) := v_N^{k+1} \text{ for all } t \in (t_N^k, t_N^{k+1}], \tag{19a}$$

$$\underline{v}_N(t) := v_N^k \text{ for all } t \in [t_N^k, t_N^{k+1}). \tag{19b}$$

Accordingly, $\overline{\mathscr{E}}$, resp. $\underline{\mathscr{E}}$, indicates that the interpolants $\overline{g_N}$ and $\overline{u_{\mathrm{Neu}}}_N$, resp. $\underline{g_N}$ and $\underline{u_{\mathrm{Neu}}}_N$ are used. In particular, thanks to Assumptions 2.1 we have for all $t \in [0, \mathsf{T}]$

$$\overline{g}_N(t) \to g(t) \text{ in } \mathbf{U} \ \ \& \ \ \overline{u_{\mathrm{Neu}}}_N(t) \to u_{\mathrm{Neu}}(t) \text{ in } L^2(\Gamma_{\mathrm{Neu}}; \mathbb{R}^d). \tag{20}$$

This puts us in the position to find the following properties of the interpolants $(\overline{u}_N, \underline{u}_N, \overline{z}_N, \underline{z}_N)$ constructed from $(u_N^k, z_N^k)_{k=0}^N$ via (19):

Theorem 4.1 (Discrete Version of (6) and Apriori Estimates) *Let the assumptions of Sect. 2 hold true and keep $N \in \mathbb{N}$ fixed. For each $k \in \{0, 1, \dots, N\}$ let (u_N^k, z_N^k) satisfy (17). Then the corresponding interpolants $(\overline{u}_N, \underline{u}_N, \overline{z}_N, \underline{z}_N)$ obtained via (19), fulfill the following discrete version of (6) for all $t \in [0, \mathsf{T}]$:*

$$\textit{for all } \tilde{u} \in \mathbf{U}: \quad \overline{\mathscr{E}}_N(t, \overline{u}_N(t), \underline{z}_N(t)) \le \overline{\mathscr{E}}_N(t, \tilde{u}, \underline{z}_N(t)), \tag{21a}$$

$$\textit{for all } \tilde{z} \in \mathbf{X}: \quad \overline{\mathscr{E}}_N(t, \overline{u}_N(t), \overline{z}_N(t)) \le \overline{\mathscr{E}}_N(t, \overline{u}_N(t), \tilde{z}) + \mathscr{R}(\tilde{z} - \overline{z}_N(t)) + \mathrm{TOL}(N), \tag{21b}$$

$$\overline{\mathscr{E}}_N(t, \overline{q}_N(t)) + \mathrm{Diss}_{\mathscr{R}}(\overline{z}_N, [0, t]) \le \overline{\mathscr{E}}_N(0, q_N^0) + \int_0^t \partial_\xi \mathscr{E}_N(\xi, \underline{q}_N(\xi))\, \mathrm{d}\xi + \mathrm{TOL}(N)N. \tag{21c}$$

In particular, there is a constant $C > 0$ such that the following bounds hold true uniformly for all $N \in \mathbb{N}$:

$$\text{for all } t \in [0, \mathsf{T}] : \ \|u_N(t)\|_{\mathbf{U}} \leq C\,, \tag{22a}$$

$$\text{for all } t \in [0, \mathsf{T}] : \ \|z_N(t)\|_{\mathbf{X}} + \|z_N(t)\|_{L^\infty(\Omega)} \leq C\,, \tag{22b}$$

$$\mathscr{R}(\overline{z}_N(\mathsf{T}) - z_N^0) \leq C \ \ \& \ \ \|\overline{z}_N\|_{BV(0,\mathsf{T};\mathbf{Z})} \leq C\,, \tag{22c}$$

where (u_N, z_N) in (22a) & (22b) stands for both $(\overline{u}_N, \overline{z}_N)$ and $(\underline{u}_N, \underline{z}_N)$.

Proof **Proof of properties** (21)**:** Taking into account the definition (19) of the interpolants $(\overline{u}_N, \underline{u}_N, \overline{z}_N, \underline{z}_N)$ we see that minimality properties (17) can be directly translated into (21a) & (21b). To find the discrete upper energy-dissipation estimate (21c) we test the minimality of u_N^k in (17a) by u_N^{k-1} and the minimality of z_N^k in (17b) by z_N^{k-1}. This results in

$$\mathscr{E}_N(t_N^k, u_N^k, z_N^{k-1}) \leq \mathscr{E}_N(t_N^k, u_N^{k-1}, z_N^{k-1})$$
$$\mathscr{E}_N(t_N^k, u_N^k, z_N^k) + \mathscr{R}(z_N^k - z_N^{k-1}) \leq \mathscr{E}_N(t_N^k, u_N^k, z_N^{k-1}) + \mathrm{TOL}(N)\,.$$

Let now $t \in (0, t_N^n]$ for some $n \leq N$. Adding the above two inequalities, adding and subtracting $\mathscr{E}_N(t_N^{k-1}, u_N^{k-1}, z_N^{k-1})$, and summing over $k \in \{1, \ldots, n\}$ we find

$$\begin{aligned}
&\mathscr{E}_N(t_N^n, u_N^n, z_N^n) + \mathscr{R}(z_N^n - z_N^0) \\
&\leq \mathscr{E}_N(t_N^0, u_N^0, z_N^0) + \sum_{k=1}^n \mathscr{E}_h(t_N^k, u_N^{k-1}, z_N^{k-1}) - \mathscr{E}_N(t_N^{k-1}, u_N^{k-1}, z_N^{k-1}) + n\mathrm{TOL}(N) \\
&= \mathscr{E}_N(t_N^0, u_N^0, z_N^0) + \sum_{k=1}^n \int_{t_N^{k-1}}^{t_N^k} \partial_\xi \mathscr{E}_N(\xi, u_N^{k-1}, z_N^{k-1})\,\mathrm{d}\xi + n\mathrm{TOL}(N)\,,
\end{aligned} \tag{23}$$

which yields (21c) for all $t \in (0, t_N^n]$ and integers $n \leq N$.
Proof of estimates (22)**:** Observe that there are constants $c_0, c_1 > 0$, such that for all $(t, u, z) \in [0, \mathsf{T}] \times \mathbf{U} \times \mathbf{Z}$ with $\mathscr{E}_N(t, u, z) < \infty$ it holds $|\partial_t \mathscr{E}_N(t, u, z)| \leq c_1(c_0 + \mathscr{E}_N(t, u, z))$. This entitles us to apply a Gronwall estimate under the time-integral in (23). Following the classical arguments for energy-dissipation inequalities in the rate-independent setting, cf., e.g., [42, Prop. 2.1.4], results in the estimates

$$c_0 + \overline{\mathscr{E}}_N(t_N^k, u_N^k, z_N^k) \leq (c_0 + \overline{\mathscr{E}}_N(0, u_N^0, z_N^0))\exp(c_1 T) \leq C\,, \tag{24a}$$

$$\mathscr{R}(z_N^k - z_N^0) \leq (c_0 + \overline{\mathscr{E}}_N(0, u_N^0, z_N^0))\exp(c_1 T) \leq C\,, \tag{24b}$$

where the uniform boundedness by $C > 0$ is due to (20) and Assumption 2.1. The estimate (22a) is then standardly obtained from the bound (24a), exploiting that $f(0) \geq a > 0$ and $\mu > 0$ by Assumption 2.1, as well as Korn's and Young's inequality. The estimate (22b) follows from the uniform boundedness of the damage gradients and the fact that $I_{[0,1]}(z_N(t)) = 0$ a.e. in Ω, ensured by (24a), whereas the first estimate in (22c) is due to (24b) and the second is a direct consequence taking into account the form of $\mathscr{R}$, see (1a). This concludes the proof of Prop. 4.1. ■

5 Numerical Experiments

We report in this section the numerical results for two two-dimensional benchmark problems taken from [1] and [38].

5.1 *Membrane with Hole*

In the sequel we specify all relevant information for the first benchmark problem from [1].

Problem Specification

We consider a body occupying a square domain with a hole around the center and which is pulled from above and below. Due to symmetry we regard only the upper right quarter of the domain. We summarize all relevant information for the first example in the following.

- **Geometry:** Length scale $L = 1$ mm;
  Domain $\Omega = (0, L)^2 \setminus \{x \in \mathbb{R}^2 :\ |x| \leq L\sqrt{2}/3\}$;
  Dirichlet boundary $\Gamma_D = ([L\sqrt{2}/3, L] \times \{0\}) \cup (\{0\} \times [L\sqrt{2}/3, L])$
- **Time horizon:** $\mathsf{T} = 1$ s
- **Load:** Dirichlet data:

$$u_D(t,x)_1 = 0 \text{ mm/s} \quad \text{if } x \in \Gamma_D^{left},$$
$$u_D(t,x)_2 = 0 \text{ mm/s} \quad \text{if } x \in \Gamma_D^{bottom};$$

Neumann data:

$$u_{\text{Neu}}(t,x) = \begin{bmatrix} 0 \ \frac{\text{N}}{\text{mm}^2\text{s}} \\ t \cdot 1 \ \frac{\text{N}}{\text{mm}^2\text{s}} \end{bmatrix} \qquad \text{if } x \in \Gamma_{\text{Neu}}^{top},$$
$$u_{\text{Neu}}(t,x) = \begin{bmatrix} 0 \ \frac{\text{N}}{\text{mm}^2\text{s}} \\ 0 \ \frac{\text{N}}{\text{mm}^2\text{s}} \end{bmatrix} \qquad \text{if } x \in \Gamma_{\text{Neu}}^{right};$$

The geometry and the applied traction are illustrated in Fig. 1.

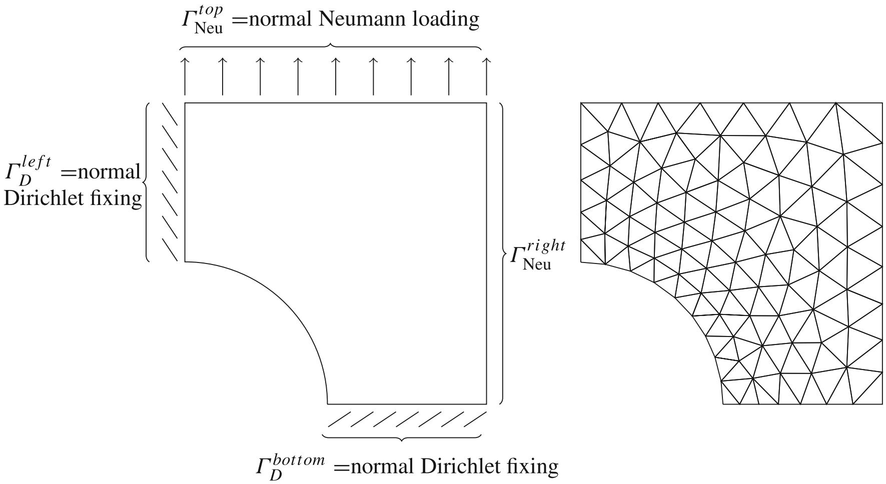
$\Gamma_{\mathrm{Neu}}^{top}$ =normal Neumann loading
Γ_{D}^{left} =normal Dirichlet fixing
$\Gamma_{\mathrm{Neu}}^{right}$
Γ_{D}^{bottom} =normal Dirichlet fixing

Fig. 1 Left: Domain Ω and illustration of applied traction for membrane with hole: the material is pulled from above. Right: Coarse triangulation ($h_{\min} = 0.055$)

- **Material parameters:** Young's modulus $E = 2900\ \mathrm{N/mm}^2$;
  Poisson's ratio $\nu = 0.4$;
  Lamé constants

$$\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)} \approx 4142.9\ \frac{\mathrm{N}}{\mathrm{mm}^2},\ \mu = \frac{E}{2(1+\nu)} \approx 1035.7\ \frac{\mathrm{N}}{\mathrm{mm}^2};$$

  The function f is chosen as $f(z) = a + (b-a)z$ with $a = 1/2, b = 1$;
  Damage toughness $\rho = 4 \cdot 10^{-4}\ \mathrm{N/mm}^2$;
  Regularization factor $\kappa = 10^{-6}\ \mathrm{N/mm}^2$
- **Initialization:** Initial stable state $u_h^0 \equiv 0$, $z_h^0 \equiv 1$.
- **Discretization:** Four triangulations $\mathcal{T}_h$ generated with `distmesh` (see [45]) with mesh sizes (in mm)

$$h \approx 0.204,\ h_{\min} \approx 0.055; \quad h \approx 0.09,\ h_{\min} \approx 0.034;$$
$$h \approx 0.054,\ h_{\min} \approx 0.016; \quad h \approx 0.029,\ h_{\min} \approx 0.008;$$

  Equidistant partition of $[0, \mathsf{T}]$ with $\Delta t = 10/(\lceil \mathsf{T}/h_{\min}^2 \rceil)$
- **Algorithm:** Algorithm 3.1 stops if $R_j \leq 10^{-6}/(2\max\{1, 1/(\tau_j h_{\min})\})$;
  $\overline{\tau} = h_{\min}^{-2}$, $\underline{\tau} = 10^{-4}$, $\delta = 0.5$, $\underline{\gamma} = 0.5$, $\overline{\gamma} = 0.999$

Aim

Since we are dealing with a BV-regularized damage model, i.e., the damage variable is allowed to jump in space, we want to investigate if the interfaces between

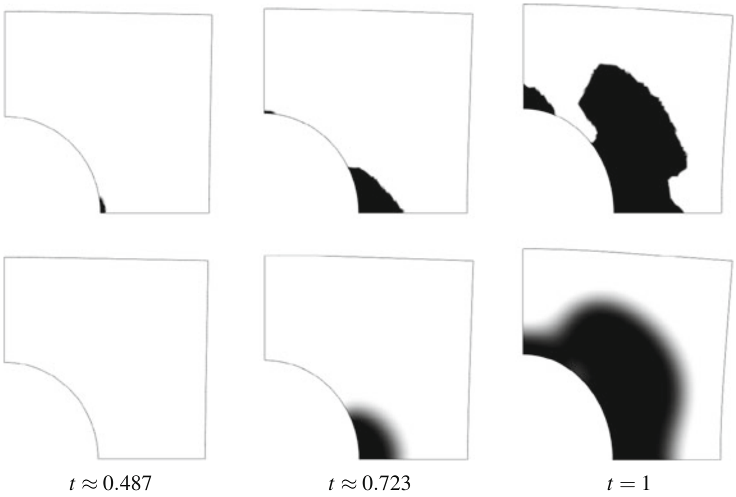

Fig. 2 Damage evolution with mesh size $h_{\min} \approx 0.008$ and time step size $\Delta t = 1/1492$. Top: BV-regularization. Bottom: Unweighted H^1-regularization. Displacements are magnified by factor 40

damaged and undamaged parts of the material are sharp at least on the scale h of the mesh resolution. We will also compare the results with an H^1-regularization, i.e., we replace $\kappa|\mathrm{D}z|(\Omega)$ by $\kappa\|\nabla z\|^2$ and by $\kappa h_{\min}\|\nabla z\|^2$ in order to investigate the influence of the chosen regularization term on the damage evolution. The dependence of the solutions on the mesh size will also be analyzed.

Results

In Fig. 2 three time steps of the damage evolution computed by Algorithm 3.2 for $h_{\min} = 0.008$ are depicted, both for the damage model with BV-regularization and unweighted H^1-regularization of the damage variable. The displacements are magnified by a factor of 40. One can clearly observe that the BV-regularization leads to sharp jumps (on the scale of h) while the transitions from undamaged ($z = 1$) to damaged ($z = 0$) parts of the material are smeared out for the H^1-regularization as it could be expected. The evolutions are more similar to each other if the H^1 regularization term is scaled with the factor $h_{\min}$ as it can be seen from Fig. 3. However, it is not clear whether the regularization term $\kappa h_{\min}\|\nabla z\|^2$ can be analytically justified, particularly with respect to the limit $h \to 0$.

In Fig. 4 we verify the energy estimate (21c) as a function of t_N^n, $n \leq N$, for three mesh sizes $h_{\min} = 0.055, 0.016, 0.008$. Obviously, the energy inequality holds and is increasing in time which is in accordance to (21c) since the inequality holds for all $t_N^k < t_N^n$, $1 \leq k \leq n \leq N$.

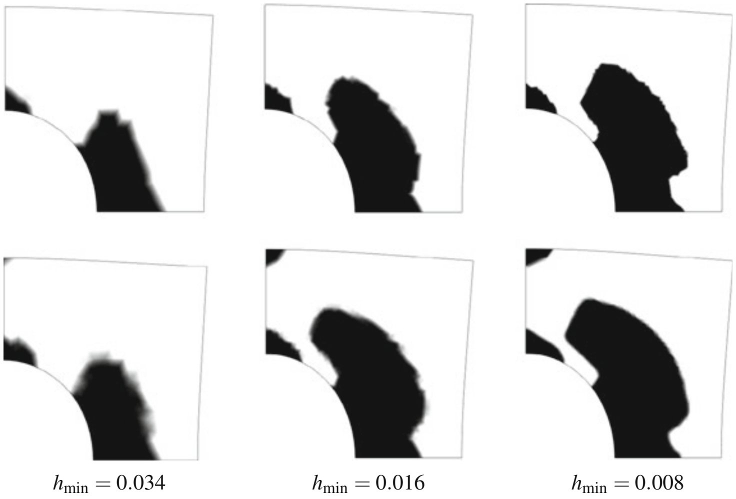

Fig. 3 Damage at $t = 1$ for different mesh sizes and time step sizes. Top: BV-regularization. Bottom: Weighted H^1-regularization with $\kappa h_{\min}\|\nabla z\|^2$. Displacements are magnified by factor 40

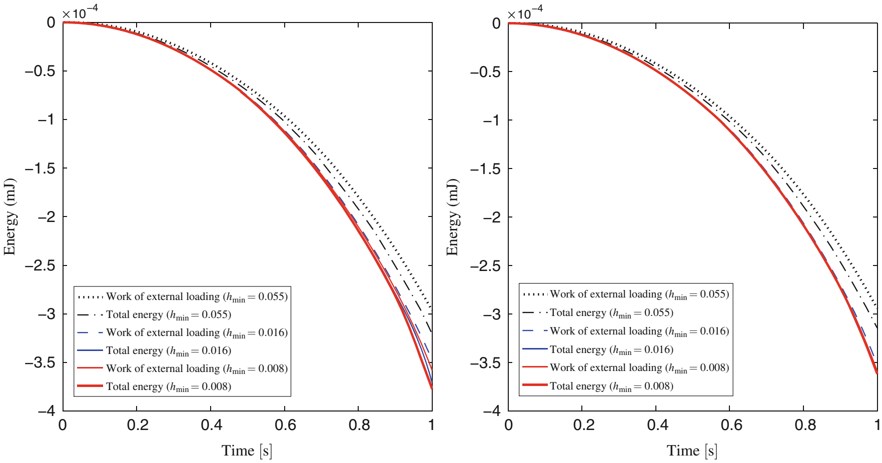

Fig. 4 Verification of energy estimate (21c) as a function of t_N^n for three different mesh sizes. Sum of stored and dissipated energy (= total energy = left-hand side of (21c)); work of external loading up to time t_N^n (= right-hand side of (21c) with $\overline{\mathscr{E}}_N(0, q_N^0) = 0$). Left: with BV-regularization; right: with H^1-regularization

5.2 *Notched Square*

The relevant information for the second test, which is taken from [38], are given below.

Problem Specification

We consider a body occupying a square domain with a notch reaching from the middle of the left edge to the center of the specimen. The specimen is pulled from above and clamped at the bottom. We summarize all relevant information for this example in the following.

- **Geometry:** Length scale $L = 1$ mm;
  Domain $\Omega = (0, L)^2 \setminus \mathrm{conv}\{(0, 0.5075), (0.5, 0.5), (0, 0.4925)\}$;
  Dirichlet boundary $\Gamma_D = ([0, L] \times \{0\}) \cup ([0, L] \times \{L\})$
- **Time horizon:** $\mathsf{T} = 1$ s
- **Load:** Dirichlet data:

$$u_D(t, x)_2 = t \cdot 0.002 \text{ mm/s} \qquad \text{if } x \in \Gamma_D^{top},$$

$$u_D(t, x) = \begin{bmatrix} 0 \text{ mm/s} \\ 0 \text{ mm/s} \end{bmatrix} \qquad \text{if } x \in \Gamma_D^{bottom};$$

Neumann data:

$$u_{\mathrm{Neu}}(t, x) = \begin{bmatrix} 0 \ \frac{\mathrm{N}}{\mathrm{mm}^2\mathrm{s}} \\ 0 \ \frac{\mathrm{N}}{\mathrm{mm}^2\mathrm{s}} \end{bmatrix} \qquad \text{if } x \in \Gamma_{\mathrm{Neu}};$$

The geometry is illustrated in Fig. 5.

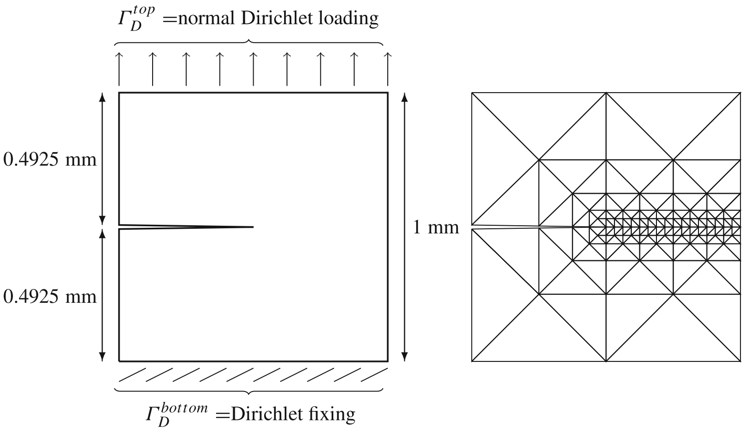
Γ_D^{top} =normal Dirichlet loading
0.4925 mm
1 mm
0.4925 mm
Γ_D^{bottom} =Dirichlet fixing

Fig. 5 Left: Domain Ω and illustration of boundary conditions for notched square: the material is pulled from above. Right: Initial locally refined mesh

- **Material parameters:** Young's modulus $E = 210$ kN/mm^2;
  Poisson's ratio $\nu = 0.3$;
  Lamé constants

$$\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)} \approx 121.15\,\frac{\text{kN}}{\text{mm}^2},\ \mu = \frac{E}{2(1+\nu)} \approx 80.77\,\frac{\text{kN}}{\text{mm}^2};$$

  The function f is chosen as $f(z) = a + (b-a)z$ with $a = 10^{-6}, b = 1$;
  Damage toughness $\rho = 2.7 \cdot 10^{-3}$ kN/mm^2;
  Regularization factor $\kappa = 10^{-7}$ kN/mm^2
- **Initialization:** Initial stable state $u_h^0 \equiv 0$, $z_h^0 \equiv 1$.
- **Discretization:** Three triangulations $\mathscr{T}_h$ generated by uniform refinement of an initial mesh refined locally in region of expected damage evolution with mesh sizes (in mm)

$$h \approx 0.25,\ h_{\min} \approx 0.0156;\ h \approx 0.125,\ h_{\min} \approx 0.0078;\ h \approx 0.0625,\ h_{\min} \approx 0.0039;$$

  Equidistant partition of $[0, \mathsf{T}]$ with $\Delta t = 10/(\lceil \mathsf{T}/h_{\min}^2 \rceil)$
- **Algorithm:** Algorithm 3.1 stops if $R_j \leq 10^{-7}/(2\max\{1, 1/(\tau_j h_{\min})\})$;
  $\overline{\tau} = h_{\min}^{-2}$, $\underline{\tau} = 10^{-3}$, $\delta = 0.5$, $\underline{\gamma} = 0.5$, $\overline{\gamma} = 0.999$

Aim

The aim of this experiment is to compare the resulting damage evolution with established numerical experiments for damage or crack propagation reported in [38, 56], which are based on a phase field approach, and to check whether our damage model yields qualitatively the same results.

Results

In Figs. 6 and 7 three snapshots of the damage evolution computed by Algorithm 3.2 for $h_{\min} \approx 0.0078$ are depicted for the damage model with BV-regularization and H^1-regularization, respectively, of the damage variable. Let us remark that the damage evolution observed in Fig. 6 qualitatively matches with the evolution reported in [38, Fig. 8] and [56, Fig. 4], i.e., the damage concentrates in a thin region around the horizontal line connecting the tip of the notch and the boundary on the right. Moreover, in contrast to the models discussed in [38, 56] the model presented in this paper is a damage model without phase field character and models by $a > 0$ only partial damage. Particularly, our model is not of Ambrosio-Tortorelli type.

In Fig. 8 the energy curves corresponding to (21c) as a function of t_N^n are depicted for three different mesh sizes. One can again observe that the energy inequality holds and that the gap is increasing in time. Furthermore, one can observe in Fig. 8 that the damage evolves relatively fast to the right boundary after the damage process has been initiated, e.g., for $h_{\min} = 0.0039$ it takes only a few milliseconds from initiation of the damage until damage reaches the boundary which is also in accordance with the observations made in [38, 56]. Note that the damage is triggered earlier for smaller mesh sizes which is on the one hand due to the singularity of

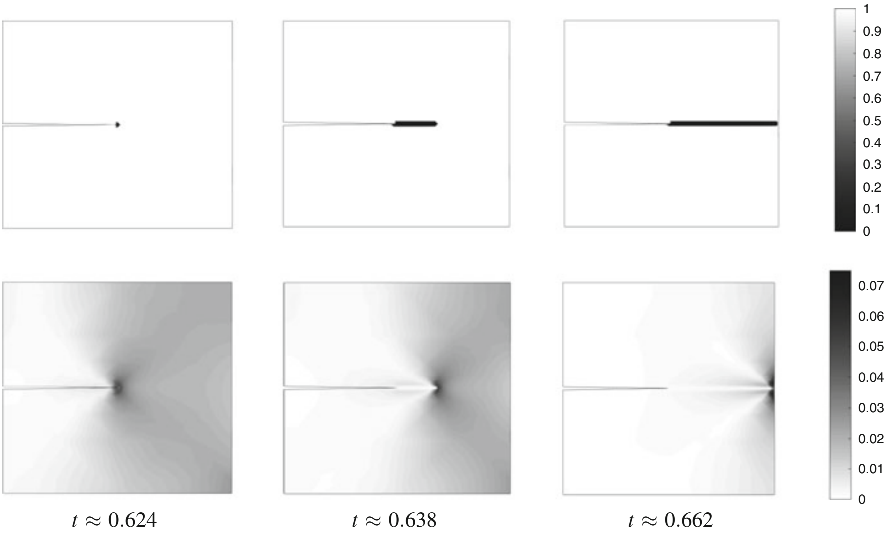

Fig. 6 BV-regularized evolution for notched square with mesh size $h_{\min} \approx 0.0078$ and time step size $\Delta t = 1/1638$. Top: Evolution of damage variable z. Bottom: Stress $\sqrt{f(z)e(u+g(t)) : \mathbb{C}e(u+g(t))}$

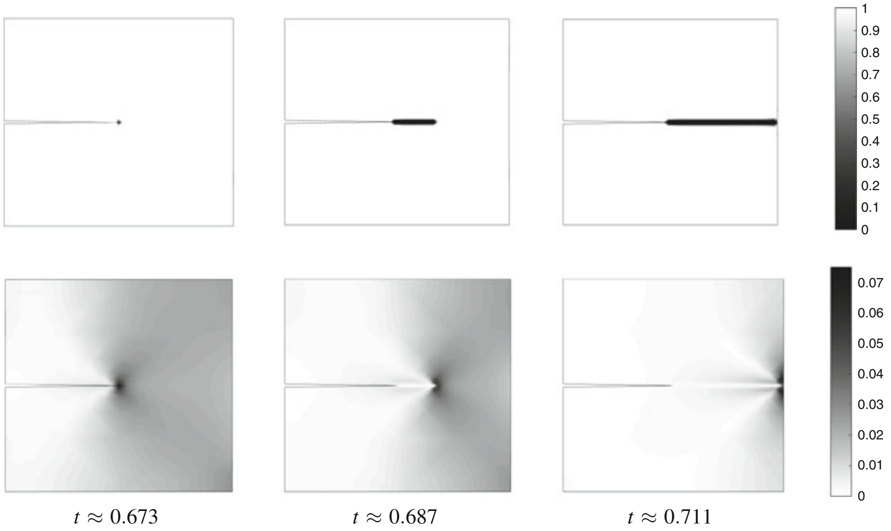

Fig. 7 H^1-regularized evolution for notched square with mesh size $h_{\min} \approx 0.0078$ and time step size $\Delta t = 1/1638$. Top: Evolution of damage variable z. Bottom: Stress $\sqrt{f(z)e(u+g(t)) : \mathbb{C}e(u+g(t))}$

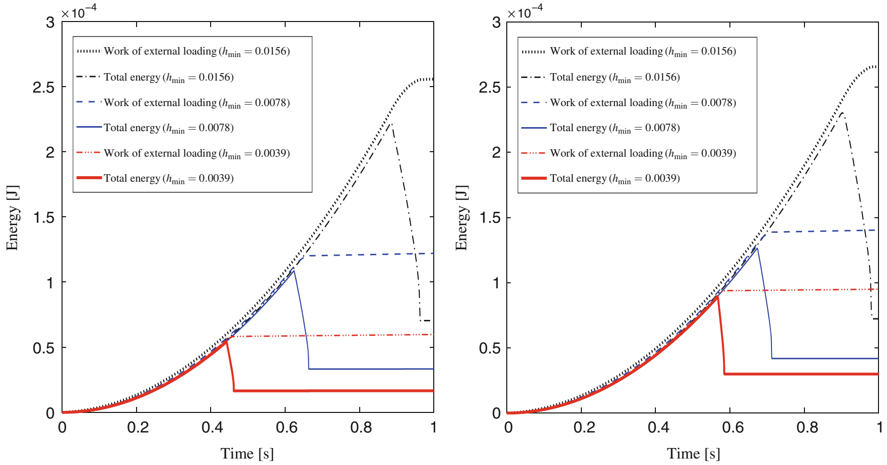

Fig. 8 Verification of energy estimate (21c) as a function of t_N^n for three different mesh sizes. Sum of stored and dissipated energy (= total energy = left-hand side of (21c)); work of external loading up to time t_N^n (= right-hand side of (21c) with $\overline{\mathscr{E}}_N(0, q_N^0) = 0$). Left: with BV-regularization; right: with H^1-regularization

the stress at the crack tip and on the other hand due to the finer partition of the time interval for smaller mesh sizes. This underlines the need for proper adaptive refinement techniques both for the space and the time variable.

6 Conclusion

The numerical experiments show that our damage model can qualitatively capture the important features of damage evolution or crack propagation already reported in [10, 38, 56] for a phase field approach and, e.g., in [44, 49, 50] for similar numerical experiments based on energetic formulations. Depending on the particular setting the BV-regularization of the damage variable can lead to transitions from damaged to undamaged zones in the material that are significantly sharper than for an H^1-regularization as it has been observed in our first experiment.

Acknowledgements This work was carried out within the project *Finite element approximation of functions of bounded variation and application to models of damage, fracture, and plasticity* within the DFG Priority Programme SPP 1748 "Reliable Simulation Techniques in Solid Mechanics. Development of Non-standard Discretisation Methods, Mechanical and Mathematical Analysis."

References

1. Alberty, J., Carstensen, C., Funken, S.A., Klose, R.: Matlab implementation of the finite element method in elasticity. Computing **69**, 239–263 (2002)
2. Ambati, M., Kruse, R., De Lorenzis, L.: A phase-field model for ductile fracture at finite strains and its experimental verification. Comput. Mech. **57** (2016)
3. Bartels, S.: Total variation minimization with finite elements: convergence and iterative solution. SIAM J. Numer. Anal. pp. 1162–1180 (2012)
4. Bartels, S.: Numerical Methods for Nonlinear Partial Differential Equations. Springer, Heidelberg (2015)
5. Bartels, S., Milicevic, M.: Alternating direction method of multipliers with variable step sizes (2017). URL https://aam.uni-freiburg.de/agba/prof/preprints/BarMil17-pre.pdf. Cited 13 Mar 2017
6. Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE Trans. Image Process. **18**, 2419–2434 (2009)
7. Bonetti, E., Schimperna, G.: Local existence for Frémond's model of damage in elastic materials. Contin. Mech. Thermodyn. **16**(4), 319–335 (2004). DOI 10.1007/s00161-003-0152-2. URL http://dx.doi.org/10.1007/s00161-003-0152-2
8. Bonetti, E., Schimperna, G., Segatti, A.: On a doubly nonlinear model for the evolution of damaging in viscoelastic materials. J. Differ. Equ. **218**(1), 91–116 (2005). DOI 10.1016/j.jde.2005.04.015. URL http://dx.doi.org/10.1016/j.jde.2005.04.015
9. Bouchitté, G., Mielke, A., Roubíček, T.: A complete-damage problem at small strain. Zeit. Angew. Math. Phys. **60**, 205–236 (2009)
10. Bourdin, B.: Numerical implementation of the variational formulation for quasi-static brittle fracture. Interfaces Free Boundaries **9**, 411–430 (2007)
11. Braides, A., Cassano, B., Garroni, A., Sarrocco, D.: Quasi-static damage evolution and homogenization: a paradigmatic case of non-commutability. Ann. Inst. H. Poincaré Anal. Non Linéaire, online (2014)
12. Brenner, S.C., Scott, L.R.: The Mathematical Theory of Finite Element Methods. Springer, New York (2008)
13. Chambolle, A.: An algorithm for total variation minimization and applications. J. Math. Imaging Vis. **20**, 89–97 (2004)
14. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**, 120–145 (2011)
15. Fiaschi, A., Knees, D., Stefanelli, U.: Young-measure quasi-static damage evolution. IMATI-Preprint 28PV10/26/0 (2010), Pavia (2010)
16. Fortin, M., Glowinski, R.: Augmented Lagrangian Methods. North-Holland, Amsterdam (1983)
17. Frémond, M.: Non-Smooth Thermomechanics. Springer, Berlin (2002)
18. Frémond, M., Nedjar, B.: Damage, gradient of damage and principle of virtual power. Int. J. Solids Struct. **33**, 1083–1103 (1996)
19. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comp. Maths. Appls. **2**, 17–40 (1976)
20. Glowinski, R.: Numerical Methods for Nonlinear Variational Problems. Springer, New York (1984)
21. Glowinski, R., Le Tallec, P.: Augmented Lagrangians and Operator-Splitting Methods in Nonlinear Mechanics. IAM, Philadelphia (1989)
22. Glowinski, R., Marroco, A.: Sur l'approximation par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires. Revue française d'automatique, informatique, recherche opérationelle. Analyse numérique **9**, 41–76 (1975)
23. Goldstein, T., O'Donoghue, B., Setzer, S., Baraniuk, R.: Fast alternating direction optimization methods. SIAM J. Imaging Sci. **7**, 1588–1623 (2014)

24. Goldstein, T., Osher, S.: The split Bregman method for L1 regularized problems. SIAM J. Imaging Sci. **2**, 323–343 (2009)
25. Gurtin, M., Francis, E.: Simple rate-independent model for damage. J. Spacecr. Rocket. **18**(3), 285–286 (1981)
26. Hackl, K., Stumpf, H.: Micromechanical concept for the analysis of damage evolution in thermo-viscoelastic and quasi-static brittle fracture. Int. J. Solids Struct. **30**, 1567–1584 (2003)
27. Halphen, B., Nguyen, Q.: Sur les matériaux standards généralisés. J. Mécanique **14**, 39–63 (1975)
28. Hanke, H., Knees, D.: A phase-field damage model based on evolving microstructure. Asymptot. Anal. **101**(3), 149–180 (2017)
29. Heinemann, C., Kraus, C.: Existence of weak solutions for Cahn–Hilliard systems coupled with elasticity and damage. Adv. Math. Sci. Appl. 321–359 (2011)
30. Heinemann, C., Kraus, C.: Complete damage in linear elastic materials – modeling, weak formulation and existence results. Calc. Var. Partial Differ. Equ. 217–250 (2015)
31. Hintermüller, M., Rautenberg, C.N., Hahn, J.: Functional-analytic and numerical issues in splitting methods for total variation-based image reconstruction. Inverse Prob. **30**(5), 055,014 (2014)
32. Kachanov, L.: On creep rupture time (in russian). Izv. Acad. Nauk. SSSR Otd. Techn. Nauk. (8), 26–31 (1958)
33. Kachanov, L.: Introduction to Continuum Damage Mechanics, 2nd edn. Mechanics of Elastic Stability. Kluwer Academic Publishers, Dordrecht (1990)
34. Knees, D., Negri, M.: Convergence of alternate minimization schemes for phase-field fracture and damage. Accepted in M3AS (2015). URL http://cvgmt.sns.it/media/doc/paper/2832/KneesNegri.pdf
35. Knees, D., Rossi, R., Zanini, C.: A vanishing viscosity approach to a rate-independent damage model. Math. Models Methods Appl. Sci. **23**(04), 565–616 (2013)
36. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**, 964–979 (1979)
37. Marigo, J., Maurini, C., Pham, K.: An overview of the modelling of fracture by gradient damage models. Meccanica **51**(12), 3107–3128 (2016)
38. Miehe, C., Hofacker, M., Welschinger, F.: A phase field model for rate-independent crack propagation: Robust algorithmic implementation based on operator splits. Comput. Methods Appl. Mech. Eng. **199**(45), 2765–2778 (2010)
39. Miehe, C., Welschinger, F., Hofacker, M.: Thermodynamically consistent phase-field models of fracture: Variational principles and multi-field FE implementations. Int. J. Numer. Meth. Eng. **83**, 1273–1311 (2010)
40. Mielke, A.: Evolution in rate-independent systems (Ch.6). In: Dafermos, C., Feireisl, E. (eds.) Handbook of Differential Equations, Evolutionary Equations, vol. 2, pp. 461–559. Elsevier B.V., Amsterdam (2005)
41. Mielke, A., Roubíček, T.: Rate-independent damage processes in nonlinear elasticity. Math. Models Methods Appl. Sci. **16**(2), 177–209 (2006)
42. Mielke, A., Roubíček, T.: Rate-independent Systems: Theory and Application. Applied Mathematical Sciences, vol. 193. Springer, New York (2015)
43. Mielke, A., Roubíček, T., Zeman, J.: Complete damage in elastic and viscoelastic media and its energetics. Comput. Methods Appl. Mech. Eng. **199**, 1242–1253 (2010). Submitted. WIAS preprint 1285
44. Mielke, A., Roubíček, T., Zeman, J.: Complete damage in elastic and viscoelastic media and its energetics. Comput. Methods Appl. Mech. Eng. **199**, 1242–1253 (2010)
45. Persson, P.O., Strang, G.: A simple mesh generator in `matlab`. SIAM Rev. **42**, 329–345 (2004)
46. Rocca, E., Rossi, R.: "entropic" solutions to a thermodynamically consistent pde system for phase transitions and damage. SIAM J. Math. Anal. **74**, 2519–2586 (2015)
47. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control. Optim. **14**, 877–898 (1976)

48. Roubíček, T., Thomas, M., Panagiotopoulos, C.: Stress-driven local-solution approach to quasistatic brittle delamination. Nonlinear Anal. Real World Appl. **22**, 645–663 (2015). DOI 10.1016/j.nonrwa.2014.09.011. URL http://dx.doi.org/10.1016/j.nonrwa.2014.09.011
49. Roubíček, T., Panagiotopoulos, C.G., Mantič, V.: Local-solution approach to quasistatic rate-independent mixed-mode delamination (2015)
50. Roubíček, T., Valdman, J.: Perfect plasticity with damage and healing at small strains, its modelling, analysis, and computer implementation. SIAM J. Appl. Math. **76**, 314–340 (2016)
51. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**, 259–268 (1992)
52. Schlüter, A., Willenbücher, A., Kuhn, C., Müller, R.: Phase field approximation of dynamic brittle fracture. Comput. Mech. **54**, 1141–1161 (2014)
53. Thomas, M.: Quasistatic damage evolution with spatial BV-regularization. Discrete Contin. Dyn. Syst. Ser. S **6**, 235–255 (2013)
54. Thomas, M., Bonetti, E., Rocca, E., Rossi, R.: A rate-independent gradient system in damage coupled with plasticity via structured strains. In: Düring, B., Schönlieb, C.B., Wolfram, M. (eds.) Gradient Flows: From Theory to Application, vol. 54, pp. 54–69. EDP Sciences (2016)
55. Thomas, M., Mielke, A.: Damage of nonlinearly elastic materials at small strain: existence and regularity results. Zeit. Angew. Math. Mech. **90**(2), 88–112 (2010)
56. Weinberg, K., Dally, T., Schuß, S., Werner, M., Bilgen, C.: Modeling and numerical simulation of crack growth and damage with a phase field approach. GAMM-Mitt. **39**(1), 55–77 (2016)
57. Wu, C., Tai, X.C.: Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and higher order models. SIAM J. Imaging Sci. **3**, 300–339 (2010)

Rigidity Effects for Antiferromagnetic Thin Films: A Prototypical Example

Andrea Braides

Abstract We consider two-dimensional discrete thin films obtained from N layers of a triangular lattice, governed by an antiferromagnetic energy. By a dimension-reduction analysis we show that, in contrast with the "total frustration" of the triangular lattice, the overall behaviour of the thin film is described by a limit interfacial energy on functions taking 2^N distinct parameters. In a sense, then the total frustration is recovered as N tends to infinity.

1 Introduction

We consider lattice energies defined on "spin functions" (i.e., functions $u = \{u_i\}$ taking the only values -1 or 1), of the form

$$-\sum_{i,j} c_{ij} u_i u_j \,, \tag{1}$$

where i, j are nodes of a (connected) portion of a lattice $\mathscr{L}$ in $\mathbb{R}^d$ and c_{ij} are interactions coefficients. In the case that $c_{ij} \geq 0$ the system is called *ferromagnetic* and its ground states are the two constant states ± 1. The overall behavior of the system when a large number of nodes are taken into account can then be described by a scaling procedure, by considering a scaling parameter $\varepsilon > 0$, a fixed parameter set Ω, and the scaled energies (obtained from the previous ones by scaling and adding constants)

$$\sum_{i,j} \varepsilon^{d-1}\, c_{ij}(u_i - u_j)^2 \tag{2}$$

A. Braides (✉)
Department of Mathematics, University of Rome Tor Vergata, Rome, Italy
e-mail: braides@mat.uniroma2.it

E. Rocca et al. (eds.), *Trends in Applications of Mathematics to Mechanics*,
Springer INdAM Series 27, https://doi.org/10.1007/978-3-319-75940-1_10

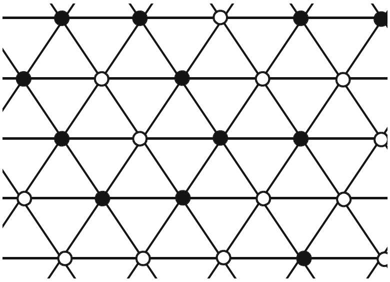

Fig. 1 A ‘disordered’ minimizer in a portion of the triangular lattice (black and white dots represent -1 and $+1$ values, respectively)

defined for i, j belonging to $\mathscr{L} \cap \frac{1}{\varepsilon}\Omega$. A discrete-to-continuum process allows to define an approximating continuum energy of interfacial type on Ω

$$\int_{\Omega\cap\partial\{u=1\}} \varphi(x, \nu_u)d\mathscr{H}^{d-1}(x),$$

where $u : \Omega \to \{-1, 1\}$ is a macroscopic parameter (the *magnetization*) defined as the limit of piecewise-constant functions u^ε defined from spin functions $\{u_i^\varepsilon\}$ as

$$u^\varepsilon(x) = u^\varepsilon_{\lfloor x/\varepsilon\rfloor} \qquad x \in \Omega$$

(up to some corrections close to $\partial\Omega$). The *surface tension* φ depends on the orientation ν_u of the interface between the two zones where $u = 1$ or $u = -1$. In many cases it is also homogeneous, and is characterized by the *Wulff shape*; i.e., the characteristic shape of minimizers with given measure.

If the system is *antiferromagnetic*; i.e., $c_{ij} \leq 0$, or a mixture of ferromagnetic and antiferromagnetic interactions, in general ground states are *frustrated*. This means that the energies in (1) cannot be minimized for each single interaction (as pictured in Fig. 1), which, in the case of antiferromagnetic coefficients would imply that $u_i = -u_j$. The simplest case of frustration is when $\mathscr{L}$ is a *triangular lattice* and we take c_{ij} different from zero only for *nearest-neighbours*, for which, for example $c_{ij} = -1$. In this case, ground states present no regularity and can arbitrarily mix the values $u_i = 1$ and $u_i = -1$. For antiferromagnetic-ferromagnetic mixtures this is “generically” not the case in the square lattice if we have a small percentage of antiferromagnetic interactions [4]. In [6] examples are shown also of mixtures of nearest-neighbour ferromagnetic and antiferromagnetic interactions in the square lattice with a similar “total frustration”. This behaviour is not present in every system with antiferromagnetic interactions. Indeed, long-range antiferromagnetic interactions, also in the square lattice, may present a finite collection of striped or checkerboard-type ground states (see e.g. [8, 9]). Using the analysis of ground states, sometimes those systems can be described in a discrete-to-continuum fashion by a surface energy defined on partitions of the underlying reference set Ω indexed by the different textures and modulated phases [6]. For a review on the subject we refer to [3].

In this paper we consider an example of thin films for spin energies. A discrete thin film is obtained by limiting the interactions to a $N\varepsilon$-neighbourhood of a $d-1$-dimensional set ω (as in [2, 5]). We then scale the energies accordingly, as

$$\sum_{i,j} \varepsilon^{d-2} c_{ij}(u_i - u_j)^2$$

(see [5]). In the simplest case of a "coordinate thin film", when ω is contained in $\mathbb{R}^{d-1} \times \{0\}$ then the sum above may be considered as performed for i, j belonging to $\mathscr{L} \cap \big(\frac{1}{\varepsilon}\omega \times [0, N]\big)$. The limit behaviour of these energies can be then described by a *dimensionally-reduced* energy of the form

$$\int_{\partial\{u=1\}} \varphi(x, \nu_u) d\mathscr{H}^{d-2}(x),$$

where the limit magnetization is interpreted as a function $u : \omega \to \{-1, 1\}$ and the form of φ takes into account also optimization of the interactions in the "vertical" direction; i.e., in the d-coordinate (for an analog thin-film theory for bulk surface energies see [7]).

In our case, we consider $d = 2$ and $\mathscr{L} = \mathbb{T}$ the regular triangular lattice; i.e., the Bravais lattice generated by $(1, 0)$ and $(1/2, \sqrt{3}/2)$. The nearest neighbours in $\mathbb{T}$ are points at distance 1; i.e., differing by $\pm(1, 0)$, $\pm(1/2, \sqrt{3}/2)$, or $\pm(-1/2, \sqrt{3}/2)$. For each $N \in \mathbb{N}$, $N > 0$, we then consider the related discrete thin film composed of N layers with underlying set an interval I; namely,

$$\Omega_{N,\varepsilon} = I \times [0, (N-1)\varepsilon\sqrt{3}/2].$$

If we consider nearest neighbour uniform anti-ferromagnetic interactions, the thin-film energy then simply reads

$$E^N_\varepsilon(u) = -\sum_{i,j}(u_i - u_j)^2, \tag{3}$$

where the sum i, j runs on nearest-neighbours in $\big(\frac{1}{\varepsilon}I\big) \times [0, (N-1)\sqrt{3}/2]$.

The simplest case is $N = 1$, when the underlying set $\Omega_{0,\varepsilon}$ reduces to $I \times \{0\}$, which can be directly identified with I. The energy E^1_ε can then be seen as a "bulk" spin energy with underlying lattice $\mathbb{Z}$, and can be reduced to a ferromagnetic energy by adding the constant 4 in each interaction in order to make the sum positive; i.e., considering

$$E^1_\varepsilon(u) = -\sum_i((u_i - u_{i-1})^2 - 4), \tag{4}$$

and by the change of variables $w_i = (-1)^i u_i$ (see also [1]). Then the thin-film limit is defined on piecewise-constant functions w on I with values in $\{-1, 1\}$ and is given by

$$F^1(w) = 4\,\#(S(w)),$$

where $S(w)$ is the discontinuity set of w. Note that the constant $w = 1$ corresponds to taking $u_i = (-1)^i$, while the constant $w = -1$ corresponds to $u_i = (-1)^{i+1}$, so that the two ground states in terms of v correspond to two variants of oscillating u (modulated phases).

As compared to the "total frustration" of the triangular lattice the case $N = 1$ already hints that a dimensional-reduction process applied to this example of antiferromagnetic interactions may give a continuum limit taking into account only a finite number of parameters. However, this case seems oversimplified since no trace of the triangular geometry of the original lattice remains. In the rest of the paper we analyze the case $N > 1$ to show how an N-dependent finite-parameter description holds.

2 Analysis of the Thin-Film Limit

We first consider more in detail the case $N = 2$, which is pictured in Fig. 2. In the notation above, the underlying thin film is

$$\Omega_{2,\varepsilon} = I \times [0, \varepsilon\sqrt{3}/2].$$

In order to simplify the notation we also introduce a non-orthogonal coordinate system as in figure, so that the points in the thin film are parameterized by

$$Z_2 := \{(n, m) : n \in \mathbb{Z}, m \in \{0, 1\}\}.$$

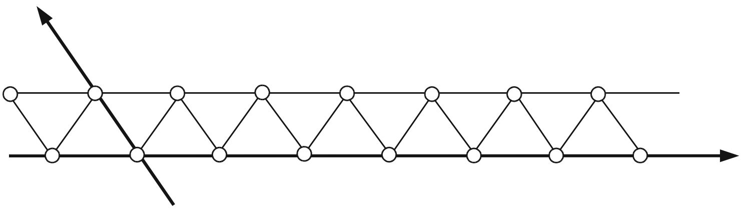

Fig. 2 Two-layer thin film with reference axes

We can write the energy as a sum of terms of the form

$$-\Big((u_{(n+1,0)}-u_{(n,0)})^2+(u_{(n,1)}-u_{(n,0)})^2+(u_{(n+1,1)}-u_{(n,1)})^2+(u_{(n+1,1)}-u_{(n,0)})^2\Big).$$

We may consider the case when the underlying interval is simply $\mathbb{R}$. In this case, we may sum on $\mathbb{Z}$ after adding a constant and regrouping the interactions as follows to avoid $+\infty-\infty$ indeterminate forms:

$$\begin{aligned}E^2_\varepsilon(u) &= -\sum_{n\in\mathbb{Z}}\Big((u_{(n+1,0)}-u_{(n,0)})^2\\&+\frac{1}{2}(u_{(n+1,1)}-u_{(n,0)})^2+\frac{1}{2}(u_{(n+1,1)}-u_{(n+1,0)})^2-6\Big)\\&-\sum_{n\in\mathbb{Z}}\Big((u_{(n+1,1)}-u_{(n,1)})^2+\frac{1}{2}(u_{(n+1,1)}-u_{(n,0)})^2+\frac{1}{2}(u_{(n,1)}-u_{(n,0)})^2-6\Big).\end{aligned}\tag{5}$$

In this way the energy is split in its contributions in each triangle. The first sum takes into account triangles with a side in the lower layer $m=0$ and the second sum takes into account triangles with a side in the upper layer $m=1$. The factor $1/2$ takes into account that non-horizontal sides belong to two neighbouring triangles. Note that not having alternate states on the horizontal (boundary) sides is more "costly" than on the others.

Note that the term

$$-\Big((u_{(n+1,0)}-u_{(n,0)})^2+\frac{1}{2}(u_{(n+1,1)}-u_{(n,0)})^2+\frac{1}{2}(u_{(n+1,1)}-u_{(n+1,0)})^2-6\Big)$$

is always non-negative, and it is zero only if

$$u_{(n+1,0)}\neq u_{(n,0)}.$$

In the same way, each term in the second sum is minimized only when $u_{(n+1,1)}\neq u_{(n,1)}$. This observation implies that ground states, with zero energy are all u that satisfy

$$\begin{aligned}u_{(n,0)}&=(-1)^n \text{ for all } n \ \text{ or } \ u_{(n,0)}=(-1)^{n+1} \text{ for all } n,\\u_{(n,1)}&=(-1)^n \text{ for all } n \ \text{ or } \ u_{(n,1)}=(-1)^{n+1} \text{ for all } n;\end{aligned}$$

i.e., with alternating values of u on the two horizontal layers. Hence, we have four ground states determined by their values at $n=0$

$$(u_{(0,0)},u_{(0,1)})\in\{-1,1\}^2=:X_2.$$

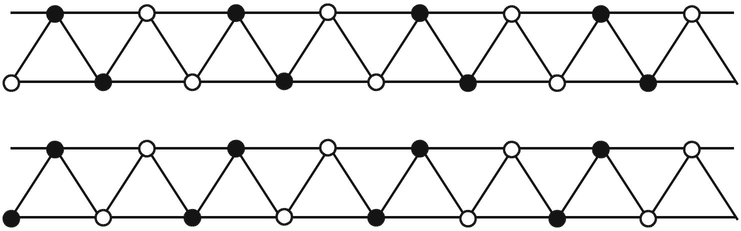

Fig. 3 A picture of ground states, with black/white circles indicating $-1/1$ values

For $x \in X_2$ we define

$$v^x : Z_2 \to \{-1, 1\}$$

as the ground state with $(v^x(0,0), v^x(0,1)) = x$.

Note that the two ground states determined by $\pm(1,1)$ (or by $\pm(-1,1)$, correspondingly), differ by a horizontal translation by $(1,0)$, while those determined by $(-1,1)$ and $(1,-1)$ are obtained by a reflection around a vertical line from $(1,1)$ and $(-1,-1)$ (see Fig. 3).

Note, moreover, that if u is a function with finite energy then there are a finite number of indices n such that u does not minimize the terms in the sum in (5). This implies that a sequence of functions with equibounded energy is precompact for the following notion of convergence.

The *discrete-to-continuum convergence* of a family of functions $u^\varepsilon : Z_2 \to \{-1, 1\}$ to a function $v : \mathbb{R} \to X_2$ with a finite number of points of discontinuity $S(v) = \{t_1, \dots, t_K\}$ is defined by the requirement that, denoted by x_j $(j = 0, \dots, K)$ the constant value of v on (t_j, t_{j+1}) (where $t_0 = -\infty$ and $t_{K+1} = +\infty$), for every $\delta > 0$ if ε is small enough then u^ε_n is equal to the ground state v^{x_j} respectively for

$$\begin{aligned} -\frac{1}{\varepsilon\delta} < n < \frac{1}{\varepsilon}(t_1 - \delta) \qquad & \text{if } j = 0 \\ \frac{1}{\varepsilon}(t_j + \delta) < n < \frac{1}{\varepsilon}(t_{j+1} - \delta) \quad & \text{if } j \in \{1, \dots, K-1\} \\ \frac{1}{\varepsilon}(t_K + \delta) < n < \frac{1}{\varepsilon\delta} \qquad & \text{if } j = K. \end{aligned}$$

This convergence may be equally stated as the convergence of the auxiliary functions $\tilde{u}_\varepsilon : \mathbb{R} \to V \cup \{(0,0)\}$ defined by

$$\tilde{u}_\varepsilon(t) = \begin{cases} x & \text{if } u^\varepsilon_j = v^x \text{ on } \left\{ \left\lfloor \frac{t}{\varepsilon} \right\rfloor, \left\lfloor \frac{t}{\varepsilon} \right\rfloor + 1 \right\} \times \{0, 1\} \\ (0,0) & \text{otherwise} \end{cases}$$

in $L^1_{\rm loc}(\mathbb{R})$. In the definition of the function $\tilde{u}_\varepsilon$ we scale the domain by ε and identify the value on two consecutive triangles (i.e., on the vertices of a unit square in the parameterization on Z_2) with the common parameter $x \in X_2$ when the corresponding u^ε coincides with v^x on those triangles. This parameter $x \in X_2$ is well defined except for a finite number of $\lfloor \frac{t}{\varepsilon} \rfloor$, so we may arbitrarily extend the definition by $(0, 0)$ on the complement.

We may describe the limit behaviour of the energies E^2_ε as defined in (5) by exhibiting a Γ-limit with respect to the convergence above, of the form

$$F^2(v) = \sum_{t \in S(v)} \varphi(v(t^-), v(t^+)), \tag{6}$$

where $t^\pm \in X_2$ are the left-hand and right-hand limit values of v at t. The energy function $\varphi(x, x')$ is obtained by computing the optimal transition between two states v^x and $v^{x'}$.

The picture in Fig. 4 describes an optimal transition when $x = (1, 1)$ and $x' = (-1, -1)$, or the converse. We may consider $v(t) = x$ for $t > 0$ and $v(t) = x'$ for $t < 0$ and $u^\varepsilon \to v$. In this case there must be some index n with a non-optimal interaction $u^\varepsilon(n, 0) = u^\varepsilon(n + 1, 0)$ and some index n' with $u^\varepsilon(n', 1) = u^\varepsilon(n' + 1, 1)$. In the picture such a u^ε is shown, optimizing all other interactions. The thick lines correspond to frustrated interactions. Computing the energy of such u^ε, which amounts just to the contributions of the two triangles highlighted in the picture, we obtain the value $\varphi((1, 1), (-1, -1)) = 4$. The same argument and a vertical symmetry argument shows that $\varphi((1, -1), (-1, 1))$ has the same value.

Similarly, in order to describe the optimal transition when $x = (1, 1)$ and $x' = (-1, 1)$ or the converse, we may remark that optimal u^ε must have $u^\varepsilon(n, 0) = u^\varepsilon(n + 1, 0)$ for some index n. In Fig. 5 we picture an optimal such u^ε, for which all interactions are optimal except one with $u^\varepsilon(n, 0) = u^\varepsilon(n+1, 0)$. The corresponding computation gives $\varphi((1, 1), (-1, 1)) = 2$.

Finally, in the case $x = (1, 1)$ and $x' = (1, -1)$, or the converse, we again note that optimal u^ε must have $u^\varepsilon(n, 1) = u^\varepsilon(n + 1, 1)$ for some index n, but there are two equivalent optimal arrangements, whether $u^\varepsilon(n, 0) = u^\varepsilon(n, 1)$ or $u^\varepsilon(n, 0) \neq$

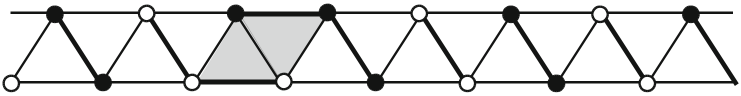

Fig. 4 An optimal transition between (1, 1) and (−1, −1)

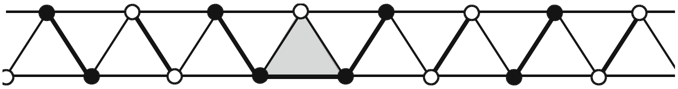

Fig. 5 An optimal transitions between (1, 1) and (−1, 1)

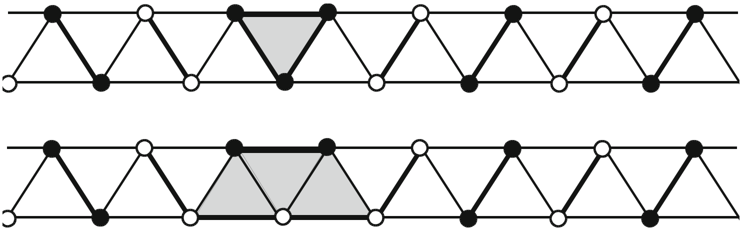

Fig. 6 Optimal transitions between (1, 1) and (1, −1)

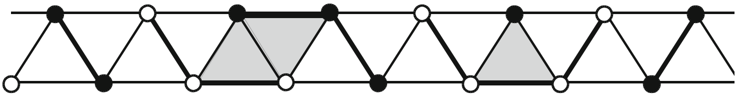

Fig. 7 Split optimal transitions between (1, 1) and (1, −1)

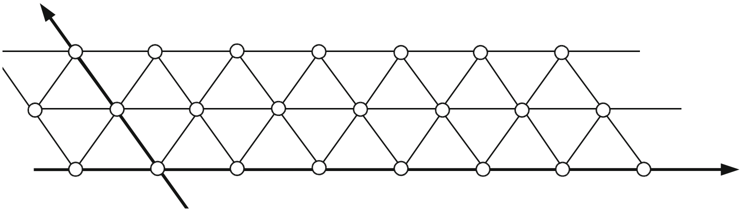

Fig. 8 Three-layer thin film with reference axes

$u^\varepsilon(n, 1)$. These two cases are pictured in Fig. 6 and both give $\varphi((1, 1), (-1, 1)) = 6$. Note that another optimal arrangement is obtained e.g. by combining the transitions between (1, 1) and (−1, 1) and between (−1, 1) and (1, −1). This corresponds to the lower case in Fig. 6 splitting the three non-optimal triangles into a pair with a common side and an isolated one (see Fig. 7). Analogously, the two joined triangles can be similarly split.

The Γ-limit result is finally obtained by superposing these constructions to obtain a recovery sequence for an arbitrary v.

Using a notation analogous to the one introduced above, we can now generalize this computation to a larger number of layers. For $N > 2$ we will not compute the energy function φ as above, but focus on its definition and in particular on its domain.

We first consider the case $N = 3$, whose underlying thin film is pictured in Fig. 8 together with the reference axes. The corresponding reference set is

$$Z_3 := \{(n, m) : n \in \mathbb{Z}, m \in \{0, 1, 2\}\}.$$

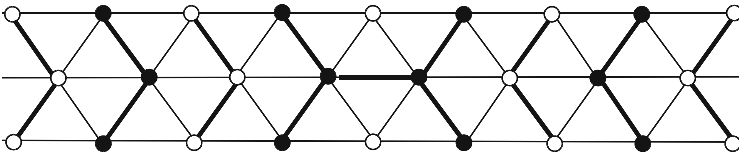

Fig. 9 A non-periodic minimizer

We can again consider the antiferromagnetic energy as a sum of the contribution of each triangle. The difference with the case $N = 2$ is that, while the energy of a triangle with a horizontal side on the top or bottom layer is as before, triangles with horizontal sides in the interior give an energy with a weight $1/2$ for all sides. For example, we have the contribution

$$-\frac{1}{2}\Big((u_{(n+1,1)} - u_{(n,1)})^2 + (u_{(n,1)} - u_{(n,0)})^2 + (u_{(n+1,1)} - u_{(n,0)})^2 - 4\Big)$$

for triangles in the lower row of triangles and a side in the middle layer of points.

For every $x = (x_0, x_1, x_2) \in \{\pm 1\}^3$ we denote the ground state given by

$$u^x(n, m) = x_m(-1)^n \text{ for all } n \in \mathbb{Z} \text{ and } m \in \{0, 1, 2\}.$$

Differently than the case $N = 2$, we note that a function u with zero energy is not necessarily one of those eight ground states, but may otherwise coincide with two of those for $n \geq M$ and for $n < -M$, respectively, for some $M \in \mathbb{N}$. Such a case is pictured in Fig. 9. Note that all functions with zero energy must have alternating values for $m = 0$ and $m = 2$. This implies that if, for example, $u(n, 2) = u(n, 1)$ for some n then the value of u is determined for $(n', 1)$ and $(n', 2)$ for all $n' \leq n$ as an alternating state. Similarly, if $u(n-1, 0) = u(n, 1)$. A symmetric argument also applies for minimizers which are determined for $n' \geq n$. This observation eventually implies that the one in Fig. 9 is the only non-periodic minimizer, up to translations.

As a consequence, we may define a convergence $u^\varepsilon \to v$ analog to the case $N = 2$, where now $v : \mathbb{R} \to X_3 := \{-1, 1\}^3$. We may describe the Γ-limit as a thin-film limit F^3 with the same form as (6), with $\varphi(x, x')$ the optimal-transition energy. The observations above show that $\varphi > 0$ except for

$$\varphi((1, -1, -1), (1, 1, -1)) = \varphi((-1, 1, 1), (-1 - 1, 1)) = 0.$$

Note that $\varphi((1, -1, -1), (1, 1, -1)) \neq \varphi((1, 1, -1), (1, -1, -1))$ so that φ is not symmetric, and that the energy F^3 is coercive even though its integrand is not strictly positive.

The two cases above carry the relevant information to treat the general case, which shows that the description of the thin-film limit needs a parameter space of increasing, but finite, cardinality; namely 2^N where N is the number of layers. We briefly sketch the argument, which generalizes what has been noticed above.

We consider a minimizer u.

1) we first note that the upper layer must be alternating; i.e., $u(n+1, N) \neq u(n, N)$ for all n
2) we either have $u(n, N) \neq u(n-1, N-1)$ for all n or $u(n_1, N) = u(n_1-1, N-1)$ for some n_1. In this case by minimality we have $u(n_1, N-1) \neq u(n_1-1, N-1)$. By Step 1 above we have $u(n_1+1, N) = u(n_1, N-1)$, so that we may proceed by induction and conclude that $u(n, N) = u(n-1, N-1)$ for all $n \geq n_1$. Hence, either u is alternating on the $N-1$-th layer, or it is alternating for $n < n_1$ and $n > n_1$.
3) proceeding in the same way we deduce that u is alternating in the $(N-2)$-th layer up to at most three indices (one less than n_1, one larger than n_1, and n_1 itself). We note that, as in Step 2, for $n > n_1$ there may exist a unique n_2 such that $u(n, N-1) = u(n-1, N-2)$ for $n \geq n_2$ and $u(n, N-1) \neq u(n-1, N-2)$ for $n < n_2$, but not the converse.
4) Proceeding by finite induction on the label of the layer, we deduce that $n \mapsto u(n, k)$ is alternating for each $k \in \{1, \dots, N\}$ up to a bounded number of n, with the bound independent of n. Moreover, in each interval of n where $n \mapsto u(n, k)$ is alternating there may exist a unique $\overline{n}$ such that $u(n, k) = u(n-1, k-1)$ for $n \geq \overline{n}$ and $u(n, k) \neq u(n-1, k-1)$ for $n < \overline{n}$, but not the converse. Moreover, $n \mapsto u(n, N)$ and $n \mapsto u(n, 0)$ are alternating.

   Note that this characterization also holds locally if we suppose that u has zero energy in an interval of n.

   From this characterization, we deduce that if u^ε is a sequence with bounded energy, then it must coincide with an alternating state on each layer up to a finite number of indices. At this point we may proceed as above. The description in the general case is summarized in the conclusions below.

3 Conclusions

We consider an infinite thin film parameterized on the set

$$\mathbb{T}_{N,\varepsilon} = \big(\mathbb{R} \times [0, (N-1)\varepsilon\sqrt{3}/2]\big) \cap \varepsilon\mathbb{T},$$

where $\mathbb{T}$ is a regular triangular lattice with one lattice vector $(1, 0)$, and the corresponding nearest-neighbour antiferromagnetic energy E^N_ε. In order to avoid indeterminate forms such energy is written as the sum of the contribution of each triangle of side-length ε contained in $\mathbb{T}_{N,\varepsilon}$, renormalized so that separately minimizing in each triangle gives zero energy. Note that the normalization is different if the triangle has one horizontal side on the upper or lower layer.

We have shown that there are 2^N distinct ground states of E_1^N, which are two-periodic in the direction $(1, 0)$. On each of the layers such ground states are alternating, so that each of these ground states u^x can be parameterized by a point x in the set

$$Z_N := \{\pm 1\}^N .$$

We may define a compact convergence of discrete functions $u^\varepsilon : \mathbb{T}_{N,\varepsilon} \to \{-1, 1\}$ to a function $v : \mathbb{R} \to Z_N$ with a finite number of discontinuities, which highlights that, up to a finite number of locations, a function u_ε with bounded energy E_ε^N coincides with a scaled version of the periodic minimizers.

With respect to this convergence the Γ-limit has the form

$$F^N(v) = \sum_{t\in S(v)} \varphi_N(v(t^-), v(t^+)),$$

where $S(v)$ is the set of discontinuity points of v. The function $\varphi_N : Z_N \times Z_N \to [0, +\infty)$ is an optimal-transition energy defined by

$$\varphi_N(x, x') = \min \Big\{ E_1^N(u) : u = u^x \text{ on } \mathbb{T}_{N,1} \cap (-\infty, -M], \\ u = u^{x'} \text{ on } \mathbb{T}_{N,1} \cap [M, +\infty), M \in \mathbb{N} \Big\}$$

(note that it suffices to take $M = N$ since we have a bound by a test function for which only at most one column of N triangles is not optimal). The energy F^N is coercive; i.e., its finiteness implies a finite number of discontinuity points of v. Note that the description above also holds for thin films with $\mathbb{R}$ substituted by a finite interval $[a, b]$, up to adding a boundary term. This extra term is not of interest since we focus on the number of limit parameters and not on the details of the energy.

The analysis above shows that the surface effects of the thin-film environment (i.e., the fact that ground states need to be alternating on the upper and lower layers due to the asymmetry of boundary sites) propagates inside the thin-film to limit the number of parameters needed to describe the limit. This rigidity effect "weakens" as the number of layers tends to infinity, as is testified by the (exponentially) diverging number of parameters. In a sense then, the "total frustration" of the triangular lattice can be seen as a limit behaviour as $N \to +\infty$.

Acknowledgements This work stems from very inspiring discussions with Roberto Alicandro and Marco Cicalese at TU Munich.

References

1. Alicandro, R., Braides, A., Cicalese, M.: Phase and anti-phase boundaries in binary discrete systems: a variational viewpoint. Netw. Heterog. Media **1**, 85–107 (2006)
2. Alicandro, R., Braides, A., Cicalese, M.: Continuum limits of discrete thin films with superlinear growth densities. Calc. Var. Partial Differ. Equ. **33**, 267–297 (2008)
3. Braides, A.: Discrete-to-continuum variational methods for lattice systems. In: Jang, S., Kim, Y., Lee, D., Yie, I. (eds.) Proceedings of the International Congress of Mathematicians August 13–21, 2014, Seoul, Korea, Vol. IV, pp. 997–1015. Kyung Moon Sa, Seoul (2014)
4. Braides, A., Causin, A., Piatnitski, A., Solci, M.: Asymptotic behaviour of ground states for mixtures of ferromagnetic and antiferromagnetic interactions in a dilute regime. Preprint 2017
5. Braides, A., Causin, A., Solci, M.: Interfacial energies on quasicrystals. IMA J. Appl. Math. **77**, 816–836 (2012)
6. Braides, A., Cicalese, M.: Interfaces, modulated phases and textures in lattice systems. Arch. Ration. Mech. Anal. **223**, 977–1017 (2017)
7. Braides, A., Fonseca, I.: Brittle thin films. Appl. Math. Optim. **44**, 299–323 (2001)
8. Giuliani, A., Lebowitz, J.L., Lieb, E.H.: Checkerboards, stripes, and corner energies in spin models with competing interactions. Phys. Rev. B **84**, 064205 (2011)
9. Seul, M., Andelman, D.: Domain shapes and patterns: the phenomenology of modulated phases. Science **267**(5197), 476–483 (1995)

Limiting Problems for a Nonstandard Viscous Cahn–Hilliard System with Dynamic Boundary Conditions

Pierluigi Colli, Gianni Gilardi, and Jürgen Sprekels

Abstract This note is concerned with a nonlinear diffusion problem of phase-field type, consisting of a parabolic system of two partial differential equations, complemented by boundary and initial conditions. The system arises from a model of two-species phase segregation on an atomic lattice and was introduced by Podio-Guidugli in Ric. Mat. **55** (2006), pp. 105–118. The two unknowns are the phase parameter and the chemical potential. In contrast to previous investigations about this PDE system, we consider here a dynamic boundary condition for the phase variable that involves the Laplace-Beltrami operator and models an additional nonconserving phase transition occurring on the surface of the domain. We are interested in some asymptotic analysis and first discuss the asymptotic limit of the system as the viscosity coefficient of the order parameter equation tends to 0: the convergence of solutions to the corresponding solutions for the limit problem is proven. Then, we study the long-time behavior of the system for both problems, with positive or zero viscosity coefficient, and characterize the omega-limit set in both cases.

P. Colli (✉) · G. Gilardi
Dipartimento di Matematica "F. Casorati", Università di Pavia, Pavia, Italy
e-mail: pierluigi.colli@unipv.it; gianni.gilardi@unipv.it

J. Sprekels
Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany

Department of Mathematics, Humboldt-Universität zu Berlin, Berlin, Germany
e-mail: juergen.sprekels@wias-berlin.de

E. Rocca et al. (eds.), *Trends in Applications of Mathematics to Mechanics*,
Springer INdAM Series 27, https://doi.org/10.1007/978-3-319-75940-1_11

1 Introduction

A recent line of research originated from the following evolutionary system of partial differential equations:

$$2\rho\,\partial_t\mu + \mu\,\partial_t\rho - \Delta\mu = 0 \quad \text{and} \quad \mu \geq 0 \tag{1.1}$$

$$-\Delta\rho + F'(\rho) = \mu \tag{1.2}$$

in $Q_\infty := \Omega \times (0, +\infty)$, where $\Omega \subset \mathbb{R}^3$ is a bounded and smooth domain with boundary Γ. The system (1.1)–(1.2) comes out from a model for phase segregation through atom rearrangement on a lattice that has been proposed by Podio-Guidugli [48]. This model (see also [12] for a detailed derivation) is a modification of the Fried–Gurtin approach to phase segregation processes (cf. [34, 41]). The order parameter ρ, which in many cases represents the (normalized) density of one of the phases, and the chemical potential μ are the unknowns of the system. Moreover, F' represents the derivative of a double-well potential F. Besides everywhere defined potentials, a typical and important example of F is the so–called *logarithmic double-well potential* given by

$$F_{log}(r) := (1+r)\ln(1+r) + (1-r)\ln(1-r) + \alpha_1(1-r^2) + \alpha_2 r, \qquad r \in (-1,1), \tag{1.3}$$

for some real coefficients α_1, α_2. Note that, if α_2 is taken null and $\alpha_1 > 1$, it turns out that F actually exhibits two wells, with a local maximum at $r = 0$. In the case when $\alpha_2 \neq 0$, then one of the two minima of F is preferred, in the sense that there is a global minimum point (positive if $\alpha_2 < 0$, negative if $\alpha_2 > 0$) of the function. As a particular feature of (1.3), observe that the derivative of the logarithmic potential becomes singular at ± 1.

About equations (1.1) and (1.2), we point out that the model developed in [48] is based on a local free energy density (in the bulk) of the form

$$\psi(\rho, \nabla\rho, \mu) = -\mu\,\rho + F(\rho) + \frac{1}{2}\,|\nabla\rho|^2. \tag{1.4}$$

From (1.4) one derives equations (1.1)–(1.2), which must be complemented with boundary and initial conditions. As far as the former are concerned, the standard boundary conditions for this class of problems are the homogeneous Neumann ones, namely

$$\partial_\nu\mu = \partial_\nu\rho = 0 \quad \text{on } \Sigma_\infty := \Gamma \times (0, +\infty), \tag{1.5}$$

where ∂_ν denotes the outward normal derivative. Combining now (1.1)–(1.2) with (1.5), we obtain a set of equations and conditions that is a variation of the celebrated Cahn–Hilliard system originally introduced in [1] and first studied mathematically in [31] (for an updated list of references on the Cahn–Hilliard system, see [42]). Nonetheless, an initial value problem for (1.1)–(1.2), (1.5) turns out to be strongly ill-posed (see [15, Subsect. 1.4], where an example is given): indeed, the related problem may have infinitely many smooth and even nonsmooth solutions. Then, two small regularizing parameters $\varepsilon > 0$ and $\delta > 0$ were introduced and considered in [12], which led to the regularized model equations

$$\big(\varepsilon + 2\rho\big)\, \partial_t \mu + \mu\, \partial_t \rho - \Delta\mu = 0\,, \tag{1.6}$$

$$\delta\, \partial_t \rho - \Delta\rho + F'(\rho) = \mu\,. \tag{1.7}$$

This regularized system has been deeply examined in [12], when both ε and δ are positive and fixed. In addition, let us underline that, while one can let ε tend to zero (see [16]) and obtain a solution to the limiting problem with $\varepsilon = 0$, it seems extremely difficult to pass to the limit as δ goes to 0. In fact, ill-posedness still holds for $\delta = 0$, even if ε is kept positive. Hence, one has to assume that δ is a fixed positive coefficient. Therefore, from now on, we take $\delta = 1$, without loss of generality. Let us point out that the long-time behavior of the solutions has been studied both with $\varepsilon > 0$ (cf. [12]) and $\varepsilon = 0$ (cf. [16]).

The system (1.6)–(1.7) constitutes a modification of the so-called *viscous* Cahn–Hilliard system (see [47] and the recent contributions[3, 20, 22] along with their references). We point out that (1.6)–(1.7) was analyzed, in the case of the boundary conditions (1.5), in the papers [12, 14, 18] concerning well-posedness, regularity, and optimal control. Later, the local free energy density (1.4) was generalized to the form

$$\psi(\rho, \nabla\rho, \mu) = -\mu\, g(\rho) + F(\rho) + \frac{1}{2}\, |\nabla\rho|^2, \tag{1.8}$$

thus putting $g(\rho)$ in place of ρ, where g is a nonnegative function on the domain of F. This leads to the system

$$\big(\varepsilon + 2g(\rho)\big)\, \partial_t \mu + \mu\, g'(\rho)\, \partial_t \rho - \Delta\mu = 0, \tag{1.9}$$

$$\partial_t \rho - \Delta\rho + F'(\rho) = \mu\, g'(\rho), \tag{1.10}$$

which is a generalization of (1.6)–(1.7) and has been studied in [13, 17] for the case $\varepsilon = 1$. Let us mention also the contribution [9] dealing with the time discretization of the problem and proving convergence results and error estimates. The related phase relaxation system (in which the diffusive term $-\Delta\rho$ disappears from (1.10)), has been dealt with in [10, 11, 19]. We also point out the recent papers [23–25], where a nonlocal version of (1.9)–(1.10)—based on the replacement of the diffusive term of (1.10) with a nonlocal operator acting on ρ—has been largely investigated, also from the side of optimal control.

Now, if we take $\varepsilon = 0$ in (1.9)–(1.10), we obtain

$$2g(\rho)\,\partial_t\mu + \mu\,g'(\rho)\,\partial_t\rho - \Delta\mu = 0 \tag{1.11}$$

$$\partial_t\rho - \Delta\rho + F'(\rho) = \mu\,g'(\rho), \tag{1.12}$$

which looks like a generalization of the viscous version of (1.1)–(1.2), where the affine function $\rho \mapsto \rho$ is replaced by a concave function $\rho \mapsto g(\rho)$, with g possessing suitable properties that are made precise in the later assumption (2.5). In particular, the new g may be symmetric and strictly concave: a possible simple choice of g satisfying (2.5) is

$$g(r) = 1 - r^2, \quad r \in [-1, 1]. \tag{1.13}$$

Note that, if one collects (1.3) and (1.13) and assumes $\alpha_2 \neq 0$, the combined function

$$-\mu g(\rho) + F_{log}(\rho) \quad \text{(which is a part of } \psi) \tag{1.14}$$

shows a global minimum in all cases, and it depends on the values of $(\alpha_1 - \mu)$ and α_2 which minimum actually occurs. Let us notice that the function in (1.14) turns out to be convex in the whole of $(-1, 1)$ for sufficiently large values of μ. On the other hand, the framework fixed by assumptions (2.5)–(2.8) allows for more general choices of g and F.

However, until now the boundary conditions (1.5), of Neumann type for both μ and ρ, have been considered in our discussion. Instead, in the present work we treat the dynamic boundary condition for ρ, i.e., we complement the above systems with

$$\partial_\nu\mu = 0 \quad \text{and} \quad \partial_\nu\rho + \partial_t\rho_\Gamma - \Delta_\Gamma\rho_\Gamma + F'_\Gamma(\rho_\Gamma) = 0 \quad \text{on } \Sigma_\infty, \tag{1.15}$$

where ρ_Γ is the trace of ρ, Δ_Γ is the Laplace-Beltrami operator on the boundary, F'_Γ is the derivative of another potential F_Γ having more or less the same behavior as F, and the right-hand side of the dynamic boundary condition equals zero, just for simplicity. Indeed, one could consider a nonzero forcing term satisfying proper assumptions, as done in [26]. Once again, we have to add initial conditions.

Thus, we are concerned with a total free energy of the system which also includes a contribution on the boundary; in fact, we postulate that a phase transition phenomenon is occurring as well on the boundary, and the physical variable on the boundary is just the trace of the phase variable in the bulk. This corresponds to a total free energy functional of the form

$$\begin{aligned}\boldsymbol{\Psi}[\rho(t), \rho_\Gamma(t), \mu(t)] &= \int_\Omega \Big[-\mu\,g(\rho) + F(\rho) + \frac{1}{2}\,|\nabla\rho|^2\Big](t) \\ &\quad + \int_\Gamma \Big[[-u_\Gamma\,\rho_\Gamma + F_\Gamma(\rho_\Gamma) + \frac{1}{2}\,|\nabla_\Gamma\rho_\Gamma|^2\Big](t), \quad t \geq 0,\end{aligned} \tag{1.16}$$

where ∇_Γ is the surface gradient and u_Γ may stand for the source term that exerts a (boundary) control on the system. From this expression of the total free energy, one recovers the PDE system resulting from equations (1.11)–(1.12) and the boundary conditions (1.15), with u_Γ in place of 0 in the right-hand side of the second condition. In relation to this, we would like to mention the contribution [27] dealing with the optimal boundary control problem for the system (1.6)–(1.7), (1.15) with $\varepsilon = 1$.

As for the dynamic boundary conditions, we would like to add some comments on the recent growing interest in the mathematical literature, either for the justification (see, e.g., [32, 33, 44]) or for the investigation of systems including dynamic boundary conditions. Without trying to be exhaustive, we point out at least the contributions [2, 4–8, 20–22, 28–30, 35–40, 43, 45, 46, 49, 50], which are concerned with various types of systems endowed with the dynamic boundary conditions for either some or all of the unknowns. Our citations mostly refer to phase-field models involving the Allen–Cahn and Cahn–Hilliard equations, whose structure is generally simpler than the one considered in the present paper.

Our aim here is investigating the long-time behavior of the full system in both the cases $\varepsilon > 0$ and $\varepsilon = 0$ (similar to [12, 16], in which the Neumann boundary conditions (1.5) were considered). More precisely, we show that the ω-limit of any trajectory in a suitable topology consists only of stationary solutions. In order to treat this problem also with $\varepsilon = 0$, we first study the asymptotics as ε tends to zero. To do that, we underline that the reasonable and somehow natural assumptions (2.5) for g along with the requirements (2.6)–(2.8) on F and F_Γ allow us to show that the variables ρ and ρ_Γ are strictly separated from the (singular) values ± 1. Indeed, we can prove this separation property and obtain the strict positivity of $g(\rho)$ as a consequence.

The paper is organized as follows: in the next section, we list our assumptions and notations and state our results, while the corresponding proofs are given in the last two sections. Precisely, in Sect. 3, we perform the asymptotic analysis as ε tends to zero and prove the well-posedness of the problem for $\varepsilon = 0$; in Sect. 4, we study the long-time behavior of the solution under the assumption $\varepsilon \geq 0$.

2 Statement of the Problem and Results

In this section, we state precise assumptions and notations and present our results. First of all, the set $\Omega \subset \mathbb{R}^3$ is assumed to be bounded, connected and smooth. As in the Introduction, ∂_ν and Δ_Γ stand for the outward normal derivative and the Laplace-Beltrami operator on the boundary Γ. Furthermore, we denote by ∇_Γ the surface gradient.

If X is a (real) Banach space, $\| \cdot \|_X$ denotes both its norm and the norm of X^3, X^* is its dual space, and ${}_{X^*}\langle \cdot , \cdot \rangle_X$ is the dual pairing between X^* and X. The only exception from this convention is given by the L^p spaces, $1 \leq p \leq \infty$, for which we use the abbreviating notation $\|\cdot\|_p$ for the norms in $L^p(\Omega)$. Furthermore, we put

$$H := L^2(\Omega)\,, \quad V := H^1(\Omega) \quad \text{and} \quad W := \{v \in H^2(\Omega) :\ \partial_\nu v = 0\}, \tag{2.1}$$

$$H_\Gamma := L^2(\Gamma) \quad \text{and} \quad V_\Gamma := H^1(\Gamma), \tag{2.2}$$

$$\mathcal{H} := H \times H_\Gamma \quad \text{and} \quad \mathcal{V} := \{(v, v_\Gamma) \in V \times V_\Gamma :\ v_\Gamma = v_{|\Gamma}\}. \tag{2.3}$$

We also set, for convenience,

$$\begin{aligned} &Q_t := \Omega \times (0,t) \quad \text{and} \quad \Sigma_t := \Gamma \times (0,t) \quad \text{for } 0 < t < +\infty, \\ &Q_\infty := \Omega \times (0,+\infty) \quad \text{and} \quad \Sigma_\infty := \Gamma \times (0,+\infty), \end{aligned} \tag{2.4}$$

and often use the shorter notations Q and Σ if $t = T$, a fixed final time $T \in (0,+\infty)$.

Now, we list our assumptions. For the structure of our system, we are given three functions $g \in C^2[-1,1]$ and $F,\ F_\Gamma \in C^2(-1,1)$ which satisfy

$$g \geq 0, \quad g'' \leq 0, \quad g'(-1) > 0 \quad \text{and} \quad g'(1) < 0, \tag{2.5}$$

$$\lim_{r\searrow -1} F'(r) = \lim_{r\searrow -1} F'_\Gamma(r) = -\infty \quad \text{and} \quad \lim_{r\nearrow 1} F'(r) = \lim_{r\nearrow 1} F'_\Gamma(r) = +\infty, \tag{2.6}$$

$$F''(r) \geq -C \quad \text{and} \quad F''_\Gamma(r) \geq -C, \quad \text{for every } r \in (-1,1), \tag{2.7}$$

$$|F'(r)| \leq \eta |F'_\Gamma(r)| + C \quad \text{for every } r \in (-1,1), \tag{2.8}$$

with some positive constants C and η.

For the initial data, we make rather strong assumptions in order to apply the results of [26] without any trouble. However, our first assumption on μ_0 could be replaced by $\mu_0 \in V$. Precisely, we assume that

$$\mu_0 \in W \quad \text{and} \quad \mu_0 \geq 0 \quad \text{in } \Omega\,; \tag{2.9}$$

$$\rho_0 \in H^2(\Omega)\,, \quad \rho_{0|\Gamma} \in H^2(\Gamma)\,, \quad \min \rho_0 > -1 \quad \text{and} \quad \max \rho_0 < 1\,. \tag{2.10}$$

At this point, we are ready to state our problem. For $\varepsilon \geq 0$, we look for a triplet (μ, ρ, ρ_Γ) satisfying the regularity requirements and solving the problem stated below. As for the regularity, we pretend that

$$\mu \in H^1(0,T;H) \cap C^0([0,T];V) \cap L^2(0,T;W), \tag{2.11}$$

$$(\rho, \rho_\Gamma) \in W^{1,\infty}(0,T;\mathcal{H}) \cap H^1(0,T;\mathcal{V}) \cap L^\infty(0,T;H^2(\Omega) \times H^2(\Gamma)), \tag{2.12}$$

$$\mu \geq 0\,, \quad -1 < \rho < 1 \quad \text{and} \quad (F'(\rho), F'_\Gamma(\rho_\Gamma)) \in L^\infty(0,T;\mathcal{H}), \tag{2.13}$$

for every finite $T > 0$, and the problem reads

$$\big(\varepsilon + 2g(\rho)\big)\partial_t\mu + \mu g'(\rho)\partial_t\rho - \Delta\mu = 0 \quad \text{a.e. in } Q_\infty\,, \tag{2.14}$$

$$\int_\Omega \partial_t\rho\, v + \int_\Gamma \partial_t\rho_\Gamma\, v_\Gamma + \int_\Omega \nabla\rho\cdot\nabla v + \int_\Gamma \nabla_\Gamma\rho_\Gamma\cdot\nabla_\Gamma v_\Gamma$$
$$+ \int_\Omega F'(\rho)v + \int_\Gamma F'_\Gamma(\rho_\Gamma)v_\Gamma = \int_\Omega \mu g'(\rho)v$$
$$\text{a.e. in } (0,+\infty) \text{ and for every } (v, v_\Gamma)\in\mathcal{V}\,, \tag{2.15}$$

$$\mu(0) = \mu_0 \quad \text{and} \quad \rho(0) = \rho_0 \quad \text{a.e. in } \Omega\,. \tag{2.16}$$

Notice that the Neumann boundary condition $\partial_\nu\mu = 0$ and the fact that ρ_Γ is the trace of ρ on Σ are contained in (2.11) and (2.12), respectively, due to the definitions (2.1)–(2.3) of the spaces involved. By accounting for the regularity conditions (2.11)–(2.13), it is clear that the variational problem (2.15) is equivalent to

$$\partial_t\rho - \Delta\rho + F'(\rho) = \mu g'(\rho) \quad \text{in } Q_\infty\,, \tag{2.17}$$

$$\partial_\nu\rho + \partial_t\rho_\Gamma - \Delta_\Gamma\rho_\Gamma + F'_\Gamma(\rho_\Gamma) = 0 \quad \text{on } \Sigma_\infty\,. \tag{2.18}$$

Moreover, it follows from standard embedding results (see, e.g., [51, Sect. 8, Cor. 4]) that $\rho\in C^0(\overline{Q})$ and thus also $\rho_\Gamma\in C^0(\overline{\Sigma})$.

Our starting point is the well-posedness result for $\varepsilon > 0$ that we state below and is already known. Indeed, recalling (2.6)–(2.7), we set

$$\widehat{\beta}(r) := F(r) - F(0) - F'(0)r + \frac{C}{2}r^2 \quad \text{for } r\in(-1,1) \quad \text{and} \quad \widehat{\pi} := F - \widehat{\beta},$$

and analogously introduce $\widehat{\beta}_\Gamma$ and $\widehat{\pi}_\Gamma$, starting from F_Γ. Then, we consider the convex and lower semicontinuous extensions of $\widehat{\beta}$ and $\widehat{\beta}_\Gamma$ to the whole of $\mathbb{R}$ and smooth extensions of $\widehat{\pi}$ and $\widehat{\pi}_\Gamma$ with bounded second derivatives. Therefore, the assumptions of [26, Thm. 2.1] are satisfied and the following well-posedness result holds true.

Theorem 1 *Assume* (2.5)–(2.8) *and* $\varepsilon > 0$ *for the structure and* (2.9)–(2.10) *for the initial data. Then problem* (2.14)–(2.16) *has a unique solution* $(\mu^\varepsilon, \rho^\varepsilon, \rho^\varepsilon_\Gamma)$ *satisfying the regularity properties* (2.11)–(2.13).

Our aim is the following: $i)$ by starting from the solution $(\mu^\varepsilon, \rho^\varepsilon, \rho^\varepsilon_\Gamma)$, we let ε tend to zero and prove that problem (2.14)–(2.16) with $\varepsilon = 0$ has a solution (μ, ρ, ρ_Γ); $ii)$ such a solution is unique; $iii)$ for $\varepsilon \geq 0$, we study the ω-limit of every trajectory.

Indeed, for i) and ii), we prove the following result in Sect. 3:

Theorem 2 *Assume* (2.5)–(2.8) *for the structure and* (2.9)–(2.10) *for the initial data. Then problem* (2.14)–(2.16) *with* $\varepsilon = 0$ *has a unique solution* (μ, ρ, ρ_Γ) *satisfying the regularity properties* (2.11)–(2.13)*. Moreover, for some constants* $\rho_*, \rho^* \in (-1, 1)$ *that depend only on the shape of the nonlinearities and on the initial data, both* (μ, ρ, ρ_Γ) *and the solution* $(\mu^\varepsilon, \rho^\varepsilon, \rho^\varepsilon_\Gamma)$ *given by Theorem 1 satisfy the separation property*

$$\rho_* \le \rho \le \rho^* \quad \textit{and} \quad \rho_* \le \rho^\varepsilon \le \rho^* \quad \textit{in } \overline{\Omega} \times [0, +\infty). \tag{2.19}$$

Finally, $(\mu^\varepsilon, \rho^\varepsilon, \rho^\varepsilon_\Gamma)$ *converges to* (μ, ρ, ρ_Γ) *in a proper topology.*

The last Sect. 4 is devoted to study the long-time behavior of the solution in both the cases $\varepsilon > 0$ and $\varepsilon = 0$. To this end, for a fixed $\varepsilon \ge 0$, we use the simpler symbol (μ, ρ, ρ_Γ) for the solution on $[0, +\infty)$ and observe that the regularity (2.11)–(2.13) on every finite time interval implies that (μ, ρ, ρ_Γ) is a continuous $(H \times \mathcal{V})$-valued function. In particular, it can be evaluated at every time t, and the following definition of ω-limit is completely meaningful:

$$\begin{aligned}\omega(\mu, \rho, \rho_\Gamma) := \Big\{ &(\mu_\omega, \rho_\omega, \rho_{\omega_\Gamma}) \in H \times \mathcal{V} : \ (\mu, \rho, \rho_\Gamma)(t_n) \to (\mu_\omega, \rho_\omega, \rho_{\omega_\Gamma}) \\ &\text{weakly in } H \times \mathcal{V} \text{ for some sequence } t_n \nearrow +\infty \Big\}.\end{aligned} \tag{2.20}$$

Besides, we consider the stationary solutions. It is immediately seen that a stationary solution is a triplet $(\mu_s, \rho_s, \rho_{s_\Gamma})$ satisfying the following conditions: the first component μ_s is a constant, and $(\rho_s, \rho_{s_\Gamma}) \in \mathcal{V}$ is a solution to the system

$$\begin{aligned}&\int_\Omega \nabla\rho_s \cdot \nabla v + \int_\Gamma \nabla_\Gamma \rho_{s_\Gamma} \cdot \nabla_\Gamma v_\Gamma + \int_\Omega F'(\rho_s) v + \int_\Gamma F'_\Gamma(\rho_{s_\Gamma}) v_\Gamma \\ &= \int_\Omega \mu_s\, g'(\rho_s) v \qquad \text{for every } (v, v_\Gamma) \in \mathcal{V}.\end{aligned} \tag{2.21}$$

In terms of a boundary value problem, the conditions $(\rho_s, \rho_{s_\Gamma}) \in \mathcal{V}$ and (2.21) mean that

$$\begin{aligned}&-\Delta\rho_s + F'(\rho_s) = \mu_s\, g'(\rho_s) \quad \text{in } \Omega, \\ &\rho_{s_\Gamma} = \rho_{s|\Gamma} \quad \text{and} \quad \partial_\nu \rho_s - \Delta_\Gamma \rho_{s_\Gamma} + F'_\Gamma(\rho_{s_\Gamma}) = 0 \quad \text{on } \Gamma.\end{aligned}$$

We prove the following result:

Theorem 3 *Assume* (2.5)–(2.8) *and* $\varepsilon \geq 0$ *for the structure and* (2.9)–(2.10) *for the initial data, and let* (μ, ρ, ρ_Γ) *be the unique solution to problem* (2.14)–(2.16) *satisfying the regularity requirements* (2.11)–(2.13)*. Then the* ω*-limit* (2.20) *is nonempty and consists only of stationary solutions. In particular, there exists a constant* μ_s *such that problem* (2.21) *has at least one solution* $(\rho_s, \rho_{s_\Gamma}) \in \mathcal{V}$.

Throughout the paper, we will repeatedly use the Young inequality

$$a\,b \leq \delta\, a^2 + \frac{1}{4\delta}\, b^2 \quad \text{for all } a, b \in \mathbb{R} \text{ and } \delta > 0, \tag{2.22}$$

as well as the Hölder inequality and the continuity of the embedding $V \subset L^p(\Omega)$ for every $p \in [1, 6]$ (since Ω is three-dimensional, bounded and smooth). Besides, this embedding is compact for $p < 6$, and also the embedding $W \subset C^0(\overline{\Omega})$ is compact. In particular, we have the compactness inequality

$$\|v\|_4 \;\leq\; \delta\,\|\nabla v\|_2 \;+\; \widetilde{C}_\delta\,\|v\|_2 \quad \text{for every } v \in H^1(\Omega) \text{ and } \delta > 0, \tag{2.23}$$

where $\widetilde{C}_\delta$ depends only on Ω and δ. We also recall some well-known estimates from trace theory and from the theory of elliptic equations we use in the sequel. For any v and v_Γ that make the right-hand sides meaningful, we have that

$$\|\partial_\nu v\|_{H^{-1/2}(\Gamma)} \leq C_\Omega\big(\|v\|_{H^1(\Omega)} + \|\Delta v\|_{L^2(\Omega)}\big)\,, \tag{2.24}$$

$$\|\partial_\nu v\|_{L^2(\Gamma)} \leq C_\Omega\big(\|v\|_{H^{3/2}(\Omega)} + \|\Delta v\|_{L^2(\Omega)}\big)\,, \tag{2.25}$$

$$\|v\|_{H^2(\Omega)} \leq C_\Omega\big(\|v_{|\Gamma}\|_{H^{3/2}(\Gamma)} + \|\Delta v\|_{L^2(\Omega)}\big)\,, \tag{2.26}$$

$$\|v\|_{H^2(\Omega)} \leq C_\Omega\big(\|v\|_{H^1(\Omega)} + \|\Delta v\|_{L^2(\Omega)}\big) \quad \text{if } \partial_\nu v = 0 \;\text{ on } \Gamma\,, \tag{2.27}$$

$$\|v_\Gamma\|_{H^2(\Gamma)} \leq C_\Omega\big(\|v_\Gamma\|_{H^1(\Gamma)} + \|\Delta_\Gamma v_\Gamma\|_{L^2(\Gamma)}\big)\,, \tag{2.28}$$

$$\|v_\Gamma\|_{H^{3/2}(\Gamma)} \leq C_\Omega\big(\|v_\Gamma\|_{H^1(\Gamma)} + \|\Delta_\Gamma v_\Gamma\|_{H^{-1/2}(\Gamma)}\big)\,, \tag{2.29}$$

with a constant $C_\Omega > 0$ that depends only on Ω.

We conclude this section by stating a general rule concerning the constants that appear in the estimates to be performed in the sequel. The small-case symbol c stands for a generic constant whose values might change from line to line and even within the same line and depends only on Ω, on the shape of the nonlinearities, and on the constants and the norms of the functions involved in the assumptions of our statements. In particular, the values of c do not depend on ε and T if the latter is considered. A small-case symbol with a subscript like c_δ (in particular, with $\delta = T$) indicates that the constant might depend on the parameter δ, in addition. On the contrary, we mark precise constants that we can refer to by using different symbols, like in (2.7)–(2.8) and (2.23)–(2.29).

3 Well-Posedness

This section is devoted to the proof of Theorem 2. First, we prove the separation properties (2.19). Then, we show uniqueness. Finally, we prove convergence for the family $\{(\mu^\varepsilon, \rho^\varepsilon, \rho_\Gamma^\varepsilon)\}$ and derive existence for the problem with $\varepsilon = 0$.

Separation We assume that $\varepsilon \geq 0$ and that (μ, ρ, ρ_Γ) is a solution to problem (2.14)–(2.16) satisfying (2.11)–(2.13). Recalling (2.10) and (2.5)–(2.7), we may choose $\rho_*, \rho^* \in (-1, 1)$ such that $\rho_* \leq \rho_0 \leq \rho^*$ and

$$\begin{aligned} &g'(r) > 0 \quad \text{and} \quad F'(r) < 0 \quad \text{for } -1 < r \leq \rho_*, \\ &g'(r) < 0 \quad \text{and} \quad F'(r) > 0 \quad \text{for } \rho^* \leq r < 1. \end{aligned}$$

Now, we show that $\rho_* \leq \rho \leq \rho^*$, using the positivity of μ (see (2.13)). In fact, we prove just the upper inequality, since the proof of the other is similar. We test (2.15), written at the time s, by $((\rho - \rho^*)^+, (\rho_\Gamma - \rho^*)^+)(s)$ and integrate over $(0, t)$ with respect to s. We have

$$\begin{aligned} &\frac{1}{2}\int_\Omega |(\rho(t) - \rho^*)^+|^2 + \frac{1}{2}\int_\Gamma |(\rho_\Gamma(t) - \rho^*)^+|^2 \\ &+ \int_{Q_t} |\nabla(\rho - \rho^*)^+|^2 + \int_{\Sigma_t} |\nabla_\Gamma(\rho_\Gamma - \rho^*)^+|^2 \\ &+ \int_{Q_t} F'(\rho)\,(\rho - \rho^*)^+ + \int_{\Sigma_t} F_\Gamma'(\rho_\Gamma)\,(\rho_\Gamma - \rho^*)^+ = \int_{Q_t} \mu g'(\rho)(\rho - \rho^*)^+ . \end{aligned}$$

All of the terms on the left-hand side are nonnegative, while the right-hand side is nonpositive. We conclude that $(\rho(t) - \rho^*)^+ = 0$ in $\overline{\Omega}$ for every $t > 0$, i.e., our assertion.

Consequence Since g, F and F_Γ are smooth on $(-1, 1)$ and (2.5) implies that g is strictly positive on $(-1, 1)$, the separation inequalities (2.19) imply the bounds

$$g(\rho) \geq g_* > 0 \quad \text{and} \quad |\Phi(\rho)| \leq C^* \quad \text{in } \overline{Q}_\infty, \quad |\Phi_\Gamma(\rho_\Gamma)| \leq C^* \quad \text{on } \overline{\Sigma}_\infty, \tag{3.1}$$

for $\Phi \in \{g, g', g'', F, F', F''\}$ and $\Phi_\Gamma \in \{F_\Gamma, F_\Gamma', F_\Gamma''\}$, and for some constants g_* and C^* that depend only on the shape of the nonlinearities and the initial datum ρ_0. In particular, they do not depend on ε.

Uniqueness We prove that the solution to problem (2.14)–(2.16) with $\varepsilon = 0$ is unique. To this end, we fix $T > 0$ and two solutions $(\mu_i, \rho_i, \rho_{i_\Gamma})$, $i = 1, 2$, and show that they coincide on $\overline{\Omega} \times [0, T]$. We set for convenience $\mu := \mu_1 - \mu_2$ and analogously define ρ and ρ_Γ. Then, we write (2.14) for both solutions and test the difference by μ. Using the identity

$$
\begin{aligned}
&\{2g(\rho_1)\partial_t\mu_1 + \mu_1 g'(\rho_1)\partial_t\rho_1 - 2g(\rho_2)\partial_t\mu_2 - \mu_2 g'(\rho_2)\partial_t\rho_2\}\mu \\
&= \partial_t\bigl(g(\rho_1)\,\mu^2\bigr) + 2\partial_t\mu_2\,\bigl(g(\rho_1) - g(\rho_2)\bigr)\,\mu + \mu_2\bigl(g'(\rho_1)\partial_t\rho_1 - g'(\rho_2)\partial_t\rho_2\bigr)\mu\,,
\end{aligned}
$$

we obtain that

$$
\begin{aligned}
&\int_\Omega g(\rho_1(t))\,|\mu(t)|^2 + \int_{Q_t} |\nabla\mu|^2 \\
&\quad = -\int_{Q_t} 2\partial_t\mu_2\,\bigl(g(\rho_1) - g(\rho_2)\bigr)\,\mu - \int_{Q_t} \mu_2\bigl(g'(\rho_1)\partial_t\rho_1 - g'(\rho_2)\partial_t\rho_2\bigr)\mu\,.
\end{aligned}
\tag{3.2}
$$

Next, we write (2.15) at the time s for both solutions, test the difference by $\partial_t(\rho, \rho_\Gamma)(s)$, and integrate over $(0,t)$ with respect to s. Then, we add $\int_{Q_t} \rho\,\partial_t\rho + \int_{\Sigma_t} \rho_\Gamma\,\partial_t\rho_\Gamma$ to both sides. We get

$$
\begin{aligned}
&\int_{Q_t} |\partial_t\rho|^2 + \int_{\Sigma_t} |\partial_t\rho_\Gamma|^2 + \frac{1}{2}\,\|\rho(t)\|_V^2 + \frac{1}{2}\,\|\rho_\Gamma(t)\|_{V_\Gamma}^2 \\
&\quad = -\int_{Q_t} \bigl(F'(\rho_1) - F'(\rho_2)\bigr)\partial_t\rho - \int_{\Sigma_t} \bigl(F'_\Gamma(\rho_{1_\Gamma}) - F'_\Gamma(\rho_{2_\Gamma})\bigr)\partial_t\rho_\Gamma \\
&\qquad + \int_{Q_t} \bigl(\mu_1 g'(\rho_1) - \mu_2 g'(\rho_2)\bigr)\partial_t\rho + \int_{Q_t} \rho\,\partial_t\rho + \int_{\Sigma_t} \rho_\Gamma\,\partial_t\rho_\Gamma\,.
\end{aligned}
\tag{3.3}
$$

At this point, we add (3.2)–(3.3) to each other and use the separation property, the first inequality in (3.1) for ρ_1, and the boundedness and the Lipschitz continuity of the nonlinearities on $[\rho_*, \rho^*]$. We find that

$$
\begin{aligned}
&g_*\int_\Omega |\mu(t)|^2 + \int_{Q_t} |\nabla\mu|^2 + \int_{Q_t} |\partial_t\rho|^2 \\
&+ \int_{\Sigma_t} |\partial_t\rho_\Gamma|^2 + \frac{1}{2}\,\|\rho(t)\|_V^2 + \frac{1}{2}\,\|\rho_\Gamma(t)\|_{V_\Gamma}^2 \\
&\le c\int_{Q_t} |\partial_t\mu_2|\,|\rho|\,|\mu| + c\int_{Q_t} \mu_2\bigl(|\partial_t\rho| + |\rho|\,|\partial_t\rho_2|\bigr)\,|\mu| \\
&\quad + c\int_{Q_t} |\rho|\,|\partial_t\rho| + c\int_{\Sigma_t} |\rho_\Gamma|\,|\partial_t\rho_\Gamma| + c\int_{Q_t} \bigl(\mu_1|\rho| + |\mu|\bigr)\,|\partial_t\rho|\,.
\end{aligned}
\tag{3.4}
$$

Many integrals on the right-hand side can be dealt with just using the Hölder and Young inequalities. Thus, we consider just the terms that need some treatment. In the next lines, we owe to the continuous embeddings $V \subset L^p(\Omega)$ for $p \in [1,6]$

and $W \subset C^0(\overline{\Omega})$, and δ is a positive parameter. We have

$$\int_{Q_t} |\partial_t \mu_2|\, |\rho|\, |\mu| \le \int_0^t \|\partial_t \mu_2(s)\|_2 \|\rho(s)\|_4 \|\mu(s)\|_4 \, ds$$
$$\le \delta \int_0^t \|\mu(s)\|_V^2 \, ds + c_\delta \int_0^t \|\partial_t \mu_2(s)\|_H^2 \|\rho(s)\|_V^2 \, ds \,,$$

and we notice that the function $s \mapsto \|\partial_t \mu_2(s)\|_H^2$ belongs to $L^1(0,T)$ by (2.11) for μ_2. We estimate the next integral as follows,

$$\int_{Q_t} \mu_2 \bigl(|\partial_t \rho| + |\rho|\, |\partial_t \rho_2|\bigr)\, |\mu|$$
$$\le \int_0^t \|\mu_2(s)\|_\infty \|\partial_t \rho(s)\|_2 \|\mu(s)\|_2 \, ds$$
$$+ c \int_0^t \|\mu_2(s)\|_6 \|\rho(s)\|_6 \|\partial_t \rho_2(s)\|_6 \|\mu(s)\|_6 \, ds$$
$$\le \delta \int_0^t \|\partial_t \rho(s)\|_H^2 \, ds + c_\delta \int_0^t \|\mu_2(s)\|_W^2 \|\mu(s)\|_H^2 \, ds$$
$$+ \delta \int_0^t \|\mu(s)\|_V^2 \, ds + c_\delta \int_0^t \|\mu_2(s)\|_V^2 \|\partial_t \rho_2(s)\|_V^2 \|\rho(s)\|_V^2 \, ds \,,$$

and we point out that the functions $s \mapsto \|\mu_2(s)\|_W^2$, $s \mapsto \|\mu_2(s)\|_V^2$, and $s \mapsto \|\partial_t \rho_2(s)\|_V^2$, belong to $L^1(0,T)$, $L^\infty(0,T)$ and $L^1(0,T)$, respectively, due to (2.11)–(2.12) for μ_2 and ρ_2. Finally, we estimate one further term. We have that

$$\int_{Q_t} \mu_1 |\rho|\, |\partial_t \rho| \le \int_0^t \|\mu_1(s)\|_4 \|\rho(s)\|_4 \|\partial_t \rho(s)\|_2 \, ds$$
$$\le \delta \int_0^t \|\partial_t \rho\|_H^2 \, ds + c_\delta \int_0^t \|\mu_1(s)\|_V^2 \|\rho(s)\|_V^2 \, ds \,,$$

where the function $s \mapsto \|\mu_1(s)\|_V^2$ belongs to $L^\infty(0,T)$. Therefore, by choosing δ small enough and coming back to (3.4), we can apply the Gronwall lemma to conclude that (μ, ρ, ρ_Γ) vanishes on $\overline{\Omega} \times [0,T]$.

Now, we show the existence of a solution to problem (2.14)–(2.16) with $\varepsilon = 0$ and prove the last sentence of the statement of Theorem 2. To do that, it suffices to establish a number of a priori estimates on the solution $(\mu^\varepsilon, \rho^\varepsilon, \rho_\Gamma^\varepsilon)$ on an arbitrarily fixed time interval $[0,T]$ and to use proper compactness results. As the uniqueness of the solution to the limiting problem is already known, it follows that the convergence properties proved below for a subsequence actually hold for the whole family. In view of the asymptotic behavior that we aim to study in the next section, we distinguish in the notation the constants that may depend on T, as explained at the end of Sect. 2. Of course, we can assume $\varepsilon \le 1$. In order to keep the length of the paper reasonable, we perform some of the next estimates just formally.

First a Priori Estimate We observe that

$$\big\{(\varepsilon + 2g(\rho^\varepsilon))\partial_t\mu^\varepsilon + \mu^\varepsilon g'(\rho^\varepsilon)\partial_t\rho^\varepsilon\big\}\mu^\varepsilon = \partial_t\big(\big(\tfrac{\varepsilon}{2} + g(\rho^\varepsilon)\big)|\mu^\varepsilon|^2\big).$$

Hence, if we multiply (2.14) by μ^ε and integrate over Q_t, we obtain that

$$\frac{\varepsilon}{2}\int_\Omega |\mu^\varepsilon(t)|^2 + \int_\Omega g(\rho^\varepsilon(t))|\mu^\varepsilon(t)|^2 + \int_{Q_t} |\nabla\mu^\varepsilon|^2 = \frac{\varepsilon}{2}\int_\Omega \mu_0^2 + \int_\Omega g(\rho_0)\mu_0^2 \,.$$

By accounting for (2.19) and (3.1), we deduce, for every $t \ge 0$, the global estimate

$$g_*\int_\Omega |\mu^\varepsilon(t)|^2 + \int_{Q_t} |\nabla\mu^\varepsilon|^2 \le \frac{1}{2}\int_\Omega \mu_0^2 + \int_\Omega g(\rho_0)\mu_0^2 = c\,. \tag{3.5}$$

Second a Priori Estimate We write (2.15) at the time s and choose the test pair $(v, v_\Gamma) = (\partial_t\rho^\varepsilon, \partial_t\rho_\Gamma^\varepsilon)(s)$, which is allowed by the regularity (2.12). Then, we integrate over $(0, t)$. Thanks to the Schwarz and Young inequalities, we have

$$\begin{aligned}
&\int_{Q_t} |\partial_t\rho^\varepsilon|^2 + \int_{\Sigma_t} |\partial_t\rho_\Gamma^\varepsilon|^2 + \frac{1}{2}\int_\Omega |\nabla\rho^\varepsilon(t)|^2 + \frac{1}{2}\int_\Gamma |\nabla_\Gamma\rho_\Gamma^\varepsilon(t)|^2 \\
&\quad + \int_\Omega F(\rho^\varepsilon(t)) + \int_\Gamma F_\Gamma(\rho_\Gamma^\varepsilon(t)) \\
&= \frac{1}{2}\int_\Omega |\nabla\rho_0|^2 + \frac{1}{2}\int_\Gamma |\nabla_\Gamma\rho_{0|\Gamma}|^2 \\
&\quad + \int_\Omega F(\rho_0) + \int_\Gamma F_\Gamma(\rho_{0|\Gamma}) + \int_{Q_t} \mu^\varepsilon g'(\rho^\varepsilon)\partial_t\rho^\varepsilon \\
&\le c + \frac{1}{2}\int_{Q_t} |\partial_t\rho^\varepsilon|^2 + c\int_{Q_t} |\mu^\varepsilon|^2.
\end{aligned}$$

Since $|\rho^\varepsilon| \le 1$, (3.5) holds, and (2.7) implies that F and F_Γ are bounded from below, we deduce that

$$\begin{aligned}
&\|(\rho^\varepsilon, \rho_\Gamma^\varepsilon)\|_{H^1(0,T;\mathcal{H})\cap L^\infty(0,T;\mathcal{V})} + \|F(\rho^\varepsilon)\|_{L^\infty(0,T;L^1(\Omega))} \\
&\qquad\qquad + \|F_\Gamma(\rho_\Gamma^\varepsilon)\|_{L^\infty(0,T;L^1(\Gamma))} \le c_T\,.
\end{aligned} \tag{3.6}$$

Third a Priori Estimate By starting from (2.17)–(2.18) and accounting for (3.1) and (3.5)–(3.6), we successively deduce a number of estimates with the help of the inequalities (2.24)–(2.29), written with $v = \rho^\varepsilon(t)$ and $v_\Gamma = \rho^\varepsilon_\Gamma(t)$ and then squared and integrated over $(0, T)$. We have

$$\begin{aligned}
&\|\Delta\rho^\varepsilon\|_{L^2(0,T;H)} \le c_T \quad \text{from (2.17)},\\
&\|\partial_\nu\rho^\varepsilon\|_{L^2(0,T;H^{-1/2}(\Gamma))} \le c_T \quad \text{from (2.24)},\\
&\|\Delta_\Gamma\rho^\varepsilon_\Gamma\|_{L^2(0,T;H^{-1/2}(\Gamma))} \le c_T \quad \text{from (2.18)},\\
&\|\rho^\varepsilon_\Gamma\|_{L^2(0,T;H^{3/2}(\Gamma))} \le c_T \quad \text{from (2.29)},\\
&\|\rho^\varepsilon\|_{L^2(0,T;H^2(\Omega))} \le c_T \quad \text{from (2.26)},\\
&\|\partial_\nu\rho^\varepsilon\|_{L^2(0,T;H_\Gamma)} \le c_T \quad \text{from (2.25)},\\
&\|\Delta_\Gamma\rho^\varepsilon_\Gamma\|_{L^2(0,T;H_\Gamma)} \le c_T \quad \text{from (2.18)},\\
&\|\rho^\varepsilon_\Gamma\|_{L^2(0,T;H^2(\Gamma))} \le c_T \quad \text{from (2.28)}.
\end{aligned}$$

In conclusion, we have proved that

$$\|\rho^\varepsilon\|_{L^2(0,T;H^2(\Omega))} + \|\rho^\varepsilon_\Gamma\|_{L^2(0,T;H^2(\Gamma))} \le c_T\,. \tag{3.7}$$

Fourth a Priori Estimate We (formally) differentiate (2.15) with respect to time and set $\zeta := \partial_t\rho^\varepsilon$ and $\zeta_\Gamma := \partial_t\rho^\varepsilon_\Gamma$, for brevity. Then we write the variational equation we obtain at the time s and test it by $(\zeta, \zeta_\Gamma)(s)$. Finally, we integrate over $(0, t)$ and add $C\int_{Q_t}|\zeta|^2 + C\int_{\Sigma_t}|\zeta_\Gamma|^2$ to both sides, where C is the constant that appears in (2.7). We obtain the identity

$$\begin{aligned}
&\frac{1}{2}\int_\Omega|\zeta(t)|^2 + \frac{1}{2}\int_\Gamma|\zeta_\Gamma(t)|^2 + \int_{Q_t}|\nabla\zeta|^2 + \int_{\Sigma_t}|\nabla_\Gamma\zeta_\Gamma|^2\\
&\quad + \int_{Q_t}\big(F''(\rho^\varepsilon) + C\big)|\zeta|^2 + \int_{Q_t}\big(F''_\Gamma(\rho^\varepsilon_\Gamma) + C\big)|\zeta_\Gamma|^2\\
&= \frac{1}{2}\int_\Omega|\zeta(0)|^2 + \frac{1}{2}\int_\Gamma|\zeta_\Gamma(0)|^2 + \int_{Q_t}\partial_t\mu^\varepsilon\, g'(\rho^\varepsilon)\zeta + \int_{Q_t}\mu^\varepsilon g''(\rho^\varepsilon)|\zeta|^2\\
&\quad + C\int_{Q_t}|\zeta|^2 + C\int_{\Sigma_t}|\zeta_\Gamma|^2\,.
\end{aligned} \tag{3.8}$$

All of the terms on the left-hand side are nonnegative, while the second volume integral over Q_t on the right-hand side is nonpositive since $\mu^\varepsilon \ge 0$ and $g'' \le 0$. It remains to find bounds for the first volume integral over Q_t on the right-hand side and for the sum of the terms that involve the initial values. We handle the latter first. To this end, we write (2.15) at the time $t = 0$ and test it by $(v, v_\Gamma) = (\zeta, \zeta_\Gamma)(0)$.

We obtain

$$\int_\Omega |\zeta(0)|^2 + \int_\Gamma |\zeta_\Gamma(0)|^2 = -\int_\Omega \nabla\rho_0 \cdot \nabla\zeta(0) - \int_\Gamma \nabla_\Gamma \rho_{0|\Gamma} \cdot \nabla_\Gamma \zeta_\Gamma(0)$$
$$- \int_\Omega F'(\rho_0)\zeta(0) - \int_\Gamma F'_\Gamma(\rho_{0|\Gamma})\zeta_\Gamma(0) + \int_\Omega \mu_0 g'(\rho_0)\zeta(0)\,. \tag{3.9}$$

On account of (2.10), we have, using Young's inequality and (2.25),

$$\begin{aligned}&-\int_\Omega \nabla\rho_0 \cdot \nabla\zeta(0) - \int_\Gamma \nabla_\Gamma \rho_{0|\Gamma} \cdot \nabla_\Gamma \zeta_\Gamma(0)\\ &= \int_\Omega \Delta\rho_0\, \zeta(0) - \int_\Gamma \bigl(\partial_\nu \rho_0 - \Delta_\Gamma \rho_{0|\Gamma}\bigr)\zeta_\Gamma(0)\\ &\le \frac{1}{4}\int_\Omega |\zeta(0)|^2 + \frac{1}{4}\int_\Gamma |\zeta_\Gamma(0)|^2 + c\,\|\rho_0\|^2_{H^2(\Omega)} + c\,\|\rho_{0|\Gamma}\|^2_{H^2(\Gamma)}\,.\end{aligned}$$

Moreover, it follows from (2.9), (2.10), (3.1), and Young's inequality that the expression in the second line of (3.9) is bounded by

$$\frac{1}{4}\int_\Omega |\zeta(0)|^2 \,+\, \frac{1}{4}\int_\Gamma |\zeta_\Gamma(0)|^2 \,+\, c\,.$$

We thus have shown that

$$\int_\Omega |\zeta(0)|^2 + \int_\Gamma |\zeta_\Gamma(0)|^2 \,\le\, c\,. \tag{3.10}$$

It remains to bound the first volume integral over Q_t in (3.8), which we denote by I. This estimate requires more effort. At first, observe that (2.14) implies that

$$\partial_t \mu^\varepsilon \,=\, \frac{1}{\varepsilon + 2g(\rho^\varepsilon)}\,\Delta\mu^\varepsilon \,-\, \frac{g'(\rho^\varepsilon)}{\varepsilon + 2g(\rho^\varepsilon)}\,\zeta\,\mu^\varepsilon\,, \tag{3.11}$$

where, thanks to (3.1), $1/(\varepsilon + 2g(\rho^\varepsilon)) \le 1/(2g^*)$ for all $\varepsilon > 0$. Now, using (3.11), we find that

$$I \,= \int_{Q_t} \frac{g'(\rho^\varepsilon)\,\zeta}{\varepsilon + 2g(\rho^\varepsilon)}\,\Delta\mu^\varepsilon \,-\, \int_{Q_t} \mu^\varepsilon\,\frac{(g'(\rho^\varepsilon))^2}{\varepsilon + 2g(\rho^\varepsilon)}\,\zeta^2 \,=:\, I_1 + I_2\,, \tag{3.12}$$

with obvious notation. The second integral is easy to handle. In fact, thanks to (3.1), (2.23), and Hölder's and Young's inequalities, we infer that

$$\begin{aligned} I_2 &\le c \int_0^t \|\mu^\varepsilon(s)\|_4 \, \|\zeta(s)\|_2 \, \|\zeta(s)\|_4 \, ds \\ &\le \frac{1}{6} \int_{Q_t} |\nabla \zeta|^2 + c \int_0^t \Bigl(1 + \|\mu^\varepsilon(s)\|_V^2\Bigr) \|\zeta(s)\|_H^2 \, ds \,, \end{aligned} \tag{3.13}$$

where we know from (3.5) that $\int_0^T \|\mu^\varepsilon(s)\|_V^2 \, ds \le c_T$. For the first integral, integration by parts and (3.1) yield that

$$\begin{aligned} I_1 &= - \int_{Q_t} \nabla \mu^\varepsilon \cdot \nabla \Bigl(\frac{g'(\rho^\varepsilon) \, \zeta}{\varepsilon + 2 g(\rho^\varepsilon)} \Bigr) \\ &\le C_1 \int_{Q_t} |\nabla \mu^\varepsilon| \, |\nabla \zeta| + C_1 \int_{Q_t} |\nabla \mu^\varepsilon| \, |\nabla \rho^\varepsilon| \, |\zeta| =: C_1 (I_{11} + I_{12}), \end{aligned} \tag{3.14}$$

with obvious notation. Clearly, owing to (3.5) and Young's inequality, we find that

$$C_1 \, I_{11} \le \frac{1}{6} \int_{Q_t} |\nabla \zeta|^2 + c \,. \tag{3.15}$$

Moreover, invoking Hölder's and Young's inequalities, the compactness inequality (2.23), as well as the continuity of the embedding $H^2(\Omega) \subset W^{1,4}(\Omega)$, we infer that

$$\begin{aligned} C_1 \, I_{12} &\le C_1 \int_0^t \|\nabla \mu^\varepsilon(s)\|_2 \, \|\nabla \rho^\varepsilon(s)\|_4 \, \|\zeta(s)\|_4 \, ds \\ &\le \frac{1}{6} \int_{Q_t} |\nabla \zeta|^2 + c \int_{Q_t} |\zeta|^2 + c \int_0^t \|\nabla \mu^\varepsilon(s)\|_2^2 \, \|\rho^\varepsilon(s)\|_{H^2(\Omega)}^2 \, ds \,. \end{aligned} \tag{3.16}$$

Notice that $\int_0^T \|\nabla \mu^\varepsilon(s)\|_2^2 \, ds \le c$ for every $T > 0$, by virtue of (3.5). We now aim to estimate $\|\rho^\varepsilon(s)\|_{H^2(\Omega)}$ in terms of ζ and ζ_Γ. To this end, we derive a chain of estimates which are each valid for almost every $s \in (0, T)$. To begin with, we deduce from (3.5) and (3.6) that

$$\|\Delta \rho^\varepsilon(s)\|_2 = \|\zeta(s) + F'(\rho^\varepsilon(s)) - \mu^\varepsilon(s) \, g'(\rho^\varepsilon(s))\|_2 \le c + \|\zeta(s)\|_2 \,. \tag{3.17}$$

Consequently, by (2.24) we have that

$$\|\partial_\nu \rho^\varepsilon(s)\|_{H^{-1/2}(\Gamma)} \le C_\Omega \bigl(\|\rho^\varepsilon(s)\|_V + \|\Delta \rho^\varepsilon(s)\|_2\bigr) \le c_T \, (1 + \|\zeta(s)\|_2) \,, \tag{3.18}$$

and (2.18), (3.1) and (3.6) imply that

$$\begin{aligned}\|\Delta_\Gamma \rho_\Gamma^\varepsilon(s)\|_{H^{-1/2}(\Gamma)} &\le \|\partial_\nu \rho^\varepsilon(s) + F_\Gamma'(\rho_\Gamma^\varepsilon(s)) + \zeta_\Gamma(s)\|_{H^{-1/2}(\Gamma)} \\ &\le c_T\, (1 + \|\zeta(s)\|_2 + \|\zeta_\Gamma(s)\|_{H_\Gamma})\,. \end{aligned} \tag{3.19}$$

But then, thanks to (2.29) and (3.6), it is clear that

$$\begin{aligned}\|\rho_\Gamma^\varepsilon(s)\|_{H^{3/2}(\Gamma)} &\le C_\Omega \left(\|\rho_\Gamma^\varepsilon(s)\|_{H^1(\Gamma)} + \|\Delta_\Gamma \rho_\Gamma^\varepsilon(s)\|_{H^{-1/2}(\Gamma)}\right) \\ &\le c_T\, (1 + \|\zeta(s)\|_2 + \|\zeta_\Gamma(s)\|_{H_\Gamma})\,, \end{aligned} \tag{3.20}$$

whence, owing to (2.26), we finally arrive at the estimate

$$\|\rho^\varepsilon(s)\|_{H^2(\Omega)} \;\le\; c_T \left(1 + \|\zeta(s)\|_H + \|\zeta_\Gamma(s)\|_{H_\Gamma}\right)\,. \tag{3.21}$$

We thus obtain from (3.16) that

$$\begin{aligned} C_1\, I_{12} \;\le\; \frac{1}{6}\int_{Q_t} |\nabla\zeta|^2 \;+\; c\int_{Q_t} |\zeta|^2 \;+\; c_T \\ +\; c_T \int_0^t \|\nabla\mu^\varepsilon(s)\|_2^2 \left(\|\zeta(s)\|_H^2 + \|\zeta_\Gamma(s)\|_{H_\Gamma}^2\right) ds\,. \end{aligned} \tag{3.22}$$

Therefore, recalling (3.8) and invoking the estimates (3.10), (3.13)–(3.16), we can apply Gronwall's lemma and conclude that

$$\|(\partial_t\rho^\varepsilon, \partial_t\rho_\Gamma^\varepsilon)\|_{L^\infty(0,T;\mathcal{H})\cap L^2(0,T;\mathcal{V})} \le c_T\,. \tag{3.23}$$

Fifth a Priori Estimate We now notice that (3.21) and (3.23) imply that

$$\|\rho^\varepsilon\|_{L^\infty(0,T;H^2(\Omega))} \;\le\; c_T\,. \tag{3.24}$$

Then we may infer from (2.25), (2.18), (2.28), in this order, the estimates

$$\begin{aligned} &\|\partial_\nu\rho^\varepsilon\|_{L^\infty(0,T;H_\Gamma)} \le c_T\,, \\ &\quad \|\Delta_\Gamma\rho_\Gamma^\varepsilon\|_{L^\infty(0,T;H_\Gamma)} \le c_T\,, \quad \|\rho_\Gamma^\varepsilon\|_{L^\infty(0,T;H^2(\Gamma))} \le c_T\,, \end{aligned}$$

so that

$$\|(\rho^\varepsilon, \rho_\Gamma^\varepsilon)\|_{L^\infty(0,T;H^2(\Omega)\times H^2(\Gamma))} \le c_T\,. \tag{3.25}$$

Sixth a Priori Estimate At this point, we can multiply (2.14) by $\partial_t \mu^\varepsilon$ and integrate over Q_t. Then, we add $\int_{Q_t} \mu^\varepsilon \partial_t \mu^\varepsilon$ to both sides. By owing to the Hölder, Sobolev and Young inequalities, we obtain

$$\begin{aligned}
&\int_{Q_t} \bigl(\varepsilon + 2g(\rho^\varepsilon)\bigr) |\partial_t \mu^\varepsilon|^2 + \frac{1}{2} \, \|\mu^\varepsilon(t)\|_V^2 \\
&= \frac{1}{2} \, \|\mu_0\|_V^2 + \int_{Q_t} \mu^\varepsilon \partial_t \mu^\varepsilon - \int_{Q_t} \mu^\varepsilon g'(\rho^\varepsilon) \partial_t \rho^\varepsilon \partial_t \mu^\varepsilon \\
&\leq c + \int_0^t \|\mu^\varepsilon(s)\|_2 \|\partial_t \mu^\varepsilon(s)\|_2 \, ds + c \int_0^t \|\mu^\varepsilon(s)\|_4 \|\partial_t \rho^\varepsilon(s)\|_4 \|\partial_t \mu^\varepsilon(s)\|_2 \, ds \\
&\leq c + g_* \int_{Q_t} |\partial_t \mu^\varepsilon|^2 + c \, \|\mu^\varepsilon\|_{L^2(0,t;H)}^2 + c \int_0^t \|\partial_t \rho^\varepsilon(s)\|_V^2 \, \|\mu^\varepsilon(s)\|_V^2 \, ds \, ,
\end{aligned}$$

where g_* is the constant introduced in (3.1). As $2g(\rho^\varepsilon) \geq 2g_*$, we may use (3.5), (3.23) and Gronwall's lemma to conclude that

$$\|\mu^\varepsilon\|_{H^1(0,T;H) \cap L^\infty(0,T;V)} \leq c_T \, . \tag{3.26}$$

By comparison in (2.14), we estimate $\Delta \mu^\varepsilon$. Hence, by applying (2.27), we derive that

$$\|\mu^\varepsilon\|_{L^2(0,T;W)} \leq c_T \, . \tag{3.27}$$

Conclusion If we collect all the previous estimates and use standard compactness results, then we have (in principle for a subsequence) that

$$\begin{aligned}
&\mu^\varepsilon \to \mu \quad \text{in } H^1(0,T;H) \cap L^\infty(0,T;V) \cap L^2(0,T;W) \, , \\
&(\rho^\varepsilon, \rho_\Gamma^\varepsilon) \to (\rho, \rho_\Gamma) \\
&\qquad \text{in } W^{1,\infty}(0,T;\mathcal{H}) \cap H^1(0,T;\mathcal{V}) \cap L^\infty(0,T;H^2(\Omega) \times H^2(\Gamma)) \, ,
\end{aligned}$$

as $\varepsilon \searrow 0$, the convergence being understood in the sense of the corresponding weak star topologies. Notice that the limiting triplet fulfills the regularity requirements (2.11)–(2.13). Next, by the compact embeddings $V \subset L^5(\Omega)$, $H^2(\Omega) \subset C^0(\overline{\Omega})$, and $H^2(\Gamma) \subset C^0(\Gamma)$, and using well-known strong compactness results (see, e.g., [51, Sect. 8, Cor. 4]), we deduce the useful strong convergence

$$\mu^\varepsilon \to \mu \quad \text{in } C^0([0,T]; L^5(\Omega)), \quad (\rho^\varepsilon, \rho_\Gamma^\varepsilon) \to (\rho, \rho_\Gamma) \quad \text{in } C^0(\overline{Q}) \times C^0(\overline{\Sigma}). \tag{3.28}$$

This allows us to deal with nonlinearities and to take the limits of the products that appear in the equations. Hence, we easily conclude that the triplet (μ, ρ, ρ_Γ) solves (2.14) and the time-integrated version of (2.15) on $(0, T)$ (which is equivalent to (2.15) itself) with $\varepsilon = 0$. Moreover, the initial conditions (2.16) easily pass to the limit in view of (3.28). This concludes the existence proof. By uniqueness, the whole family $\{(\mu^\varepsilon, \rho^\varepsilon, \rho_\Gamma^\varepsilon)\}$ converges to (μ, ρ, ρ_Γ) in the above topology as $\varepsilon \searrow 0$. □

4 Long-Time Behavior

This section is devoted to the proof of Theorem 3. In the sequel, it is understood that $\varepsilon \in [0, 1]$ is fixed and that (μ, ρ, ρ_Γ) is the unique solution to problem (2.14)–(2.16) given by Theorems 1 and 2 in the two cases $\varepsilon > 0$ and $\varepsilon = 0$, respectively. First of all, we have to show that the ω-limit (2.20) is nonempty. This necessitates proper a priori estimates on the whole half-line $\{t \geq 0\}$.

First Global Estimate From (3.5), we immediately deduce that

$$\|\mu\|_{L^\infty(0,+\infty;H)} \leq c \quad \text{and} \quad \int_{Q_\infty} |\nabla \mu|^2 \leq c\,. \tag{4.1}$$

Second Global Estimate We start by rearranging (2.14) as follows:

$$\mu g'(\rho)\partial_t \rho = \partial_t\big((\varepsilon + 2g(\rho))\mu\big) - \Delta\mu\,. \tag{4.2}$$

Now, we test (2.15), written at the time s, by $\partial_t(\rho, \rho_\Gamma)(s)$, integrate over $(0, t)$ and replace the right-hand side with the help of (4.2). We obtain the identity

$$\begin{aligned}
&\int_{Q_t} |\partial_t \rho|^2 + \int_{\Sigma_t} |\partial_t \rho_\Gamma|^2 + \frac{1}{2}\int_\Omega |\nabla\rho(t)|^2 + \frac{1}{2}\int_\Gamma |\nabla_\Gamma \rho_\Gamma(t)|^2 \\
&\quad + \int_\Omega F(\rho(t)) + \int_\Gamma F_\Gamma(\rho_\Gamma(t)) \\
&= \frac{1}{2}\int_\Omega |\nabla\rho_0|^2 + \frac{1}{2}\int_\Gamma |\nabla_\Gamma \rho_{0|\Gamma}|^2 + \int_\Omega F(\rho_0) + \int_\Gamma F_\Gamma(\rho_{0|\Gamma}) + \int_{Q_t} \mu g'(\rho)\partial_t\rho \\
&= c + \int_\Omega \big(\varepsilon + 2g(\rho(t))\big)\mu(t) - \int_\Omega \big(\varepsilon + 2g(\rho_0)\big)\mu_0 - \int_{Q_t} \Delta\mu\,.
\end{aligned}$$

The last integral vanishes since $\partial_\nu \mu = 0$. By recalling that F and F_Γ are bounded from below and that $|\rho| \leq 1$, and using (4.1), we deduce that

$$\|(\rho, \rho_\Gamma)\|_{L^\infty(0,+\infty;\mathcal{V})} \leq c\,, \quad \int_{Q_\infty} |\partial_t\rho|^2 \leq c \quad \text{and} \quad \int_{\Sigma_\infty} |\partial_t \rho_\Gamma|^2 \leq c\,. \tag{4.3}$$

First Conclusion The first inequalities of (4.1) and (4.3), along with the continuity of (μ, ρ, ρ_Γ) from $[0, +\infty)$ to $H \times \mathcal{V}$, ensure that the ω-limit (2.20) is nonempty. Namely, every divergent sequence of times contains a subsequence $t_n \nearrow +\infty$ such that $(\mu, \rho, \rho_\Gamma)(t_n)$ converges weakly in $H \times \mathcal{V}$.

After establishing the first part of Theorem 3, we prove the second one. Thus, we pick any element $(\mu_\omega, \rho_\omega, \rho_{\omega\Gamma})$ of the ω-limit (2.20) and show that it is a stationary solution of our problem, i.e., that μ_ω is a constant μ_s and that the pair $(\rho_\omega, \rho_{\omega\Gamma})$ coincides with a solution $(\rho_s, \rho_{s_\Gamma})$ to problem (2.21). To this end, we fix a sequence $t_n \nearrow +\infty$ such that

$$(\mu, \rho, \rho_\Gamma)(t_n) \to (\mu_\omega, \rho_\omega, \rho_{\omega\Gamma}) \quad \text{weakly in } H \times \mathcal{V} \tag{4.4}$$

and study the behavior of the solution on the time interval $[t_n, t_n + T]$ with a fixed $T > 0$. For convenience, we shift everything to $[0, T]$ by introducing $(\mu^n, \rho^n, \rho^n_\Gamma) : [0, T] \to H \times \mathcal{V}$ as follows

$$\begin{aligned} &\mu^n(t) := \mu(t_n + t), \quad \rho^n(t) := \rho(t_n + t) \\ &\text{and} \quad \rho^n_\Gamma(t) := \rho_\Gamma(t_n + t) \quad \text{for } t \in [0, T]. \end{aligned} \tag{4.5}$$

As T is fixed once and for all, we do not care on the dependence of the constants on T even in the notation, and write Q and Σ for Q_T and Σ_T, respectively. The inequalities (4.1) and (4.3) imply that

$$\|(\mu^n, \rho^n, \rho^n_\Gamma)\|_{L^\infty(0,T;H\times\mathcal{V})} \le c\,, \tag{4.6}$$

$$\lim_{n\to\infty} \Big(\int_Q |\nabla \mu^n|^2 + \int_Q |\partial_t \rho^n|^2 + \int_\Sigma |\partial_t \rho^n_\Gamma|^2 \Big) = 0\,. \tag{4.7}$$

The bound (4.6) yields a convergent subsequence in the weak star topology. If we still label it by the index n to simplify the notation, we have

$$(\mu^n, \rho^n, \rho^n_\Gamma) \to (\mu^\infty, \rho^\infty, \rho^\infty_\Gamma) \quad \text{weakly star in } L^\infty(0, T; H \times \mathcal{V}). \tag{4.8}$$

Now, we aim to improve the quality of the convergence. Thus, we derive further estimates.

First Auxiliary Estimate A partial use of (4.7) provides a bound, namely

$$\|\mu^n\|_{L^2(0,T;V)} + \|(\partial_t \rho^n, \partial_t \rho^n_\Gamma)\|_{L^2(0,T;\mathcal{H})} \le c\,. \tag{4.9}$$

Second Auxiliary Estimate We can repeat the argument that led to (3.7) and arrive at

$$\|(\rho^n, \rho^n_\Gamma)\|_{L^2(0,T;H^2(\Omega)\times H^2(\Gamma))} \le c\,. \tag{4.10}$$

Third Auxiliary Estimate We recall that μ^n and the space derivatives $D_i\rho^n$ and $D_i g(\rho^n) = g'(\rho^n) D_i\rho^n$ are bounded in

$$L^\infty(0,T;H) \cap L^2(0,T;L^6(\Omega)),$$

by (4.6), (4.9), (4.10), and the continuous embedding $V \subset L^6(\Omega)$. On the other hand, the continuous embedding

$$L^\infty(0,T;H) \cap L^2(0,T;L^6(\Omega)) \subset L^4(0,T;L^3(\Omega)) \cap L^6(0,T;L^{18/7}(\Omega))$$

holds true, by virtue of the interpolation inequalities. Therefore, we conclude that

$$\|\mu^n\|_{L^4(0,T;L^3(\Omega))} + \|\nabla\rho^n\|_{L^4(0,T;L^3(\Omega))} + \|\nabla g(\rho^n)\|_{L^6(0,T;L^{18/7}(\Omega))} \le c\,. \tag{4.11}$$

Fourth Auxiliary Estimate We want to improve the convergence of μ^n. However, we cannot multiply (2.14) by $\partial_t\mu$ since we do not have any information on $\nabla\mu(t_n)$. Therefore, we derive an estimate for $\partial_t\mu^n$ in a dual space. By recalling that $g(\rho) \ge g_*$ (see (3.1)), we divide both sides of (2.14) by $\varepsilon + 2g(\rho)$. Then, we take an arbitrary test function $v \in L^4(0,T;V)$, multiply the equality we obtain by v, integrate over $\Omega \times (t_n, t_n+T)$ and rearrange. We get

$$\int_Q \partial_t\mu^n\, v = -\int_Q \frac{\mu^n g'(\rho^n)\partial_t\rho^n v}{\varepsilon + 2g(\rho^n)} + \int_Q \Delta\mu^n\, \frac{v}{\varepsilon + 2g(\rho^n)}\,,$$

and we now treat the terms on the right-hand side separately. The first one is handled using Hölder's inequality, namely,

$$-\int_Q \frac{\mu^n g'(\rho^n)\partial_t\rho^n v}{\varepsilon + 2g(\rho^n)} \le c\|\mu^n\|_{L^4(0,T;L^3(\Omega))}\|\partial_t\rho^n\|_{L^2(0,T;L^2(\Omega))}\|v\|_{L^4(0,T;L^6(\Omega))}\,.$$

We integrate the other term by parts and use the Hölder, Sobolev and Young inequalities as follows:

$$\int_Q \Delta\mu^n\, \frac{v}{\varepsilon + 2g(\rho^n)} = -\int_Q \nabla\mu^n \cdot \frac{(\varepsilon + 2g(\rho^n))\nabla v - 2v g'(\rho^n)\nabla\rho^n}{(\varepsilon + 2g(\rho^n))^2}$$

$$\le c\|\nabla\mu^n\|_{L^2(0,T;H)}\|v\|_{L^2(0,T;V)} + c\int_0^T \|\nabla\mu^n(s)\|_2\|v(s)\|_6\|\nabla\rho^n(s)\|_3\, ds$$

$$\le c\|\nabla\mu^n\|_{L^2(0,T;H)}\big(\|v\|_{L^2(0,T;V)} + \|\nabla\rho^n\|_{L^4(0,T;L^3(\Omega))}\|v\|_{L^4(0,T;L^6(\Omega))}\big)\,.$$

Therefore, we have for every $v \in L^4(0,T;V)$

$$\int_Q \partial_t \mu^n \, v \le c \|\mu^n\|_{L^4(0,T;L^3(\Omega))} \|\partial_t \rho^n\|_{L^2(0,T;H)} \|v\|_{L^4(0,T;V)}$$
$$+ c \|\nabla \mu^n\|_{L^2(0,T;H)} \big(1 + \|\nabla \rho^n\|_{L^4(0,T;L^3(\Omega))}\big) \|v\|_{L^4(0,T;V)} \,.$$

Hence, on account of (4.1), (4.3) and (4.11), we conclude that

$$\|\partial_t \mu^n\|_{L^{4/3}(0,T;V^*)} \le c \,. \tag{4.12}$$

Conclusion By recalling the estimates (4.9)–(4.10) and (4.12), we see that the convergence (4.8) can be improved as follows:

$$\mu^n \to \mu^\infty \quad \text{in } W^{1,4/3}(0,T;V^*) \cap L^\infty(0,T;H) \cap L^2(0,T;V) \,,$$
$$(\rho^n, \rho^n_\Gamma) \to (\rho^\infty, \rho^\infty_\Gamma)$$
$$\text{in } H^1(0,T;\mathcal{H}) \cap L^\infty(0,T;\mathcal{V}) \cap L^2(0,T;H^2(\Omega) \times H^2(\Gamma)) \,,$$

all in the sense of the corresponding weak star topologies. Now, we prove that the limiting triple $(\mu^\infty, \rho^\infty, \rho^\infty_\Gamma)$ solves problem (2.14)–(2.15), the first equation being understood in a generalized sense. By [51, Sect. 8, Cor. 4] and the compact embeddings $H^2(\Omega) \subset V \subset H \subset V^*$ and $H^2(\Gamma) \subset V_\Gamma \subset H_\Gamma$, we also have (for a not relabeled subsequence)

$$\mu^n \to \mu^\infty \quad \text{strongly in } C^0([0,T];V^*) \cap L^2(0,T;H) \text{ and a.e. in } Q, \tag{4.13}$$
$$(\rho^n, \rho^n_\Gamma) \to (\rho^\infty, \rho^\infty_\Gamma)$$
$$\text{strongly in } C^0([0,T];\mathcal{H}) \cap L^2(0,T;\mathcal{V}) \text{ and a.e. on } Q \times \Sigma, \tag{4.14}$$
$$\nabla g(\rho^n) = g'(\rho^n)\nabla\rho^n \to g'(\rho^\infty)\nabla\rho^\infty = \nabla g(\rho^\infty) \quad \text{a.e. in } Q \,. \tag{4.15}$$

It follows that $(F'(\rho^n), F'_\Gamma(\rho^n_\Gamma))$ strongly converges to $(F'(\rho^\infty), F'_\Gamma(\rho^\infty_\Gamma))$ in $L^\infty(0,T;\mathcal{H})$, just by Lipschitz continuity. This allows us to conclude that $(\rho^\infty, \rho^\infty_\Gamma)$ solves the time-integrated version of (2.15), thus equation (2.15) itself. As for (2.14), we recall (4.11) and notice that $4 < 6$ and $2 < 18/7$. Then, with the help of (4.15) and the Egorov theorem, we deduce that

$$\nabla g(\rho^n) \to \nabla g(\rho^\infty) \quad \text{strongly in } (L^4(0,T;L^2(\Omega)))^3, \quad \text{whence}$$
$$g(\rho^n) \to g(\rho^\infty) \quad \text{strongly in } L^4(0,T;V).$$

Therefore, if we assume that $v \in L^\infty(0,T;W^{1,\infty}(\Omega))$, we have that

$$g(\rho^n)v \to g(\rho^\infty)v \quad \text{strongly in } L^4(0,T;V), \qquad \text{whence}$$
$$\int_Q \big(\varepsilon + 2g(\rho^n)\big)\partial_t\mu^n\, v \to {}_{L^{4/3}(0,T;V^*)}\langle \partial_t\mu^\infty, \big(\varepsilon + 2g(\rho^\infty)\big)v\rangle_{L^4(0,T;V)} .$$

On the other hand, from the convergence almost everywhere, we also have

$$g'(\rho^n) \to g'(\rho^\infty) \quad \text{strongly in } L^4(0,T;L^6(\Omega)),$$

since $g'(\rho^n)$ is bounded in $L^\infty(Q)$. Moreover, (4.11) implies that μ^n converges to μ^∞ weakly in $L^4(0,T;L^3(\Omega))$. On the other hand, (4.7) yields the strong convergence of $\partial_t\rho^n$ to 0 in $L^2(0,T;H)$ (by the way, 0 must coincide with $\partial_t\rho^\infty$). We deduce that

$$\mu^n g'(\rho^n)\partial_t\rho^n \to \mu^\infty g'(\rho^\infty)\partial_t\rho^\infty \quad \text{weakly in } L^1(Q) .$$

Therefore, we conclude that

$$\begin{aligned}&{}_{L^{4/3}(0,T;V^*)}\langle \partial_t\mu^\infty, \big(\varepsilon + 2g(\rho^\infty)\big)v\rangle_{L^4(0,T;V)}\\ &\quad + \int_Q \mu^\infty g'(\rho^\infty)\partial_t\rho^\infty\, v + \int_Q \nabla\mu^\infty\cdot\nabla v = 0 \end{aligned} \tag{4.16}$$

for every $v \in L^\infty(0,T;W^{1,\infty}(\Omega))$. On the other hand, we know that $\mu^\infty \in L^4(0,T;L^3(\Omega))$ by (4.11) and that $\partial_t\rho^\infty \in L^2(0,T;H)$. Since g' is bounded and the continuous embedding $V \subset L^6(\Omega)$ implies $L^{6/5}(\Omega) \subset V^*$, we also have that

$$\mu^\infty g'(\rho^\infty)\partial_t\rho^\infty \in L^{4/3}(0,T;L^{6/5}(\Omega)) \subset L^{4/3}(0,T;V^*) .$$

Hence, by a simple density argument, we see that the variational equation (4.16) also holds true for every $v \in L^4(0,T;V)$. At this point, we observe that (4.7) implies that

$$\nabla\mu^\infty = 0\,, \quad \partial_t\rho^\infty = 0 \quad \text{and} \quad \partial_t\rho_\Gamma^\infty = 0\,. \tag{4.17}$$

In particular, (4.16) reduces to

$${}_{L^{4/3}(0,T;V^*)}\langle \partial_t\mu^\infty, \big(\varepsilon + 2g(\rho^\infty)\big)v\rangle_{L^4(0,T;V)} = 0 \quad \text{for every } v \in L^4(0,T;V)$$

and we easily infer that $\partial_t\mu^\infty = 0$. Indeed, the inequality $g(\rho^n) \ge g_*$ for every n implies $g(\rho^\infty) \ge g_*$. Thus, every $\varphi \in C_c^\infty(Q)$ can be written as $\varphi = (\varepsilon + 2g(\rho^\infty))v$ for some $v \in L^4(0,T;V)$ since $\nabla\rho^\infty \in (L^4(0,T;H))^3$ by (4.11). Therefore, $\partial_t\mu^\infty$ actually vanishes and we conclude that μ^∞ takes a constant value μ_s.

From (4.17) we also deduce that $(\rho^\infty, \rho_\Gamma^\infty)$ is a time-independent pair $(\rho_s, \rho_{s_\Gamma})$, so that (2.15) reduces to (2.21). Finally, we show that $(\mu_s, \rho_s, \rho_{s_\Gamma}) = (\mu_\omega, \rho_\omega, \rho_{\omega_\Gamma})$. Indeed, (4.13) and (4.14) imply that

$$(\mu^n, (\rho^n, \rho_\Gamma^n)) \to (\mu^\infty, (\rho^\infty, \rho_\Gamma^\infty)) \quad \text{strongly in } C^0([0,T]; V^*) \times C^0([0,T]; \mathcal{H}),$$

and we infer that

$$(\mu, (\rho, \rho_\Gamma))(t_n) = (\mu^n, (\rho^n, \rho_\Gamma^n))(0) \to (\mu^\infty, (\rho^\infty, \rho_\Gamma^\infty))(0) = (\mu_s, (\rho_s, \rho_{s_\Gamma}))$$
$$\text{weakly in } V^* \times \mathcal{H}.$$

By comparing with (4.4), we conclude that $(\mu_s, \rho_s, \rho_{s_\Gamma}) = (\mu_\omega, \rho_\omega, \rho_{\omega_\Gamma})$, and the proof is complete. □

Acknowledgements PC and GG gratefully acknowledge some financial support from the MIUR-PRIN Grant 2015PA5MP7 "Calculus of Variations", the GNAMPA (Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni) of INdAM (Istituto Nazionale di Alta Matematica) and the IMATI—C.N.R. Pavia.

References

1. Cahn, J.W., Hilliard, J.E.: Free energy of a nonuniform system I. Interfacial free energy. J. Chem. Phys. **2**, 258–267 (1958)
2. Calatroni, L., Colli, P.: Global solution to the Allen–Cahn equation with singular potentials and dynamic boundary conditions. Nonlinear Anal. **79**, 12–27 (2013)
3. Cavaterra, C., Grasselli, M., Wu, H.: Non-isothermal viscous Cahn–Hilliard equation with inertial term and dynamic boundary conditions. Commun. Pure Appl. Anal. **13**, 1855–1890 (2014)
4. Cherfils, L., Gatti, S., Miranville, A.: A variational approach to a Cahn-Hilliard model in a domain with nonpermeable walls. J. Math. Sci. (N.Y.) **189**, 604–636 (2013)
5. Chill, R., Fašangová, E., Prüss, J.: Convergence to steady states of solutions of the Cahn-Hilliard equation with dynamic boundary conditions. Math. Nachr. **279**, 1448–1462 (2006)
6. Colli, P., Fukao, T.: The Allen–Cahn equation with dynamic boundary conditions and mass constraints. Math. Methods Appl. Sci. **38**, 3950–3967 (2015)
7. Colli, P., Fukao, T.: Cahn–Hilliard equation with dynamic boundary conditions and mass constraint on the boundary. J. Math. Anal. Appl. **429**, 1190–1213 (2015)
8. Colli, P., Fukao, T.: Equation and dynamic boundary condition of Cahn–Hilliard type with singular potentials. Nonlinear Anal. **127**, 413–433 (2015)
9. Colli, P., Gilardi, G., Krejčí, P., Podio-Guidugli, P., Sprekels, J.: Analysis of a time discretization scheme for a nonstandard viscous Cahn–Hilliard system. ESAIM Math. Model. Numer. Anal. **48**, 1061–1087 (2014)
10. Colli, P., Gilardi, G., Krejčí, P., Sprekels, J.: A vanishing diffusion limit in a nonstandard system of phase field equations. Evol. Equ. Control Theory **3**, 257–275 (2014)
11. Colli, P., Gilardi, G., Krejčí, P., Sprekels, J.: A continuous dependence result for a nonstandard system of phase field equations. Math. Methods Appl. Sci. **37**, 1318–1324 (2014)
12. Colli, P., Gilardi, G., Podio-Guidugli, P., Sprekels, J.: Well-posedness and long-time behaviour for a nonstandard viscous Cahn-Hilliard system. SIAM J. Appl. Math. **71**, 1849–1870 (2011)

13. Colli, P., Gilardi, G., Podio-Guidugli, P., Sprekels, J.: Global existence for a strongly coupled Cahn-Hilliard system with viscosity. Boll. Unione Mat. Ital. (9) **5**, 495–513 (2012)
14. Colli, P., Gilardi, G., Podio-Guidugli, P., Sprekels, J.: Distributed optimal control of a nonstandard system of phase field equations. Contin. Mech. Thermodyn. **24**, 437–459 (2012)
15. Colli, P., Gilardi, G., Podio-Guidugli, P., Sprekels, J.: Continuous dependence for a nonstandard Cahn-Hilliard system with nonlinear atom mobility. Rend. Sem. Mat. Univ. Pol. Torino **70**, 27–52 (2012)
16. Colli, P., Gilardi, G., Podio-Guidugli, P., Sprekels, J.: An asymptotic analysis for a nonstandard Cahn-Hilliard system with viscosity. Discrete Contin. Dyn. Syst. Ser. S **6**, 353–368 (2013)
17. Colli, P., Gilardi, G., Podio-Guidugli, P., Sprekels, J.: Global existence and uniqueness for a singular/degenerate Cahn-Hilliard system with viscosity. J. Differ. Equ. **254**, 4217–4244 (2013)
18. Colli, P., Gilardi, G., Sprekels, J.: Analysis and optimal boundary control of a nonstandard system of phase field equations. Milan J. Math. **80**, 119–149 (2012)
19. Colli, P., Gilardi, G., Sprekels, J.: Regularity of the solution to a nonstandard system of phase field equations. Rend. Cl. Sci. Mat. Nat. **147**, 3–19 (2013)
20. Colli, P., Gilardi, G., Sprekels, J.: On the Cahn–Hilliard equation with dynamic boundary conditions and a dominating boundary potential. J. Math. Anal. Appl. **419**, 972–994 (2014)
21. Colli, P., Gilardi, G., Sprekels, J.: A boundary control problem for the pure Cahn–Hilliard equation with dynamic boundary conditions. Adv. Nonlinear Anal. **4**, 311–325 (2015)
22. Colli, P., Gilardi, G., Sprekels, J.: A boundary control problem for the viscous Cahn–Hilliard equation with dynamic boundary conditions. Appl. Math. Optim. **73**, 195–225 (2016)
23. Colli, P., Gilardi, G., Sprekels, J.: On an application of Tikhonov's fixed point theorem to a nonlocal Cahn-Hilliard type system modeling phase separation. J. Differ. Equ. **260**, 7940–7964 (2016)
24. Colli, P., Gilardi, G., Sprekels, J.: Distributed optimal control of a nonstandard nonlocal phase field system. AIMS Math. **1**, 225–260 (2016)
25. Colli, P., Gilardi, G., Sprekels, J.: Distributed optimal control of a nonstandard nonlocal phase field system with double obstacle potential. Evol. Equ. Control Theory **6**, 35–58 (2017)
26. Colli, P., Gilardi, G., Sprekels, J.: Global existence for a nonstandard viscous Cahn–Hilliard system with dynamic boundary condition. SIAM J. Math. Anal. **49**, 1732–1760 (2017)
27. Colli, P., Gilardi, G., Sprekels, J.: Optimal boundary control of a nonstandard viscous Cahn–Hilliard system with dynamic boundary condition. Nonlinear Anal. **170**, 171–196 (2018)
28. Colli, P., Sprekels, J.: Optimal control of an Allen–Cahn equation with singular potentials and dynamic boundary condition. SIAM J. Control Optim. **53**, 213–234 (2015)
29. Conti, M., Gatti, S., Miranville, A.: Attractors for a Caginalp model with a logarithmic potential and coupled dynamic boundary conditions. Anal. Appl. (Singap.) **11**, 1350024, 31 pp. (2013)
30. Conti, M., Gatti, S., Miranville, A.: Multi-component Cahn–Hilliard systems with dynamic boundary conditions. Nonlinear Anal. Real World Appl. **25**, 137–166 (2015)
31. Elliott, C.M., Zheng, S.: On the Cahn–Hilliard equation. Arch. Ration. Mech. Anal. **96**, 339–357 (1986)
32. Fischer, H.P., Maass, Ph., Dieterich, W.: Novel surface modes in spinodal decomposition. Phys. Rev. Lett. **79**, 893–896 (1997)
33. Fischer, H.P., Maass, Ph., Dieterich, W.: Diverging time and length scales of spinodal decomposition modes in thin flows. Europhys. Lett. **42**, 49–54 (1998)
34. Fried, E., Gurtin, M.E.: Continuum theory of thermally induced phase transitions based on an order parameter. Phys. D **68**, 326–343 (1993)
35. Gal, C.G., Grasselli, M.: The non-isothermal Allen-Cahn equation with dynamic boundary conditions. Discrete Contin. Dyn. Syst. **22**, 1009–1040 (2008)
36. Gal, C.G., Warma, M.: Well posedness and the global attractor of some quasi-linear parabolic equations with nonlinear dynamic boundary conditions. Differ. Integr. Equ. **23**, 327–358 (2010)
37. Gilardi, G., Miranville, A., Schimperna, G.: On the Cahn-Hilliard equation with irregular potentials and dynamic boundary conditions. Commun. Pure. Appl. Anal. **8**, 881–912 (2009)

38. Gilardi, G., Miranville, A., Schimperna, G.: Long-time behavior of the Cahn–Hilliard equation with irregular potentials and dynamic boundary conditions. Chin. Ann. Math. Ser. B **31**, 679–712 (2010)
39. Goldstein, G.R., Miranville, A.: A Cahn-Hilliard-Gurtin model with dynamic boundary conditions. Discrete Contin. Dyn. Syst. Ser. S **6**, 387–400 (2013)
40. Goldstein, G.R., Miranville, A., Schimperna, G.: A Cahn–Hilliard model in a domain with non-permeable walls. Phys. D **240**, 754–766 (2011)
41. Gurtin, M.E.: Generalized Ginzburg–Landau and Cahn–Hilliard equations based on a micro-force balance. Phys. D **92**, 178–192 (1996)
42. Heida, M.: Existence of solutions for two types of generalized versions of the Cahn–Hilliard equation. Appl. Math. **60**, 51–90 (2015)
43. Israel, H.: Long time behavior of an Allen-Cahn type equation with a singular potential and dynamic boundary conditions. J. Appl. Anal. Comput. **2**, 29–56 (2012)
44. Liero, M.: Passing from bulk to bulk-surface evolution in the Allen-Cahn equation. NoDEA Nonlinear Differ. Equ. Appl. **20**, 919–942 (2013)
45. Miranville, A., Rocca, E., Schimperna, G., Segatti, A.: The Penrose-Fife phase-field model with coupled dynamic boundary conditions. Discrete Contin. Dyn. Syst. **34**, 4259–4290 (2014)
46. Miranville, A., Zelik, S.: Exponential attractors for the Cahn-Hilliard equation with dynamic boundary conditions. Math. Methods Appl. Sci. **28**, 709–735 (2005)
47. Novick-Cohen, A.: On the viscous Cahn-Hilliard equation. In: Material instabilities in continuum mechanics (Edinburgh, 1985–1986), pp. 329–342. Oxford Sci. Publ., Oxford Univ. Press, New York (1988)
48. Podio-Guidugli, P.: Models of phase segregation and diffusion of atomic species on a lattice. Ric. Mat. **55**, 105–118 (2006)
49. Prüss, J., Racke, R., Zheng, S.: Maximal regularity and asymptotic behavior of solutions for the Cahn-Hilliard equation with dynamic boundary conditions. Ann. Mat. Pura Appl. (4) **185**, 627–648 (2006)
50. Racke, R., Zheng, S.: The Cahn-Hilliard equation with dynamic boundary conditions. Adv. Differ. Equ. **8**, 83–110 (2003)
51. Simon, J.: Compact sets in the space $L^p(0, T; B)$. Ann. Mat. Pura Appl. (4) **146**, 65–96 (1987)

On a Cahn–Hilliard–Darcy System for Tumour Growth with Solution Dependent Source Terms

Harald Garcke and Kei Fong Lam

Abstract We study the existence of weak solutions to a mixture model for tumour growth that consists of a Cahn–Hilliard–Darcy system coupled with an elliptic reaction-diffusion equation. The Darcy law gives rise to an elliptic equation for the pressure that is coupled to the convective Cahn–Hilliard equation through convective and source terms. Both Dirichlet and Robin boundary conditions are considered for the pressure variable, which allow for the source terms to be dependent on the solution variables.

1 Introduction

At the fundamental level, cancer involves the unregulated growth of tissue inside the human body, which are caused by many biological and chemical mechanisms that take place at multiple spatial and temporal scales. In order to understand how these multiscale mechanisms are driving the progression of the cancer cells, whose dynamics may be too complex to be approached by experimental techniques, mathematical modelling can be used to provide a tractable description of the dynamics that isolate the key mechanisms and guide specific experiments.

We focus on the subclass of models for tumour growth known as diffuse interface models. These are continuum models that capture the macroscopic dynamics of the morphological changes of the tumour. For the simplest situation where there are only tumour cells and host cells in the presence of a nutrient, the model equations consists of a Cahn–Hilliard equation coupled to a reaction-diffusion equation for the nutrient. By treating the tumour and host cells as inertia-less fluids, a Darcy system can be appended to the Cahn–Hilliard equation, leading to a Cahn–Hilliard–Darcy system. For details regarding the diffuse interface models for tumour growth we refer the reader to [3, 6, 7, 16, 18, 21] and the references therein.

H. Garcke (✉) · K. F. Lam
Universität Regensburg, Regensburg, Germany
e-mail: harald.garcke@ur.de; kei-fong.lam@ur.de

E. Rocca et al. (eds.), *Trends in Applications of Mathematics to Mechanics*,
Springer INdAM Series 27, https://doi.org/10.1007/978-3-319-75940-1_12

Our interest lies in providing analytical results for these models, namely in establishing the existence of weak solutions to the model equations. Below, we introduce the Cahn–Hilliard–Darcy model to be studied: Let $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, be a bounded domain with boundary Γ, and denote, for $T > 0$, $Q := \Omega \times (0, T)$ and $\Sigma := \Gamma \times (0, T)$. We study the following elliptic-parabolic system:

$$\operatorname{div} \mathbf{v} = \Gamma_{\mathbf{v}}(\varphi, \sigma) \quad \text{in } Q, \tag{1a}$$

$$\partial_t \varphi + \operatorname{div}(\varphi \mathbf{v}) = \operatorname{div}(m(\varphi)\nabla\mu) + \Gamma_{\varphi}(\varphi, \sigma) \quad \text{in } Q, \tag{1b}$$

$$\mu = A\Psi'(\varphi) - B\Delta\varphi - \chi\sigma \quad \text{in } Q, \tag{1c}$$

$$0 = \Delta\sigma - h(\varphi)\sigma \quad \text{in } Q, \tag{1d}$$

$$\partial_{\mathbf{n}}\varphi = 0, \quad \sigma = 1 \quad \text{on } \Sigma, \tag{1e}$$

$$\varphi(0) = \varphi_0 \quad \text{in } \Omega, \tag{1f}$$

where $\partial_{\mathbf{n}} f := \nabla f \cdot \mathbf{n}$ is the normal derivative of f on the boundary Γ, with unit normal $\mathbf{n}$, and in this work, we focus on the following variants of Darcy's law and the boundary conditions

$$\mathbf{v} = -K(\nabla q + \varphi\nabla(\mu + \chi\sigma)) \text{ in } Q, \quad q = 0, \quad m(\varphi)\partial_{\mathbf{n}}\mu = \varphi\mathbf{v}\cdot\mathbf{n} \text{ on } \Sigma, \tag{2a}$$

$$\mathbf{v} = -K(\nabla p - (\mu + \chi\sigma)\nabla\varphi) \text{ in } Q, \quad \mu = 0, \quad K\partial_{\mathbf{n}}p = a(g - p) \text{ on } \Sigma, \tag{2b}$$

$$\mathbf{v} = -K(\nabla p - (\mu + \chi\sigma)\nabla\varphi) \text{ in } Q, \quad \partial_{\mathbf{n}}\mu = 0, \quad K\partial_{\mathbf{n}}p = a(g - p) \text{ on } \Sigma, \tag{2c}$$

for some positive constant a and prescribed function g. In (1), $\mathbf{v}$ denotes the volume-averaged velocity of the cell mixture, σ denotes the concentration of the nutrient, φ denotes the difference in volume fractions, with $\{\varphi = 1\}$ representing unmixed tumour tissue, and $\{\varphi = -1\}$ representing the host tissue, and μ denotes the chemical potential for φ.

The positive constant K is the permeability of the mixture, $m(\varphi)$ is a positive mobility for φ. The constant parameter $\chi \geq 0$ regulates the chemotaxis effect (see [16] for more details), $\Psi'(\cdot)$ is the derivative of a potential function $\Psi(\cdot)$ that has two equal minima at ±1, A and B denote two positive constants related to the thickness of the diffuse interface and the surface tension, $h(\varphi)$ is an interpolation function that satisfies $h(-1) = 0$ and $h(1) = 1$.

In (2), both p and q denote the pressure. The Darcy law in (2a) with pressure q can be obtained from the Darcy law in (2b) and (2c) with pressure p by setting $q = p - (\mu + \chi\sigma)\varphi$. The source terms $\Gamma_{\mathbf{v}}$ and Γ_{φ} model, for instance, the growth of the tumour and its effect on the velocity field. We refer to [16, §2.5] for a discussion regarding the choices for the source terms Γ_{φ}, $\Gamma_{\mathbf{v}}$.

We now compare the model (1) with other models studied in the literature.

1. In the absence of velocity, i.e., setting $\mathbf{v} = 0$ in (1b) and neglecting (1a), we obtain a elliptic-parabolic system that couples a Cahn–Hilliard equation with source term and an elliptic equation for the nutrient. A similar system has been studied by the authors in [12] with Dirichlet boundary conditions for φ, μ, σ. For systems where (1d) has an additional $\partial_t \sigma$ on the left-hand side, the well-posedness of solutions have been studied in [5, 11, 14, 15] for particular choices of the source term Γ_φ. We also mention the work of [8] for the analysis of a system of equations similar to (1) with $\chi = 0$.
2. In the case $\sigma = 0$, (1) with the Darcy law (2b) reduces to a Cahn–Hilliard–Darcy system, and well-posedness results have been established in [20] for $\Gamma_\mathbf{v} = \Gamma_\varphi = 0$ and $\partial_\mathbf{n} p = \partial_\mathbf{n} \mu = 0$ on Σ, and in [19] for prescribed source terms $\Gamma_\mathbf{v} = \Gamma_\varphi \neq 0$ and $\partial_\mathbf{n} p = \partial_\mathbf{n} \mu = 0$ on Σ. In [2] a related system, known as the Cahn–Hilliard–Brinkman system, is studied, which features an additional term $-\nu \Delta \mathbf{v}$ on the left-hand side of the Darcy law (2b), but with $\Gamma_\mathbf{v} = \Gamma_\varphi = 0$. Analogously, (1) without σ and the Darcy law (2a) with boundary conditions $\partial_\mathbf{n} p = \partial_\mathbf{n} \mu = \partial_\mathbf{n} \varphi = 0$ on Σ has been studied in [10]. For strong solutions to the Cahn–Hilliard–Darcy system on the d-dimensional torus, $d = 2, 3$, we refer the reader to [23, 24].
3. In [13], the authors established the global existence of weak solutions to (1) with the Darcy law (2b) that features the following convection-reaction-diffusion equation for σ:

$$\partial_t \sigma + \operatorname{div}(\sigma \mathbf{v}) = \Delta \sigma - \chi \Delta \varphi - S,$$

   with a prescribed source term $\Gamma_\mathbf{v}$ and source terms Γ_φ, S that depend on φ, σ and μ with at most linear growth, along with the boundary conditions $\partial_\mathbf{n} \mu = \partial_\mathbf{n} \varphi = \partial_\mathbf{n} p = 0$ and a Robin boundary condition for σ.

For the analyses performed on Cahn–Hilliard–Darcy systems in the literature, many have considered Neumann boundary conditions. However, a feature of the Neumann conditions for p and φ is that

$$\int_\Omega \Gamma_\mathbf{v} \, \mathrm{dx} = \int_\Omega \operatorname{div} \mathbf{v} \, \mathrm{dx} = \int_\Gamma \mathbf{v} \cdot \mathbf{n} \, \mathrm{d}\Gamma = \int_\Gamma -K \partial_\mathbf{n} p + K(\mu + \chi \sigma) \partial_\mathbf{n} \varphi \, \mathrm{d}\Gamma = 0,$$

that is, the source term $\Gamma_\mathbf{v}$ necessarily has zero mean. For source terms $\Gamma_\mathbf{v}$ that depend on φ and σ, this property may not be satisfied in general. To allow for source terms that need not have zero mean, one method is to prescribe alternate boundary conditions for the pressure, see for example [4, §2.2.9] and [16, §2.4.4].

In this work, we consider analysing the model with a Dirichlet boundary condition and also a Robin boundary condition for the pressure. Then, the source term $\Gamma_\mathbf{v}$ does not need to fulfil the zero mean condition. However, it turns out that in the derivation of a priori estimates for the model, we encounter the following:

- For the natural boundary condition $\partial_\mathbf{n} \mu = 0$ and the Robin boundary condition $K \partial_\mathbf{n} p = a(g - p)$ on Σ, we have to restrict our analysis to potentials Ψ that have quadratic growth (Theorem 2.3).

- To consider potentials with polynomial growth of order larger than two, we need to prescribe the boundary conditions (2a) and (2b) for the chemical potential μ (Theorems 2.1 and 2.2).

Let us briefly motivate the choices in (2a) and (2b). Due to the quasi-static nature of the nutrient equation (1d), we do not obtain a natural energy identity for the system (1) in contrast to the models studied in [13, 14, 16]. For simplicity, let $m(\varphi) = 1$, $K = 1$ and consider testing (1b) with $\mu + \chi\sigma$, (1c) with $\partial_t \varphi$, the Darcy law (2b) with $\mathbf{v}$. Integrating by parts and upon adding leads to

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{dt}} \int_{\Omega} A\Psi(\varphi) + \frac{B}{2} |\nabla\varphi|^2 \,\mathrm{dx} + \int_{\Omega} |\nabla\mu|^2 + |\mathbf{v}|^2 \,\mathrm{dx} \\ = \int_{\Omega} -\chi\nabla\mu\cdot\nabla\sigma + \Gamma_{\mathbf{v}}(p - \varphi(\mu + \chi\sigma)) + \Gamma_{\varphi}(\mu + \chi\sigma)\,\mathrm{dx} \\ + \int_{\Gamma} \partial_{\mathbf{n}}\mu(\mu + \chi\sigma) - p\mathbf{v}\cdot\mathbf{n}\,\mathrm{d}\Gamma.\end{aligned} \tag{3}$$

If we prescribe the boundary conditions $\partial_{\mathbf{n}}\mu = 0$ and $-\mathbf{v}\cdot\mathbf{n} = \partial_{\mathbf{n}} p = a(g - p)$, i.e., the boundary conditions in (2c), then the boundary term in (3) poses no difficulties. The main difficulty in obtaining a priori estimates from (3) is to control the source terms $\Gamma_{\mathbf{v}}\mu\varphi$ and $\Gamma_{\varphi}\mu$ with the left-hand side of (3). In the absence of any previous a priori estimates, to control terms involving μ by the term $\|\nabla\mu\|_{L^2(\Omega)}^2$ on the left-hand side via the Poincaré inequality, an estimate of the square of the mean of μ is needed. As observed in [14], this leads to a restriction to quadratic growth assumptions for the potential Ψ.

Furthermore, new difficulties arises in estimating the source term $\Gamma_{\mathbf{v}} p$ if we do not prescribe a Neumann boundary condition for p. The methodology used in [13, 19] to obtain an estimate for $\|p\|_{L^2(\Omega)}$ relies on the assumption that $\Gamma_{\mathbf{v}}$ is prescribed and has zero mean, and $\partial_{\mathbf{n}} p = 0$ on Σ. The arguments in [13, 19] seem not to be applicable for our present setting (see Remark 1 below), where $\Gamma_{\mathbf{v}}$ is dependent on φ and σ, and a Robin boundary condition is prescribed for p. This motivates the choice of a Dirichlet condition for μ to handle the source term $\Gamma_{\mathbf{v}}\varphi\mu$ and $\Gamma_{\varphi}\mu$, and as we will see later in Sect. 4 (specifically (32)), the Dirichlet boundary condition for μ is needed to obtain an L^2-estimate for p.

Alternatively, we may consider the discussion in [13, §8] regarding reformulations of the Darcy law. Choosing $q = p - \varphi(\mu + \chi\sigma)$ leads to the Darcy law variant in (2a). A similar testing procedure leads to

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{dt}} \int_{\Omega} A\Psi(\varphi) + \frac{B}{2} |\nabla\varphi|^2 \,\mathrm{dx} + \int_{\Omega} |\nabla\mu|^2 + |\mathbf{v}|^2 \,\mathrm{dx} \\ = \int_{\Omega} -\chi\nabla\mu\cdot\nabla\sigma + \Gamma_{\mathbf{v}} q + \Gamma_{\varphi}(\mu + \chi\sigma)\,\mathrm{dx} \\ + \int_{\Gamma} (\partial_{\mathbf{n}}\mu - \varphi\mathbf{v}\cdot\mathbf{n})(\mu + \chi\sigma) - q\mathbf{v}\cdot\mathbf{n}\,\mathrm{d}\Gamma.\end{aligned} \tag{4}$$

Here we observed that the source term involving $\Gamma_{\mathbf{v}}$ simplifies to just $\Gamma_{\mathbf{v}} q$, and in exchange, we see the appearance of $(q + \varphi\mu + \chi\varphi\sigma)\mathbf{v} \cdot \mathbf{n}$ appearing in the boundary term. Comparing to the previous set-up with (2b), we have shifted the problematic terms to the boundary integral. Choosing $\mathbf{v} \cdot \mathbf{n} = 0$ on Σ is not desirable, as equation (1a) would the imply that $\Gamma_{\mathbf{v}}(\varphi, \sigma)$ must have zero mean. We may instead consider the boundary conditions

$$\partial_{\mathbf{n}}\mu = 0, \quad \mathbf{v} \cdot \mathbf{n} = -\partial_{\mathbf{n}} q - \chi\varphi\partial_{\mathbf{n}}\sigma = a(q + \varphi(\mu + \chi\sigma)) \text{ on } \Sigma,$$

then the boundary term in (4) poses no additional difficulties in obtaining a priori estimate. In exchange, obtaining an estimate for $\|q\|_{L^2(\Omega)}$ to deal with the source term $\Gamma_{\mathbf{v}} q$ becomes more involved, as the variational formulation for the pressure system now reads as

$$\int_\Omega \nabla q \cdot \nabla\zeta \, \mathrm{dx} + \int_\Gamma aq\zeta \, \mathrm{d}\Gamma = \int_\Omega \Gamma_{\mathbf{v}}\zeta - \varphi\nabla(\mu + \chi\sigma) \cdot \nabla\zeta \, \mathrm{dx} - \int_\Gamma a\varphi(\mu + \chi\sigma)\zeta \, \mathrm{d}\Gamma$$

for a test function ζ. Estimates for q will now involve an estimate for $\|\varphi\mu\|_{L^2(\Gamma)}$, and this is more difficult to control than $\|\varphi\mu\|_{L^2(\Omega)}$. This motivates the choice of a Dirichlet condition for q and the boundary condition $\partial_{\mathbf{n}}\mu = \varphi\mathbf{v} \cdot \mathbf{n}$ to eliminate the boundary term in (4).

This paper is organized as follows. In Sect. 2 we state the main assumptions and the main results. In Sect. 3 we outline the existence proof by first studying a parabolic-regularized variant of (1)–(2a) where we add $\theta\partial_t\sigma$ to the left-hand side of (1d) for $\theta \in (0, 1]$ and replace σ with $\mathcal{T}(\sigma)$ in $\Gamma_{\mathbf{v}}$ and Γ_φ, where $\mathcal{T}$ is a cut-off operator. The a priori estimates necessary for a Galerkin approximation to the parabolic-regularized problem is then derived, with which the weak existence for the original problem can be attained by passing to the limit $\theta \to 0$. The analogous a priori estimates for the Robin boundary conditions (2b) and (2c) are specified in Sects. 4 and 5, respectively.

Notation For convenience, we will often use the notation $L^p := L^p(\Omega)$ and $W^{k,p} := W^{k,p}(\Omega)$ for any $p \in [1, \infty]$, $k > 0$ to denote the standard Lebesgue spaces and Sobolev spaces equipped with the norms $\|\cdot\|_{L^p}$ and $\|\cdot\|_{W^{k,p}}$. In the case $p = 2$ we use $H^k := W^{k,2}$ and the norm $\|\cdot\|_{H^k}$. Due to the Dirichlet boundary condition for σ and μ, we denote the space H^1_0 as the completion of $C^\infty_c(\Omega)$ with respect to the H^1 norm. We will use the isometric isomorphism $L^p(Q) \cong L^p(0, T; L^p)$ and $L^p(\Sigma) \cong L^p(0, T; L^p(\Gamma))$ for any $p \in [1, \infty)$. Moreover, the dual space of a Banach space X will be denoted by X^*, and the duality pairing between X and X^* is denoted by $\langle\cdot, \cdot\rangle_X$. We denote the dual space to H^1_0 as H^{-1}. For $d = 2$ or 3, let $\mathrm{d}\Gamma$ denote integration with respect to the $(d - 1)$ dimensional Hausdorff measure on Γ, and we denote $\mathbb{R}^d$-valued functions in boldface. For convenience, we will often use the notation

$$\int_Q f := \int_0^T \int_\Omega f \,\mathrm{dx}\,\mathrm{dt}, \quad \int_{\Omega_t} f := \int_0^t \int_\Omega f \,\mathrm{dx}\,\mathrm{ds}, \quad \int_{\Gamma_t} f := \int_0^t \int_\Gamma f \,\mathrm{d}\Gamma\,\mathrm{ds}$$

for any $f \in L^1(Q)$ and for any $t \in (0, T]$.

Useful Preliminaries For convenience, we recall the *Poincaré inequality*: There exist a positive constant C_p depending only on Ω such that

$$\left\| f - \overline{f} \right\|_{L^r} \leq C_p \|\nabla f\|_{L^r} \text{ for all } f \in W^{1,r}, 1 \leq r \leq \infty, \tag{5}$$

where $\overline{f} := \frac{1}{|\Omega|} \int_\Omega f \,\mathrm{dx}$ denotes the mean of f. Furthermore, we have

$$\|f\|_{L^2} \leq C_p \left(\|\nabla f\|_{L^2} + \|f\|_{L^2(\Gamma)} \right) \text{ for } f \in H^1, \tag{6}$$

$$\|f\|_{L^2} \leq C_p \|\nabla f\|_{L^2} \qquad \text{for } f \in H^1_0. \tag{7}$$

The *Gagliardo–Nirenberg interpolation inequality* in dimension d (see [9, Theorem 2.1] and [1, Theorem 5.8]): Let Ω be a bounded domain with Lipschitz boundary, and $f \in W^{m,r} \cap L^q$, $1 \leq q, r \leq \infty$. For any integer j, $0 \leq j < m$, suppose there is $\alpha \in \mathbb{R}$ such that

$$\frac{1}{p} = \frac{j}{d} + \left(\frac{1}{r} - \frac{m}{d} \right) \alpha + \frac{1-\alpha}{q}, \quad \frac{j}{m} \leq \alpha \leq 1.$$

If $r \in (1, \infty)$ and $m - j - \frac{d}{r}$ is a non-negative integer, we in addition assume $\alpha \neq 1$. Under these assumptions, there exists a positive constant C depending only on Ω, m, j, q, r, and α such that

$$\|D^j f\|_{L^p} \leq C \|f\|_{W^{m,r}}^{\alpha} \|f\|_{L^q}^{1-\alpha}. \tag{8}$$

For $f \in L^2$, $g \in L^2(\Gamma)$, and $\beta > 0$, let $u \in H^1$, $w \in H^1_0$ be the unique solutions to the elliptic problems

$$\begin{aligned} -\Delta w &= f \text{ in } \Omega, \quad w = 0 && \text{on } \Gamma, \\ -\Delta u &= f \text{ in } \Omega, \quad \partial_{\mathbf{n}} u + \beta u = g && \text{on } \Gamma. \end{aligned}$$

We use the notation $u = (-\Delta_R)^{-1}(f, \beta, g)$ and $w = (-\Delta_D)^{-1}(f)$. Furthermore, if in addition $g \in H^{\frac{1}{2}}(\Gamma)$ and Γ is a C^2-boundary, then by elliptic regularity theory [17, Thm. 2.4.2.6] and [17, Thm. 2.4.2.5], it holds that $w \in H^2 \cap H^1_0$ and $u \in H^2$ with

$$\|w\|_{H^2} \leq C \|f\|_{L^2}, \quad \|u\|_{H^2} \leq C \left(\|f\|_{L^2} + \|g\|_{H^{\frac{1}{2}}(\Gamma)} \right).$$

2 Assumptions and Main Results

Assumption 2.1

(A1) $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, is a bounded domain with C^3-boundary Γ. The positive constants a, T, A, B, χ, K are fixed. The function $g \in L^2(\Sigma)$ and the initial condition $\varphi_0 \in H^1$ are prescribed.

(A2) The mobility $m \in C^0(\mathbb{R})$ satisfies $0 < m_0 \leq m(s) \leq m_1$ for all $s \in \mathbb{R}$. The function $h \in C^0(\mathbb{R})$ is non-negative and is bounded above by 1.

(A3) The potential $\Psi \in C^2(\mathbb{R})$ is non-negative and, for $r \in [0, 2]$ and for all $s \in \mathbb{R}$, there exist positive constants C_1, C_2, C_3 and C_4 such that

$$\Psi(s) \geq C_1 |s|^2 - C_2, \quad \left|\Psi''(s)\right| \leq C_3 \left(1 + |s|^r\right), \quad \left|\Psi'(s)\right| \leq C_4 \left(1 + \Psi(s)\right).$$

(A4) The source terms $\Gamma_{\mathbf{v}}$ and $\Gamma_{\boldsymbol{\varphi}}$ are of the form

$$\Gamma_{\mathbf{v}}(\varphi, \sigma) = b_{\mathbf{v}}(\varphi)\sigma + f_{\mathbf{v}}(\varphi), \quad \Gamma_{\varphi}(\varphi, \sigma) = b_{\varphi}(\varphi)\sigma + f_{\varphi}(\varphi),$$

where $b_{\mathbf{v}}, b_{\varphi}, f_{\mathbf{v}}, f_{\varphi}$ are bounded and continuous functions.

We first give the results to the problem (1), (2a).

Definition 2.1 We call a quintuple $(\varphi, \mu, \sigma, \mathbf{v}, q)$ a weak solution to (1), (2a) if

$$\varphi \in L^\infty(0, T; H^1) \cap L^2(0, T; H^3) \cap W^{1,\frac{8}{5}}(0, T; (H^1)^*), \quad \mathbf{v} \in L^2(Q),$$

$$\sigma \in (1 + L^2(0, T; H_0^1)), \quad \mu \in L^2(0, T; H^1), \quad q \in L^{\frac{8}{5}}(0, T; H_0^1),$$

and satisfies $\varphi(0) = \varphi_0$, $0 \leq \sigma \leq 1$ a.e. in Q, and

$$0 = \langle \partial_t \varphi, \zeta \rangle_{H^1} + \int_\Omega m(\varphi)\nabla\mu \cdot \nabla\zeta - \varphi\mathbf{v} \cdot \nabla\zeta - \Gamma_\varphi(\varphi, \sigma)\zeta \,\mathrm{dx}, \tag{9a}$$

$$0 = \int_\Omega (\mu + \chi\sigma)\zeta - A\Psi'(\varphi)\zeta - B\nabla\varphi \cdot \nabla\zeta \,\mathrm{dx}, \tag{9b}$$

$$0 = \int_\Omega \nabla\sigma \cdot \nabla\xi + h(\varphi)\sigma\xi \,\mathrm{dx}, \tag{9c}$$

$$0 = \int_\Omega K\nabla q \cdot \nabla\xi - \Gamma_{\mathbf{v}}(\varphi, \sigma)\xi + K\varphi\nabla(\mu + \chi\sigma) \cdot \nabla\xi \,\mathrm{dx}, \tag{9d}$$

$$0 = \int_\Omega \mathbf{v} \cdot \mathbf{y} + K\nabla q \cdot \mathbf{y} + K\varphi\nabla(\mu + \chi\sigma) \cdot \mathbf{y} \,\mathrm{dx}, \tag{9e}$$

for a.e. $t \in (0, T)$ and all $\zeta \in H^1$, $\xi \in H_0^1$, $\mathbf{y} \in L^2$.

Theorem 2.1 *Under Assumption 2.1, there exists a weak solution to* (1), (2a) *in the sense of Definition 2.1.*

For the problem (1), (2b) we have the following.

Definition 2.2 We call a quintuple $(\varphi, \mu, \sigma, \mathbf{v}, p)$ a weak solution to (1), (2b) if

$$\varphi \in L^\infty(0,T;H^1) \cap L^2(0,T;H^3) \cap W^{1,\frac{8}{5}}(0,T;H^{-1}), \quad \mathbf{v} \in L^2(Q),$$
$$\sigma \in (1 + L^2(0,T;H^1_0)), \quad \mu \in L^2(0,T;H^1_0), \quad p \in L^{\frac{8}{5}}(0,T;H^1),\ p|_\Sigma \in L^2(\Sigma),$$

and satisfies $\varphi(0) = \varphi_0$, $0 \leq \sigma \leq 1$ a.e. in Q, (9b), (9c), and

$$0 = \langle \partial_t \varphi, \xi \rangle_{H^1_0} + \int_\Omega m(\varphi)\nabla\mu \cdot \nabla\xi - \varphi\mathbf{v} \cdot \nabla\xi - \Gamma_{\mathbf{v}}(\varphi,\sigma)\xi \,\mathrm{d}x, \tag{10a}$$

$$0 = \int_\Omega K\nabla p \cdot \nabla\zeta - \Gamma_{\mathbf{v}}(\varphi,\sigma)\zeta - K(\mu + \chi\sigma)\nabla\varphi \cdot \nabla\zeta \,\mathrm{d}x + \int_\Gamma a(p-g)\zeta \,\mathrm{d}\Gamma, \tag{10b}$$

$$0 = \int_\Omega \mathbf{v} \cdot \mathbf{y} + K\nabla p \cdot \mathbf{y} - K(\mu + \chi\sigma)\nabla\varphi \cdot \mathbf{y} \,\mathrm{d}x, \tag{10c}$$

for a.e. $t \in (0,T)$ and all $\zeta \in H^1$, $\xi \in H^1_0$, $\mathbf{y} \in L^2$.

Theorem 2.2 *Under Assumption 2.1, there exists a weak solution to* (1), (2b) *in the sense of Definition 2.2.*

Analogously for the problem (1), (2c) we have the following.

Definition 2.3 We call a quintuple $(\varphi, \mu, \sigma, \mathbf{v}, p)$ a weak solution to (1), (2c) if

$$\varphi \in L^\infty(0,T;H^1) \cap L^2(0,T;H^3) \cap W^{1,\frac{8}{5}}(0,T;(H^1)^*), \quad \mathbf{v} \in L^2(Q),$$
$$\sigma \in (1 + L^2(0,T;H^1_0)), \quad \mu \in L^2(0,T;H^1), \quad p \in L^{\frac{8}{5}}(0,T;H^1),\ p|_\Sigma \in L^2(\Sigma),$$

and satisfies $\varphi(0) = \varphi_0$, $0 \leq \sigma \leq 1$ a.e. in Q, (9b), (9c), (10b) and (10c) and

$$0 = \langle \partial_t \varphi, \zeta \rangle_{H^1} + \int_\Omega m(\varphi)\nabla\mu \cdot \nabla\zeta + \nabla\varphi \cdot \mathbf{v}\zeta + \Gamma_{\mathbf{v}}(\varphi,\sigma)\varphi\zeta - \Gamma_\varphi(\varphi,\sigma)\zeta \,\mathrm{d}x, \tag{11}$$

for a.e. $t \in (0,T)$ and all $\zeta \in H^1$, $\xi \in H^1_0$, $\mathbf{y} \in L^2$.

Theorem 2.3 *Under Assumption 2.1, with* (A3) *replaced by*

$$\Psi(s) \geq C_1 |s|^2 - C_2, \quad |\Psi''(s)| \leq C_3 \quad \forall s \in \mathbb{R}, \tag{12}$$

for some positive constants C_1, C_2, C_3 , there exists a weak solution to (1)*,* (2c) *in the sense of Definition 2.3.*

We use the fact that $H^1 \subset\subset L^2 \subset (H^1)^*$, $H^1 \subset\subset L^2 \subset H^{-1}$, and [22, §8, Cor. 4] to deduce that $\varphi \in C^0([0,T];L^2)$ in all cases, and thus $\varphi(0)$ makes sense as a function in L^2. This implies that the initial condition φ_0 is attained in all cases.

3 Dirichlet Boundary Conditions for the Pressure

We show the existence of weak solutions to (1), (2a) by means of a Galerkin approximation, and first consider a regularisation of (1), (2a), where (1d) is replaced with

$$\theta \partial_t \sigma - \Delta\sigma + h(\varphi)\sigma = 0 \text{ in } Q, \quad \sigma = 1 \text{ on } \Sigma, \quad \sigma(0) = \sigma_0 \text{ in } \Omega \tag{13}$$

for some $\theta \in (0,1]$, and $\sigma_0 \in L^2(\Omega)$. Furthermore, we introduce a cut-off operator $\mathscr{T}(s) := \max(0, \min(1,s))$ and replace the source terms with

$$\Gamma_{\mathbf{v}}(\varphi,\sigma) = b_{\mathbf{v}}(\varphi)\mathscr{T}(\sigma) + f_{\mathbf{v}}(\varphi), \quad \Gamma_{\varphi}(\varphi,\sigma) = b_{\varphi}(\varphi)\mathscr{T}(\sigma) + f_{\varphi}(\varphi).$$

The procedure is to first use a Galerkin approximation to deduce the existence of a weak solution quintuple $(\varphi^\theta, \mu^\theta, \sigma^\theta, \mathbf{v}^\theta, q^\theta)$ to the regularized problem, and subsequently employ a weak comparison principle at the continuous level to show $0 \le \sigma^\theta \le 1$ a.e. in Q, so that the cut-off operator $\mathscr{T}$ can then be neglected. Then, we pass to the limit $\theta \to 0$ to obtain the existence of a weak solution to (1), (2a).

Below we will derive the necessary a priori estimates to prove existence of weak solutions to the regularized problem

$$\begin{aligned}
\operatorname{div} \mathbf{v} &= b_{\mathbf{v}}(\varphi)\mathscr{T}(\sigma) + f_{\mathbf{v}}(\varphi) && \text{in } Q, && \text{(14a)}\\
\mathbf{v} &= -K(\nabla q + \varphi\nabla(\mu + \chi\sigma)) && \text{in } Q, && \text{(14b)}\\
\partial_t\varphi + \operatorname{div}(\varphi\mathbf{v}) &= \operatorname{div}(m(\varphi)\nabla\mu) + b_{\varphi}(\varphi)\mathscr{T}(\sigma) + f_{\varphi}(\varphi) && \text{in } Q, && \text{(14c)}\\
\mu &= A\Psi'(\varphi) - B\Delta\varphi - \chi\sigma && \text{in } Q, && \text{(14d)}\\
\theta\partial_t\sigma &= \Delta\sigma - h(\varphi)\sigma && \text{in } Q, && \text{(14e)}\\
\partial_{\mathbf{n}}\varphi = 0, \quad m(\varphi)\partial_{\mathbf{n}}\mu &= \varphi\mathbf{v}\cdot\mathbf{n}, \quad q = 0, \quad \sigma = 1 \text{ on } \Sigma, && && \text{(14f)}\\
\varphi(0) = \varphi_0, \quad \sigma(0) &= \sigma_0 && \text{in } \Omega, && \text{(14g)}
\end{aligned}$$

with an initial condition $0 \le \sigma_0 \le 1$ a.e. in Ω.

Lemma 1 *Under Assumption 2.1 and $0 \le \sigma_0 \le 1$ a.e. in Ω, for any $\theta \in (0,1]$, there exists a weak solution quintuple $(\varphi^\theta, \mu^\theta, \sigma^\theta, \mathbf{v}^\theta, q^\theta)$ in the sense of Definition 2.1 with additionally $\sigma^\theta \in H^1(0,T;H^{-1})$, $\sigma^\theta(0) = \sigma_0$ a.e. in Ω,* *and* (9c) *is replaced by*

$$0 = \langle \theta \partial_t \sigma^\theta, \xi \rangle_{H_0^1} + \int_\Omega \nabla \sigma^\theta \cdot \nabla \xi + h(\varphi^\theta)\sigma^\theta \xi \,\mathrm{dx} \quad \forall \xi \in H_0^1. \tag{15}$$

Furthermore, there exists a positive constant C not depending on $\theta, \varphi^\theta, \mu^\theta, \sigma^\theta$, $\mathbf{v}^\theta, q^\theta$ such that

$$\begin{aligned}&\|\Psi(\varphi^\theta)\|_{L^\infty(0,T;L^1)} + \|\Psi'(\varphi^\theta)\|_{L^2(0,T;H^1)} + \|\varphi^\theta\|_{L^\infty(0,T;H^1)\cap L^2(0,T;H^3)} \\ &\quad + \|\mu^\theta\|_{L^2(0,T;H^1)} + \|\mathbf{v}^\theta\|_{L^2(Q)} + \|q^\theta\|_{L^{\frac{8}{5}}(0,T;H_0^1)} + \|\partial_t \varphi^\theta\|_{L^{\frac{8}{5}}(0,T;(H^1)^*)} \\ &\quad + \|\sigma^\theta\|_{L^2(0,T;H^1)} + \|\theta \partial_t \sigma^\theta\|_{L^2(0,T;H^{-1})} \leq C.\end{aligned} \tag{16}$$

Proof The details regarding the existence of Galerkin solutions via the theory of ODEs can be found in [13, 19], and so we will omit the details and focus only on the a priori estimates. In the following, C denotes a positive constant not depending on $(\varphi, \mu, \sigma, \mathbf{v}, q)$ and θ, and may vary from line to line.

At the Galerkin level, we may replace duality pairings in (9a) and (15) with L^2-inner products. For convenience let us reuse the variables $\varphi, \mu, \sigma, \mathbf{v}, q$ as the Galerkin solutions. Let $Z > 0$ be a constant yet to be specified, then substituting $\xi = Z(\sigma - 1)$ in (15), $\zeta = \partial_t \varphi$ in (9b), $\zeta = \mu + \chi\sigma$ in (9a), $\mathbf{y} = K^{-1}\mathbf{v}$ in (9e) and summing leads to

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{dt}} \int_\Omega A\Psi(\varphi) + \frac{B}{2}|\nabla\varphi|^2 + \frac{Z}{2}\theta|\sigma - 1|^2 \,\mathrm{dx} \\ &\qquad + \int_\Omega m(\varphi)|\nabla\mu|^2 + \frac{1}{K}|\mathbf{v}|^2 + Z|\nabla\sigma|^2 + Zh(\varphi)|\sigma|^2 \,\mathrm{dx} \\ &\quad = \int_\Omega -m(\varphi)\chi\nabla\mu \cdot \nabla\sigma + \Gamma_\varphi(\mu + \chi\sigma) + \Gamma_{\mathbf{v}} q + Zh(\varphi)\sigma \,\mathrm{dx}.\end{aligned} \tag{17}$$

Similarly to [12] we estimate terms on the right-hand side involving σ by $C_1\|\nabla\sigma\|_{L^2}^2 + C_2 Z$ through the use of the Poincaré inequality, where C_1, C_2 are positive constants such that C_1 is independent of Z. Thanks to the cutoff operator and the boundedness of $f_{\mathbf{v}}$ and f_φ, we see that

$$\begin{aligned}&\left|\int_\Omega \Gamma_\varphi(\mu + \chi\sigma) + \Gamma_{\mathbf{v}} q \,\mathrm{dx}\right| \\ &\quad \leq C\left(1 + \|\mu - \overline{\mu}\|_{L^1} + |\overline{\mu}|_{L^1} + \|q\|_{L^2} + \|\sigma - 1\|_{L^2}\right) \\ &\quad \leq C(1 + |\overline{\mu}| + \|q\|_{L^2}) + \frac{m_0}{4}\|\nabla\mu\|_{L^2}^2 + \|\nabla\sigma\|_{L^2}^2,\end{aligned} \tag{18}$$

where we have used the Poincaré inequality (5) with $r = 1$ and Young's inequality. From substituting $\zeta = 1$ in (9b) and using (A3), we find that

$$|\overline{\mu}| \leq C(1 + \|\sigma - 1\|_{L^2} + \|\Psi'(\varphi)\|_{L^1}) \leq C\left(1 + \|\Psi(\varphi)\|_{L^1} + \|\nabla\sigma\|_{L^2}\right). \tag{19}$$

To obtain an estimate of $\|q\|_{L^2}$, we look at the pressure system, whose weak formulation is given by (9d). Let $f := (-\Delta_D)^{-1}(q/K)$, so that

$$\int_\Omega K\nabla f \cdot \nabla\phi \,\mathrm{dx} = \int_\Omega q\phi \,\mathrm{dx} \text{ for all } \phi \in H_0^1.$$

Substituting $\xi = f$ in (9d) and $\phi = q$ in the above leads to

$$\begin{aligned}\|q\|_{L^2}^2 = \int_\Omega K\nabla q \cdot \nabla f \,\mathrm{dx} &= \int_\Omega \Gamma_{\mathbf{v}} f - K\varphi\nabla(\mu + \chi\sigma) \cdot \nabla f \,\mathrm{dx} \\ &\leq \|\Gamma_{\mathbf{v}}\|_{L^2}\|f\|_{L^2} + K\|\varphi\nabla(\mu + \chi\sigma)\|_{L^{\frac{6}{5}}}\|\nabla f\|_{L^6} \\ &\leq C\left(1 + \|\varphi\|_{L^3}\|\nabla(\mu + \chi\sigma)\|_{L^2}\right)\|f\|_{H^2}.\end{aligned}$$

Using the elliptic regularity estimate $\|f\|_{H^2} \leq C\|q\|_{L^2}$, we find that

$$\begin{aligned}\|q\|_{L^2} &\leq C\left(1 + \|\varphi\|_{H^1}\|\nabla(\mu + \chi\sigma)\|_{L^2}\right) \\ &\leq \frac{m_0}{4}\|\nabla\mu\|_{L^2}^2 + \|\nabla\sigma\|_{L^2}^2 + C\left(1 + \|\Psi(\varphi)\|_{L^1} + \|\nabla\varphi\|_{L^2}^2\right),\end{aligned} \tag{20}$$

where we have used the Sobolev embedding $H^1 \subset L^3$ and (A3). Then, substituting the estimates (19), (20) into (18), we find that the right-hand side of (17) can be estimated as

$$\begin{aligned}|\mathrm{RHS}| &\leq \frac{m_0}{4}\|\nabla\mu\|_{L^2}^2 + \frac{\chi^2 m_1^2}{m_0}\|\nabla\sigma\|_{L^2}^2 + Z\|\sigma - 1\|_{L^1} + Z \\ &\quad + C\left(1 + \|\sigma - 1\|_{L^2} + \|\mu - \overline{\mu}\|_{L^1} + |\overline{\mu}| + \|q\|_{L^2}\right) \\ &\leq \frac{3m_0}{4}\|\nabla\mu\|_{L^2}^2 + \left(\frac{\chi^2 m_1^2}{m_0} + 4\right)\|\nabla\sigma\|_{L^2}^2 \\ &\quad + C\left(1 + Z^2 + \|\Psi(\varphi)\|_{L^1} + \|\nabla\varphi\|_{L^2}^2\right).\end{aligned}$$

Neglecting the non-negative term $Zh(\varphi)\,|\sigma|^2$ on the left-hand side of (17) and choosing $Z > \frac{\chi^2 m_1^2}{m_0} + 4$ yields the differential inequality

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{dt}}\left(\|\Psi(\varphi)\|_{L^1} + \|\nabla\varphi\|_{L^2}^2 + \theta\|\sigma - 1\|_{L^2}^2\right) - C\left(\|\Psi(\varphi)\|_{L^1} + \|\nabla\varphi\|_{L^2}^2\right) \\ &\quad + \|\nabla\mu\|_{L^2}^2 + \|\mathbf{v}\|_{L^2}^2 + \|\nabla\sigma\|_{L^2}^2 \leq C.\end{aligned}$$

By (A1), (A3) and the Sobolev embedding $H^1 \subset L^6$, it holds that $\Psi(\varphi_0) \in L^1$. Hence, by an application of Gronwall's inequality we obtain

$$\begin{aligned}\sup_{t\in(0,T]} &\left(\|\Psi(\varphi(t))\|_{L^1} + \|\nabla\varphi(t)\|_{L^2}^2 + \theta\|\sigma(t)-1\|_{L^2}^2 \right) \\ &+ \|\nabla\mu\|_{L^2(Q)}^2 + \|\mathbf{v}\|_{L^2(Q)}^2 + \|\nabla\sigma\|_{L^2(Q)}^2 \leq C,\end{aligned}$$

where we have also used that $\theta\|\sigma_0 - 1\|_{L^2}^2 \leq \|\sigma_0 - 1\|_{L^2}^2$ as $\theta \in (0,1]$. Then, using (19) and (A3) and the Poincaré inequality for φ and μ yields

$$\begin{aligned}\sup_{t\in(0,T]} &\left(\|\Psi(\varphi(t))\|_{L^1} + \|\varphi(t)\|_{H^1}^2 + \theta\|\sigma\|_{L^2}^2 \right) \\ &+ \|\mu\|_{L^2(0,T;H^1)}^2 + \|\mathbf{v}\|_{L^2(Q)}^2 + \|\sigma\|_{L^2(0,T;H^1)}^2 \leq C.\end{aligned} \tag{21}$$

Next, looking at (9b) as an elliptic equation for φ, and using that the potential Ψ has polynomial growth of order less than 6, we employ the bootstrapping argument in [12, §3.3] and in [13, §4.2] to deduce that

$$\|\Psi'(\varphi)\|_{L^2(0,T;H^1)} + \|\varphi\|_{L^2(0,T;H^3)} \leq C. \tag{22}$$

Then, substituting $\xi = q$ in (9d) and the Poincaré inequality (7) gives

$$\begin{aligned} K\|\nabla q\|_{L^2}^2 &\leq \|\Gamma_{\mathbf{v}}\|_{L^2}\|q\|_{L^2} + K\|\varphi\nabla(\mu+\chi\sigma)\|_{L^2}\|\nabla q\|_{L^2} \\ &\leq C + \frac{K}{2}\|\nabla q\|_{L^2}^2 + C\|\varphi\|_{L^\infty}^2\|\nabla(\mu+\chi\sigma)\|_{L^2}^2 \\ &\leq C + \frac{K}{2}\|\nabla q\|_{L^2}^2 + C\|\varphi\|_{L^\infty(0,T;L^6)}^{\frac{3}{2}}\|\varphi\|_{H^3}^{\frac{1}{2}}\|\nabla(\mu+\chi\sigma)\|_{L^2}^2, \end{aligned}$$

where we have also used the Gagliardo–Nirenburg inequality (8) in three dimensions. Thus we obtain

$$\begin{aligned} \int_0^T \|q\|_{H^1}^{\frac{8}{5}}\,\mathrm{dt} &\leq C\left(1 + \|\varphi\|_{L^\infty(0,T;H^1)}^{\frac{6}{5}} \int_0^T \|\varphi\|_{H^3}^{\frac{2}{5}}\|\nabla(\mu+\chi\sigma)\|_{L^2}^{\frac{8}{5}}\,\mathrm{dt}\right) \\ &\leq C\left(1 + \|\varphi\|_{L^2(0,T;H^3)}^{\frac{2}{5}}\|\nabla(\mu+\chi\sigma)\|_{L^2(Q)}^{\frac{8}{5}}\right) \leq C. \end{aligned} \tag{23}$$

Lastly, we see that for any $\zeta \in L^{\frac{8}{3}}(0,T;H^1)$,

$$\begin{aligned} \left|\int_Q \varphi\mathbf{v}\cdot\nabla\zeta\right| &\leq \int_0^T \|\varphi\|_{L^\infty}\|\mathbf{v}\|_{L^2}\|\nabla\zeta\|_{L^2}\,\mathrm{dt} \\ &\leq C\|\varphi\|_{L^\infty(0,T;H^1)}^{\frac{3}{4}}\|\mathbf{v}\|_{L^2(Q)}\|\varphi\|_{L^2(0,T;H^3)}^{\frac{1}{4}}\|\zeta\|_{L^{\frac{8}{3}}(0,T;H^1)} \leq C\|\zeta\|_{L^{\frac{8}{3}}(0,T;H^1)}, \end{aligned} \tag{24}$$

and so from (9a), we obtain

$$\|\partial_t\varphi\|_{L^{\frac{8}{5}}(0,T;(H^1)^*)} \leq C\left(1+\|\nabla\mu\|_{L^2(Q)}+\|\operatorname{div}(\varphi\mathbf{v})\|_{L^{\frac{8}{5}}(0,T;(H^1)^*)}\right) \leq C. \tag{25}$$

Similarly, from (15) we see that

$$\|\theta\partial_t\sigma\|_{L^2(0,T;H^{-1})} \leq \|\sigma\|_{L^2(0,T;H^1)} \leq C. \tag{26}$$

The a priori estimates (21), (22), (23), (25) and (26) are sufficient to deduce the existence of a weak solution quadruple $(\varphi^\theta, \mu^\theta, \sigma^\theta, \mathbf{v}^\theta, q^\theta)$ to (14) with the regularities stated in Lemma 1 which satisfies (9a), (9b), (9d), (9e), and (15) for a.e. $t \in (0,T)$ and all $\zeta \in H^1$, $\mathbf{y} \in L^2$, $\xi \in H^1_0$. We refer the reader to [13] for the details in passing to the limit. Let us just mention that thanks to boundedness in $L^2(0,T;H^1_0) \cap H^1(0,T;H^{-1})$ and [22, §8, Cor. 4] the Galerkin approximations for σ converges strongly in $L^2(Q)$ and hence also a.e. in Q. Furthermore, the estimate (16) is obtained by passing to the limit in the a priori estimates (21), (22), (23), (25) and (26) for the Galerkin approximation and using weak/weak* lower semi-continuity of the norms.

To complete the proof, it remains to show that $0 \leq \sigma^\theta \leq 1$ a.e. in Q by means of a weak comparison principle. For this we substitute $\xi = (\sigma^\theta - 1)_+ := \max(\sigma^\theta - 1, 0)$ and $\xi = (\sigma^\theta)_- := \max(-\sigma^\theta, 0)$ in (15), and note that due to the boundary condition $\sigma^\theta = 1$ on Σ, necessarily $(\sigma^\theta - 1)_+, (\sigma^\theta)_- \in H^1_0$. The former yields

$$\begin{aligned}&\frac{\theta}{2}\frac{\mathrm{d}}{\mathrm{dt}}\|(\sigma^\theta - 1)_+\|^2_{L^2}\\ &\quad = -\|\nabla(\sigma^\theta - 1)_+\|^2_{L^2} - \int_\Omega h(\varphi)\left|(\sigma^\theta - 1)_+\right|^2 + h(\varphi)(\sigma^\theta - 1)_+ \,\mathrm{dx} \leq 0,\end{aligned}$$

and the latter yields

$$\frac{\theta}{2}\frac{\mathrm{d}}{\mathrm{dt}}\|(\sigma^\theta)_-\|^2_{L^2} = -\|\nabla(\sigma^\theta)_-\|^2_{L^2} - \int_\Omega h(\varphi)\left|(\sigma^\theta)_-\right|^2 \mathrm{dx} \leq 0.$$

From both inequalities we infer that for any $t \in (0,T)$,

$$\|(\sigma^\theta(t) - 1)_+\|^2_{L^2} \leq \|(\sigma_0 - 1)_+\|^2_{L^2} = 0, \quad \|(\sigma^\theta(t))_-\|^2_{L^2} \leq \|(\sigma_0)_-\|^2_{L^2} = 0,$$

as $0 \leq \sigma_0 \leq 1$ a.e. in Ω. This yields that $0 \leq \sigma^\theta \leq 1$ a.e. in Q.

□

At this point, we can neglect the cut-off operator $\mathscr{T}$ in (14) and now pass to the limit $\theta \to 0$. By virtue of (16) we have boundedness of $(\varphi^\theta, \mu^\theta, \sigma^\theta, \mathbf{v}^\theta, q^\theta)$ in the Bochner spaces stated in Definition 2.1. Denoting the limit functions as

$(\varphi, \mu, \sigma, \mathbf{v}, q)$, it is a standard argument to show that the above quintuple is a weak solution of (1)–(2a) in the sense of Definition 2.1, and thus we omit the details.

4 Robin Boundary Conditions for the Pressure

To prove Theorem 2.2 for the system (1)–(2b), it suffices to prove the existence of a weak solution $(\varphi^\theta, \mu^\theta, \sigma^\theta, \mathbf{v}^\theta, p^\theta)$ to the regularized problem consisting of (14a), (14c), (14d), (14e) and

$$\mathbf{v} = -K\left(\nabla p - (\mu + \chi\sigma)\nabla\varphi\right) \quad \text{in } Q, \tag{27}$$

along with the initial-boundary conditions

$$\begin{aligned} \partial_{\mathbf{n}}\varphi = 0, \quad \mu = 0, \quad K\partial_{\mathbf{n}} p = a(g - p), \quad \sigma = 1 \text{ on } \Sigma, \\ \varphi(0) = \varphi_0, \quad \sigma(0) = \sigma_0 \text{ in } \Omega, \end{aligned}$$

and then pass to the limit $\theta \to 0$. We focus on obtaining a priori estimates for the regularized problem and omit the argument for $\theta \to 0$ as it follows straightforwardly from the a priori estimates.

Lemma 2 *Under Assumption 2.1 and* $0 \le \sigma_0 \le 1$ *a.e. in* Ω*, for any* $\theta \in (0, 1]$*, there exists a weak solution quintuple* $(\varphi^\theta, \mu^\theta, \sigma^\theta, \mathbf{v}^\theta, p^\theta)$ *in the sense of Definition 2.2 with additionally* $\sigma^\theta \in H^1(0, T; H^{-1})$*,* $\sigma^\theta(0) = \sigma_0$ *a.e. in* Ω*, and (9c) is replaced by (15). Furthermore, there exists a positive constant C not depending on* $\theta, \varphi^\theta, \mu^\theta, \sigma^\theta, \mathbf{v}^\theta, p^\theta$ *such that*

$$\begin{aligned} &\|\Psi(\varphi^\theta)\|_{L^\infty(0,T;L^1)} + \|\Psi'(\varphi^\theta)\|_{L^2(0,T;H^1)} + \|\varphi^\theta\|_{L^\infty(0,T;H^1)\cap L^2(0,T;H^3)} \\ &\quad + \|\mu^\theta\|_{L^2(0,T;H^1)} + \|\mathbf{v}^\theta\|_{L^2(Q)} + \|p^\theta\|_{L^{\frac{8}{5}}(0,T;H^1)} + \|\partial_t\varphi^\theta\|_{L^{\frac{8}{5}}(0,T;H^{-1})} \\ &\quad + \|\sigma^\theta\|_{L^2(0,T;H^1)} + \|\theta\partial_t\sigma^\theta\|_{L^2(0,T;H^{-1})} + \|p^\theta\|_{L^2(\Sigma)} \le C. \end{aligned} \tag{28}$$

Proof Once again we will only derive the a priori estimates. Substituting $\xi = Z(\sigma - 1)$ in (15) for some constant $Z > 0$ yet to be determined, $\zeta = \partial_t\varphi$ in (9b), $\xi = \mu + \chi(\sigma - 1)$ in (10a), $\mathbf{y} = K^{-1}\mathbf{v}$ in (10c), and summing leads to

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{dt}}\int_\Omega A\Psi(\varphi) + \frac{B}{2}|\nabla\varphi|^2 - \chi\varphi + \frac{Z}{2}\theta|\sigma - 1|^2 \,\mathrm{dx} \\ &\quad + \int_\Omega m(\varphi)|\nabla\mu|^2 + \frac{1}{K}|\mathbf{v}|^2 + Z|\nabla\sigma|^2 + Zh(\varphi)|\sigma|^2 \,\mathrm{dx} + a\|p\|_{L^2(\Gamma)}^2 \end{aligned}$$

$$= \int_\Omega -\chi m(\varphi)\nabla\mu \cdot \nabla\sigma + \Gamma_\varphi(\mu + \chi(\sigma-1)) + Zh(\varphi)\sigma \,\mathrm{dx}$$
$$+ \int_\Omega p\Gamma_{\mathbf{v}} + \varphi\mathbf{v}\cdot\nabla(\mu+\chi(\sigma-1)) + (\mu+\chi\sigma)\nabla\varphi\cdot\mathbf{v}\,\mathrm{dx} + \int_\Gamma agp\,\mathrm{d}\Gamma. \quad (29)$$

Using that $(\mu + \chi(\sigma - 1)) = 0$ on Γ and the product rule, we have

$$\int_\Omega \varphi\mathbf{v}\cdot\nabla(\mu+\chi(\sigma-1)) + (\mu+\chi\sigma)\nabla\varphi\cdot\mathbf{v}\,\mathrm{dx}$$
$$= \int_\Omega \chi\mathbf{v}\cdot\nabla\varphi - \Gamma_{\mathbf{v}}\varphi(\mu+\chi(\sigma-1))\,\mathrm{dx}.$$

Thus, we obtain the following identity from integrating (29) in time

$$\int_\Omega \left(A\Psi(\varphi) + \frac{B}{2}|\nabla\varphi|^2 - \chi\varphi + \frac{Z}{2}\theta\,|\sigma-1|^2\right)(t)\,\mathrm{dx}$$
$$+ \int_{\Omega_t}\left(m(\varphi)\,|\nabla\mu|^2 + \frac{1}{K}|\mathbf{v}|^2 + Z\,|\nabla\sigma|^2 + Zh(\varphi)\,|\sigma|^2\right) + \int_{\Gamma_t} a\,|p|^2$$
$$= \int_{\Omega_t}(-\chi m(\varphi)\nabla\mu\cdot\nabla\sigma + \chi\nabla\varphi\cdot\mathbf{v} + Zh(\varphi)\sigma) + \int_{\Gamma_t} agp$$
$$+ \int_{\Omega_t}\big(\Gamma_{\mathbf{v}}(p - \varphi(\mu+\chi(\sigma-1))) + \Gamma_\varphi(\mu+\chi(\sigma-1))\big)$$
$$+ \int_\Omega\left(A\Psi(\varphi_0) + \frac{B}{2}|\nabla\varphi_0|^2 - \chi\varphi_0 + \frac{Z}{2}\theta\,|\sigma_0-1|^2\right)\mathrm{dx} =: I_1 + I_2 + I_3. \quad (30)$$

Note that by (A3) and the fact that $\theta \in (0, 1]$, the third term I_3 on the right-hand side of (30) is bounded, and by Young's inequality

$$\left|\int_\Omega \chi\varphi\,\mathrm{dx}\right| \leq \chi\,|\Omega|^{\frac{1}{2}}\,\|\varphi\|_{L^2} \leq \frac{A}{2C_1}\|\varphi\|_{L^2}^2 + C \leq \frac{A}{2}\|\Psi(\varphi)\|_{L^1} + C,$$

which implies that

$$\int_\Omega (A\Psi(\varphi) - \chi\varphi)\,(t)\,\mathrm{dx} \geq \frac{A}{2}\|\Psi(\varphi(t))\|_{L^1} - C.$$

Next, for I_1, using the Poincaré inequality in L^1 on $(\sigma - 1)$, Hölder's inequality and Young's inequality, we have

$$|I_1| \leq \frac{m_0}{4}\|\nabla\mu\|_{L^2(Q)}^2 + \left(\frac{\chi^2 m_1^2}{m_0} + 1\right)\|\nabla\sigma\|_{L^2(Q)}^2 + \frac{1}{2K}\|\mathbf{v}\|_{L^2(Q)}^2 + \frac{a}{2}\|p\|_{L^2(\Sigma)}^2$$
$$+ C\left(1 + Z^2 + \|\nabla\varphi\|_{L^2(Q)}^2 + \|g\|_{L^2(\Sigma)}^2\right).$$

It remains to estimate I_2, and we first obtain an estimate on $\|p\|_{L^2}$ by looking at the pressure system, whose weak formulation is given by (10b). Let $f := (-\Delta_R)^{-1}(p/K, a/K, 0)$, so that

$$\int_\Omega K\nabla f \cdot \nabla\phi \,\mathrm{dx} + \int_\Gamma af\phi \,\mathrm{d}\Gamma = \int_\Omega p\phi \,\mathrm{dx} \text{ for all } \phi \in H^1.$$

Substituting $\zeta = f$ in (10b) and $\phi = p$ in the above leads to

$$\begin{aligned}\|p\|_{L^2}^2 &= \int_\Omega \Gamma_{\mathbf{v}} f + K(\mu + \chi\sigma)\nabla\varphi \cdot \nabla f \,\mathrm{dx} + \int_\Gamma agf \,\mathrm{d}\Gamma \\ &\leq \|\Gamma_{\mathbf{v}}\|_{L^2}\|f\|_{L^2} + K\|(\mu + \chi\sigma)\nabla\varphi\|_{L^{\frac{6}{5}}}\|\nabla f\|_{L^6} + a\|g\|_{L^2(\Gamma)}\|f\|_{L^2(\Gamma)} \\ &\leq C\left(1 + \|g\|_{L^2(\Gamma)} + \|(\mu + \chi\sigma)\nabla\varphi\|_{L^{\frac{6}{5}}}\right)\|f\|_{H^2}.\end{aligned} \tag{31}$$

Using the elliptic regularity estimate $\|f\|_{H^2} \leq C\|p\|_{L^2}$, we obtain, analogous to (20),

$$\begin{aligned}\|p\|_{L^2} &\leq C\left(1 + \|g\|_{L^2(\Gamma)} + \|(\mu + \chi\sigma)\nabla\varphi\|_{L^{\frac{6}{5}}}\right) \\ &\leq C\left(1 + \|g\|_{L^2(\Gamma)} + \|\mu + \chi\sigma\|_{L^6}\|\nabla\varphi\|_{L^{\frac{3}{2}}}\right) \\ &\leq C\left(1 + \|g\|_{L^2(\Gamma)} + \left(1 + \|\nabla\mu\|_{L^2} + \|\nabla\sigma\|_{L^2}\right)\|\nabla\varphi\|_{L^{\frac{3}{2}}}\right),\end{aligned} \tag{32}$$

where we have applied the Poincaré inequality (7) to μ and $\sigma - 1$, and the Sobolev embedding $H^1 \subset L^6$. Using the boundedness of $\Gamma_{\mathbf{v}}$ and Γ_φ, (A3) and $\|p\|_{L^1(Q)} \leq C\|p\|_{L^1(0,T;L^2)}$, we see that

$$\begin{aligned}|I_2| &\leq C\left(1 + \|p\|_{L^1(Q)} + \left(1 + \|\varphi\|_{L^2(Q)}\right)\left(\|\mu\|_{L^2(Q)} + \|\sigma - 1\|_{L^2(Q)}\right)\right) \\ &\leq C\left(1 + \|g\|_{L^2(\Sigma)} + \|\nabla\varphi\|_{L^2(Q)}^2 + \|\varphi\|_{L^2(Q)}^2\right) + \frac{m_0}{4}\|\nabla\mu\|_{L^2(Q)}^2 + \|\nabla\sigma\|_{L^2(Q)}^2 \\ &\leq C\left(1 + \|\Psi(\varphi)\|_{L^1(Q)} + \|\nabla\varphi\|_{L^2(Q)}^2 + \|g\|_{L^2(\Sigma)}^2\right) + \frac{m_0}{4}\|\nabla\mu\|_{L^2(Q)}^2 + \|\nabla\sigma\|_{L^2(Q)}^2.\end{aligned}$$

Thus, choosing $Z > \frac{\chi^2 m_1^2}{m_0^2} + 2$, we obtain from (30) the inequality

$$\begin{aligned}&\left(\|\Psi(\varphi(t))\|_{L^1} + \|\nabla\varphi(t)\|_{L^2}^2 + \theta\|\sigma(t) - 1\|_{L^2}^2\right) \\ &\quad + \|\nabla\mu\|_{L^2(Q)}^2 + \|\mathbf{v}\|_{L^2(Q)}^2 + \|\nabla\sigma\|_{L^2(Q)}^2 + \|p\|_{L^2(\Sigma)}^2 \\ &\leq C\left(1 + \|g\|_{L^2(\Sigma)}^2 + \|\Psi(\varphi)\|_{L^1(Q)} + \|\nabla\varphi\|_{L^2(Q)}^2\right),\end{aligned}$$

for all $t \in (0, T]$. Applying the integral version of Gronwall's inequality [14, Lem. 3.1], we obtain

$$\begin{aligned}\sup_{t\in(0,T]} &\left(\|\Psi(\varphi(t))\|_{L^1} + \|\nabla\varphi(t)\|_{L^2}^2 + \theta\|\sigma(t) - 1\|_{L^2}^2\right) \\ &+ \|\nabla\mu\|_{L^2(Q)}^2 + \|\nabla\sigma\|_{L^2(Q)}^2 + \|\mathbf{v}\|_{L^2(Q)}^2 + \|p\|_{L^2(\Sigma)}^2 \leq C.\end{aligned} \tag{33}$$

Then, using (A3) and the Poincaré inequality for μ and σ, this yields

$$\begin{aligned}\sup_{t\in(0,T]} &\left(\|\Psi(\varphi(t))\|_{L^1} + \|\varphi(t)\|_{H^1}^2 + \theta\|\sigma(t) - 1\|_{L^2}^2\right) \\ &+ \|\mu\|_{L^2(0,T;H^1)}^2 + \|\sigma\|_{L^2(0,T;H^1)}^2 + \|\mathbf{v}\|_{L^2(Q)}^2 + \|p\|_{L^2(\Sigma)}^2 \leq C.\end{aligned} \tag{34}$$

Analogous to the Dirichlet case, a bootstrapping argument akin to [12, §3.3] and [13, §4.2] leads to the estimate

$$\|\Psi'(\varphi)\|_{L^2(0,T;H^1)} + \|\varphi\|_{L^2(0,T;H^3)} \leq C. \tag{35}$$

Then, from (10b) and the Poincaré inequality (6), it holds that

$$\begin{aligned}K\|\nabla p\|_{L^2}^2 + \frac{a}{2}\|p\|_{L^2(\Gamma)}^2 &\leq \|\Gamma_{\mathbf{v}}\|_{L^2}\|p\|_{L^2} + K\|(\mu + \chi\sigma)\nabla\varphi\|_{L^2}\|\nabla p\|_{L^2} + \frac{a}{2}\|g\|_{L^2(\Gamma)}^2 \\ &\leq C\left(1 + \|g\|_{L^2(\Gamma)}^2\right) + \frac{K}{2}\|\nabla p\|_{L^2}^2 + \frac{a}{4}\|p\|_{L^2(\Gamma)}^2 + K\|(\mu + \chi\sigma)\nabla\varphi\|_{L^2}^2,\end{aligned}$$

which implies that

$$\|p\|_{H^1} \leq C\left(1 + \|g\|_{L^2(\Gamma)} + \|(\mu + \chi\sigma)\nabla\varphi\|_{L^2}\right). \tag{36}$$

By the Gagliardo–Nirenburg inequality (8) for $d = 3$, we see that

$$\|\nabla\varphi\|_{L^3} \leq C\|\varphi\|_{H^3}^{\frac{1}{4}}\|\varphi\|_{L^6}^{\frac{3}{4}}, \tag{37}$$

and thus $(\mu + \chi\sigma)\nabla\varphi \in L^{\frac{8}{5}}(0, T; L^2)$. From (36) this implies that

$$\|p\|_{L^{\frac{8}{5}}(0,T;H^1)} \leq C. \tag{38}$$

Analogous to (24), for $\xi \in L^{\frac{8}{3}}(0, T; H_0^1)$, using that $\varphi \in L^\infty(0, T; H^1) \cap L^2(0, T; H^3)$ and $\mathbf{v} \in L^2(Q)$ leads to

$$\left|\int_Q \varphi\mathbf{v}\cdot\nabla\xi\right| \leq C\|\xi\|_{L^{\frac{8}{3}}(0,T;H_0^1)},$$

which in turn gives

$$\|\partial_t \varphi\|_{L^{\frac{8}{5}}(0,T;H^{-1})} \leq C \tag{39}$$

by the inspection of (10a). Similarly, by inspection of (15), the a priori estimate (26) is also valid.

The a priori estimates (26), (34), (35), (38) and (39), together with a Galerkin approximation are sufficient to deduce the existence of a weak solution quintuple $(\varphi^\theta, \mu^\theta, \sigma^\theta, \mathbf{v}^\theta, p^\theta)$ satisfying the assertions of Lemma 2. Once again, (28) follows from weak/weak* lower semi-continuity of the norms, and the assertion $0 \leq \sigma^\theta \leq 1$ a.e. in Q follows from a weak comparison principle as in the proof of Lemma 1.

□

Remark 1 The necessity of a Dirichlet condition for μ is due to the fact that we cannot control $\|\mu\nabla\varphi\|_{L^{\frac{6}{5}}}$ in (32) simply with the left-hand side of (30) if we assume $\partial_{\mathbf{n}}\mu = 0$ on Σ. One could consider the splitting

$$\begin{aligned}\|\mu\nabla\varphi\|_{L^{\frac{6}{5}}} &\leq \|(\mu - \overline{\mu})\nabla\varphi\|_{L^{\frac{6}{5}}} + |\overline{\mu}|\,\|\nabla\varphi\|_{L^{\frac{6}{5}}} \leq \|\mu - \overline{\mu}\|_{L^6}\|\nabla\varphi\|_{L^{\frac{3}{2}}} + |\overline{\mu}|\,\|\nabla\varphi\|_{L^{\frac{6}{5}}} \\ &\leq C\|\nabla\mu\|_{L^2}\|\nabla\varphi\|_{L^{\frac{3}{2}}} + C\left(1 + \|\sigma - 1\|_{L^2} + \|\Psi'(\varphi)\|_{L^1}\right)\|\nabla\varphi\|_{L^{\frac{6}{5}}},\end{aligned}$$

and in order to control the second term, it is desirable to have an estimate of the form

$$\|\Psi'(\varphi)\|_{L^1}^2 \leq C\left(1 + \|\Psi(\varphi)\|_{L^1}\right).$$

This leads to the situation encountered in [14] and restricts Ψ to have quadratic growth. Furthermore, the ansatz in [13, 19] is to consider the splitting

$$\begin{aligned}\left|\int_\Omega \Gamma_{\mathbf{v}}(p - \mu\varphi)\,\mathrm{dx}\right| &= \left|\int_\Omega \Gamma_{\mathbf{v}}(p - \overline{\mu}\varphi) + \Gamma_{\mathbf{v}}(\overline{\mu} - \mu)\varphi\,\mathrm{dx}\right| \\ &\leq \left|\int_\Omega \Gamma_{\mathbf{v}}(p - \overline{\mu}\varphi)\,\mathrm{dx}\right| + C\|\nabla\mu\|_{L^2}\|\varphi\|_{L^2}.\end{aligned}$$

If p satisfies the Darcy law (2b) with the boundary condition $\partial_{\mathbf{n}} p = 0$ on Σ, and if $\Gamma_{\mathbf{v}}$ has zero mean, then we can write

$$p = (-\Delta_N)^{-1}\left(\Gamma_{\mathbf{v}}/K - \operatorname{div}\left((\mu - \overline{\mu} + \chi\sigma)\nabla\varphi\right) - \overline{\mu}\operatorname{div}\left(\nabla(\varphi - \overline{\varphi})\right)\right),$$

where for $f \in L^2$ with $\overline{f} = \frac{1}{|\Omega|}\int_\Omega f\,\mathrm{dx} = 0$, we denote $u := (-\Delta_N)^{-1}(f) \in H^1$ as the unique weak solution to

$$-\Delta u = f \text{ in } \Omega, \quad \partial_{\mathbf{n}} u = 0 \text{ on } \Gamma \text{ with } \overline{u} = 0.$$

A short calculation shows that

$$-(-\Delta_N)^{-1}(\operatorname{div}(\overline{\mu}\nabla(\varphi-\overline{\varphi})))=\overline{\mu}(\varphi-\overline{\varphi}),$$

and so

$$\int_\Omega \Gamma_{\mathbf{v}}(p-\overline{\mu}\varphi)\,\mathrm{dx}=\int_\Omega \Gamma_{\mathbf{v}}\left((-\Delta_N)^{-1}\left(\Gamma_{\mathbf{v}}/K-\operatorname{div}((\mu-\overline{\mu}+\chi\sigma)\nabla\varphi)\right)\right)-\Gamma_{\mathbf{v}}\overline{\mu}\,\overline{\varphi}\,\mathrm{dx}.$$

In [13, 19], $\Gamma_{\mathbf{v}}$ has zero mean, and so the last term on the right-hand side vanishes, but this is not the case in our present setting, and thus the approach of [13, 19] seems not to give any advantage in deriving a priori estimates.

5 Neumann Boundary Conditions for the Chemical Potential

In this section, let us state an analogous result to Lemma 2 for the regularized problem consisting of (14a), (14c), (14d), (14e) and (27), but now we consider the boundary conditions

$$\partial_{\mathbf{n}}\varphi=\partial_{\mathbf{n}}\mu=0,\quad K\partial_{\mathbf{n}}p=a(g-p)\text{ on }\Sigma, \tag{40}$$

and (12) instead of (A3). The assertion is formulated as follows.

Lemma 3 *Under Assumption 2.1* (*with* (12) *instead of* (A3)) *and* $0\le\sigma_0\le 1$ *a.e. in* Ω*, for any* $\theta\in(0,1]$*, there exists a weak solution quintuple* $(\varphi^\theta,\mu^\theta,\sigma^\theta,\mathbf{v}^\theta,p^\theta)$ *in the sense of Definition 2.3 with additionally* $\sigma^\theta\in H^1(0,T;H^{-1})$*,* $\sigma^\theta(0)=\sigma_0$ *a.e. in* Ω*, and* (9c) *is replaced by* (15)*. Furthermore, there exists a positive constant* C *not depending on* $\theta,\varphi^\theta,\mu^\theta,\sigma^\theta,\mathbf{v}^\theta,p^\theta$ *such that*

$$\begin{aligned}&\|\Psi(\varphi^\theta)\|_{L^\infty(0,T;L^1)}+\|\Psi'(\varphi^\theta)\|_{L^2(0,T;H^1)}+\|\varphi^\theta\|_{L^\infty(0,T;H^1)\cap L^2(0,T;H^3)}\\&\quad+\|\mu^\theta\|_{L^2(0,T;H^1)}+\|\mathbf{v}^\theta\|_{L^2(Q)}+\|p^\theta\|_{L^{\frac{8}{5}}(0,T;H^1)}+\|\partial_t\varphi^\theta\|_{L^{\frac{8}{5}}(0,T;(H^1)^*)}\\&\quad+\|\sigma^\theta\|_{L^2(0,T;H^1)}+\|\theta\partial_t\sigma^\theta\|_{L^2(0,T;H^{-1})}+\|p^\theta\|_{L^2(\Sigma)}\le C.\end{aligned}\tag{41}$$

Proof Once again we will only derive the a priori estimates and omit the details of the Galerkin approximation. Substituting $\zeta=\mu+\chi\sigma$ into (11), and upon adding with the equalities obtained from substituting $\xi=Z(\sigma-1)$ in (15), $\zeta=\partial_t\varphi$ in (9b) and $\mathbf{y}=K^{-1}\mathbf{v}$ in (10c) we have

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{dt}}\int_\Omega A\Psi(\varphi)+\frac{B}{2}|\nabla\varphi|^2+\frac{Z}{2}\theta\,|\sigma-1|^2\,\mathrm{dx}\\&\quad+\int_\Omega m(\varphi)\,|\nabla\mu|^2+\frac{1}{K}|\mathbf{v}|^2+Z\,|\nabla\sigma|^2+Zh(\varphi)\,|\sigma|^2\,\mathrm{dx}+\int_\Gamma a\,|p|^2\,\mathrm{d}\Gamma\end{aligned}$$

$$\begin{aligned} &= \int_\Omega -\chi m(\varphi)\nabla\mu\cdot\nabla\sigma + \Gamma_\varphi(\mu+\chi\sigma) + \Gamma_{\mathbf{v}}(p-\varphi(\mu+\chi\sigma))\,\mathrm{dx} \\ &\quad + \int_\Omega Zh(\varphi)\sigma\,\mathrm{dx} + \int_\Gamma agp\,\mathrm{d}\Gamma. \end{aligned} \tag{42}$$

For the terms $-\chi m(\varphi)\nabla\mu\cdot\nabla\sigma$ and $Zh(\varphi)\sigma$, as well as the boundary term agp on the right-hand side we use Hölder's inequality, Young's inequality and the Poincaré inequality applied to $(\sigma-1)$ to obtain

$$\begin{aligned} &\left| \int_\Omega -\chi m(\varphi)\nabla\mu\cdot\nabla\sigma + Zh(\varphi)(\sigma-1+1)\,\mathrm{dx} + \int_\Gamma agp\,\mathrm{d}\Gamma \right| \\ &\quad \leq \frac{m_0}{4}\|\nabla\mu\|_{L^2}^2 + \left(\frac{\chi^2 m_1^2}{m_0^2}+1\right)\|\nabla\sigma\|_{L^2}^2 + \frac{a}{2}\|p\|_{L^2(\Sigma)}^2 + \frac{a}{2}\|g\|_{L^2(\Sigma)}^2 + C(1+Z^2). \end{aligned}$$

Since the pressure p satisfies the same Poisson equation, by following the computations in Sect. 4 and the discussion in Remark 1, we obtain

$$\begin{aligned} \|p\|_{L^2} &\leq C\left(1+\|g\|_{L^2(\Gamma)} + \|(\mu+\chi(\sigma-1+1))\nabla\varphi\|_{L^{\frac{6}{5}}}\right) \\ &\leq C\left(1+\|g\|_{L^2(\Gamma)} + \left(\|\nabla\mu\|_{L^2}+\|\nabla\sigma\|_{L^2}\right)\|\nabla\varphi\|_{L^{\frac{3}{2}}} + (1+|\overline{\mu}|)\|\nabla\varphi\|_{L^{\frac{6}{5}}}\right). \end{aligned}$$

Substituting $\zeta = 1$ in (9b), we can estimate the mean of μ by

$$|\overline{\mu}| \leq C\left(\|\sigma\|_{L^2} + \|\Psi'(\varphi)\|_{L^1}\right), \tag{43}$$

and so by Young's inequality and the boundedness of $\Gamma_{\mathbf{v}}$, we see that

$$\begin{aligned} |X| &:= \left|\int_\Omega \Gamma_{\mathbf{v}}(p-\varphi(\mu-\overline{\mu}) - \varphi(\overline{\mu}+\chi\sigma))\,\mathrm{dx}\right| \\ &\leq C\left(\|p\|_{L^2} + \|\varphi\|_{L^2}\|\nabla\mu\|_{L^2} + \left(\|\sigma\|_{L^2}+\|\Psi'(\varphi)\|_{L^1}\right)\|\varphi\|_{L^2}\right) \\ &\leq C\left(1+\|g\|_{L^2(\Gamma)} + \left(1+\|\nabla\mu\|_{L^2}+\|\nabla\sigma\|_{L^2}+\|\Psi'(\varphi)\|_{L^1}\right)\|\varphi\|_{H^1}\right) \\ &\leq \frac{m_0}{4}\|\nabla\mu\|_{L^2}^2 + \|\nabla\sigma\|_{L^2}^2 + C\left(1+\|g\|_{L^2(\Gamma)} + \|\Psi'(\varphi)\|_{L^1}^2 + \|\varphi\|_{L^2}^2 + \|\nabla\varphi\|_{L^2}^2\right). \end{aligned}$$

Using that Ψ has quadratic growth, we can find positive constants C_4, C_5 such that

$$\left|\Psi'(s)\right| \leq C_4\,|s| + C_5 \quad \forall s\in\mathbb{R},$$

and so by (12)

$$\|\Psi'(\varphi)\|_{L^1}^2 \leq C\left(1+\|\varphi\|_{L^2}^2\right) \leq C\left(1+\|\Psi(\varphi)\|_{L^1}\right). \tag{44}$$

This implies that

$$|X| \leq \frac{m_0}{4}\|\nabla\mu\|_{L^2}^2 + \|\nabla\sigma\|_{L^2}^2 + C\left(1+\|g\|_{L^2(\Gamma)}^2 + \|\Psi(\varphi)\|_{L^1} + \|\nabla\varphi\|_{L^2}^2\right).$$

In a similar fashion, the second term on the right-hand side of (42) can be estimated as

$$\begin{aligned}\left|\int_\Omega \Gamma_\varphi(\mu-\overline{\mu}+\overline{\mu}+\chi(\sigma-1+1))\,\mathrm{dx}\right| &\leq C\left(1+|\overline{\mu}|+\|\nabla\mu\|_{L^2}+\|\nabla\sigma\|_{L^2}\right)\\ &\leq \frac{m_0}{4}\|\nabla\mu\|_{L^2}^2+\|\nabla\sigma\|_{L^2}^2 + C\left(1+\|\Psi(\varphi)\|_{L^1}\right),\end{aligned}$$

and we obtain from (42)

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{dt}}\int_\Omega A\Psi(\varphi)+\frac{B}{2}|\nabla\varphi|^2+\frac{Z}{2}\theta\,|\sigma-1|^2\,\mathrm{dx}\\ &\quad+\frac{m_0}{4}\|\nabla\mu\|_{L^2}^2+\frac{1}{K}\|\mathbf{v}\|_{L^2}^2+\left(Z-\frac{\chi^2 m_1^2}{m_0^2}-3\right)\|\nabla\sigma\|_{L^2}^2+\frac{a}{2}\|p\|_{L^2(\Gamma)}^2\\ &\leq C\left(1+Z^2+\|g\|_{L^2(\Gamma)}^2+\|\Psi(\varphi)\|_{L^1}+\|\nabla\varphi\|_{L^2}^2\right).\end{aligned}$$

Applying Gronwall's inequality leads to (33), and the a priori estimate (34) follows by applying (43), (44) and the Poincaré inequality (5) for μ and $\sigma - 1$. The other a priori estimates (35), (38) follow from a similar argument. For the time derivative $\partial_t\varphi$, we note that $\nabla\varphi\cdot\mathbf{v}\in L^{\frac{8}{5}}(0,T;(H^1)^*)$ by (37), and so from (11) it holds that

$$\|\partial_t\varphi\|_{L^{\frac{8}{5}}(0,T;(H^1)^*)} \leq C. \tag{45}$$

Together with (26), the a priori estimates (34), (35), (38) and (45), and a Galerkin approximation are sufficient to deduce the existence of a weak solution quintuple $(\varphi^\theta,\mu^\theta,\sigma^\theta,\mathbf{v}^\theta,p^\theta)$ satisfying the assertions of Lemma 3. Furthermore, by weak/weak* lower semi-continuity of the norms we obtain the estimate (41), and by a weak comparison principle, it also holds that $0\leq\sigma^\theta\leq 1$ a.e. in Q.

□

For the proof of Theorem 2.3 we pass to the limit $\theta\to 0$, using the estimate (41). We omit the details as it is a standard argument.

References

1. Adams, R.A., Fournier, J.J.F.: Sobolev spaces. Pure and Applied Mathematics, vol. 140, 2nd edn. Elsevier/Academic Press, Amsterdam (2003)
2. Bosia, S., Conti, M., Grasselli, M.: On the Cahn–Hilliard–Brinkman system. Commun. Math. Sci. **13**(6), 1541–1567 (2015)
3. Chen, Y., Wise, S.M., Shenoy, V.B., Lowengrub, J.S.: A stable scheme for a nonlinear, multiphase tumor growth model with an elastic membrane. Int. J. Numer. Method Biomed. Eng. **30**, 726–754 (2014)
4. Chen, Z., Huan, G., Ma, Y.: Computational Methods for Multiphase Flows in Porous Media. Society for Industrial and Applied Mathematics, Philadelphia (2006)
5. Colli, P., Gilardi, G., Hilhorst, D.: On a Cahn–Hilliard type phase field model related to tumor growth. Discrete Contin. Dyn. Syst. **35**(6), 2423–2442 (2015)
6. Cristini, V., Li, X., Lowengrub, J.S., Wise, S.M.: Nonlinear simulations of solid tumor growth using a mixture model: Invasion and branching. J. Math. Biol. **58**, 723–763 (2009)
7. Cristini, V., Lowengrub, J.: Multiscale Modeling of Cancer: An Integrated Experimental and Mathematical Modeling Approach. Cambridge University Press, Cambridge (2010)
8. Dai, M., Feireisl, E., Rocca, E., Schimperna, G., Schonbek, M.: Analysis of a diffuse interface model for multispecies tumor growth. Nonlinearity **30**(4), 1639–1658 (2017)
9. DiBenedetto, E.: Degenerate Parabolic Equations. Universitext. Springer, New York (1993)
10. Feng, X., Wise, S.M.: Analysis of a Darcy-Cahn-Hilliard diffuse interface model for the Hele-Shaw flow and its fully discrete finite element approximation. SIAM J. Numer. Anal. **50**, 1320–1343 (2012)
11. Frigeri, S., Grasselli, M., Rocca, E.: On a diffuse interface model of tumor growth. Eur. J. Appl. Math. **26**, 215–243 (2015)
12. Garcke, H., Lam, K.F.: Analysis of a Cahn–Hilliard system with non-zero Dirichlet conditions modeling tumor growth with chemotaxis. Discrete Contin. Dyn. Syst. **37**(8), 4277–4308 (2017)
13. Garcke, H., Lam, K.F.: Global weak solutions and asymptotic limits of a Cahn–Hilliard–Darcy system modelling tumour growth. AIMS Math. **1**(3), 318–360 (2016)
14. Garcke, H., Lam, K.F.: Well-posedness of a Cahn–Hilliard–Darcy system modelling tumour growth with chemotaxis and active transport. Eur. J. Appl. Math. **28**(2), 284–316 (2017)
15. Garcke, H., Lam, K.F., Rocca, E.: Optimal control of treatment time in a diffuse interface model of tumor growth. Appl. Math. Optim. (2017, to be appear). DOI:10.1007/s00245–017-9414-4
16. Garcke, H., Lam, K.F., Sitka, E., Styles, V.: A Cahn–Hilliard–Darcy model for tumour growth with chemotaxis and active transport. Math. Models Methods Appl. Sci. **26**(6), 1095–1148 (2016)
17. Grisvard, P.: Elliptic Problems on Nonsmooth Domains. Volume Monographs and Studies in Mathematics, Vol 24. Pitman, Boston (1985)
18. Hawkins-Daarud, A., van der Zee, K.G., Oden, J.T.: Numerical simulation of a thermodynamically consistent four-species tumor growth model. Int. J. Numer. Method Biomed. Eng. **28**, 3–24 (2012)
19. Jiang, J., Wu, H., Zheng, S.: Well-posedness and long-time behavior of a non-autonomous Cahn–Hilliard–Darcy system with mass source modeling tumor growth. J. Differ. Equ. **259**(7), 3032–3077 (2015)
20. Lowengrub, J.S., Titi, E., Zhao, K.: Analysis of a mixture model of tumor growth. Eur. J. Appl. Math. **24**, 691–734 (2013)
21. Oden, J.T., Hawkins, A., Prudhomme, S.: General diffuse-interface theories and an approach to predictive tumor growth modeling. Math. Models Methods Appl. Sci. **58**, 723–763 (2010)
22. Simon, J.: Compact sets in space $L^p(0, T; B)$. Ann. Mat. Pura Appl. **146**(1), 65–96 (1986)
23. Wang, X., Wu, H.: Long-time behavior for the Hele–Shaw–Cahn–Hilliard system. Asymptot. Anal. **78**(4), 217–245 (2012)
24. Wang, X., Zhang, Z.: Well-posedness of the Hele–Shaw–Cahn–Hilliard system. Ann. Inst. H. Poincaré Anal. Non Linéaire. **30**(3), 367–384 (2013)

Molecular Extended Thermodynamics of a Rarefied Polyatomic Gas

Tommaso Ruggeri

Abstract Extended Thermodynamics can be considered as a theory of *continuum with structure* because there are new field variables with respect to the classical approach and they are dictated at mesoscopic level by the kinetic theory. In this survey I present some recent results on the so called *Molecular Extended Thermodynamics* (MET) in which the macroscopic fields are related to the moments of a distribution function that for polyatomic gas contains an extra variable taking into account the internal degrees of freedom of a molecule. The closure is obtained via the variational procedure of the *Maximum Entropy Principle* (MEP). Particular attention will be paid on the simple model of MET with six independent fields, i.e., the mass density, the velocity, the temperature and the dynamic pressure, without adopting near-equilibrium approximation. The model obtained is the simplest example of non-linear dissipative fluid after the ideal case of Euler. The system is symmetric hyperbolic with the convex entropy density and the K-condition is satisfied. Therefore, in contrast to the Euler case, there exist global smooth solutions provided that the initial data are sufficiently smooth.

1 Continuum and Kinetic Approaches of a Non-Equilibrium Gas

The study of nonequilibrium phenomena in gases is particularly important from a theoretical point of view and also from a viewpoint of many possible practical applications. We have two complementary approaches to study rarefied gases, namely the *continuum approach* and the *kinetic approach*.

The continuum model consists in the description of the system by means of macroscopic equations (e.g., fluid-dynamic equations) obtained on the basis of

T. Ruggeri (✉)
Department of Mathematics and Alma Mater Research Center of Applied Mathematics AM^2,
University of Bologna, Bologna, Italy
e-mail: tommaso.ruggeri@unibo.it

E. Rocca et al. (eds.), *Trends in Applications of Mathematics to Mechanics*,
Springer INdAM Series 27, https://doi.org/10.1007/978-3-319-75940-1_13

conservation laws and appropriate constitutive equations. A typical example is the *thermodynamics of irreversible processes* (TIP). The applicability of this classical macroscopic theory is, however, inherently restricted to a nonequilibrium state characterized by a small *Knudsen number* K_n, which is a measure to what extent the gas is rarefied:

$$K_n = \frac{\text{mean free path of molecule}}{\text{macroscopic characteristic length}}.$$

The approach based on the kinetic theory postulates that the state of a gas can be described by the velocity distribution function. The evolution of the distribution function is governed by the Boltzmann equation. The kinetic theory is applicable to a nonequilibrium state characterized by a large K_n, and the transport coefficients naturally emerge from the theory itself. Therefore the range of the applicability of the Boltzmann equation is limited to rarefied gases.

The Rational Extended Thermodynamics theory (RET) [1], which is a generalization of the TIP theory, also belongs to the continuum approach but is applicable to a nonequilibrium state with larger K_n. In a sense, RET is a sort of bridge between TIP and the kinetic theory. An interesting point to be noticed is that, in the case of rarefied gases, there exists a common applicability range of the RET theory and the kinetic theory. Therefore, in such a range, the results from the two theories should be consistent with each other. Because of this, we can expect that the kinetic-theoretical considerations can motivate us at mesoscopic level to establish the mathematical structure of the RET theory.

2 Extended Thermodynamics of Rarefied Monatomic Gases

The kinetic theory describes a state of a rarefied gas by using the phase density (velocity distribution function) $f(\mathbf{x}, t, \mathbf{c})$, where $f(\mathbf{x}, t, \mathbf{c})d\mathbf{c}$ is the number density of (monatomic) molecules at the point $\mathbf{x}$ and time t that have velocities between $\mathbf{c}$ and $\mathbf{c} + d\mathbf{c}$. Time-evolution of the phase density is governed by the Boltzmann equation:

$$\partial_t f + c_i \, \partial_i f = Q, \tag{1}$$

where the right-hand side, the collision term, describes the effect of collisions between molecules. Here

$$\partial_t \equiv \frac{\partial}{\partial t} \quad \text{and} \quad \partial_i \equiv \frac{\partial}{\partial x_i},$$

and as usual we omit the symbol of sum over repeated italic indexes between 1 to 3. Most macroscopic thermodynamic quantities are identified as the moments of the phase density

$$F = \int_{\mathbb{R}^3} f d\mathbf{c}, \qquad F_{k_1 k_2 \cdots k_j} = \int_{\mathbb{R}^3} f c_{k_1} c_{k_2} \cdots c_{k_j} d\mathbf{c}, \quad (j = 1, \dots) \tag{2}$$

and due to the Boltzmann equation (1), the moments satisfy an infinite hierarchy of balance laws in which the flux in one equation becomes the density in the next one:

$$\begin{array}{l}
\partial_t F + \partial_i F_i = 0 \\
\qquad \swarrow \\
\partial_t F_{k_1} + \partial_i F_{ik_1} = 0 \\
\qquad \swarrow \\
\partial_t F_{k_1 k_2} + \partial_i F_{ik_1 k_2} = P_{<k_1 k_2>} \\
\qquad \swarrow \\
\partial_t F_{k_1 k_2 k_3} + \partial_i F_{ik_1 k_2 k_3} = P_{k_1 k_2 k_3} \\
\vdots \\
\partial_t F_{k_1 k_2 \dots k_N} + \partial_i F_{ik_1 k_2 \dots k_N} = P_{k_1 k_2 \dots k_N} \\
\vdots
\end{array} \tag{3}$$

where

$$P_{k_1 k_2 \cdots k_j} = \int_{\mathbb{R}^3} Q c_{k_1} c_{k_2} \cdots c_{k_j} d\boldsymbol{c}.$$

As $P_{kk} = 0$, we notice that the first five equations are exactly the conservation laws, and correspond to the conservation laws of mass, momentum and energy (except for the factor 2) of continuum thermomechanics. For this reason we have the expression, in particular, for the flux of (3)$_2$:

$$F_{ik} = \rho v_i v_k - t_{ik}, \tag{4}$$

where ρ is the mass density, v_i the velocity, and t_{ij} denotes the stress tensor:

$$t_{ik} = -p\delta_{ik} + \sigma_{ik}, \quad \sigma_{ik} = -\Pi\delta_{ik} + \sigma_{<ik>}$$

with p, Π, and $\sigma_{<ik>}$ being, respectively, the equilibrium pressure, the dynamical (non-equilibrium) pressure, and the shear viscous deviatoric tensor $\sigma_{<ik>}$. While the trace of the density in (3)$_3$ denotes, except for the factor 2, the total energy:

$$F_{ll} = 2\rho\varepsilon + \rho v^2, \tag{5}$$

where ε is the specific internal energy. As a consequence from the trace of (4) and (5), we have the relationship $3(p + \Pi) = 2\rho\varepsilon$. As Π is a nonequilibrium quantity that vanishes in equilibrium, we obtain

$$p = \frac{2}{3}\rho\varepsilon \quad \text{and} \quad \Pi \equiv 0. \tag{6}$$

Then the gas under consideration is indeed monatomic, and the dynamic pressure vanishes identically. This is a strong limitation on the kinetic theory. It is valid only for rarefied monatomic gases with viscous stress tensor σ_{ij} that must be deviatoric, i.e., traceless: $\Pi \equiv 0$.

When we cut the hierarchy at the density with tensor of rank N, we have the problem of closure because the last flux and the production terms are not in the list of the densities. The first idea of RET [1] was to view the truncated system as a phenomenological system of continuum mechanics and then we consider the new quantities as local constitutive functions of the densities:

$$\begin{aligned}
&F_{k_1k_2\ldots k_Nk_{N+1}} \equiv F_{k_1k_2\ldots k_Nk_{N+1}}\left(F, F_{k_1}, F_{k_1k_2}, \ldots F_{k_1k_2\ldots k_N}\right),\\
&P_{<k_1k_2>} \equiv P_{<k_1k_2>}\left(F, F_{k_1}, F_{k_1k_2}, \ldots F_{k_1k_2\ldots k_N}\right),\\
&P_{k_1k_2\ldots k_j} \equiv P_{k_1k_2\ldots k_j}\left(F, F_{k_1}, F_{k_1k_2}, \ldots F_{k_1k_2\ldots k_N}\right), \quad 3 \le j \le N.
\end{aligned} \tag{7}$$

According with the continuum theory, the restrictions on the constitutive equations come only from *universal principles*, i.e.: *Entropy principle*, *Objectivity Principle* and *Causality and Stability* (convexity of the entropy).

The most interesting physical cases was the 13 fields theory in classical framework [2] and the 14 fields in the context of relativistic fluids [3]. In both cases the previous universal principles are enough to determine completely the form of the constitutive equations (7) at least in a theory not so far from a equilibrium state (linear with respect the non-equilibrium variables).

3 Closure via the Maximum Entropy Principle and Molecular Extended Thermodynamics of Monatomic Gases

If the number of moments increases, it becomes to be too difficult to adopt the pure continuum approach for a system with such a large number of field variables. Therefore it is necessary to recall that the field variables are the moments of a distribution function. To obtain the closure of the balance equations of the moments truncated at some tensorial order N, we adopt the *maximum entropy principle* (MEP). This is the procedure of the so-called *molecular extended thermodynamics* (molecular RET) [4].

The principle of maximum entropy has its root in statistical mechanics. It is developed by Jaynes [5] in the context of the theory of information basing on the Shannon entropy. Nowadays the importance of MEP is recognized fully due to the numerous applications in many fields [6], for example, in the field of computer graphics. MEP states that the probability distribution that represents the current state of knowledge in the best way is the one with the largest entropy. Another way of stating this is as follows: take precisely stated prior data or testable information about a probability distribution function. Then consider the set of all trial probability distributions that would encode the prior data. Of those, one with maximal information entropy is the proper distribution, according to this principle.

Concerning the applicability of MEP in nonequilibrium thermodynamics, this was originally by the observation made by Kogan [7] that Grad's distribution [8] function maximizes the entropy. The MEP was proposed in RET for the first time by Dreyer [9]. In this way the 13-moment theory closure can be obtained in three different ways: phenomenological RET, Grad kinetic method, and MEP. A remarkable point is that all closures are equivalent to each other! The MEP procedure was then generalized by Müller and Ruggeri to the case of any number of moments in the first edition of their book proving that the closed system is symmetric hyperbolic [4]. In MET the complete equivalence between the closures via the entropy principle and via the MEP was finally proved by Boillat and Ruggeri in [10].

In the case of monatomic gases, we can define the moments (2) using a multi-index:

$$F_A = \begin{cases} F & \text{for } A = 0 \\ F_{k_1 k_2 \cdots k_A} & \text{for } 1 \leq A \leq N, \end{cases}$$

and in this way the truncated system (3) at the tensorial order N can be rewritten in a simple form:

$$\partial_t F_A + \partial_i F_{iA} = P_A, \qquad A = 0, \dots N \tag{8}$$

with

$$F_A = m \int_{\mathbb{R}^3} c_A \, f \, d\mathbf{c}, \quad F_{iA} = m \int_{\mathbb{R}^3} c_i c_A \, f \, d\mathbf{c}, \quad P_A = m \int_{\mathbb{R}^3} c_A \, Q \, d\mathbf{c} \tag{9}$$

and

$$c_A = \begin{cases} 1 & \text{for } A = 0 \\ c_{k_1} c_{k_2} \cdots c_{k_A} & \text{for } 1 \leq A \leq N. \end{cases}$$

The variational problem, from which the distribution function f render the entropy

$$h = -k_B \int_{\mathbb{R}^3} f \log f \; d\mathbf{c}$$

(k_B is the Boltzmann constant) maximum for the prescribed moments, is obtained through the functional (we omit the symbol of sum from 0 to N in the repeated capital indexes $A, B, \dots$):

$$\mathcal{L}_N\left(f\right) = -k_B \int_{\mathbb{R}^3} f \log f \; d\mathbf{c} + u'_A \left(F_A - m \int_{\mathbb{R}^3} c_A \, f \, d\mathbf{c} \right),$$

where u'_A are the Lagrange multipliers:

$$u'_A = \begin{cases} u' & \text{for } A = 0 \\ u'_{k_1 k_2 \cdots k_A} & \text{for } 1 \leq A \leq N. \end{cases}$$

The distribution function f_N which maximizes the functional $\mathcal{L}_N$ is given by [1, 4, 11, 12]:

$$f_N = \exp\left(-1 - \frac{m}{k}\chi_N\right), \qquad \chi_N = u'_A c_A. \tag{10}$$

In an equilibrium state, (10) reduces to the Maxwellian distribution function $f^{(M)}$. Then, the system may be rewritten as follows:

$$J_{AB}\partial_t u'_B + J_{iAB}\partial_i u'_B = P_A(u'_C), \qquad A = 0, \dots, N \tag{11}$$

where

$$J_{AB}\left(u'_C\right) = -\frac{m^2}{k_B}\int_{\mathbb{R}^3} f_N \, c_A c_B \, d\mathbf{c}, \qquad J_{iAB}\left(u'_C\right) = -\frac{m^2}{k_B}\int_{\mathbb{R}^3} f_N \, c_i c_A c_B \, d\mathbf{c}.$$

Because of the fact that the matrices J_{AB}, J_{iAB} are symmetric with respect to the multi-index A, B and J_{AB} is definite negative, the system (11) is symmetric hyperbolic [1, 4, 11] and the Lagrange multipliers coincide with the *main field* according with the general theory of systems of balance laws with a convex entropy density [13–17]. We observe that f_N is not a solution of the Boltzmann equation. But we have the conjecture (open problem) that, for $N \to \infty$, f_N tends to a solution of the Boltzmann equation.

4 Convergence Problem and Approximation Near an Equilibrium State

All results explained above are valid also for a case far from equilibrium provided that the integrals in (9) are convergent. The problem of the convergence of the moments is one of the main questions in a far-from-equilibrium case. In particular the index of truncation N must be even [11, 18]. This implies, in particular, that a theory with 13 moments is not allowed when far from equilibrium! Moreover, if the conjecture that the distribution function f_N, when $N \to \infty$, tends to the distribution function f that satisfies the Boltzmann equation is true, we need another convergence requirement for χ given in (11). These problems were studied by Boillat and Ruggeri [11].

To bypass the question of convergence of integrals, the distribution function obtained as the solution of the variational problem is considered only in the neighborhood of a local equilibrium state, and we formally expand the distribution function (10) as the perturbation of the Maxwellian distribution $f^{(M)}$:

$$f_N \approx f^{(M)} \left(1 - \frac{m}{k_B} \tilde{u}'_A c_A\right), \quad \tilde{u}'_A = u'_A - u'^E_A, \tag{12}$$

where u'^E_A are the main field components evaluated in the local equilibrium state. More high expansion was considered in the paper [19].

This is a big limitation of the theory because the theory is valid only near equilibrium and hyperbolicity exists only in some small domain of the configuration space near equilibrium. Notice that f_N given by (12) is not always positive!

5 ET Beyond the Monatomic Gas: Polyatomic Gas

The previous ET theory, being strictly connected with the kinetic theory, suffers from nearly the same limitations as the Boltzmann equation.

In the case of polyatomic gases, on the other hand, the rotational and vibrational degrees of freedom of a molecule, which are not present in monatomic gases, come into play [20], and in the case of dense gases, as the average distance between the constituent molecules is finite, the interaction between the molecules cannot be neglected. From a mathematical standpoint, these effects are responsible for intrinsic changes in the structure of the system of field equations. Single hierarchy of field equations as in the case of monatomic gases is no longer valid. In particular, the internal specific energy is no longer related to the pressure in a simple way.

After several tentative theories, a satisfactory *14-field* ET theory for dense gases and for rarefied polyatomic ones, was recently developed by Arima, Taniguchi, Ruggeri and Sugiyama [21]. This theory adopts two parallel hierarchies (binary hierarchy) for the independent fields: the mass density, the velocity, the internal energy, the shear stress, the dynamic pressure and the heat flux. One hierarchy consists of balance equations for the mass density, the momentum density and the momentum flux (*momentum-like* hierarchy), and the other one consists of balance equations for the energy density and the energy flux (*energy-like* hierarchy):

$$
\begin{aligned}
&\partial_t F + \partial_i F_i = 0, \\
&\partial_t F_{k_1} + \partial_i F_{ik_1} = 0, \\
&\partial_t F_{k_1k_2} + \partial_i F_{ik_1k_2} = P_{k_1k_2}, \qquad && \partial_t G_{kk} + \partial_i G_{ikk} = 0, \\
& && \partial_t G_{kkk_1} + \partial_i G_{kkik_1} = Q_{kkk_1}.
\end{aligned}
\tag{13}
$$

These hierarchies cannot merge with each other in contrast to the case of rarefied monatomic gases because the specific internal energy (the intrinsic part of the energy density) is no longer related to the pressure (one of the intrinsic parts of the momentum flux).

By means of the closure procedure of the ET theory, the constitutive equations are determined explicitly by the thermal and caloric equations of state. For example, let us consider the particular case of rarefied polyatomic gases with the thermal and caloric equations of state given by (polytropic gas)

$$
p = \frac{k_B}{m}\rho T \quad \text{and} \quad \varepsilon = \frac{D}{2}\frac{k_B}{m}T, \qquad (D = 3 + f^i)
\tag{14}
$$

where m is the atomic mass, T the absolute temperature, and the constant D is related to the degrees of freedom of a molecule given by the sum of the space dimension 3 for the translational motion and the contribution from the internal degrees of freedom $f^i (\geq 0)$. For monatomic gases, $D = 3$ (see $(6)_1$).

Concerning the kinetic counterpart, a crucial step towards the development of the theory of rarefied polyatomic gases was made by Borgnakke and Larsen [22]. The distribution function is assumed to depend on an additional continuous variable representing the energy of the internal modes of a molecule in order to take into account the exchange of energy (other than translational one) in binary collisions. This model was initially used for Monte Carlo simulations of polyatomic gases, and later it was applied to the derivation of the generalized Boltzmann equation by Bourgat, Desvillettes, Le Tallec and Perthame [23].

As a consequence of the introduction of one additional parameter I, the velocity distribution function $f(t, \mathbf{x}, \boldsymbol{c}, I)$ is defined on the extended domain $[0, \infty) \times R^3 \times R^3 \times [0, \infty)$. Its rate of change is determined by the Boltzmann equation which has the same form as the one of monatomic gases (1) but the collision integral $Q(f)$ takes into account the influence of the internal degrees of freedom through the collisional cross section.

Pavić, Ruggeri and Simić proved [24] [1] that, by means of the MEP, the kinetic model for rarefied polyatomic gases presented in [22] and [23] yields appropriate macroscopic balance laws. This is a natural generalization of the classical procedure

[1] There are some typos in the paper [24] that were corrected in the Chapter 12 of the book [12].

of MEP from monatomic gases to polyatomic gases. They considered the case of 14 moments, and showed the complete agreement with the binary hierarchy (21). The moments are defined by

$$\begin{pmatrix} F \\ F_{i_1} \\ F_{i_1 i_2} \end{pmatrix} = \int_{\mathbb{R}^3} \int_0^\infty m \begin{pmatrix} 1 \\ c_{i_1} \\ c_{i_1} c_{i_2} \end{pmatrix} f(t, \mathbf{x}, \boldsymbol{c}, I)\, \varphi(I)\, dI\, d\boldsymbol{c},$$

$$\begin{pmatrix} G_{pp} \\ G_{ppk_1} \end{pmatrix} = \int_{\mathbb{R}^3} \int_0^\infty m \begin{pmatrix} c^2 + 2\frac{I}{m} \\ (c^2 + 2\frac{I}{m})\, c_{k_1} \end{pmatrix} f(t, \mathbf{x}, \boldsymbol{c}, I)\, \varphi(I)\, dI\, d\boldsymbol{c},$$

$$\begin{pmatrix} P_{k_1 k_2} \\ Q_{kkk_j} \end{pmatrix} = \int_{\mathbb{R}^3} \int_0^\infty m \begin{pmatrix} c_{k_1} c_{k_2} \\ (c^2 + 2\frac{I}{m})\, c_{k_1} \end{pmatrix} Q\, \varphi(I)\, dI\, d\boldsymbol{c}.$$

The weighting function $\varphi(I)$ is determined in such a way that it recovers the caloric equation of state in equilibrium for polyatomic gases. It can be shown that $\varphi(I) = I^\alpha$ leads to an appropriate caloric equation for polytropic gas (14) provided that

$$\alpha = \frac{D-5}{2}. \tag{15}$$

Therefore, also for rarefied polyatomic gases, the three closure procedures (ET, MEP and Grad) give the same result as in the monatomic case!

5.1 ET of Polyatomic Rarefied Gases with Many Moments

In the case of many moments, by using similar notations as in (8)

$$F_A = \begin{cases} F & \text{for } A = 0 \\ F_{k_1 k_2 \cdots k_A} & \text{for } 1 \le A \le N, \end{cases} \qquad G_{llA'} = \begin{cases} G_{ll} & \text{for } A' = 0 \\ G_{llk_1 k_2 \cdots k_{A'}} & \text{for } 1 \le A' \le M, \end{cases}$$

$$P_A = \begin{cases} 0 & \text{for } A = 0, 1 \\ P_{k_1 k_2 \cdots k_A} & \text{for } 2 \le A \le N, \end{cases} \qquad Q_{llA'} = \begin{cases} 0 & \text{for } A' = 0 \\ Q_{llk_1 k_2 \cdots k_{A'}} & \text{for } 1 \le A' \le M, \end{cases}$$

the system of moments can be rewritten in the form of a binary hierarchy:

$$\partial_t F_A + \partial_i F_{iA} = P_A, \qquad (A = 0, \dots, N),$$

$$\partial_t G_{llA'} + \partial_i G_{illA'} = Q_{llA'}, \qquad (A' = 0, \dots, M),$$

with

$$F_A = m \int_{\mathbb{R}^3} \int_0^\infty c_A\, f\, \varphi(I)\, dI\, d\mathbf{c}, \quad F_{iA} = m \int_{\mathbb{R}^3} \int_0^\infty c_i c_A\, f\, \varphi(I)\, dI\, d\mathbf{c},$$

$$P_A = m \int_{\mathbb{R}^3} \int_0^\infty c_A\, Q\, \varphi(I)\, dI\, d\mathbf{c},$$

$$G_{llA'} = m \int_{\mathbb{R}^3} \int_0^\infty \left(c^2 + \frac{2I}{m}\right) c_{A'}\, f\, \varphi(I) dI d\mathbf{c},$$

$$G_{lliA'} = m \int_{\mathbb{R}^3} \int_0^\infty \left(c^2 + \frac{2I}{m}\right) c_i c_{A'}\, f\, \varphi(I) dI d\mathbf{c},$$

$$Q_{llA'} = m \int_{\mathbb{R}^3} \int_0^\infty \left(c^2 + \frac{2I}{m}\right) c_{A'}\, Q\, \varphi(I) dI d\mathbf{c},$$

$$c_A = \begin{cases} 1 & \text{for } A = 0 \\ c_{k_1} c_{k_2} \cdots c_{k_A} & \text{for } 1 \le A \le N, \end{cases} \qquad c_{A'} = \begin{cases} 1 & \text{for } A' = 0 \\ c_{k_1} c_{k_2} \cdots c_{k_{A'}} & \text{for } 1 \le A' \le M. \end{cases}$$

The variational problem, from which the distribution function $f_{(N,M)}$ maximizes the entropy

$$h = -k_B \int_{\mathbb{R}^3} \int_0^\infty f \log f\, \varphi(I)\, dI\, d\mathbf{c}, \tag{16}$$

is connected to the functional:

$$\begin{aligned} \mathcal{L}_{(N,M)}(f) = &- k_B \int_{\mathbb{R}^3} \int_0^\infty f \log f\, \varphi(I)\, dI\, d\mathbf{c} \\ &+ u'_A \left(F_A - m \int_{\mathbb{R}^3} \int_0^\infty c_A\, f\, \varphi(I)\, dI\, d\mathbf{c} \right) + \\ &+ v'_{A'} \left(G_{llA'} - m \int_{\mathbb{R}^3} \int_0^\infty \left(c^2 + \frac{2I}{m}\right) c_{A'}\, f\, \varphi(I) dI d\mathbf{c} \right), \end{aligned}$$

where u'_A and $v'_{A'}$ are the Lagrange multipliers:

$$u'_A = \begin{cases} u' & \text{for } A = 0 \\ u'_{k_1 k_2 \cdots k_A} & \text{for } 1 \le A \le N, \end{cases}, \qquad v'_{A'} = \begin{cases} v' & \text{for } A' = 0 \\ v'_{k_1 k_2 \cdots k_{A'}} & \text{for } 1 \le A' \le M. \end{cases}$$

The distribution function $f_{(N,M)}$ which maximizes the functional $\mathcal{L}_{(N,M)}$ is given by

$$f_{(N,M)} = \exp\left(-1 - \frac{m}{k} \chi_{(N,M)}\right), \qquad \chi_{(N,M)} = u'_A c_A + \left(c^2 + \frac{2I}{m}\right) v'_{A'} c_{A'}.$$

Then, the system may be rewritten as follows:

$$\begin{pmatrix} J^0_{AB} & J^1_{AB'} \\ J^1_{A'B} & J^2_{A'B'} \end{pmatrix} \partial_t \begin{pmatrix} u'_B \\ v'_{B'} \end{pmatrix} + \begin{pmatrix} J^0_{iAB} & J^1_{iAB'} \\ J^1_{iA'B} & J^2_{iA'B'} \end{pmatrix} \partial_i \begin{pmatrix} u'_B \\ v'_{B'} \end{pmatrix} = \begin{pmatrix} P_A \\ Q_{llA'} \end{pmatrix}, \tag{17}$$

where

$$J^0_{AB} = -\frac{m^2}{k} \int_{\mathbb{R}^3} \int_0^\infty f\, c_A c_B \varphi(I)\, dI d\mathbf{c},$$

$$J^0_{iAB} = -\frac{m^2}{k} \int_{\mathbb{R}^3} \int_0^\infty f\, c_i c_A c_B \varphi(I)\, dI d\mathbf{c},$$

$$J^1_{AB'} = -\frac{m^2}{k} \int_{\mathbb{R}^3} \int_0^\infty f\, c_A c_{B'} \left(c^2 + \frac{2I}{m}\right) \varphi(I)\, dI d\mathbf{c},$$

$$J^1_{iAB'} = -\frac{m^2}{k} \int_{\mathbb{R}^3} \int_0^\infty f\, c_i c_A c_{B'} \left(c^2 + \frac{2I}{m}\right) \varphi(I)\, dI d\mathbf{c},$$

$$J^2_{iA'B'} = -\frac{m^2}{k} \int_{\mathbb{R}^3} \int_0^\infty f\, c_i c_{A'} c_{B'} \left(c^2 + \frac{2I}{m}\right)^2 \varphi(I)\, dI d\mathbf{c}.$$

Also in this case the closed system is symmetric hyperbolic [12, 25], and the theory of monatomic gases is a singular limit of the theory of polyatomic gases [26].

In the present case we have in principle two index of truncation M and N. In the paper [25], the following two theorems are proved:

Theorem 1 *The differential system is Galilean invariant if and only if* $M \leq N - 1$.

Theorem 2 *If* $M < N - 1$, *all characteristic velocities are independent of the internal degrees of freedom* D *and coincide with the ones of* F*-hierarchy of monatomic gases with the truncation order* N.

The requirement that the system is Galilean invariant and the characteristic velocities are functions of D leads to the relationship $M = N - 1$. According with this result, the most interesting cases are the Euler system $N = 1$, $M = 0$ and the system with 14 fields that describes the ET of dissipative fluids in the presence of viscosity and heat conduction $N = 2$, $M = 1$.

Also in the case of polyatomic gases, we have the same problematic concerning the convergence of the integrals. In particular, not only the Grad theory of monatomic gases but also the theory with 14 moments are invalid in the case far from equilibrium!

Therefore as in the monatomic gas case, the distribution function obtained as a solution of the variational problem is expanded in the neighborhood of a local equilibrium state:

$$f \approx f^{(E)} \left[1 - \frac{m}{k} \left(\tilde{u}'_A c_A + \left(c^2 + \frac{2I}{m}\right) \tilde{v}'_{A'} c_{A'}\right)\right], \quad \tilde{u}'_A = u'_A - u'^E_A, \; \tilde{v}'_{A'} = v'_{A'} - v'^E_{A'},$$

where u'^E_A and $v'^E_{A'}$ are the main field components evaluated in the local equilibrium state. The equilibrium distribution function is given by [12, 24]

$$f^{(E)} = \frac{\rho}{m\,A(T)}\left(\frac{m}{2\pi k_B T}\right)^{3/2} \exp\left\{-\frac{1}{k_B T}\left(\frac{1}{2}mC^2 + I\right)\right\}, \tag{18}$$

where

$$A(T) = \int_0^\infty \exp\left(-\frac{I}{k_B T}\right)\varphi(I)dI. \tag{19}$$

This generalizes the Maxwellian distribution function in the case of polyatomic gases, which was obtained first with different arguments in [23]. In the polytropic case, (19) becomes

$$A(T) = (k_B T)^{1+\alpha}\Gamma(1+\alpha),$$

with α related with D through (15), and Γ denotes the Gamma function.

As an example we write down the differential closed system of 14 fields [12, 21]:

$$\begin{aligned}
&\dot{\rho} + \rho\frac{\partial v_k}{\partial x_k} = 0,\\
&\rho\dot{v}_i + \frac{\partial p}{\partial x_i} + \frac{\partial \Pi}{\partial x_i} - \frac{\partial \sigma_{\langle ij\rangle}}{\partial x_j} = 0,\\
&\dot{T} + \frac{2}{D\frac{k_B}{m}\rho}(p+\Pi)\frac{\partial v_k}{\partial x_k} - \frac{2}{D\frac{k_B}{m}\rho}\frac{\partial v_i}{\partial x_k}\sigma_{\langle ik\rangle} + \frac{2}{D\frac{k_B}{m}\rho}\frac{\partial q_k}{\partial x_k} = 0,\\
&\dot{\sigma}_{\langle ij\rangle} + \sigma_{\langle ij\rangle}\frac{\partial v_k}{\partial x_k} - 2\Pi\frac{\partial v_{\langle i}}{\partial x_{j\rangle}} + 2\frac{\partial v_{\langle i}}{\partial x_k}\sigma_{\langle j\rangle k\rangle} - \frac{4}{D+2}\frac{\partial q_{\langle i}}{\partial x_{j\rangle}} - 2p\frac{\partial v_{\langle i}}{\partial x_{j\rangle}} = -\frac{1}{\tau_\sigma}\sigma_{\langle ij\rangle},\\
&\dot{\Pi} + \frac{5D-6}{3D}\Pi\frac{\partial v_k}{\partial x_k} - \frac{2(D-3)}{3D}\frac{\partial v_{\langle i}}{\partial x_{k\rangle}}\sigma_{\langle ik\rangle} + \frac{4(D-3)}{3D(D+2)}\frac{\partial q_k}{\partial x_k} + \frac{2(D-3)}{3D}p\frac{\partial v_k}{\partial x_k} = -\frac{1}{\tau_\Pi}\Pi,\\
&\dot{q}_i + \frac{D+4}{D+2}q_i\frac{\partial v_k}{\partial x_k} + \frac{2}{D+2}q_k\frac{\partial v_k}{\partial x_i} + \frac{D+4}{D+2}q_k\frac{\partial v_i}{\partial x_k}\\
&\qquad + \frac{k_B}{m}T\frac{\partial \Pi}{\partial x_i} - \frac{k_B}{m}T\frac{\partial \sigma_{\langle ik\rangle}}{\partial x_k} + \Pi\left[-\frac{\frac{k_B}{m}T}{\rho}\frac{\partial \rho}{\partial x_i} + \frac{D+2}{2}\frac{k_B}{m}\frac{\partial T}{\partial x_i} - \frac{1}{\rho}\frac{\partial \Pi}{\partial x_i} + \frac{1}{\rho}\frac{\partial \sigma_{\langle ik\rangle}}{\partial x_k}\right]\\
&\qquad - \sigma_{\langle ik\rangle}\left[-\frac{\frac{k_B}{m}T}{\rho}\frac{\partial \rho}{\partial x_k} + \frac{D+2}{2}\frac{k_B}{m}\frac{\partial T}{\partial x_k} - \frac{1}{\rho}\frac{\partial \Pi}{\partial x_k} + \frac{1}{\rho}\frac{\partial \sigma_{\langle pk\rangle}}{\partial x_p}\right] + \frac{D+2}{2}\left(\frac{k_B}{m}\right)^2\rho T\frac{\partial T}{\partial x_i}\\
&\qquad = -\frac{1}{\tau_q}q_i,
\end{aligned} \tag{20}$$

where τ_σ, τ_Π and τ_q are relaxation times. In the present case the thermal and caloric equations of state are given by (14) and the dot indicate the material derivative:

$$\dot{} = \frac{\partial}{\partial t} + v_i \frac{\partial}{\partial x_i}.$$

If we apply the so-called *Maxwellian iteration* [27] (a sort of Chapman-Enskog formal expansion with respect the relaxation times) then $(20)_{3,4}$ converges to the Navier-Stokes constitutive equations, while $(20)_5$ reduces to the Fourier law [12]. For this reason the relaxations times τ_σ, τ_Π, and τ_q are connected, respectively, with the shear viscosity, bulk viscosity, and heat conductivity. We conclude that the Navier-Stokes-Fourier parabolic system of TIP is an approximation of the previous hyperbolic system when the relaxation times are small. The reader who is interested in how the usual constitutive equations (Navier-Stokes', Fourier's, Fick's, Darcy's) are approximated from the hyperbolic balance laws when some relaxation times are negligible can read the paper [28].

A relativistic theory with 14 fields was recently given by Pennisi and Ruggeri [29].

6 The 6-Moment Case and Non-Linear Closure

The 14-field theory gives us a complete phenomenological model but its differential system is rather complex and the closure is in any way limited within near equilibrium. Let us consider now a simplified theory (ET$_6$) with 6 independent field-variables (ρ, v_i, T, Π). This simplified theory preserves the main physical properties of the more complex theory of 14 variables, in particular, when the bulk viscosity plays more important role than the shear viscosity and the heat conductivity. ET$_6$ has another advantage to offer us a more affordable hyperbolic partial differential system. In fact, it is the simplest system that takes into account a dissipation mechanism after the Euler system of perfect fluids. In the present case we have

$$\begin{aligned}
&\frac{\partial F}{\partial t} + \frac{\partial F_i}{\partial x_i} = 0, \\
&\frac{\partial F_j}{\partial t} + \frac{\partial F_{ji}}{\partial x_i} = 0, \\
&\frac{\partial F_{ll}}{\partial t} + \frac{\partial F_{lli}}{\partial x_i} = P_{ll}, \qquad \frac{\partial G_{ll}}{\partial t} + \frac{\partial G_{lli}}{\partial x_i} = 0,
\end{aligned} \tag{21}$$

where $(21)_{1,2,4}$ represent the conservation laws of mass, momentum and energy provided that $F = \rho$, $F_i = \rho v_i$, $F_{ij} = \rho v_i v_j + (p + \Pi)\delta_{ij}$, $G_{ll} = \rho v_l v_l + 2\rho\varepsilon$, and $G_{lli} = (\rho v_l v_l + 2\rho\varepsilon + 2p + 2\Pi)v_i$ with p and ε being, respectively, the pressure and the specific internal energy. The phenomenological ET$_6$ was studied in the papers [12, 30, 31].

In the molecular approach we have

$$\begin{pmatrix} F \\ F_i \\ F_{ll} \end{pmatrix} = \begin{pmatrix} \rho \\ \rho v_i \\ \rho v^2 + 3(p + \Pi) \end{pmatrix} = \int_{\mathbb{R}^3} \int_0^\infty m \begin{pmatrix} 1 \\ c_i \\ c^2 \end{pmatrix} f \ \varphi(I)\, dI\, d\mathbf{c} \tag{22}$$

and

$$G_{ll} = \rho v^2 + 2\rho\varepsilon = \int_{\mathbb{R}^3} \int_0^\infty m(c^2 + 2I/m) f \ \varphi(I)\, dI\, d\mathbf{c}, \tag{23}$$

while the production term is given by

$$P_{ll} = m \int_{\mathbb{R}^3} \int_0^\infty c^2 Q \ \varphi(I)\, dI\, d\mathbf{c}. \tag{24}$$

Note that the internal energy density can be divided into the translational part ε_K and the part of the internal degrees of freedom ε_I:

$$\begin{aligned} \rho\varepsilon_K &= \int_{\mathbb{R}^3} \int_0^\infty \frac{1}{2} m C^2 f(t, \mathbf{x}, \mathbf{C}, I)\varphi(I)\, dI\, d\mathbf{C}, \\ \rho\varepsilon_I &= \int_{\mathbb{R}^3} \int_0^\infty I f(t, \mathbf{x}, \mathbf{C}, I)\varphi(I)\, dI\, d\mathbf{C}, \end{aligned} \tag{25}$$

where we have introduced the peculiar velocity:

$$\mathbf{C} \equiv (C_i), \qquad C_i = c_i - v_i. \tag{26}$$

6.1 Molecular ET$_6$ for a Polytropic Gas

The MEP in the nonlinear polytropic ET$_6$ gives the following distribution function f that maximizes the entropy (16) under the constraints (22), (23)

$$\begin{aligned} f_{\text{Poly}} = {} & \frac{\rho}{m\,(k_B T)^{1+\alpha} \Gamma(1+\alpha)} \left(\frac{m}{2\pi k_B T} \frac{1}{1 + \frac{\Pi}{p}} \right)^{3/2} \left(\frac{1}{1 - \frac{3}{2(1+\alpha)} \frac{\Pi}{p}} \right)^{1+\alpha} \\ & \times \exp\left\{ -\frac{1}{k_B T} \left(\frac{1}{2} m C^2 \left(\frac{1}{1 + \frac{\Pi}{p}} \right) + I \left(\frac{1}{1 - \frac{3}{2(1+\alpha)} \frac{\Pi}{p}} \right) \right) \right\}. \end{aligned}$$

The proof was given in the paper [32]. It is important to remark that the distribution function is non-linear in the dynamical pressure in contrast to the usual closure of moment theory in which the non-equilibrium distribution function is a linear perturbation of the equilibrium one. The closed system and the non-equilibrium entropy thus obtained [32] are exactly the same as the ones obtained by the phenomenological approach [12, 31].

6.2 Molecular ET_6 for a Non-Polytropic Gas

In the case of ideal non-polytropic gases the specific heat $c_v = d\varepsilon(T)/dT$ is, in general, a nonlinear function of the temperature and the caloric and thermal equations of state read:

$$\varepsilon \equiv \varepsilon(T), \qquad p = \frac{k_B}{m}\rho T. \tag{27}$$

As c_v can be measured by experiments as a function of the temperature T we can obtain the specific internal energy ε as

$$\varepsilon(T) = \frac{k_B}{m}\int_{T_0}^{T} \hat{c}_v(T')\,\mathrm{d}T', \tag{28}$$

where $\hat{c}_v = (m/k_B)c_v$ is the dimensionless specific heat and T_0 is an inessential reference temperature.

From (25), inserting the equilibrium distribution (18) and taking into account (19), we obtain the internal energy at equilibrium due to the internal motion:

$$\varepsilon_I(T) = \frac{k_B}{m}T^2\frac{d\log A(T)}{dT}, \qquad \varepsilon_I = \varepsilon - \varepsilon_K, \tag{29}$$

with ε_K given by

$$\varepsilon_K = \frac{3}{2}\frac{k_B}{m}T.$$

Therefore if we know the caloric equation of state (28) we know from $(29)_2$ ε_I and therefore from $(29)_1$ we can obtain $A(T)$:

$$A(T) = A_0 \exp\left(\frac{m}{k_B}\int_{T_0}^{T}\frac{\varepsilon_I(T')}{T'^2}dT'\right), \tag{30}$$

where A_0 and T_0 are inessential constants. As was observed in [33], the function A is, according to (19), the Laplace transform of φ:

$$A(T) = \mathcal{L}_u\left[\varphi(I)\right](s), \qquad s = \frac{1}{k_B T},$$

and then we can obtain the weighting function φ as the inverse Laplace transform of A:

$$\varphi(I) = \mathcal{L}_u^{-1}\left[A(T)\right](I), \qquad T = \frac{1}{k_B s}.$$

Bisi, Ruggeri and Spiga [34] proved the following theorem about the nonequilibrium distribution function:

Theorem 3 *The distribution function that maximizes the entropy* (16) *under the constraints* (22) *and* (23) *has the form:*

$$f_{\text{Non-Poly}} = \frac{\rho}{m A(\Theta)}\left(\frac{m}{2\pi k_B T}\,\frac{1}{1+\frac{\Pi}{p}}\right)^{3/2} \exp\left\{-\frac{1}{k_B T}\left(\frac{1}{2} m C^2 \left(\frac{1}{1+\frac{\Pi}{p}}\right) + I\frac{T}{\Theta}\right)\right\}, \tag{31}$$

where the nonequilibrium temperature Θ is related to the dynamical pressure Π and the temperature T through the relation:

$$\frac{\varepsilon_I(T) - \varepsilon_I(\Theta)}{\varepsilon_K(T)} = \frac{\Pi}{p},$$

and $A(\Theta)$ is the function (30) *evaluated at the temperature Θ:*

$$A(\Theta) = A_0 \exp\left(\frac{m}{k_B}\int_{T_0}^{\Theta} \frac{\varepsilon_I(T')}{T'^2} dT'\right).$$

All the moments are convergent and the bounded solutions satisfy the inequalities:

$$-1 < \frac{\Pi}{p} < \frac{\varepsilon_I(T)}{\varepsilon_K(T)}. \tag{32}$$

The distribution function is non-linear in the dynamical pressure and is positive.

The proof of this theorem is given in [34]. In the polytropic case the non-equilibrium distribution function (31) reduces to the expression (26).

6.3 Closure and Field Equations

Substituting (31) into the fluxes we obtain the closed system of ET$_6$:

$$\begin{aligned}
&\frac{\partial \rho}{\partial t}+\frac{\partial}{\partial x_i}(\rho v_i)=0,\\
&\frac{\partial (\rho v_j)}{\partial t}+\frac{\partial}{\partial x_i}\left[(p+\Pi)\delta_{ij}+\rho v_i v_j\right]=0,\\
&\frac{\partial}{\partial t}(2\rho\varepsilon+\rho v^2)+\frac{\partial}{\partial x_i}\left\{\left[2(p+\Pi)+2\rho\varepsilon+\rho v^2\right]v_i\right\}=0,\\
&\frac{\partial}{\partial t}\left[3(p+\Pi)-2\rho\varepsilon\right]+\frac{\partial}{\partial x_i}\left\{\left[3(p+\Pi)-2\rho\varepsilon\right]v_i\right\}=P_{ll}.
\end{aligned} \tag{33}$$

Concerning the production term P_{ll}, the main problem is that, in order to have explicit expression of the production (see (24)), we need a model for the collision term, which is, in general, not easy to obtain in the case of polyatomic gases. In the case of a BGK model we have:

$$P_{ll}=-3\frac{\Pi}{\tau}.$$

The system (33) with the thermal and caloric equations of state (27) is a closed system for the 6 unknowns (ρ, v_i, T, Π), provided that we know the collision term in (33)$_4$. These results are in perfect agreement with the results derived from the phenomenological theory [31]. The differential system is symmetric hyperbolic for any possible field and the bounded solutions satisfy automatically the inequalities (32).

6.4 Entropy Density and Main Field

Concerning the entropy density (16) it is possible to obtain the following explicit expression [34]:

$$\begin{aligned}
&k=\frac{h-h^{\mathsf{eq}}}{\rho}=\int_T^{\Theta}\frac{\varepsilon_I(T')}{T'^2}\,dT'+\frac{3}{2}\frac{k_B}{m}\log(1+Z)+\frac{\varepsilon_I(\Theta)}{\Theta}-\frac{\varepsilon_I(T)}{T}\\
&Z=\frac{\Pi}{p}=\frac{\varepsilon_I(T)-\varepsilon_I(\Theta)}{\varepsilon_K(T)},
\end{aligned} \tag{34}$$

where h^{eq} is the equilibrium entropy solution of the equilibrium Gibbs equation:

$$T d\left(\frac{h^{\mathsf{eq}}}{\rho}\right)=d\varepsilon-\frac{p}{\rho^2}d\rho.$$

The function k is a convex function and have a global maximum at the equilibrium state. It is also interesting to see that expressions (34) coincide with those obtained by the phenomenological ET approach [31].

By using the results given in [34] with some algebra, it is possible to prove that the Lagrange multipliers have the following expressions:

$$\begin{aligned}
\lambda &= -\frac{g}{T} + \int_T^{\Theta} \frac{\varepsilon_I(u)}{u^2} du - \frac{3}{2}\frac{k_B}{m} \ln(1+Z) + \frac{v^2}{2T}\frac{1}{1+Z}, \\
\lambda_i &= -\frac{v_i}{T}\frac{1}{1+Z}, \\
\mu_{ll} &= \frac{1}{2\Theta}, \\
\lambda_{ll} &= -\frac{1}{2T}\left(\frac{1}{1+Z} - \frac{T}{\Theta}\right).
\end{aligned} \tag{35}$$

According with the general theory, the Lagrange multipliers (35) coincide with the components of the main field for which the system (33) becomes to be symmetric hyperbolic in the form (17) [11, 12]. Notice that, in equilibrium where $\Pi = 0$ we have $Z = 0$ and $\Theta = T$ (see $(34)_2$), then the first five components of the main field (36) coincide with those obtained by Godunov for the Euler fluid [13]:

$$\lambda|_E = -\frac{1}{T}\left(g - \frac{v^2}{2}\right), \quad \lambda_i|_E = -\frac{v_i}{T}, \quad \mu_{ll}|_E = \frac{1}{2T},$$

while $\lambda_{ll}|_E = 0$ according to the fact that the Euler fluid is a *principal subsystem* of the 6-moment system. In the polytropic case,

$$\varepsilon_I(u) = \frac{D-3}{2}\frac{k_B}{m}u$$

and the expression (34) become the ones obtained in [32]:

$$k = \frac{k_B \rho}{2m} \ln\left((1+Z)^3 \left(1 - \frac{3}{D-3}Z\right)^{D-3}\right), \qquad Z = \frac{\Pi}{p},$$

and the main field reduces to

$$\begin{aligned}
&\lambda = -\frac{g}{T} + k + \frac{v^2}{2T}\frac{1}{1+Z}, \quad \lambda_i = -\frac{v_i}{T}\frac{1}{1+Z}, \quad \mu_{ll} = \frac{1}{2T}\left(1 - \frac{3}{D-3}Z\right)^{-1}, \\
&\lambda_{ll} = -\frac{1}{2T}\frac{D}{D-3}Z\,(1+Z)^{-1}\left(1 - \frac{3}{D-3}Z\right)^{-1}.
\end{aligned} \tag{36}$$

7 Comparison Between the ET6 Theory and Meixner's Theory

The field equations (33) with the linear production can be rewritten by using the material derivative in the following simple form [12]:

$$\begin{aligned}
&\dot{\rho} + \rho \,\mathrm{div}\, \mathbf{v} = 0, \\
&\rho \dot{v}_i + \frac{\partial}{\partial x_i}(p + \Pi) = 0, \\
&\rho \dot{\varepsilon} + (p + \Pi) \mathrm{div}\, \mathbf{v} = 0, \\
&\tau \dot{\Pi} + \left(\nu + \tau \frac{5D - 6}{3D} \Pi \right) \mathrm{div}\, \mathbf{v} = -\Pi,
\end{aligned} \tag{37}$$

where the bulk viscosity $\nu \propto D - 3$. When $D \to 3$ (monatomic gas) the previous system has the same solution as that of the Euler fluid provided $\Pi(\mathbf{x}, 0) = 0$ [12]. In [12, 30, 31] it was proved that the system (37) coincides with the well-known Meixner theory with one internal variable [35, 36] and the hidden variable is strictly related to the dynamical pressure Π.

Finally we note that, in the parabolic limit case where $\tau \to 0$, the system (37) reduces to a simplified version of Navier-Stokes system for compressible fluids:

$$\begin{aligned}
&\dot{\rho} + \rho \,\mathrm{div}\, \mathbf{v} = 0, \\
&\rho \dot{v}_i + \frac{\partial}{\partial x_i}(p + \Pi) = 0, \\
&\rho \dot{\varepsilon} + (p + \Pi) \mathrm{div}\, \mathbf{v} = 0, \\
&\nu \,\mathrm{div}\, \mathbf{v} = -\Pi,
\end{aligned}$$

and the qualitative analysis of this parabolic system was studied in same papers, e.g. in [37, 38].

8 Qualitative Analysis

In the general theory of hyperbolic conservation laws and hyperbolic-parabolic conservation laws, the existence of a strictly convex entropy function, which is a generalization of the physical entropy, is a basic condition for the well-posedness. However, in the general case, and even for arbitrarily small and smooth initial data, there is no global continuation for these smooth solutions, which may develop singularities, shocks, or blow up in finite time, see for instance [39].

On the other hand, in many physical examples, thanks to the interplay between the dissipation due to the source term and the hyperbolicity there exist global smooth solutions for a suitable set of initial data.

In physical dissipative case, the hyperbolic systems are of mixed type, some equations are conservation laws and other ones are real balance laws, i.e., we are in the case in which

$$\mathbf{u}_t + \partial_i \mathbf{F}^i(\mathbf{u}) = \mathbf{F}(\mathbf{u})$$

with

$$\mathbf{F}(\mathbf{u}) \equiv \begin{pmatrix} 0 \\ \mathbf{g}(\mathbf{u}) \end{pmatrix}; \qquad \mathbf{g} \in \mathbb{R}^{N-M}.$$

In this case the coupling condition, which is discovered for the first time by Kawashima and Shizuta (K-condition) [40] such that the dissipation in the second block has an effect also on the first block of equation, plays a very important role in this case for the global existence of smooth solutions.

In fact, if the system of balance law is endowed with a convex entropy law, and it is dissipative, then the K-condition becomes a sufficient condition for the existence of global smooth solutions provided that the initial data are sufficiently smooth (Hanouzet and Natalini [41], Wen-An Yong [42], Bianchini, Hanouzet and Natalini [43]):

Theorem 4 (Global Existence) *Assume that the system of balance laws is strictly dissipative and the K-condition is satisfied. Then there exists $\delta > 0$, such that, if $\|\mathbf{u}(x,0)\|_2 \leq \delta$, there is a unique global smooth solution, which verifies*

$$\mathbf{u} \in \mathcal{C}^0\left([0,\infty);\ H^2(\mathbb{R}) \cap \mathcal{C}^1\left([0,\infty); H^1(\mathbb{R})\right)\right).$$

Moreover Ruggeri and Serre [44] proved in the one-dimensional case that the constant states are stable:

Theorem 5 (Stability of Constant State) *Under natural hypotheses of strongly convex entropy, strict dissipativeness, genuine coupling and "zero mass" initial for the perturbation of the equilibrium variables, the constant solution stabilizes*

$$\|\mathbf{u}(t)\|_2 = O\left(t^{-1/2}\right).$$

Lou and Ruggeri [45] observed that the weaker K-condition in which we require the K-condition only for the right eigenvectors corresponding to genuine nonlinear is a necessary (but not sufficient) condition for the global existence of smooth solutions. In [12, 32, 46] it was proved that ET theories satisfy the hypothesis of the previous theorems and therefore there exist global solutions provided initial data are sufficiently smooth.

9 Shock Wave Structure in a Rarefied Polyatomic Gas

As an application of the previous thermodynamic models, let us consider a shock wave propagating in a polyatomic gas. The shock wave structure in a rarefied polyatomic gas is, under some conditions, quite different from the shock wave structure in a rarefied monatomic gas due to the presence of the microscopic internal modes in a polyatomic molecule such as the rotational and vibrational modes. For examples: (1) The shock wave thickness in a rarefied monatomic gas is of the order of the mean free path. On the other hand, owing to the slow relaxation process involving the internal modes, the thickness of a shock wave in a rarefied polyatomic gas is several orders larger than the mean free path. (2) As the Mach number increases from unity, the profile of the shock wave structure in a polyatomic rarefied gas changes from the nearly symmetric profile (Type A) to the asymmetric profile (Type B), and then changes further to the profile composed of thin and thick layers (Type C)

Schematic profiles of the mass density are shown in Fig. 1. Such change of the shock wave profile with the Mach number cannot be observed in a monatomic gas. In order to explain the shock wave structure in a rarefied polyatomic gas, there have been two well-known approaches. One was proposed by Bethe and Teller and the other is proposed by Gilbarg and Paolucci. Although the Bethe-Teller theory can describe qualitatively the shock wave structure of Type C, its theoretical basis is not clear enough. The Gilbarg-Paolucci theory, on the other hand, cannot explain asymmetric shock wave structure (Type B) nor thin layer (Type C).

Recently it was shown that the ET_{14} [47] and also ET_6 [48] theories can describe the shock wave structure of all Types A to C in a rarefied polyatomic gas. This new result indicates clearly the usefulness of the ET theory for the analysis of shock wave phenomena.

Other interesting and successful applications of RET in polyatomic gas show good agreement with experiments concerning the dispersion relation in the high frequency limit, and in the light scattering problem (see [12] and reference therein).

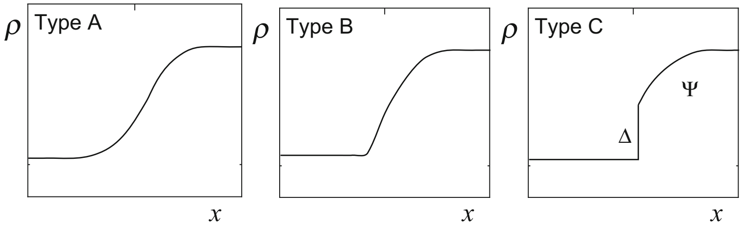

Fig. 1 Schematic representation of three types of the shock wave structure in a rarefied polyatomic gas, where ρ and x are the mass density and the position, respectively. As the Mach number increases from unity, the profile of the shock wave structure changes from Type A to Type B, and then to Type C that consists of the thin layer Φ and the thick layer Ψ

References

1. Müller, I., Ruggeri, T.: Rational Extended Thermodynamics, 2nd edn. Springer Tracts in Natural Philosophy, vol. 37. Springer, New York (1998)
2. Liu, I.-S., Müller, I.: Extended thermodynamics of classical and degenerate ideal gases. Arch. Rat. Mech. Anal. **83**, 285–332 (1983)
3. Liu, I.-S., Müller, I., Ruggeri, T.: Relativistic thermodynamics of gases. Ann. Phys. **169**, 191–219 (1986)
4. Müller, I., Ruggeri, T.: Extended Thermodynamics. Springer Tracts in Natural Philosophy, vol. 37. Springer, New York (1993)
5. Jaynes, E.T.: Information theory and statistical mechanics. Phys. Rev. **106**, 620–630 (1957); Jaynes, E.T.: Information theory and statistical mechanics II. Phys. Rev. **108**, 171–190 (1957)
6. Kapur, J.N.: Maximum Entropy Models in Science and Engineering. Wiley, New York (1989)
7. Kogan, M.N.: Rarefied Gas Dynamics. Plenum Press, New York (1969)
8. Grad, H.: On the kinetic theory of rarefied gases. Comm. Pure Appl. Math. **2**, 331–407 (1949)
9. Dreyer, W.: Maximization of the entropy in non-equilibrium. J. Phys. A: Math. Gen. **20**, 6505–6517 (1987)
10. Boillat, G., Ruggeri, T.: Moment equations in the kinetic theory of gases and wave velocities. Continuum Mech. Thermodyn. **9**, 205–212 (1997)
11. Boillat, G., Ruggeri, T.: Moment equations in the kinetic theory of gases and wave velocities. Continuum Mech. Thermodyn. **9**, 205–212 (1997)
12. Ruggeri, T., Sugiyama, M.: Rational Extended Thermodynamics beyond the Monatomic Gas. Springer, Cham-Heidelbergh-New York-Dorderecht-London (2015)
13. Godunov, S.K.: An interesting class of quasilinear systems. Sov. Math. 2, 947 (1961)
14. Boillat, G.: Sur l'existence et la recherche d'équations de conservation supplémentaires pour les systémes hyperboliques. C. R. Acad. Sci. Paris A **278**, 909–912 (1974)
15. Ruggeri, T., Strumia, A.: Main field and convex covariant density for quasi-linear hyperbolic systems. Relativistic fluid dynamics. Ann. Inst. H. Poincaré Sect. A **34**, 65–84 (1981)
16. Ruggeri, T.: Galilean invariance and entropy principle for systems of balance laws. The structure of extended thermodynamics. Continuum Mech. Thermodyn. **1**, 3–20 (1989)
17. Boillat, G., Ruggeri, T.: Hyperbolic principal subsystems: entropy convexity and subcharacteristic conditions. Arch. Rat. Mech. Anal. **137**, 305–320 (1997)
18. Levermore, C.D.: Moment closure hierarchies for kinetic theories. J. Stat. Phys. **83**, 1021 (1996)
19. Brini, F., Ruggeri, T.: Entropy principle for the moment systems of degree α associated to the Boltzmann equation. Critical derivatives and non controllable boundary data. Continuum Mech. Thermodyn. **14**, 165 (2002)
20. Kremer, G.M.: An Introduction to the Boltzmann Equation and Transport Processes in Gases. Springer, Berlin (2010)
21. Arima, T., Taniguchi, S., Ruggeri, T., Sugiyama, M.: Extended thermodynamics of dense gases. Continum Mech. Thermodyn. **24**, 271–292 (2012)
22. Borgnakke, C., Larsen, P.S.: Statistical collision model for Monte Carlo simulation of polyatomic gas mixture. J. Comput. Phys. **18**, 405–420 (1975)
23. Bourgat, J.-F., Desvillettes, L., Le Tallec, P., Perthame, B.: Microreversible collisions for polyatomic gases. Eur. J. Mech. B/Fluids **13**, 237–254 (1994)
24. Pavić, M., Ruggeri, T., Simić, S.: Maximum entropy principle for rarefied polyatomic gases. Physica A **392**, 1302–1317 (2013)
25. Arima, T., Mentrelli, A., Ruggeri, T.: Molecular extended thermodynamics of rarefied polyatomic gases and wave velocities for increasing number of moments. Ann. Phys. **345**, 111–140 (2014)
26. Arima, T., Ruggeri, T., Sugiyama, M., Taniguchi, S.: Monatomic gas as a singular limit of polyatomic gas in molecular extended thermodynamics with many moments. Ann. Phys. **372**, 83–109 (2016)

27. Ikenberry, E., Truesdell, C.: On the pressure and the flux of energy in a gas according to Maxwell's kinetic theory. J. Rat. Mech. Anal. **5**, 1–54 (1956)
28. Ruggeri, T.: Can constitutive relations be represented by non-local equations? Quart. Appl. Math. **70**, 597–611 (2012)
29. Pennisi, S., Ruggeri, T.: Relativistic extended thermodynamics of rarefied polyatomic gas. Ann. Phys. **377**, 414–445 (2017)
30. Arima, T., Taniguchi, S., Ruggeri, T., Sugiyama, M.: Extended thermodynamics of real gases with dynamic pressure: An extension of Meixner's theory. Phys. Lett. A **376**, 2799–2803 (2012)
31. Arima, T., Ruggeri, T., Sugiyama, M., Taniguchi, S.: Nonlinear extended thermodynamics of real gases with 6 fields. Int. J. Non-Linear Mech. **72**, 6–15 (2015)
32. Ruggeri, T.: Non-linear maximum entropy principle for a polyatomic gas subject to the dynamic pressure. Bull. Inst. Math. Acad. Sinica (New Series) **11**(1), 1–22 (2016)
33. Arima, T., Ruggeri, T., Sugiyama, M., Taniguchi, S.: Recent results on nonlinear extended thermodynamics of real gases with six fields Part I: general theory. Ric. Mat. **65**, 263–277 (2016)
34. Bisi, M., Ruggeri, T., Spiga, G.: Dynamical pressure in a polyatomic gas: Interplay between kinetic theory and extended thermodynamic. Kinetic and related models (KRM) **11**, 71–95 (2018)
35. Meixner, J.: Absorption und dispersion des schalles in gasen mit chemisch reagierenden und anregbaren komponenten. I. Teil. Ann. Physik **43**, 470 (1943)
36. Meixner, J.: Allgemeine theorie der schallabsorption in gasen und flussigkeiten unter berucksichtigung der transporterscheinungen. Acoustica **2**, 101 (1952)
37. Secchi, P.: Existence theorems for compressible viscous fluid having zero shear viscosity, Rend. Sem. Padova **70**, 73–102 (1983)
38. Frid, H., Shelukhin, V.: Vanishing shear viscosity in the equations of compressible fluids for the flows with the cylinder symmetry. SIAM J. Math. Anal. **31**(5), 1144–1156 (2000)
39. Dafermos, C.M.: Hyperbolic Conservation Laws in Continuum Physics. Grundlehren der mathematischen Wissenschaften, vol. 325, 3rd edn. Springer, Berlin Heidelberg (2010)
40. Kawashima, S., Shizuta, Y.: Systems of equations of hyperbolic-parabolic type with applications to the discrete Boltzmann equation. Hokkaido Math. J. **14**, 249–275 (1985)
41. Hanouzet, B., Natalini, R.: Global existence of smooth solutions for partially dissipative hyperbolic systems with a convex entropy. Arch. Rat. Mech. Anal. **169**, 89–117 (2003)
42. Yong, W.-A.: Entropy and global existence for hyperbolic balance laws. Arch. Rat. Mech. Anal. **172**(2), 247–266 (2004)
43. Bianchini, S., Hanouzet, B., Natalini, R.: Asymptotic behavior of smooth solutions for partially dissipative hyperbolic systems with a convex entropy. Comm. Pure Appl. Math., **60**, 1559–1622 (2007)
44. Ruggeri, T., Serre, D.: Stability of constant equilibrium state for dissipative balance laws system with a convex entropy. Quart. Appl. Math **62**(1), 163–179 (2004)
45. Lou, J., Ruggeri, T.: Acceleration waves and weak Shizuta-Kawashima condition. Suppl. Rend. Circ. Mat. Palermo. Non Linear Hyperbolic Fields and Waves. A tribute to Guy Boillat, Series II, Suppl. **78**, 187–200 (2006)
46. Ruggeri, T.: Entropy Principle and Global Existence of Smooth Solutions in Extended Thermodynamics. In: Hyperbolic Problems: Theory, Numerics, Applications, Vol. II, pp. 267–274. Yokohama Publishers Inc. (2006)
47. Taniguchi, S., Arima, T., Ruggeri, T., Sugiyama, M.: Thermodynamic theory of the shock wave structure in a rarefied polyatomic gas: Beyond the Bethe-Teller theory., Phys. Rev. E **89** 013025-1–013025-11 (2014)
48. Taniguchi, S., Arima, T., Ruggeri, T., Sugiyama, M.: Effect of dynamic pressure on the shock wave structure in a rarefied polyatomic gas. Phys. Fluids **26**, 016103-1–016103-15 (2014)

A Comparison of Two Settings for Stochastic Integration with Respect to Lévy Processes in Infinite Dimensions

Justin Cyr, Sisi Tang, and Roger Temam

Abstract We review two settings for stochastic integration with respect to infinite dimensional Lévy processes. We relate notions of stochastic integration with respect to square-integrable Lévy martingales, compound Poisson processes, Poisson random measures and compensated Poisson random measures. We use the Lévy-Khinchin decomposition to decompose stochastic integrals with respect to general, non-square-integrable Lévy processes into a Riemann integral and stochastic integrals with respect to a Wiener process, Poisson random measure and compensated Poisson random measure. Besides its intrinsic interest this review article is also meant as a step toward new studies in stochastic partial differential equations with Lévy noise.

1 Introduction

In this article we present in a synthetic form results on stochastic integration with respect to Lévy processes that are available in scattered form in the literature. In particular this article makes a synthesis between the presentation in the book [15] by Peszat and Zabczyk and the presentation in the book [10] by Ikeda and Watanabe. The presentation in the book by Peszat and Zabczyk is more intuitive, but the presentation in the book Ikeda and Watanabe is better technically suited for treating stochastic partial differential equations (SPDEs). More precisely, we would say that our article could help one who is familiar with SPDEs with Wiener noise transition to the Lévy noise case. The framework presented by Peszat and Zabczyk should be more intuitive than the Ikeda and Watanabe framework to one who is already acquainted with stochastic integration with respect to a Wiener process. In the Ikeda and Watanabe setting, the compensated Poisson random measures would probably seem abstract to someone who is only familiar with Wiener processes. It is also hard

J. Cyr · S. Tang · R. Temam (✉)
Department of Mathematics, Indiana University, Bloomington, IN, USA
e-mail: jrcyr@indiana.edu; sisitang@indiana.edu; temam@indiana.edu

E. Rocca et al. (eds.), *Trends in Applications of Mathematics to Mechanics*,
Springer INdAM Series 27, https://doi.org/10.1007/978-3-319-75940-1_14

to see in Ikeda and Watanabe's book how the compensated Poisson random measure is actually related to integration with respect to a Lévy process. Our article makes the argument that the Ikeda and Watanabe setting is better suited for the common SPDE tools, e.g. the Itô formula and Burkholder-Davis-Gundy (BDG) inequality, than is the Peszat and Zabczyk setting. Our article should help to bridge the gap from the intuitive setting of Peszat and Zabczyk, which provides many important results, to the setting of Ikeda and Watanabe, which is better technically suited for treating SPDEs.

Stochastic partial differential equations with Wiener noise have been studied extensively during the last four decades. In these models of PDEs with random forcing a stochastic term influences the system continuously in time. One may also wish to study stochastic partial differential equations in which stochastic terms also influence the system impulsively at random discrete times. Instead of Wiener noise, one should use noise arising from a stochastic process that has jump discontinuities. In this article we consider stochastic integration using Lévy processes as a source of noise with jump discontinuities. A Hilbert space-valued stochastic process $(L(t))_{t\geq 0}$ is called a Lévy process if L has independent, stationary increments, $t \mapsto L(t)$ is continuous in probability and $L(0) = 0$ almost surely (see Definition 2.1). For comparison, a Hilbert space-valued stochastic process $(W(t))_{t\geq 0}$ is a Wiener process if and only if W is a Lévy process with the additional property that the map $t \mapsto W(t)$ is continuous almost surely. We will review definitions and basic properties of Lévy processes and Wiener processes in Sect. 2. A Lévy process L need not be continuous a.s. in general, however every Lévy process admits a càdlàg version, i.e., a right-continuous version with left-hand limits (see Theorem 2.2). The main qualitative difference between a Wiener process and a general Lévy process is the possibility of jump discontinuities in a Lévy process. Almost surely, on every compact interval, a Lévy process may have finitely many jump discontinuities of size larger than any fixed positive number. The distinction between Wiener processes and general Lévy processes is expressed quantitatively by the Lévy-Khinchin decomposition. The decomposition (see Theorems 2.15 and 5.1 below) asserts that every Lévy process L can be decomposed in the form

$$L(t) = at + W(t) + P_0(t) + \sum_{n=1}^{\infty} \widehat{P}_n(t), \tag{1}$$

where a is a deterministic vector, $W, P_0, P_1, P_2, \ldots$ are independent Lévy processes, W is a Wiener process, for $n \geq 0$ each P_n is a type of pure jump Lévy process known as a compound Poisson process (see Definition 2.9) and $\widehat{P}_n(t) := P_n(t) - t \cdot \mathbf{E}[P_n(1)]$ is the associated compensated compound Poisson processes for $n \geq 1$ (see Definition 2.12).

In order to incorporate noise from a Lévy process into a stochastic partial differential equation one must employ some notion of stochastic integration with respect to a Lévy process. The main references are the books [10] and [15], which present different notions of stochastic integration with respect to Lévy processes based on the decomposition (1). In the book [15], Peszat and Zabczyk present a notion of stochastic integration with respect to square-integrable Lévy processes that are also martingales. This setting includes Wiener processes and the construction of the stochastic integral with respect to a general square-integrable Lévy martingale is much the same as it is for Wiener processes. We will review stochastic integration with respect to square-integrable Lévy martingales in Sect. 3. If $a = 0$ and $P_0 = 0$ in the Lévy-Khinchin decomposition (1), then L is a square-integrable Lévy martingale (see Proposition 2.11, Lemma 2.14 and Theorem 2.15). So one is left to define integration with respect to the remaining terms at and P_0 in (1). Stochastic integration with respect to the drift term at in (1) can be defined as a Bochner integral, almost surely. The compound Poisson process P_0 in (1) is not a square-integrable Lévy martingale, in general. Peszat and Zabczyk present a notion for stochastic integration with respect to a compound Poisson process P_0 using a localization argument wherein stochastic integration is performed up to stopping times before which P_0 agrees with a square-integrable compound Poisson process. We review Peszat and Zabczyk's presentation of stochastic integration with respect to compound Poisson processes in Sect. 6. In the book [10], Ikeda and Watanabe represent noise from the Wiener process W in (1) in exactly the same way as Peszat and Zabczyk. In contrast with Peszat and Zabczyk, Ikeda and Watanabe represent noise from the compound Poisson process P_0 in (1) using stochastic integration with respect to its associated Poisson random measure (see Definition 4.3). Ikeda and Watanabe represent noise from the process $\sum_{n=1}^{\infty} \widehat{P}_n$ in (1) using stochastic integration with respect to the compensated Poisson random measure associated to the process $\sum_{n=1}^{\infty} \widehat{P}_n$ (see Definition 4.8). In Sect. 2.1 we review the manner in which a Lévy process naturally gives rise to a Poisson random measure (also known as its jump measure, see Definition 2.16). In Sect. 4 we will review definitions and basic properties of Poisson random measures as well as Ikeda and Watanabe's presentation of stochastic integration with respect to Poisson random measures and compensated Poisson random measures (see Theorem 4.7).

Both of the settings for representing Lévy noise found in the books of Peszat and Zabczyk as well as Ikeda and Watanabe have been employed in models of stochastic partial differential equations with Lévy noise. For instance, see [7] and [15] itself for examples of stochastic partial differential equations that represent Lévy noise using the setting presented by Peszat and Zabczyk. See [1, 4, 14] for examples that represent Lévy noise using the setting presented by Ikeda and Watanabe. In order to compare the articles listed above, it is desirable to understand how the setting presented by Peszat and Zabczyk is related to the setting presented by Ikeda and Watanabe. This is one of our main motivations here. We show here that the setting for representing the Lévy noise presented by Peszat and Zabczyk can be converted to a special case of the setting presented by Ikeda and Watanabe (see equation (130)). Our more specific motivations are to apply two common tools in SPDEs, the Itô formula and Burkholder-Davis-Gundy inequality, to solutions of SDEs with Lévy noise in the setting presented by Peszat and Zabczyk. To illustrate the application of these tools we consider a simple SDE with Lévy noise in the

Peszat and Zabczyk framework:

$$\begin{cases} \mathrm{d}X = F\,\mathrm{d}t + \Psi\,\mathrm{d}M \\ X(0) = X_0. \end{cases} \tag{2}$$

On the right-hand side of equation (2), M is a martingale as well as a square-integrable Lévy process and Ψ belongs to the space of integrands for stochastic integration with respect to M. The space of integrands and the stochastic integral with respect to M will be defined in Sect. 3. Real-valued smooth functions of the solution X to (2) can be analyzed using the Itô formula, which is stated below. There are several equivalent ways to state the Itô formula. The most convenient for us is Theorem D.2 in [15].

Theorem 1.1 *Let $Y = N + A$ be an H-valued semimartingale, where N is an H-valued L^2-martingale and A has paths of finite variation. Let $\psi\colon H \to \mathbb{R}$ be a C^2 function such that ψ, $D\psi$ and $D^2\psi$ are uniformly continuous on bounded subsets of H. Then for each $t \geq 0$ we have*

$$\begin{aligned} \psi(Y(t)) =&\psi(Y(0))+\int_0^t (D\psi(Y(s-)), \mathrm{d}Y(s))_H + \frac{1}{2}\int_0^t D^2\psi(Y(s-))\,\mathrm{d}[[N,N]]_s^c \\ &+ \sum_{s\in(0,t]} \big(\Delta(\psi(Y(s)) - \big(D\psi(Y(s-)\big), \Delta Y(s)\big)_H\big) \end{aligned} \tag{3}$$

P-*a.s.*

On the right-hand side of (3), $\Delta Y(s) := Y(s) - Y(s-)$ denotes the jump of Y at time s and $[[N, N]]^c$ denotes the continuous part of the so-called tensor quadratic variation of N, which will be defined in Sect. 2.2. The solution X to (2) is of the form $X = N + A$ as in Theorem 1.1 with $N(t) = \int_0^t \Psi(s)\,\mathrm{d}M(s)$ and $A(t) = X_0 + \int_0^t F(s)\,\mathrm{d}s$. When applying the Itô formula to X we would like to simplify the right-hand side of (3) as explicitly as possible in terms of the coefficients F and Ψ in the original equation (2). This requires computing the jumps of X and raises the natural question

Question 1.2 What are the jumps of the stochastic integral $\big(\int_0^t \Psi(s)\,\mathrm{d}M(s)\big)_{t\geq 0}$?

Applying the Itô formula to X also requires expressing $[[N, N]]^c$ explicitly in terms of Ψ; what this entails will become more clear as we introduce additional background information in Sect. 2.2.

When making a priori estimates for SPDEs one is often tasked with estimating quantities of the form

$$\mathbf{E}\Big(\sup_{t\in[0,T]} \Big|\int_0^t \Psi(s)\,\mathrm{d}M(s)\Big|_H^p\Big),$$

where M and Ψ are still as in equation (2) and $1 \leq p < \infty$. The Burkholder-Davis-Gundy inequality asserts that this expectation is bounded by a constant times $\mathbf{E}\big[\int_0^{\cdot} \Psi(s)\,\mathrm{d}M(s)\big]_T^{p/2}$, where $\big[\int_0^{\cdot} \Psi(s)\,\mathrm{d}M(s)\big]$ denotes the quadratic variation of the stochastic integral $\big(\int_0^t \Psi(s)\,\mathrm{d}M(s)\big)_{t\geq 0}$ (see Sect. 2.2).

Question 1.3 What is the quadratic variation of $\big(\int_0^t \Psi(s)\,\mathrm{d}M(s)\big)_{t\geq 0}$?

The answers to Questions 1.2 and 1.3 are not explicit in the setting presented by Peszat and Zabczyk. However, the setting presented by Ikeda and Watanabe does provide an answer to these questions. We will address Questions 1.2 and 1.3 in Sect. 5.3.

This article is organized as follows. In Sect. 2 we recall probabilistic preliminaries. Fundamental properties and examples of Lévy processes are given in Sect. 2.1. In Sect. 2.2 we recall additional concepts from probability, such as martingales and their quadratic variation and angle bracket processes. In Sect. 3 we review Peszat and Zabczyk's presentation of stochastic integration with respect to a square-integrable Lévy martingale. In Sect. 3.1 we recall further properties of square-integrable Lévy martingales and introduce technical measurability assumptions. In Sect. 3.2 we review the construction of the stochastic integral with respect to a square-integrable Lévy martingale. In Sect. 4 we review Ikeda and Watanabe's presentation of stochastic integration with respect to Poisson random measures and compensated Poisson random measures. Background information on Poisson random measures is given in Sect. 4.1 and stochastic integration is treated in Sect. 4.2. Sections 2, 3 and 4 serve only to gather the definitions, notation and basic properties of stochastic integration presented by Peszat and Zabczyk as well as Ikeda and Watanabe that are required to compare the two settings in subsequent sections. In Sect. 5 we consider square-integrable Lévy martingales, i.e. L as in (1) with $a = 0$ and $P_0 = 0$, and compare Peszat and Zabczyk's presentation of stochastic integration to Ikeda and Watanabe's in this case. We begin in Sect. 5.1 by applying the Lévy-Khinchin decomposition to a square-integrable Lévy martingale. The heart of the comparison between the notions of stochastic integration presented by Peszat and Zabczyk versus Ikeda and Watanabe lies in Sect. 5.2. In that subsection we show that stochastic integration with respect to the process $\sum_{n=1}^{\infty} \widehat{P}_n$ in (1) as presented by Peszat and Zabczyk is a special case of stochastic integration with respect to the compensated Poisson random measure of $\sum_{n=1}^{\infty} \widehat{P}_n$ as presented by Ikeda and Watanabe (see Proposition 5.14). In Sect. 5.3 we summarize the relationship between the notions of stochastic integration presented by Peszat and Zabczyk versus Ikeda and Watanabe for square-integrable Lévy martingales. We also show that stochastic integration with respect to a square-integrable Lévy martingale, as presented by Peszat and Zabczyk, can be realized in the setting of stochastic integration with respect to Lévy noise presented in [10] (see Theorem 5.18). We consider the case of square-integrable Lévy martingales first in order to devote separate attention to reviewing Peszat and Zabczyk's presentation of stochastic integration with respect to a compound Poisson process P_0 in Sect. 6. We begin in Sect. 6.1 with the preliminary step of defining stochastic integration with respect to square-integrable compound Poisson processes (see Definition 6.7). In Sect. 6.2 we review the construction by localization of the stochastic integral with respect to a compound Poisson process P_0 as presented by Peszat and Zabczyk (see Definition 6.20). We adopt a more abstract framework for the construction of the stochastic integral with respect to P_0 by localization than do Peszat and Zabczyk. We need to use this abstract setting in order to address questions related to stochastic integration with respect to compound Poisson processes that are not treated by Peszat and Zabczyk. We will address such questions in Sect. 6 as they arise. In Sect. 6.3 we will compare the two notions of stochastic integration with respect to a square-integrable compound Poisson process given in Definitions 6.7 and 6.20 (see Proposition 6.30). Finally, in Sect. 6.4 we summarize the relationship between Peszat and Zabczyk's presentation and Ikeda and Watanabe's presentation of stochastic integration with Lévy noise. In equation (130) we show how to convert stochastic integrals with general, non-square-integrable, Lévy noise in the setting of Peszat and Zabczyk to the setting of Ikeda and Watanabe.

The framework presented by Ikeda and Watanabe that we further develop here is particularly suitable to study SPDEs with Lévy noise. Some applications will be given in [5] and in future works. Also in future works we will investigate the properties of SPDEs with Lévy noise as presented in this article as Markov processes, as well as the associated transition semigroups and generators. Some related remarks are made in Sect. 5.3, see [15].

2 Probabilistic Preliminaries

We now recall concepts from probability theory that will play major roles in the rest of the article. We begin by defining Lévy processes and introducing fundamental examples of Lévy processes. We then discuss Hilbert space-valued martingales and the notion of quadratic variation for such processes.

2.1 Lévy Processes

In this section $(\Omega, \mathscr{F}, \mathbf{P})$ is a probability space with expectation denoted by $\mathbf{E}$ and U is a real, separable Hilbert space with Borel σ-field denoted by $\mathscr{B}(U)$.

Definition 2.1 A U-valued *Lévy process* is a stochastic process $L = (L(t))_{t\geq 0}$ taking values in U that satisfies the following properties:

- (stationary increments) If $0 \leq s < t$, $0 \leq s' < t'$ and $t - s = t' - s'$, then $L(t) - L(s) \overset{\mathscr{D}}{=} L(t') - L(s')$. Here "$\overset{\mathscr{D}}{=}$" denotes equality in law and means that $\mathbf{P}[L(t) - L(s) \in \Gamma] = \mathbf{P}[L(t') - L(s') \in \Gamma]$ for every $\Gamma \in \mathscr{B}(U)$.

- (independent increments) For every sequence of points $0 \leq t_0 < t_1 < \cdots < t_n$, the random variables $L(t_1) - L(t_0), L(t_2) - L(t_1), \ldots, L(t_n) - L(t_{n-1})$ are independent; i.e., for all $\Gamma_1, \ldots, \Gamma_n \in \mathscr{B}(U)$ we have

$$\mathbf{P}\Big[\bigcap_{i=1}^{n}\{L(t_i) - L(t_{i-1}) \in \Gamma_i\}\Big] = \prod_{i=1}^{n}\mathbf{P}[L(t_i) - L(t_{i-1}) \in \Gamma_i].$$

- (stochastic continuity) For every $t_0 \geq 0$ one has $L(t) \to L(t_0)$ in probability (as U-valued random variables) as $t \to t_0$, $t > 0$; i.e., for every $\varepsilon > 0$ one has

$$\lim_{\substack{t\to t_0\\ t>0}} \mathbf{P}[|L(t) - L(t_0)|_U > \varepsilon] = 0.$$

- The process starts at $0 \in U$; i.e., $\mathbf{P}[L(0) = 0] = 1$.

A fundamental property of Lévy processes is that they admit càdlàg modifications. See Theorem 4.3 in [15] for a proof.

Theorem 2.2 *Every Lévy process L has a modification with càdlàg sample paths, i.e., there exists a Lévy process $\widetilde{L}$ such that $\mathbf{P}[L(t) = \widetilde{L}(t)] = 1$ for every $t \geq 0$ and for $\mathbf{P}$-a.e. $\omega \in \Omega$ the function $t \mapsto \widetilde{L}(\omega, t)$ is càdlàg from $[0, \infty) \to U$, i.e., this function is right continuous:*

$$\lim_{t\to t_0^+} \widetilde{L}(\omega, t) = \widetilde{L}(\omega, t_0) \qquad \text{for every } t_0 \geq 0$$

and has left-hand limits: $\lim_{t\to t_0^-} \widetilde{L}(\omega, t)$ exists for every $t_0 > 0$.

Below we recall the foundational examples of Lévy processes: the Wiener process, Poisson process, compound Poisson process, and compensated compound Poisson process.

Definition 2.3 An integrable U-valued mean-zero Lévy process W whose sample paths are continuous a.s. is called a *Wiener process*.

Although integrability and path continuity are the only extra conditions that distinguish Wiener processes from other Lévy processes a priori, there is much more that can be said about Wiener processes. Well-known basic properties of Wiener processes are summarized below; see Theorem 4.20 in [15]. This theorem guarantees that Definition 2.3 coincides with another commonly used definition of Wiener process, c.f. Definition 2.1.9 in [16], in which the stochastic continuity condition is replaced by a Gaussian condition on the distribution of increments.

Theorem 2.4 *Let W be a U-valued Wiener process. Then*

i) (square integrable) $\mathbf{E}|W(t)|_U^2 < \infty$ *for all* $t \geq 0$.
ii) (Gaussian) For all $t_1, \dots, t_n \geq 0$ *and* $x_1, \dots, x_n \in U$ *the random vector*

$$((W(t_1), x_1)_U, \dots, (W(t_n), x_n)_U)$$

has a mean-zero multivariate normal distribution on $\mathbb{R}^n$.

Remark 2.5 There is a bijective correspondence between the space $L_1^+(U)$ of bounded, symmetric, nonnegative, trace class (also called nuclear) linear operators on U and the laws of U-valued Wiener processes. Let $Q \in L_1^+(U)$. Then Q is a positive, compact operator. By the spectral theorem there exists an orthonormal basis (ONB) $(u_n)_{n=1}^\infty$ of U consisting of eigenvectors of Q with corresponding (nonnegative) eigenvalues $(\gamma_n)_{n=1}^\infty$. Let $(\beta_n)_{n=1}^\infty$ be a sequence of independent identically distributed (i.i.d.) standard real-valued Brownian motions and define

$$W(t) := \sum_{n=1}^{\infty} \sqrt{\gamma}_n \beta_n(t) u_n. \tag{4}$$

This series converges in $L^2(\Omega, \mathscr{F}, \mathbf{P}; U)$ (because $\sum_{n=1}^\infty \gamma_n = \mathrm{Tr}(Q) < \infty$), and it converges a.s. in the space $C([0, T]; U)$ (see Theorem 4.3 in [6] for a proof).

On the other hand, every Wiener process has this form—in the sense that given a Wiener process W, there exists a $Q \in L_1^+(U)$ such that (4) holds with $\beta_n(t) := \gamma_n^{-1/2} (W(t), u_n)_U$, which are i.i.d. standard Brownian motions. Furthermore, for $t_1, \dots, t_n \geq 0$ the mean-zero normally distributed random vector

$$((W(t_1), x_1)_U, \dots, (W(t_n), x_n)_U)$$

has covariance matrix $\Sigma = [t_i \wedge t_j \left(Qx_i, x_j\right)_U]_{i,j=1}^n$.

We now give prototypical examples of Lévy processes possessing jump discontinuities.

Definition 2.6 A *Poisson process* with *intensity* (or rate) $\lambda > 0$, is a real-valued Lévy process $\Pi = (\Pi(t),\ t \geq 0)$ such that $\Pi(t)$ has a Poisson distribution with mean λt for every $t \geq 0$; i.e.,

$$\mathbf{P}[\Pi(t) = k] = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \qquad \text{for each } k \in \mathbb{N} := \{0, 1, 2, \dots\}.$$

Proposition 2.8 below describes the structure of Poisson processes. In order to state this result we recall the exponential distribution.

Definition 2.7 For each $\lambda > 0$ the *exponential distribution with rate* λ, $\lambda > 0$, is the probability measure $\lambda e^{-\lambda x}\chi_{(0,\infty)}(x)\,\mathrm{d}x$ on $\mathbb{R}$. We denote the exponential distribution with rate λ by $\operatorname{Exp}(\lambda)$.

Proposition 2.8 *i) Let* $(X_n)_{n=1}^{\infty}$ *be a sequence of i.i.d.* $\operatorname{Exp}(\lambda)$ *random variables and define*

$$\Pi(t) := \max\left\{k \in \mathbb{N} : \sum_{j=1}^{k} X_j \leq t\right\}. \tag{5}$$

Then $\Pi(t)$ *is finite a.s. and defines a Poisson process with rate* λ.

ii) Conversely, if Π *is a Poisson process, then there exist i.i.d.* $\operatorname{Exp}(\lambda)$ *random variables* $(X_n)_{n=1}^{\infty}$ *such that* (5) *holds. Furthermore,* Π *only has jumps of size* 1, *i.e.,*

$$\mathbf{P}(\Pi(t) - \Pi(t-) \in \{0, 1\}) = 1, \qquad \text{for all } t \geq 0.$$

For a proof see Proposition 4.9 in [15]. We can think of a Poisson process as follows: imagine that a sequence of events is occurring (for instance, customers arriving at a queue) and that the times between consecutive events are i.i.d. $\operatorname{Exp}(\lambda)$ random variables. In this context, X_1 is the time of the first event, X_2 is the time between the first and second events, X_3 is the time between the second and third event, etc. The random variable $\Pi(t)$ counts the number of events that occur during the time interval $(0, t]$. The increment $\Pi(t) - \Pi(s)$ counts the number of events that occur in $(s, t]$.

We introduce the Hilbert space-valued generalization of the Poisson process next.

Definition 2.9 Let μ be a finite Borel measure on a Hilbert space U with $\mu(\{0\}) = 0$. A *compound Poisson process* (abbreviated CPP) with *Lévy measure* (or jump intensity measure) μ is a Lévy process P with càdlàg sample paths such that

$$\mathbf{P}[P(t) \in \Gamma] = e^{-\mu(U)t}\sum_{j=0}^{\infty}\frac{t^j}{j!}\mu^{*j}(\Gamma), \qquad \text{for all } t \geq 0,\ \Gamma \in \mathscr{B}(U).$$

In the definition above μ^{*j} denotes the convolution $\mu^{*j} := \mu * \mu * \cdots * \mu$, j times, for $j \geq 1$ and $\mu^{*0} := \delta_0$. Here we use δ_u to denote the probability measure concentrated at the point $u \in U$. Observe that a Poisson process with intensity $\lambda > 0$ is a compound Poisson process with Lévy measure $\mu := \lambda\delta_1$ on $U := \mathbb{R}$. Indeed, $\lambda\delta_1(U) = \lambda$ and $(\lambda\delta_1)^{*k} = \lambda^k\delta_k$.

The next theorem says that a compound Poisson process is a sum of a random number of i.i.d. random variables with law $\frac{1}{\mu(U)}\mu$ and the number of random variables in the sum is determined by a Poisson process. See Theorem 4.15 in [15] for a proof.

Theorem 2.10 *In the setting of Definition 2.9 let $\lambda := \mu(U)$. Then the following statements hold.*

i) Let $(Z_n)_{n=1}^{\infty}$ be i.i.d. U-valued random variables with law $\lambda^{-1}\mu$ and let Π be a Poisson process with intensity λ that is independent of $(Z_n)_{n=1}^{\infty}$. Then

$$P(t) := \sum_{j=1}^{\Pi(t)} Z_j \tag{6}$$

is a compound Poisson process with Lévy measure μ.

ii) Conversely, if P is a compound Poisson process with Lévy measure μ, then there exist i.i.d. U-valued random variables $(Z_n)_{n=1}^{\infty}$ with law $\lambda^{-1}\mu$ and an independent Poisson process Π such that (6) holds.

Integrability properties of compound Poisson processes are given below. See Proposition 4.18 in [15] for a proof.

Notation Let $(X(t))_{t\geq 0}$ be a stochastic process taking values in a Hilbert space U. If $\mathbf{E}|X(t)|_U < \infty$ for every $t \geq 0$, then we say that X is *integrable*. Similarly, if $\mathbf{E}|X(t)|_U^2 < \infty$ for every $t \geq 0$, then we say that X is *square-integrable*.

Proposition 2.11 *Let P be a compound Poisson process with Lévy measure μ. Then*

i) P is integrable if and only if $\int_U |y|_U \, \mathrm{d}\mu(y) < \infty$. In this case

$$\mathbf{E}P(t) = t \int_U y \, \mathrm{d}\mu(y).$$

ii) P is square-integrable if and only if

$$\int_U |y|^2 \, \mathrm{d}\mu(y) < \infty. \tag{7}$$

Definition 2.12 If P is an integrable CPP with Lévy measure μ, then we can define $\widehat{P}(t) := P(t) - \mathbf{E}P(t) = P(t) - t\int_U y \, \mathrm{d}\mu(y)$. The process $\widehat{P}$ is called a *compensated compound Poisson process* (abbreviated CCPP) and satisfies $\mathbf{E}[\widehat{P}(t)] = 0$ for every $t \geq 0$. Since each CCPP $\widehat{P}$ is not constant between its jump times (except in the trivial case where $\mu = 0$ and there are no jumps), $\widehat{P}$ is not itself a CPP. Instead, a CCPP is a different type of Lévy process with jump discontinuities that changes linearly as a function of time between its jumps. Note that $\widehat{P}$ is square-integrable if and only if (7) holds.

In Theorem 2.15 below we recall the Lévy-Khinchin decomposition, which says that every Lévy process is a sum of a deterministic linear growth term, a Wiener process, a compound Poisson process and compensated compound Poisson processes. In order to state this result we must first describe how a Lévy process gives rise to the Lévy measures of its compound Poisson process parts. Let L be a U-valued Lévy process. In what follows we use χ_A to denote the indicator function of a set A, i.e., $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ if $x \notin A$. Let A be a Borel subset of U that is *separated from 0*, i.e., $0 \notin \overline{A}$. Define the $\mathbb{N}$-valued stochastic process

$$\pi_A(t) := \sum_{s\in(0,t]} \chi_A(\varDelta L(s)) = \#\{s \in (0,t] : \varDelta L(s) \in A\}, \qquad \text{for } t > 0, \tag{8}$$

where $\varDelta L(s) := L(s)-L(s-)$ is the jump process of L. The fact that A is separated from 0 and L is càdlàg implies that $\pi_A(t) < \infty$ a.s. for each t. Here is a sketch of the idea (cf. Lemma 2.3.4 in [2]): if $\pi_A(t') = \infty$, then by compactness of $[0,t']$ we can find $t \le t'$ and a sequence $t_n \to t$ such that $\varDelta L(t_n) \in A$. Let $B(0, 2\varepsilon_0) \subseteq A^c$, then we can find $s_n < t_n$ with $(t_n - s_n) \to 0$ such that $|L(t_n) - L(s_n)|_U > \varepsilon_0$. This means that L has a discontinuity of the second kind at t (either left or right hand limit does not exist, depending on whether $t_n \downarrow t$ or $t_n \uparrow t$, along a subsequence. Since L has càdlàg paths a.s. we conclude that $\mathbf{P}[\pi_A(t') = \infty] = 0$. It turns out that $(\pi_A(t))_{t\ge0}$ is a Poisson process (see Proposition 4.9 (iv) in [15]). Let us denote its intensity by $\nu(A)$, i.e.,

$$\begin{aligned}\nu(A) &:= \mathbf{E}[\pi_A(1)] = \mathbf{E}[\#\{s \in (0,1] : \varDelta L(s) \in A\}] \\ &= \tfrac{1}{t}\mathbf{E}[\pi_A(t)] = \tfrac{1}{t}\mathbf{E}[\#\{s \in (0,t] : \varDelta L(s) \in A\}] \qquad \text{for all } t > 0.\end{aligned} \tag{9}$$

The formula $\nu(A) = \mathbf{E}\pi_A(1)$ still makes sense even if A is not separated from 0 but is still Borel measurable and does not contain 0. Using Tonelli's theorem we see that ν is countably additive, so ν is a Borel measure on $U \setminus \{0\}$.

Definition 2.13 The Borel measure ν on $U \setminus \{0\}$ constructed above is known as the *Lévy measure* of the Lévy process L.

The following additional properties of the Lévy measure are proved in [15] prior to Theorem 4.23.

Lemma 2.14 *Let L be a U-valued Lévy process with Lévy measure ν. Then ν satisfies*

$$\int_U (|y|_U^2 \wedge 1)\,\mathrm{d}\nu(y) < \infty. \tag{10}$$

Let $A \in \mathscr{B}(U)$ be separated from 0, *then*

$$L_A(t) := \sum_{s\in(0,t]} \chi_A(\varDelta L(s))\varDelta L(s)$$

is a compound Poisson process with Lévy measure $\nu|_A$.

In Lemma 2.14 and throughout this article we use $\nu|_A$ to denote the measure ν restricted to the σ-algebra of Borel subsets of A.

We are now able to state the Lévy-Khinchin decomposition. See Theorem 4.23 in [15] for a proof.

Theorem 2.15 *Let L be a U-valued Lévy process with Lévy measure ν. Given a sequence $(r_n)_{n=0}^{\infty}$ with $r_n \downarrow 0$ define $A_0 := \{y \in U : |y|_U \geq r_0\}$ and $A_n := \{y \in U : r_{n+1} \leq |y|_U < r_n\}$. Then the following statements hold.*

i) The compound Poisson processes $\left(L_{A_n}\right)_{n=0}^{\infty}$ are independent.
ii) There exists a vector $a \in U$ and a Wiener process W that is independent of $\left(L_{A_n}\right)_{n=0}^{\infty}$ such that

$$L(t) = at + W(t) + L_{A_0}(t) + \sum_{n=1}^{\infty} \widehat{L}_{A_n}(t) \tag{11}$$

and, with probability 1, the series on the right-hand side of (11) converges uniformly on compact subsets of $[0, \infty)$.

The processes $(\pi_A(t))_{t>0}$ defined in (8) are also of great importance. It is clear that for $0 < t < t'$ we have

$$\pi_A(t') - \pi_A(t) = \#\{s \in (t, t'] : \varDelta L(s) \in A\}.$$

There exists a unique random measure π on $(0, \infty) \times (U \setminus \{0\})$ with the property that

$$\pi((t, t'] \times A) = \pi_A(t') - \pi_A(t) \tag{12}$$

for all $0 < t < t'$ and every set $A \in \mathscr{B}(U \setminus \{0\})$ that is separated from zero, namely the random counting measure

$$\pi := \sum_{\substack{s>0 \\ \varDelta L(s)\neq 0}} \delta_{(s, \varDelta L(s))}. \tag{13}$$

See [11] for a proof.

Definition 2.16 The random measure π on $(0, \infty) \times (U \setminus \{0\})$ defined in (13) is called the *jump measure* of L.

The jump measure π plays an important role in the setting of Ikeda and Watanabe by representing the jump part of a Lévy process L in the theory of stochastic integration. We will give a general account of Ikeda and Watanabe's presentation of stochastic integration with Lévy noise in Sect. 4.

2.2 Martingales and Quadratic Variation

In this subsection we recall Hilbert space-valued martingales and the associated notion of quadratic variation. These notions appear in the Burkholder-Davis-Gundy inequality, which is frequently used to study SPDEs. In this work we are interested in stochastic processes formed by stochastic integration with respect to Lévy processes. In particular, in Sect. 5 we treat stochastic integration with respect to square-integrable Lévy martingales. The stochastic integral takes values in a real, separable Hilbert space H which may be different from the space U where the Lévy noise takes its values. As we will see in Theorem 3.12 and Theorem 4.7, these stochastic integrals are also square-integrable martingales. For this reason we restrict our treatment of quadratic variation to square-integrable Hilbert space-valued martingales. However, the notion of quadratic variation can be defined for more general processes known as semimartingales. We follow [13] as our main reference in this subsection. We begin by recalling the definitions of martingales and stopping times.

Fix a separable, real Hilbert space H and a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$. That is, $(\mathscr{F}_t)_{t\geq 0}$ is an increasing family of σ-fields on Ω that are all contained in $\mathscr{F}$.

Definition 2.17 An H-valued stochastic process $(M(t))_{t\geq 0}$ is *adapted* to the filtration $(\mathscr{F}_t)_{t\geq 0}$ if for every $t \geq 0$, $M(t)$ is a measurable function from $(\Omega, \mathscr{F}_t) \to (H, \mathscr{B}(H))$. The process $(M(t))_{t\geq 0}$ is called an *$\mathscr{F}_t$-martingale* (or just martingale when the filtration is clear) if it is adapted, integrable and satisfies the martingale property:

$$\mathbf{E}[M(t) \mid \mathscr{F}_s] = M(s) \quad \mathbf{P}\text{-a.s.} \qquad \text{for all } t \geq s \geq 0. \tag{14}$$

By the defining property of conditional expectation, the martingale property (14) is equivalent to the condition that for every $\Gamma \in \mathscr{F}_s$ we have

$$\int_\Gamma M(t)\,\mathrm{d}\mathbf{P} = \int_\Gamma M(s)\,\mathrm{d}\mathbf{P} \tag{15}$$

as H-valued Bochner integrals. If M is a real-valued, adapted, integrable process and if instead of equality in (14) and (15) we have the inequality $\geq$, then M is called a *submartingale*.

Remark 2.18 Let $1 \leq p < \infty$ and suppose that M is an H-valued $\mathscr{F}_t$-martingale such that $\mathbf{E}|M(t)|_H^p < \infty$ for every $t \geq 0$, then the real-valued process $|M(t)|_H^p$ is an $\mathscr{F}_t$-submartingale. This follows from Jensen's inequality for conditional expectation; see Theorem 3.35 in [15] for a proof.

As mentioned above, the stochastic integrals that we construct in Theorem 3.12 and Theorem 4.7 satisfy the integrability property in Remark 2.18 with $p = 2$, so this case will be our focus.

Definition 2.19 The space of right-continuous, H-valued $\mathscr{F}_t$-martingales M with the property that $\mathbf{E}|M(t)|_H^2 < \infty$ for every $t \geq 0$ is denoted by $\mathscr{M}^2(H)$; it is a Fréchet space under the seminorms $M \mapsto (\mathbf{E}|M(t)|_H^2)^{1/2}$ for $t \geq 0$. For each fixed $T \geq 0$ we denote by $\mathscr{M}_T^2(H)$ the space of restrictions of elements of $\mathscr{M}^2(H)$ to the interval $[0, T]$. For each $M \in \mathscr{M}^2(H)$ and $t \geq s \geq 0$ we have $\mathbf{E}|M(s)|_H^2 \leq \mathbf{E}|M(t)|_H^2$ by Remark 2.18. From this and the martingale property (14) it follows that $\mathscr{M}_T^2(H)$ is a Hilbert space which is isometrically isomorphic to $L^2(\Omega, \mathscr{F}_T, \mathbf{P}; H)$.

The notion of stopping time, defined below, will be used frequently.

Definition 2.20 A nonnegative random variable τ is said to be an *$\mathscr{F}_t$-stopping time* if $\{\tau \leq t\} \in \mathscr{F}_t$ for every $t \geq 0$. The prefix $\mathscr{F}_t$ is often omitted when there is no confusion.

The result below is the basis for the definition of quadratic variation. See Theorem 2.5 in [13] for a proof.

Theorem 2.21 *Let $M, N \in \mathscr{M}^2(H)$. For every $t \geq 0$ and every sequence $(\Pi^n)_{n=1}^\infty$ of increasing sequences $\Pi^n := \{0 = t_1^n < t_2^n < t_3^n \cdots\}$ in $[0, \infty)$ such that*

i) $\lim_{k\to\infty} t_k^n = \infty$, *for every n and*
ii) $\lim_{n\to\infty} \sup_k (t_{k+1}^n - t_k^n) = 0$,

the random variables

$$S_n(t, M, N) := \sum_{k=1}^{\infty} \big(M(t \wedge t_{k+1}^n) - M(t \wedge t_k^n), N(t \wedge t_{k+1}^n) - N(t \wedge t_k^n)\big)_H \tag{16}$$

converge in $L^1(\Omega, \mathscr{F}_t, \mathbf{P})$ as $n \to \infty$. Moreover, the limit does not depend on the sequence $(\Pi^n)_{n=1}^\infty$ and, as a process, the limit is $\mathscr{F}_t$-adapted and a.s. has right-continuous paths of finite variation (i.e., bounded variation on each compact interval).

Definition 2.22 Let $M, N \in \mathscr{M}^2(H)$. For each $t \geq 0$ we denote by $[M, N]_t$ the limit in $L^1(\Omega, \mathscr{F}_t, \mathbf{P})$ of the random variables $(S_n(t, M, N))_{n=1}^\infty$ from (16). The process $[M, N]$ is called the *mutual quadratic variation* of M and N. When $M = N$ we simply write $[M] := [M, M]$ and call this process the *quadratic variation* of M.

Our interest in the notion of quadratic variation comes from the Burkholder-Davis-Gundy (BDG) inequality which is stated next for elements of $\mathscr{M}^2(H)$.

Theorem 2.23 *For every $1 \leq p < \infty$ there exists a constant $C_p > 0$ such that for every càdlàg $M \in \mathscr{M}^2(H)$ with $M(0) = 0$ and for every $\mathscr{F}_t$-stopping time τ one has*

$$C_p^{-1}\mathbf{E}[M]_\tau^{p/2} \leq \mathbf{E} \sup_{t\in[0,\tau]} |M(t)|_H^p \leq C_p\mathbf{E}[M]_\tau^{p/2}, \tag{17}$$

with the understanding that each term appearing in the inequality is finite if and only if the others are finite.

For a proof see, e.g., [12]. The fact that the constant C_p does not depend on the martingale M or the stopping time τ is a key part of the conclusion. The upper bound in the right-hand inequality in (17) is used very frequently in the treatment of stochastic partial differential equations. Since the constant C_p does not depend on M or τ, one can employ the BDG inequality when making a priori estimates for stochastic partial differential equations.

It is often difficult to determine the quadratic variation of a martingale $M \in \mathscr{M}^2(H)$ by computing the limit in L^1 of the sequence $(S_n(t, M, M))_{n=1}^{\infty}$ in (16). Below we recall a direct sum decomposition of the space $\mathscr{M}_T^2(H)$ and the notion of angle bracket process which will aid in the computation of $[M]$ in the special cases that we are interested in, namely, where M is a process formed by stochastic integration. We state the decomposition theorem first; for a proof see Theorems 17.7 and 20.2 in [13].

Theorem 2.24 *For each $T \geq 0$ let $\mathscr{M}_T^{2,c}(H)$ denote the subspace of continuous martingales in $\mathscr{M}_T^2(H)$. Then $\mathscr{M}_T^{2,c}(H)$ is closed in $\mathscr{M}_T^2(H)$ and its orthogonal complement is the closure of the space of martingales in $\mathscr{M}_T^2(H)$ that start at* 0 *and have bounded variation on* $[0, T]$ *a.s.*

Definition 2.25 We denote by $\mathscr{M}_T^{2,d}(H)$ the closure of the space of bounded variation martingales in $\mathscr{M}_T^2(H)$ that start at 0. Theorem 2.24 asserts that every $M \in \mathscr{M}_T^2(H)$ can be written uniquely as

$$M = M^c + M^d, \tag{18}$$

where $M^c \in \mathscr{M}_T^{2,c}(H)$ and $M^d \in \mathscr{M}_T^{2,d}(H)$. We call M^c the *continuous part* of M and we call M^d the *purely discontinuous part* of M.

Before defining the angle bracket process of an H-valued L^2-martingale we recall the notion of predictability, which will be used extensively to define the stochastic integrals in Sects. 3 and 4.

Definition 2.26 Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$ be a filtered probability space and let $T > 0$. The σ-field on $\Omega \times [0, T]$ generated by sets of the form

$$A \times (s, t], \qquad A \in \mathscr{F}_s,\ 0 \leq s < t \leq T,$$

is called the *predictable σ-field* and is denoted by $\mathscr{P}_{[0,T]}$. Functions on $\Omega \times [0, T]$ that are $\mathscr{P}_{[0,T]}$-measurable are called *predictable*.

As we will see in Sect. 3, the predictability assumption is crucial for developing the entire theory of stochastic integration. At the moment we require predictability simply in order to define the angle bracket process of a martingale. The existence of the angle bracket process is obtained through the following result, which is an application of the Doob-Meyer decomposition theorem (see [13]).

Theorem 2.27 *Let $M, N \in \mathscr{M}^2(H)$. Then there exists a unique real-valued, finite variation, predictable process V with $V(0) = 0$ such that $(M, N)_H - V$ is a martingale.*

Definition 2.28 For $M, N \in \mathscr{M}^2(H)$ we denote by $\langle M, N\rangle$ the unique real-valued, finite variation, predictable process V starting from 0 produced by Theorem 2.27. When $M = N$ we simply write $\langle M\rangle := \langle M, M\rangle$ and call this process the *angle bracket* of M. The names Meyer process, predictable-variation process and first increasing process are also used to refer to $\langle M\rangle$.

The next result gives a relationship between the mutual quadratic variation $[M, N]$ and the angle bracket $\langle M, N\rangle$ between two processes $M, N \in \mathscr{M}^2(H)$.

Theorem 2.29 *Let $T > 0$ and let $M, N \in \mathscr{M}_T^2(H)$. For every $t \in [0, T]$ we have*

$$[M, N]_t = \langle M^c, N^c\rangle_t + \sum_{s\in(0,t]} (\varDelta M(s), \varDelta N(s))_H \quad \text{a.s.} \tag{19}$$

and the series on the right-hand side is summable a.s. In particular, for every $M \in \mathscr{M}_T^2(H)$ and $t \in [0, T]$ we have

$$[M]_t = [M^c]_t + [M^d]_t. \tag{20}$$

Proof See Theorem 20.5 and Corollary 18.9 in [13] for a proof of (19). Equation (20) follows from the bilinearity of mutual quadratic variation and formula (19) because the continuous part of M has no jumps and $(M^d)^c = 0$. □

We now recall the notion of tensor quadratic variation, which was required earlier in the statement of the Itô formula (Theorem 1.1). Before doing so we recall some definitions related to Hilbert-Schmidt operators.

Notation Let U and H be real, separable Hilbert spaces. We denote by $L_2(U, H)$ the space of Hilbert-Schmidt operators from U to H, i.e., the space of bounded linear operators $\varPhi : U \to H$ with the property that $\sum_{k=1}^{\infty} |\varPhi u_k|_H^2 < \infty$ for some ONB $(u_k)_{k=1}^{\infty}$ of U. When equipped with the inner product

$$(\varPhi, \varPsi)_{L_2(U,H)} := \sum_{k=1}^{\infty} (\varPhi u_k, \varPsi u_k)_H ,$$

the space $L_2(U, H)$ becomes a Hilbert space. Furthermore, the inner product defined above does not depend on the choice of the orthonormal basis of U. For vectors $u \in U$ and $h \in H$ we denote by $h \otimes u$ the linear map from $U \to H$ defined by $(h \otimes u)(v) := (v, u)_U\, h$. It is easy to see that $h \otimes u \in L_2(U, H)$.

Definition 2.30 Let $M \in \mathscr{M}^2(H)$ and let $(e_k)_{k=1}^{\infty}$ be an orthonormal basis of H. For each positive integer k the process $M^k(t) := (e_k, M(t))_H$ belongs to $\mathscr{M}^2(\mathbb{R})$. The *tensor quadratic variation* of M is the $L_2(H, H)$-valued process

$$[[M, M]]_t := \sum_{k,j=1}^{\infty} [M^k, M^j]_t (e_k \otimes e_j).$$

This sum converges in the space $L_2(H, H)$ for each $t \geq 0$ a.s. and does not depend on the choice of the orthonormal basis $(e_k)_{k=1}^{\infty}$ (see Theorem 26.11 in [13]). The continuous part of $[[M, M]]$ is defined to be the $L_2(H, H)$-valued process

$$[[M, M]]_t^c := \sum_{k,j=1}^{\infty} [(M^k)^c, (M^j)^c]_t (e_k \otimes e_j).$$

3 Stochastic Integration with Respect to Square-Integrable Lévy Martingales

In this section we review Peszat and Zabczyk's presentation of stochastic integration with respect to square-integrable Lévy processes that are also martingales. We begin by collecting additional properties of square-integrable Lévy martingales. The main reference for this section is Chapter 8 of the book [15] by Peszat and Zabczyk.

3.1 Square-Integrable Lévy Martingales

The following measurability property plays a crucial role in the construction of the stochastic integral in this section.

Definition 3.1 Let L be a stochastic process on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$ taking values in a real, separable Hilbert space U. We say that L is an *$\mathscr{F}_t$-Lévy process* if L is a Lévy process, L is adapted to $(\mathscr{F}_t)_{t\geq 0}$ and

$$L(t) - L(s) \text{ is independent of } \mathscr{F}_s \text{ for all } t \geq s \geq 0. \tag{21}$$

We will say that W is an $\mathscr{F}_t$-Wiener process if W is both a Wiener process and an $\mathscr{F}_t$-Lévy process. We will use the terms $\mathscr{F}_t$-Poisson process and $\mathscr{F}_t$-compound Poisson process in the same manner.

Remark 3.2 Let L be an integrable, mean-zero U-valued Lévy process on a probability space $(\Omega, \mathscr{F}, \mathbf{P})$. It is easy to see that L is a martingale with respect to its natural filtration $\widetilde{\mathscr{F}}_t := \sigma(L(s) : s \leq t)$ because L has independent increments. In the theory of stochastic integration one typically works with a filtration that is complete, i.e., $\mathscr{F}_0$ contains the **P**-null sets, and right-continuous, i.e., $\mathscr{F}_t = \bigcap_{s>t} \mathscr{F}_s$ for all $t \geq 0$. The natural filtration of L may not be complete or right-continuous, however it can always be enlarged to a filtration $(\mathscr{F}_t)_{t\geq 0}$ that is complete and right-continuous and in such a way that L is an $\mathscr{F}_t$-Lévy process. This is proved in Proposition 2.1.13 of [16] when L is a Wiener process but their argument applies whenever L is right-continuous and has independent increments.

In the remainder of Sect. 3 we work with a process M on $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$ taking values in a real, separable Hilbert space U that satisfies the following assumption.

Assumption 3.3 The U-valued stochastic process M is a square-integrable, mean-zero $\mathscr{F}_t$-Lévy process.

The independence condition (21) implies that M is an $\mathscr{F}_t$-martingale. Assumption 3.3 is stronger than assuming that M is both a square-integrable Lévy process and an $\mathscr{F}_t$-martingale. We have been using the term "square-integrable Lévy martingale" up until now just to delay stating the more technical condition (21) that is a part of Assumption 3.3. In the context of stochastic integration we will only consider square-integrable Lévy martingales and filtrations that also satisfy Assumption 3.3. Difficulties that are treated in Chapter 8 of [15] for general martingales do not arise for Lévy processes. So this section is a shorter and simpler version of Chapter 8 of [15] where the results are more involved in their statement and in their proof.

Theorem 3.4 *Let M satisfy Assumption 3.3. Then the following statements hold:*

i) There exists a bounded, linear, symmetric, positive operator $Q : U \to U$ such that

$$\mathbf{E}\big((M(t), x)_U \, (M(s), y)_U \big) = (t \wedge s)\, (Qx, y)_U \quad \text{for all } t, s \geq 0, \ \text{for all } x, y \in U. \tag{22}$$

ii) Furthermore, Q is of trace class and

$$\mathbf{E}\,(M(t), SM(t))_U = t\mathrm{Tr}(SQ), \qquad \text{for all } S \in L(U, U),$$

and in particular $\mathbf{E}|M(t)|_U^2 = t\mathrm{Tr}Q$.

In the statement of Theorem 3.4 and in what follows we denote the space of bounded linear operators from a Hilbert space U to a Hilbert space H by $L(U, H)$. See Theorem 4.44 in [15] for a proof of Theorem 3.4, which is valid even when M is just a square-integrable Lévy martingale.

Definition 3.5 Let M satisfy Assumption 3.3. The positive, trace class operator Q defined by (22) is called the *covariance operator* of M. Recall that we write $L_1^+(U)$ for the space of all positive, trace class operators on U. So, Theorem 3.4 says that $Q \in L_1^+(U)$.

3.2 Integration with Respect to Square-Integrable Lévy Martingales

We are now ready to define stochastic integration with respect to a square-integrable Lévy martingale M that satisfies Assumption 3.3. Fix another real separable Hilbert space H. The processes that we will integrate with respect to M will be $L(U, H)$-valued. We begin by defining the stochastic integral of certain step functions called simple processes.

Definition 3.6 We denote by $\mathscr{S}(U, H)$ the space of all $L(U, H)$-valued stochastic processes Ψ of the form

$$\Psi(\omega, s) = \sum_{j=0}^{m-1} \chi_{A_j}(\omega)\chi_{(t_j,t_{j+1}]}(s)\Phi_j, \tag{23}$$

where $0 = t_0 < t_1 < \cdots < t_m$, $A_j \in \mathscr{F}_{t_j}$, and $\Phi_j \in L(U, H)$. The elements of $\mathscr{S}(U, H)$ are called *simple processes.*

Remark 3.7 We emphasize that the space $\mathscr{S}(U, H)$ of simple processes depends on the filtration $(\mathscr{F}_t)_{t\geq 0}$ through the assumption that $A_j \in \mathscr{F}_{t_j}$. This condition means that each simple process $\Psi \in \mathscr{S}(U, H)$ is predictable in the sense of Definition 2.26. Predictability of simple processes is crucial in the proof of the isometric formula in Proposition 3.9 below, which is what allows the notion of stochastic integration to be extended beyond simple processes.

Definition 3.8 For a simple process $\Psi \in \mathscr{S}(U, H)$, we define the *stochastic integral* of Ψ with respect to M by

$$\int_0^t \Psi(s)\,\mathrm{d}M(s) := I_t^M(\Psi) := \sum_{j=0}^{m-1} \chi_{A_j}\Phi_j(M(t_{j+1}\wedge t) - M(t_j \wedge t)) \qquad \text{for all } t \geq 0.$$

Thus, the stochastic integral $I_t^M(\Psi)$ is an H-valued stochastic process.

The basic isometric formula, also called the Itô isometry, is stated below. Recall that $L_2(U, H)$ denotes the space of Hilbert-Schmidt operators from U to H.

Proposition 3.9 *For every $\Psi \in \mathscr{S}(U, H)$ and $t \geq 0$ we have*

$$\mathbf{E}|I_t^M(\Psi)|_H^2 = \mathbf{E}\int_0^t ||\Psi(s)Q^{1/2}||_{L_2(U,H)}^2\,\mathrm{d}s, \tag{24}$$

where Q is the covariance operator of M.

Note that the right-hand side of (24) is finite because $||\Phi Q^{1/2}||_{L_2(U,H)}^2 \leq ||\Phi||_{L(U,H)} \cdot \mathrm{Tr}(Q) < \infty$ for every $\Phi \in L(U, H)$. The independence condition (21) in Definition 3.1 plays a crucial role in the proof of Proposition 3.9. See Proposition 8.6 in [15] for a proof of Proposition 3.9.

Next we would like to extend the stochastic integration map I_T^M to a larger space of integrands that contains the simple processes. Before that we must construct the completion (or closure) of $\mathscr{S}(U, H)$.

Definition 3.10 Let $Q \in L_1^+(U)$. We define an inner product on the space $U_0 := Q^{1/2}(U)$ by

$$(x, y)_{U_0} := \big(Q^{-1/2}x, Q^{-1/2}y\big)_U \qquad \text{for all } x, y \in Q^{1/2}(U), \tag{25}$$

where $Q^{-1/2} \colon Q^{1/2}(U) \to \mathscr{N}(Q^{1/2})^{\perp}$ is the pseudoinverse of $Q^{1/2}$. The restriction $Q^{1/2}|_{\mathscr{N}(Q^{1/2})^{\perp}}$ is a bijection onto its image U_0 and $Q^{-1/2}$ is defined to be the inverse of this mapping. Although the subspace $Q^{1/2}(U)$ is typically not closed in U, the map $Q^{1/2}|_{\mathscr{N}(Q^{1/2})^{\perp}} \colon \mathscr{N}(Q^{1/2})^{\perp} \to U_0$ is an isometric isomorphism under the inner product in (25), so U_0 is complete. We will occasionally write $Q^{1/2}(U)$ instead of U_0 for the range of $Q^{1/2}$ endowed with the inner product defined in (25) in order to make the dependence on the operator Q explicit in the notation.

If $Q \in L_1^+(U)$, then Q is a symmetric compact operator, so there exists an orthonormal basis $(u_n)_{n=1}^{\infty}$ of U consisting of eigenvectors of Q. It is easy to see that the nonzero terms in $\big(Q^{1/2}u_n\big)_{n=1}^{\infty}$ form an orthonormal basis of U_0. Therefore, the Itô isometry in (24) can be restated in the equivalent form

$$\mathbf{E}|I_t^M(\Psi)|_H^2 = \mathbf{E}\int_0^t \big|\big|\Psi(s)\big|\big|_{L_2(U_0,H)}^2 \,\mathrm{d}s, \qquad \text{for all } \Psi \in \mathscr{S}(U, H),\ t \geq 0. \tag{26}$$

When $t = T$, the right-hand side of (26) is the norm squared in the space

$$\mathscr{X}_T := L^2\big(\Omega \times [0, T], \mathscr{F} \otimes \mathscr{B}([0, T]),\ \mathrm{d}\mathbf{P} \otimes \mathrm{d}t;\ L_2(U_0, H)\big).$$

We observed above that every $\Psi \in \mathscr{S}(U, H)$ belongs to the space $\mathscr{X}_T$. We will continue to use $\mathscr{S}(U, H)$ to denote the space of equivalence classes of simple processes in the space $\mathscr{X}_T$. Note that if $\Psi, \Phi \in \mathscr{S}(U, H)$ and $\Psi = \Phi$ in the space $\mathscr{X}_T$, it does not necessarily follow that Ψ and Φ are equal in $L(U, H)$, $\mathrm{d}\mathbf{P}\,\mathrm{d}t$-a.e. Instead, $\Psi = \Phi$ in $\mathscr{X}_T$ means only that $\Psi Q^{1/2} = \Phi Q^{1/2}$, $\mathrm{d}\mathbf{P}\,\mathrm{d}t$-a.e. That is, Ψ and Φ do not necessarily agree on all of U, but they do agree on the range of $Q^{1/2}$, $\mathrm{d}\mathbf{P}\,\mathrm{d}t$-a.e. Equation (26) shows that I_t^M is well-defined on the space $\mathscr{S}(U, H)$ viewed as equivalence classes in $\mathscr{X}_T$, i.e., if $\Psi, \Phi \in \mathscr{S}(U, H)$ and $\Psi Q^{1/2} = \Phi Q^{1/2}$, $\mathrm{d}\mathbf{P}\,\mathrm{d}t$-a.e., then $I_t^M(\Psi) = I_t^M(\Phi)$ in $L^2(\Omega; H)$. We can now extend $I_T^M \colon \mathscr{S}(U, H) \to L^2(\Omega; H)$ uniquely to an isometry on the closure of $\mathscr{S}(U, H)$ in the space $\mathscr{X}_T$. The resulting isometry is the stochastic integral with respect to M. Before stating the general properties of the stochastic integral with respect to M we pause to identify the closure of $\mathscr{S}(U, H)$ in the space $\mathscr{X}_T$; see Lemma 8.13 in [15] for a proof.

Lemma 3.11 *The closure of $\mathscr{S}(U, H)$ in the space $\mathscr{X}_T$ is the subspace of predictable processes in $\mathscr{X}_T$. In other words, $\mathscr{S}(U, H)$ is dense in the space*

$$\mathbf{L}^2_{U_0,T}(H) := L^2(\Omega \times [0, T], \mathscr{P}_{[0,T]}, \mathrm{d}\mathbf{P} \otimes \mathrm{d}t; L_2(U_0, H)), \tag{27}$$

where $\mathscr{P}_{[0,T]}$ is the σ-field of predictable sets (see Definition 2.26).

While M does not appear explicitly in the notation $\mathbf{L}^2_{U_0,T}(H)$, note that the space $\mathbf{L}^2_{U_0,T}(H)$ depends on the law of M through U_0. Note that the space $\mathbf{L}^2_{U_0,T}(H)$ of integrands for stochastic integration with respect to M depends on the filtration $(\mathscr{F}_t)_{t\geq 0}$ through the requirement of predictability. We gather the main facts about the stochastic integral with respect to M below.

Theorem 3.12 *Let M be a square-integrable, mean-zero, U-valued $\mathscr{F}_t$-Lévy process. Then the following statements hold.*

i) For every $t \in [0, T]$, $I^M_t : \mathbf{L}^2_{U_0,t}(H) \to L^2(\Omega; H)$ is an isometry, i.e.,

$$\mathbf{E}\left(I^M_t(\Psi), I^M_t(\Phi)\right)_H = \mathbf{E}\int_0^t (\Psi(s), \Phi(s))_{L_2(U_0,H)}\, \mathrm{d}s,$$

and $\mathbf{E}|I^M_t(\Psi)|^2_H = \mathbf{E}\int_0^t ||\Psi(s)||^2_{L_2(U_0,H)}\, \mathrm{d}s$ for all $\Psi, \Phi \in \mathbf{L}^2_{U_0,t}(H)$.

ii) For every $\Psi \in \mathbf{L}^2_{U_0,T}(H)$ the process $\left(I^M_t(\Psi)\right)_{t\in[0,T]}$ is a square-integrable H-valued martingale that begins at 0.

iii) For every $\Psi \in \mathbf{L}^2_{U_0,T}(H)$ the angle bracket of $\left(I^M_t(\Psi)\right)_{t\in[0,T]}$ is given by the formula

$$\left\langle I^M(\Psi)\right\rangle_t = \int_0^t ||\Psi(s)||^2_{L_2(U_0,H)}\, \mathrm{d}s. \tag{28}$$

iv) Let $A \in L(H, V)$ where V is a real, separable Hilbert space. For every $\Psi \in \mathbf{L}^2_{U_0,T}(H)$ we have $A\Psi \in \mathbf{L}^2_{U_0,T}(V)$ and $AI^M_t(\Psi) = I^M_t(A\Psi)$. That is, bounded operators can be passed inside the stochastic integral.

Proof Statement i) follows from the construction of I^M_t via Proposition 3.9. The remaining statements hold for simple processes and extend to the case of $\Psi \in \mathbf{L}^2_{U_0,T}(H)$ because $\mathscr{S}(U, H)$ is dense in $\mathbf{L}^2_{U_0,T}(H)$. □

As a first example we consider stochastic integration with respect to a Wiener process.

Example 3.13 Let $M = W$ be a U-valued Wiener process. Theorem 2.4 shows that W is square-integrable, so it has a covariance operator $Q \in L^+_1(U)$. In addition, since W is an $\mathscr{F}_t$-Wiener process with respect to its natural filtration, we see that W satisfies Assumption 3.3. We denote the space of integrands for stochastic integration with respect to W by $\mathbf{L}^2_{U_0,T}(H)$. Since W is continuous a.s., $\int_0^t \Psi(s)\, \mathrm{d}W(s)$ is continuous a.s. when $\Psi \in \mathscr{S}(U, H)$. In fact, the square-integrable H-valued martingale $\left(I_t^W(\Psi)\right)_{t\in[0,T]}$ is continuous a.s. for every integrand $\Psi \in \mathbf{L}^2_{U_0,T}(H)$; see [16]. In particular, Theorem 2.29 implies that the quadratic variation of $I^W(\Psi)$ is equal to its angle bracket. So (28) gives

$$[I^W(\Psi)]_t = \int_0^t ||\Psi(s)||^2_{L_2(U_0,H)}\,\mathrm{d}s. \tag{29}$$

The upper bound in the BDG inequality (Theorem 2.23) for stochastic integrals with respect to W takes the following form: for every $1 \leq p < \infty$ there exists a constant $C_p \in (0,\infty)$ such that for every $\mathscr{F}_t$-stopping time τ and every $\Psi \in \mathbf{L}^2_{U_0,T}(H)$ we have

$$\mathbf{E}\sup_{t\in[0,\tau]}\left|\int_0^t \Psi(s)\,\mathrm{d}W(s)\right|_H^p \leq C_p\mathbf{E}\Big(\int_0^\tau ||\Psi(s)||^2_{L_2(U_0,H)}\,\mathrm{d}s\Big)^{p/2}. \tag{30}$$

Example 3.14 Let P be a square-integrable U-valued compound Poisson process. Since P is integrable we can define the compensated compound Poisson process $\widehat{P}(t) := P(t) - \mathbf{E}[P(t)]$. It is clear that $\widehat{P}$ is a mean-zero Lévy process, so $\widehat{P}$ is an $\mathscr{F}_t$-compound Poisson process with respect to its natural filtration. Therefore, $\widehat{P}$ satisfies Assumption 3.3 and the stochastic integral $I^{\widehat{P}}$ can be defined in the sense above. We will take a closer look at stochastic integration with respect to $\widehat{P}$ in Sect. 5.2.

The result below will be used to define stochastic integration with respect to non-square-integrable compound Poisson processes by localization in Sect. 6.2.

Lemma 3.15 *Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$ be a filtered probability space, let M be a U-valued Lévy process satisfying Assumption 3.3 and let Q be the covariance operator of M. Let τ be an $\mathscr{F}_t$-stopping time such that $\mathbf{P}[\tau \leq T] = 1$. Then*

i) $\Psi \mapsto \chi_{[0,\tau]}\Psi$ is a continuous linear map sending $\mathbf{L}^2_{Q^{1/2}(U),T}(H) \to \mathbf{L}^2_{Q^{1/2}(U),T}(H)$.
ii) If τ takes finitely many values $\mathbf{P}$-a.s., then for every $\Psi \in \mathscr{S}(U,H)$ the processes $\chi_{[0,\tau]}\Psi$ and $\chi_{(\tau,T]}\Psi$ also belong to $\mathscr{S}(U,H)$.
iii) If $(\tau_n)_{n=1}^\infty$ is a sequence of stopping times such that $\tau_n \leq T$ and $\tau_n \downarrow \tau$, then for every $\Psi \in \mathbf{L}^2_{Q^{1/2}(U),T}(H)$ we have $\chi_{[0,\tau_n]}\Psi \to \chi_{[0,\tau]}\Psi$ in $\mathbf{L}^2_{Q^{1/2}(U),T}(H)$.
iv) For every $\Psi \in \mathscr{S}(U,H)$ and all $t \in [0,T]$ we have

$$\int_0^t \chi_{[0,\tau]}(s)\Psi(s)\,\mathrm{d}M(s) = \int_0^{t\wedge\tau} \Psi(s)M(s). \tag{31}$$

Proof *i)* Since τ is a stopping time the set $A := \{(\omega,s) \in \Omega\times[0,T] : s \leq \tau(\omega)\}$ is predictable. Indeed, we have

$$A^c = \bigcup_{q\in\mathbb{Q}\cap(0,T)} (\{\tau \leq q\}\times(q,T]) \qquad \in \mathscr{P}_{[0,T]}.$$

So $\chi_{[0,\tau(\omega)]}(s)\Psi(\omega,s) = \chi_A(\omega,s)\Psi(\omega,s)$ is predictable for every process Ψ in the space $\mathbf{L}^2_{Q^{1/2}(U),T}(H)$. It is clear that multiplication by $\chi_{[0,\tau]}$ is linear and bounded on $\mathbf{L}^2_{Q^{1/2}(U),T}(H)$ with norm less than or equal to 1.

ii) (cf. Lemma 2.3.9 in [16]) Let $\Psi \in \mathscr{S}(U,H)$ and write Ψ as in (23). Since τ takes finitely many values a.s. we can write $\tau(\omega) = \sum_{n=0}^{N} a_n \chi_{\{\tau=a_n\}}(\omega)$ for some constants $0 < a_0 < a_1 < \cdots < a_N \le T$. Since τ is a stopping time we have $\{\tau = a_n\} \in \mathscr{F}_{a_n}$ for each n. The process $\chi_{(\tau,T]}\Psi$ belongs to $\mathscr{S}(U,H)$ because

$$\chi_{(\tau,T]}(s)\Psi(\omega,s) = \sum_{j=0}^{m-1}\sum_{n=0}^{N} \chi_{A_j}(\omega)\chi_{(t_j,t_{j+1}]}(s)\chi_{\{\tau=a_n\}}(\omega)\chi_{(a_n,T]}(s)\Phi_j$$

$$= \sum_{j=0}^{m-1}\sum_{n=0}^{N} \chi_{A_j\cap\{\tau=a_n\}}(\omega)\chi_{(t_j\vee a_n,t_{j+1}\vee a_n]}(s)\Phi_j. \qquad (32)$$

We obtain the second line above using the fact that

$$(t_j,t_{j+1}] \cap (a_n,T] = \begin{cases} \varnothing & \text{if } t_{j+1} \le a_n \\ (a_n,t_{j+1}] & \text{if } t_j \le a_n \le t_{j+1} \\ (t_j,t_{j+1}] & \text{if } a_n \le t_j \end{cases}.$$

Since $A_j \cap \{\tau = a_n\} \in \mathscr{F}_{t_j\vee a_n}$ we see that $\chi_{(\tau,T]}\Psi \in \mathscr{S}(U,H)$. Next, we see that $\chi_{[0,\tau]}\Psi$ is the difference between two simple processes, namely $\chi_{[0,\tau]}\Psi = \Psi - \chi_{(\tau,T]}\Psi$, so $\chi_{[0,\tau]}\Psi \in \mathscr{S}(U,H)$ as well.

iii) Let $(\tau_n)_{n=1}^{\infty}$ be stopping times such that $\tau_n \le T$ and $\tau_n \downarrow \tau$ a.s. For $\Psi \in \mathbf{L}^2_{Q^{1/2}(U),T}(H)$ we have

$$\mathbf{E}\int_0^T \left|\left|\chi_{[0,\tau_n]}(s)\Psi(s) - \chi_{[0,\tau]}(s)\Psi(s)\right|\right|^2_{L_2(Q^{1/2}(U),H)}\,\mathrm{d}s$$

$$= \mathbf{E}\int_{\tau}^{\tau_n} ||\Psi(s)||^2_{L_2(Q^{1/2}(U),H)}\,\mathrm{d}s,$$

and the right-hand side of the equation above tends to 0 by the dominated convergence theorem.

iv) Let $\Psi \in \mathscr{S}(U,H)$ and write Ψ as in (23). First, assume that τ takes finitely many values $\mathbf{P}$-a.s. and write $\tau = \sum_{n=0}^{N} a_n\chi_{\{\tau=a_n\}}$ for constants $0 < a_0 < a_1 < \cdots < a_N \le T$. Since $\chi_{[0,\tau]}\Psi = \Psi - \chi_{(\tau,T]}\Psi$ we can compute $I_t^M(\chi_{[0,\tau]}\Psi)$

using linearity of the stochastic integral. Using (32) we obtain

$$
\begin{aligned}
I_t^M(\chi_{[0,\tau]}\Psi) = I_t^M(\Psi) - I_t^M(\chi_{(\tau,T]}\Psi)& \\
&= \sum_{j=0}^{m-1} \chi_{A_j}\Phi_j\big(M(t_{j+1}\wedge t) - M(t_j\wedge t)\big) \\
&\quad - \sum_{j=0}^{m-1}\sum_{n=0}^{N} \chi_{A_j\cap\{\tau=a_n\}}\Phi_j\big(M((t_{j+1}\vee a_n)\wedge t) - M((t_j\vee a_n)\wedge t)\big) \\
&= \sum_{j=0}^{m-1} \chi_{A_j}\Phi_j\big(M(t_{j+1}\wedge t) - M(t_j\wedge t)\big) \\
&\quad - \sum_{j=0}^{m-1} \chi_{A_j}\Phi_j\big(M((t_{j+1}\vee \tau)\wedge t) - M((t_j\vee \tau)\wedge t)\big).
\end{aligned}
$$

In each case $\tau \le t_j$, and $t_j < \tau \le t_{j+1}$, and $t_{j+1} < \tau$, there is cancellation in the last line above and the expression simplifies to

$$
I_t^M(\chi_{[0,\tau]}\Psi) = \sum_{j=0}^{m-1} \chi_{A_j}\Phi_j\big(M(t_{j+1}\wedge\tau\wedge t) - M(t_j\wedge\tau\wedge t)\big),
$$

which is the same as $I_{t\wedge\tau}^M(\Psi)$. Now we use a limiting argument to extend to the case where τ may take infinitely many values with positive probability but is still bounded by T a.s. There exist stopping times $(\tau_n)_{n=1}^{\infty}$ that take finitely many values, are bounded by T and decrease to τ a.s., for instance,

$$
\tau_n := \begin{cases} T\frac{k+1}{2^n} & \text{if } T\frac{k}{2^n} < \tau \le T\frac{k+1}{2^n} \text{ for some } 0 \le k \le 2^n - 1 \\ 0 & \text{otherwise.} \end{cases}
$$

Since the Lévy process M has right continuous sample paths and Ψ is a simple process it is easy to see that

$$
I_{t\wedge\tau_n}^M(\Psi) \to I_{t\wedge\tau}^M(\Psi) \quad \text{in } H \text{ a.s.} \tag{33}
$$

At the same time we have $\chi_{[0,\tau_n]}\Psi \to \chi_{[0,\tau]}\Psi$ in the space $\mathbf{L}^2_{Q^{1/2}(U),T}(H)$ by part iii), so $I_t^M(\chi_{[0,\tau_n]}\Psi) \to I_t^M(\chi_{[0,\tau]}\Psi)$ in $L^2(\Omega; H)$ by Theorem 3.12. By passing to a subsequence that converges in H a.s. and using (33) we find that $I_{t\wedge\tau}^M(\Psi) = I_t^M(\chi_{[0,\tau]}\Psi)$ in H a.s. □

4 Stochastic Integration with Respect to Poisson Random Measures

In this section we introduce the notion of stationary Poisson point process and the theory of stochastic integration with respect to the induced Poisson random measure and compensated Poisson random measure. This is a part of the setting of stochastic integration with Lévy noise as presented by Ikeda and Watanabe. The main references for this section are [10] and [4]. We also mention the article [17], which gives a comprehensive treatment of stochastic integration of Banach space-valued functions with respect to compensated Poisson random measures, and the book [3], which presents the related notion of stochastic integration with respect to martingale measures. We will restrict our attention to the case of stochastic integration of functions taking values in a real, separable Hilbert space. We will not discuss martingale measures but will only discuss how the corresponding theory of stochastic integration is related to stochastic integration with respect to compensated Poisson random measures (see Remark 4.9). We begin by introducing the notion of Poisson point process, then we introduce stochastic integration.

4.1 Poisson Point Processes

We begin with some preliminary notions leading to the definition of Poisson point processes. Below we use the notation $\mathbb{N} := \{0, 1, 2, \ldots\}$ and $\overline{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$. We continue to work on a fixed probability space $(\Omega, \mathscr{F}, \mathbf{P})$.

Definition 4.1 Let $(Z, \mathscr{Z})$ be a measurable space. A *point function* on Z is a partial function $\alpha\colon (0, \infty) \rightharpoonup Z$ whose domain $\mathcal{D}(\alpha) \subset (0, \infty)$ is at most countable. A point function α naturally induces an $\overline{\mathbb{N}}$-valued measure N_α on $(0, \infty) \times Z$ via

$$N_\alpha(\Gamma) := \#\{t \in \mathcal{D}(\alpha) : (t, \alpha(t)) \in \Gamma\} \qquad \text{for all } \Gamma \in \mathscr{B}(0, \infty) \otimes \mathscr{Z}.$$

Let Π_Z be the set of all point functions on Z. Let $\mathscr{Q}$ be the σ-field on Π_Z generated by sets of the form $S(\Gamma, k) := \{\alpha \in \Pi_Z : N_\alpha(\Gamma) = k\}$ over all $\Gamma \in \mathscr{B}(0, \infty) \otimes \mathscr{Z}$ and all $k \in \mathbb{N}$. A Z-valued *point process* on $(\Omega, \mathscr{F}, \mathbf{P})$ is simply a function $\Xi : \Omega \to \Pi_Z$ that is measurable from $(\Omega, \mathscr{F})$ to $(\Pi_Z, \mathscr{Q})$.

Lemma 4.2 *A function $\Xi : \Omega \to \Pi_Z$ is a Z-valued point process if and only if for every $\Gamma \in \mathscr{B}(0, \infty) \otimes \mathscr{Z}$ the function $N_\Xi(\Gamma)\colon \Omega \to \overline{\mathbb{N}}$ is an $\mathscr{F}$-measurable random variable.*

Proof A standard argument shows that Ξ is measurable from $(\Omega, \mathscr{F})$ to $(\Pi_Z, \mathscr{Q})$ if and only if $\Xi^{-1}(S(\Gamma, k)) \in \mathscr{F}$ for every $\Gamma \in \mathscr{B}(0, \infty) \otimes \mathscr{Z}$ and $k \in \mathbb{N}$. For such Γ and k note that

$$\Xi^{-1}(S(\Gamma, k)) = \{\omega \in \Omega : N_{\Xi(\omega)}(\Gamma) = k\} = \{N_\Xi(\Gamma) = k\}.$$

Also, since $\{N_\Xi(\Gamma) = \infty\} = \Omega \setminus \bigcup_{k\in\mathbb{N}}\{N_\Xi(\Gamma) = k\}$ we see that $N_\Xi(\Gamma)$ is $\mathscr{F}$-measurable if and only if $\{N_\Xi(\Gamma) = k\} \in \mathscr{F}$ for each $k \in \mathbb{N}$. Therefore, Ξ is a Z-valued point process if and only if $\Xi^{-1}(S(\Gamma, k)) \in \mathscr{F}$ for every $k \in \mathbb{N}$ and every Borel set $\Gamma \subseteq (0, \infty) \times Z$. This occurs if and only if $N_\Xi(\Gamma)$ is $\mathscr{F}$-measurable. □

Definition 4.3 Let $(E, \mathscr{E}, \lambda)$ be a σ-finite measure space, that is, there exists $(E_n)_{n=1}^{\infty} \subset \mathscr{E}$ such $E = \bigcup_{n=1}^{\infty} E_n$ and $\lambda(E_n) < \infty$ for each n. A function π from Ω to the set of $\overline{\mathbb{N}}$-valued measures on $(E, \mathscr{E})$ is called a *Poisson random measure* with *intensity measure* λ if for every $\Gamma \in \mathscr{E}$ the $\overline{\mathbb{N}}$-valued random variable $\pi(\Gamma)$ has a Poisson distribution with mean $\lambda(\Gamma)$ (possibly ∞) and if for every collection of pairwise disjoint sets $\Gamma_1, \ldots, \Gamma_m \in \mathscr{E}$ the random variables $\pi(\Gamma_1), \ldots, \pi(\Gamma_m)$ are independent.

Definition 4.4 A Z-valued point process Ξ is called a *stationary Poisson point process* if there exists a σ-finite measure ν on $(Z, \mathscr{Z})$ such that N_Ξ is a Poisson random measure on $(0, \infty) \times Z$ with intensity measure $\mathrm{d}t \otimes \mathrm{d}\nu$ (that is to say, if and only if N_Ξ is a stationary Poisson random measure on $(0, \infty) \times Z$). Let $(\mathscr{F}_t)_{t\geq 0}$ be a filtration contained in $\mathscr{F}$. We say that Ξ is a stationary $\mathscr{F}_t$*-Poisson point process* if Ξ is a stationary Poisson point process with the additional property that for every $A \in \mathscr{Z}$ with $\nu(A) < \infty$ the $\mathbb{N}$-valued process $(N_\Xi((0, t] \times A))_{t\geq 0}$ is an $\mathscr{F}_t$-Poisson process.

Remark 4.5 Suppose that $\Xi : \Omega \to \Pi_Z$ is a function (with no measurability assumptions a priori) such that N_Ξ is a stationary Poisson random measure on $(0, \infty) \times Z$. The definition of Poisson random measure says that for each $\Gamma \in \mathscr{B}(0, \infty) \otimes \mathscr{Z}$, $N_\Xi(\Gamma)$ is a Poisson random variable. By Lemma 4.2 it follows that Ξ is measurable from $(\Omega, \mathscr{F})$ to $(\Pi_Z, \mathscr{Q})$, so Ξ is a stationary Poisson point process on Z. Therefore, to show that $\Xi : \Omega \to \Pi_Z$ is a stationary Poisson point process it is sufficient to show that N_Ξ is a stationary Poisson random measure.

Example 4.6 Let L be a U-valued Lévy process and let π be the jump measure of L as in Definition 2.16. We will show that π is a Poisson random measure on $(0, \infty) \times (U \setminus \{0\})$. Since L has independent increments it follows that the random variables $\pi(\Gamma_1), \ldots, \pi(\Gamma_m)$ are independent when $\Gamma_1, \ldots, \Gamma_m \in \mathscr{B}(0, \infty) \otimes \mathscr{B}(U \setminus \{0\})$ are pairwise disjoint. Recall that for each $A \in \mathscr{B}(U \setminus \{0\})$ that is separated from zero the process $(\pi_A(t))_{t\geq 0}$ defined in (8) is a Poisson process with rate $\nu(A)$, where ν is the Lévy measure of L. Therefore, by (12) it follows that the random variable $\pi((t, t'] \times A)$ has a Poisson distribution with mean $(t' - t)\nu(A)$ for all $t' > t > 0$. A routine σ-field argument, using the fact that a sum of independent Poisson random variables with means $\lambda_1, \lambda_2, \ldots$ has a Poisson distribution with mean $\sum_{j=1}^{\infty} \lambda_j$, shows that $\pi(\Gamma)$ has a Poisson distribution with mean $\int_\Gamma \mathrm{d}t \otimes \mathrm{d}\nu$ for every $\Gamma \in \mathscr{B}(0, \infty) \otimes \mathscr{B}(U \setminus \{0\})$. This shows that π is a Poisson random measure on $(0, \infty) \times (U \setminus \{0\})$ with intensity measure $\mathrm{d}t \otimes \mathrm{d}\nu$.

The stochastic integrals that Rüdiger considers in [17] are more general than the stochastic integrals with respect to compensated Poisson random measures that we consider here for two reasons. First, as we have already mentioned, he considers integrands taking values in separable Banach spaces while we consider integrands taking values in separable Hilbert spaces. Second, each Poisson random measure that he considers arises as in Example 4.6 from the jumps of a càdlàg process X having all of the properties of an $\mathscr{F}_t$-Lévy process except, possibly, for the stationary increments property (see Definition 2.1). Since our goal is to study stochastic integration with respect to Lévy processes we will instead consider Poisson random measures arising as in Example 4.6 from the jumps of a càdlàg $\mathscr{F}_t$-Lévy process.

4.2 Integration with Respect to Poisson Random Measures

Let Ξ be a stationary Poisson point process on Z with intensity measure ν and let H be a separable, real Hilbert space. We consider the following spaces of functions for integration with respect to the Poisson random measure N_Ξ induced by Ξ. For $q \in [1,\infty]$ we introduce the notation

$$\mathbf{F}^q_{\nu,T}(H) := L^q(\Omega \times [0,T] \times Z, \mathscr{P}_{[0,T]} \otimes \mathscr{Z}, \mathrm{d}\mathbf{P} \otimes \mathrm{d}t \otimes \mathrm{d}\nu; H). \tag{34}$$

For the purpose of stochastic integration we will only be interested in the spaces $\mathbf{F}^1_{\nu,T}(H)$ and $\mathbf{F}^2_{\nu,T}(H)$. Below we gather the basic facts in [10] (see also [4]) about integration of functions in these spaces with respect to N_Ξ.

Theorem 4.7 *Let Ξ be a stationary $\mathscr{F}_t$-Poisson point process on a measurable space $(Z,\mathscr{Z})$ with intensity measure ν. Then the following statements hold.*

i) (Integrands in $\mathbf{F}^1_{\nu,T}(H)$) Let $f \in \mathbf{F}^1_{\nu,T}(H)$. Then the following statements hold.

a) $\mathbf{E}\int_{(0,t]}\int_Z |f(s,z)|_H \,\mathrm{d}N_\Xi = \mathbf{E}\int_0^t\int_Z |f(s,z)|_H \,\mathrm{d}\nu\,\mathrm{d}s < \infty$ for every $t \in [0,T]$.

b) For each $t \in [0,T]$ the H-valued integral $\int_{(0,t]}\int_Z f(s,z)\,\mathrm{d}N_\Xi$ exists a.s. and is equal to the absolutely convergent sum $\sum_{s\in(0,t]} f(s,\Xi(s))$.

c) For each $t \in [0,T]$ we have $\mathbf{E}\int_{(0,t]}\int_Z f(s,z)\,\mathrm{d}N_\Xi = \mathbf{E}\int_0^t\int_Z f(s,z)\,\mathrm{d}s\,\mathrm{d}\nu$.

ii) (Integrands in $\mathbf{F}^2_{\nu,T}(H)$)

a) For $f \in \mathbf{F}^1_{\nu,T}(H) \cap \mathbf{F}^2_{\nu,T}(H)$ define

$$\int_{(0,t]}\int_Z f(s,z)\,\mathrm{d}\widehat{N}_\Xi := \int_{(0,t]}\int_Z f(s,z)\,\mathrm{d}N_\Xi - \int_0^t\int_Z f(s,z)\,\mathrm{d}\nu\,\mathrm{d}s. \tag{35}$$

Then the process $\big(\int_{(0,t]}\int_Z f(s,z)\,\mathrm{d}\widehat{N}_\Xi\big)_{t\in[0,T]}$ *belongs to the space* $\mathscr{M}_T^2(H)$ *of square-integrable* H*-valued martingales on* $[0,T]$ *and*

$$\mathbf{E}\Big|\int_{(0,t]}\int_Z f(s,z)\,\mathrm{d}\widehat{N}_\Xi\Big|_H^2 = \mathbf{E}\int_0^t\int_Z |f(s,z)|_H^2\,\mathrm{d}\nu\,\mathrm{d}s. \tag{36}$$

b) $\mathbf{F}^1_{\nu,T}(H)\cap\mathbf{F}^2_{\nu,T}(H)$ *is dense in* $\mathbf{F}^2_{\nu,T}(H)$.

c) Given $f\in\mathbf{F}^2_{\nu,T}(H)$ *let* $(f_n)_{n=1}^\infty$ *be a sequence in* $\mathbf{F}^1_{\nu,T}(H)\cap\mathbf{F}^2_{\nu,T}(H)$ *that converges to* f *in* $\mathbf{F}^2_{\nu,T}(H)$. *By ii a) the sequence* $\big(\int_{(0,t]}\int_Z f_n(s,z)\,\mathrm{d}\widehat{N}_\Xi\big)_{n=1}^\infty$ *is Cauchy in the space* $\mathscr{M}_T^2(H)$. *Furthermore, the limit does not depend on the particular sequence* $(f_n)_{n=1}^\infty$. *Therefore, we can define the* H*-valued process* $\big(\int_{(0,t]}\int_Z f(s,z)\,\mathrm{d}\widehat{N}_\Xi\big)_{t\in[0,T]}$ *to be the limit of any such sequence. By construction, this is a square-integrable* H*-valued martingale and equation* (36) *continues to hold.*

Definition 4.8 We refer to $\widehat{N}_\Xi$ as the *compensated Poisson random measure*. For $T>0$ we define a map $I_T^{\widehat{N}_\Xi}:\mathbf{F}^2_{\nu,T}(H)\to L^2(\Omega;H)$ by

$$I_T^{\widehat{N}_\Xi}(f) := \int_{(0,T]}\int_Z f(s,z)\,\mathrm{d}\widehat{N}_\Xi \qquad \text{for all } f\in\mathbf{F}^2_{\nu,T}(H). \tag{37}$$

Equation (36) says that $I_T^{\widehat{N}_\Xi}$ is an isometry between these spaces.

Remark 4.9 We mention that because functions in $\mathbf{F}^2_{\nu,T}(H)$ are assumed to be $\mathscr{P}_{[0,T]}\otimes\mathscr{Z}$-measurable, the stochastic integration map $I_T^{\widehat{N}_\Xi}$ defined above is an example of what Rüdiger calls the *simple-2 integral* in [17]. By generalizing the notion of stochastic integration with respect to martingale measures as presented in [3], it is possible to extend $I_T^{\widehat{N}_\Xi}$ to the larger class of functions

$$\begin{aligned}\{f\in L^2(\Omega\times[0,T]\times Z,&\mathscr{F}_T\otimes\mathscr{B}([0,T]\times Z),\ \mathrm{d}\mathbf{P}\otimes\mathrm{d}t\otimes\mathrm{d}\nu;H)\\ &: f(\cdot,t,z)\text{ is }\mathscr{F}_t\text{-measurable }\forall(t,z)\in[0,T]\times Z\}.\end{aligned}$$

This is what Rüdiger refers to as the *strong-2 integral* in [17].

Remark 4.10 We emphasize here that stochastic integration with respect to the Poisson random measure N_Ξ is only defined for integrands in $\mathbf{F}^1_{\nu,T}(H)$, while stochastic integration with respect the compensated Poisson random measure $\widehat{N}_\Xi$ is only defined for integrands in $\mathbf{F}^2_{\nu,T}(H)$. The formula

$$\begin{aligned}\int_{(0,t]}\int_Z f(s,z)\,\mathrm{d}\widehat{N}_\Xi &= \int_{(0,t]}\int_Z f(s,z)\,\mathrm{d}N_\Xi - \int_0^t\int_Z f(s,z)\,\mathrm{d}\nu\,\mathrm{d}s \qquad (38)\\ &= \sum_{s\in(0,t]} f(s,\Xi(s)) - \int_0^t\int_Z f(s,z)\,\mathrm{d}\nu\,\mathrm{d}s\end{aligned}$$

is only valid for integrands $f \in \mathbf{F}^1_{\nu,T}(H) \cap \mathbf{F}^2_{\nu,T}(H)$. Indeed, the left-hand side of (38) is only defined when $f \in \mathbf{F}^2_{\nu,T}(H)$ while the right-hand side of (38) is only defined when $f \in \mathbf{F}^1_{\nu,T}(H)$. For an arbitrary $f \in \mathbf{F}^2_{\nu,T}(H)$, the process $\big(\int_{(0,t]}\int_Z f(s,z)\,\mathrm{d}\widehat{N}_\varXi\big)_{t\in[0,T]}$ is defined as a limit, so (38) may not hold.

Remark 4.10 Because of (38) it is tempting to believe that $\widehat{N}_\varXi$ is a random signed measure given by $\widehat{N}_\varXi = N_\varXi - \mathrm{d}t \otimes \mathrm{d}\nu$. This is incorrect because the set function $N_\varXi - \mathrm{d}t \otimes \mathrm{d}\nu$ is undefined on sets $\varGamma \in \mathscr{B}(0,\infty) \otimes \mathscr{Z}$ with the property that $\int_\varGamma \mathrm{d}t\,\mathrm{d}\nu = \infty$. Since the Lebesgue measure on $(0,\infty)$ is not finite such sets $\varGamma$ always exist, even when ν is a finite measure.

Corollary 4.12 *Let $\varXi$ be a stationary $\mathscr{F}_t$-Poisson point process on a measurable space $(Z,\mathscr{Z})$ with intensity measure ν. If $\nu(Z) < \infty$, then $\mathbf{F}^2_{\nu,T}(H) \subseteq \mathbf{F}^1_{\nu,T}(H)$ and the inclusion is continuous. In particular,* (38) *holds for all $f \in \mathbf{F}^2_{\nu,T}(H)$.*

Proof Since $\nu(Z) < \infty$ this follows from the Cauchy-Schwarz inequality. □

Remark 4.13 For every $f \in \mathbf{F}^2_{\nu,T}(H)$, the stochastic integral $\big(I_t^{\widehat{N}_\varXi}(f)\big)_{t\in[0,T]} = \big(\int_{(0,t]}\int_Z f(s,z)\,\mathrm{d}\widehat{N}_\varXi\big)_{t\in[0,T]}$ is purely discontinuous and starts from 0, i.e., $I^{\widehat{N}_\varXi}(f) \in \mathscr{M}_T^{2,d}(H)$ (see Definition 2.25). To show this, suppose first that $f \in \mathbf{F}^1_{\nu,T}(H) \cap \mathbf{F}^2_{\nu,T}(H)$. In this case $I^{\widehat{N}_\varXi}(f)$ is given by formula (38), so it clearly starts from 0 and its jumps are summable by Theorem 4.7. This means that $I^{\widehat{N}_\varXi}(f)$ has bounded variation on $[0,T]$. Since $\mathscr{M}_T^{2,d}(H)$ is the closure of the space of bounded variation martingales in $\mathscr{M}_T^2(H)$ that start at 0, the general case where $f \in \mathbf{F}^2_{\nu,T}(H)$ follows from the construction of $I^{\widehat{N}_\varXi}(f)$ in Theorem 4.7.

The next result gives the quadratic variation of an H-valued stochastic integral with respect to the compensated Poisson random measure $\widehat{N}_\varXi$.

Theorem 4.14 *Let $\varXi$ be a stationary $\mathscr{F}_t$-Poisson point process on a measurable space $(Z,\mathscr{Z})$ with intensity measure ν and let $N_\varXi$ denote the associated Poisson random measure. For every $f \in \mathbf{F}^2_{\nu,T}(H)$ the quadratic variation of the H-valued martingale $\big(I_t^{\widehat{N}_\varXi}(f)\big)_{t\in[0,T]}$ is given by*

$$[I^{\widehat{N}_\varXi}(f)]_t = \int_{(0,t]}\int_Z |f(s,z)|_H^2\,\mathrm{d}N_\varXi(s,z). \tag{39}$$

See Theorem 8.23 in [15] for a proof. While the proof of Theorem 8.23 in [15] is written for real-valued integrands instead of H-valued integrands it is easy to see that the proof remains valid for H-valued integrands when products of real numbers are replaced by inner products in H. With (39) in hand the upper bound in the BDG inequality from Theorem 2.23 takes the following form for stochastic integrals with respect to the compensated Poisson random measure $\widehat{N}_\varXi$.

Corollary 4.15 *Let Ξ be a stationary $\mathscr{F}_t$-Poisson point process on a measurable space $(Z, \mathscr{Z})$ with intensity measure ν and let N_Ξ denote the associated Poisson random measure. For every $1 \leq p < \infty$ there exists a constant $C_p \in (0, \infty)$ such that for every $\mathscr{F}_t$-stopping time τ and for every $f \in \mathbf{F}^2_{\nu,T}(H)$ we have*

$$\mathbf{E} \sup_{t\in[0,\tau]} \left| \int_{(0,t]} \int_Z f(s,z)\, \mathrm{d}\widehat{N}_\Xi(s,z) \right|_H^p \leq C_p \mathbf{E}\Big(\int_{(0,\tau]} \int_Z |f(s,z)|_H^2 \, \mathrm{d}N_\Xi(s,z) \Big)^{p/2}. \tag{40}$$

In the next result we show that the jump measure π of a Lévy process is the type of Poisson random measure considered in the setting presented by Ikeda and Watanabe. Specifically, we show that the jumps of a Lévy process form a stationary Poisson point process and that the induced Poisson random measure coincides with the jump measure π.

Proposition 4.16 *Let L be a U-valued $\mathscr{F}_t$-Lévy process. Then the jumps of L induce a stationary $\mathscr{F}_t$-Poisson point process Ξ on U. Furthermore, the Poisson random measure N_Ξ is the jump measure of L and its intensity measure is $\mathrm{d}t \otimes \mathrm{d}\nu$, where ν is the Lévy measure of L.*

Proof Let ν be the Lévy measure of L and let $\big(T_j\big)_{j=1}^\infty$ be the jump times[1] of L. For each $\omega \in \Omega$ define a U-valued point function $\Xi_\omega \colon (0, \infty) \rightharpoonup U$ by

$$\mathcal{D}(\Xi_\omega) := \{T_1(\omega), T_2(\omega), \ldots\}, \qquad \Xi_\omega(T_j(\omega)) := \Delta L(T_j(\omega)) \quad \text{for every } j \geq 1.$$

That is, the domain of Ξ is the set of jump times of L and Ξ sends each jump time T_j to the value of the jump $\Delta L(T_j)$ occurring at that time. The Poisson random measure N_Ξ induced by Ξ is precisely the jump measure $\pi := \sum_{j=1}^\infty \delta_{(T_j, \Delta L(T_j))}$ of L. We know from Example 4.6 that π is a Poisson random measure on $(0, \infty) \times (U \setminus \{0\})$ with intensity measure $\mathrm{d}t \otimes \mathrm{d}\nu$, so it follows from Remark 4.5 that Ξ is a stationary Poisson point process on U. For $t' \geq t \geq 0$ and $A \in \mathscr{B}(U \setminus \{0\})$ it is intuitively clear that the random variable $\pi((t, t'] \times A)$ is measurable with respect to the σ-field $\mathscr{G}_{t,t'} := \sigma(L(s) - L(t), s \in [t, t'])$ because one can determine how many jumps of L that occur during the time interval $(t, t']$ lie in A based on the increments of L on $(t, t']$. It is indeed true that $\pi((t, t'] \times A)$ is $\mathscr{G}_{t,t'}$-measurable (see Lemma 13.5 in [11] and the related result Lemma 5.3 below). Since L is an $\mathscr{F}_t$-Lévy process it follows that Ξ is a stationary $\mathscr{F}_t$-Poisson point process. □

[1] Since L is càdlàg it is possible for jump times of L to accumulate, provided that the sizes of the accumulating jumps tend to zero. Although the jumps of L can be enumerated it may not be possible to enumerate them in increasing order. Thus, one should not assume that $T_j(\omega) < T_{j+1}(\omega)$.

5 Comparing Lévy Noise: The Square-Integrable Case

In Sect. 3 we considered stochastic integration with respect to square-integrable Lévy martingales M in the setting presented by Peszat and Zabczyk. A drawback to the abstract construction of the stochastic integral with respect to M in Theorem 3.12 is that it is often difficult to make explicit computations when the need arises. Specifically, *it is not straightforward to find the jumps of a stochastic integral with respect to M or its quadratic variation* and these are required when using the Itô formula and BDG inequality to treat SPDEs. We have identified these issues in Sect. 1 where we posed Question 1.2, "What are the jumps of $\big(I_t^M(\Psi)\big)_{t\geq 0}$?" and Question 1.3, "What is the quadratic variation of $\big(I_t^M(\Psi)\big)_{t\geq 0}$?" In order to answer these questions we identify the continuous and purely discontinuous parts of M (see Definition 2.25) in Sect. 5.1 with the help of the Lévy-Khinchin decomposition. As we will see in Lemma 5.1 below, the continuous part of M is a Wiener process W and the purely discontinuous part of M, which we will denote by $\mathscr{L}$, has the form $\mathscr{L} = \sum_{n=0}^{\infty} \widehat{P}_n$, where $(P_n)_{n=0}^{\infty}$ are independent CPPs as in (1) (see Theorem 2.15 for an explicit construction of these CPPs in terms of M). From the decomposition $M = W + \mathscr{L}$ we show that the stochastic integral with respect to M decomposes according to the formal rule $\mathrm{d}M = \mathrm{d}W + \mathrm{d}\mathscr{L}$ (see Lemma 5.6). In Sect. 5.2 we show that the stochastic integral with respect to the process $\mathscr{L}$ can be expressed according to the formal rule $\mathrm{d}\mathscr{L} = \mathrm{d}\widehat{\pi}$, where $\widehat{\pi}$ is the compensated jump measure of M (see Proposition 5.14). Finally, in Sect. 5.3 we show how the stochastic integral with respect to a square-integrable Lévy martingale as presented by Peszat and Zabczyk can be realized in the setting presented by Ikeda and Watanabe and we answer Questions 1.2 and 1.3.

5.1 *The Lévy-Khinchin Decomposition*

We begin by restating the Lévy-Khinchin decomposition (Theorem 2.15) in the special case of a square-integrable, Lévy martingale M and stating additional results about the structure of the covariance operator of M.

Lemma 5.1 *Let M be a square-integrable, Lévy martingale M taking values in a real, separable Hilbert space U with Lévy measure ν. Then the following statements hold:*

i) There exists a U-valued Wiener process W and a sequence of U-valued square-integrable compound Poisson processes $(P_n)_{n=1}^{\infty}$ such that all of these processes are independent and

$$M(t) = W(t) + \sum_{n=1}^{\infty} \widehat{P}_n(t) \tag{41}$$

in U, where a.s. the series converges uniformly in t on compact subsets of $[0, \infty)$. *Conversely, every process M of the form in* (41) *is a square-integrable Lévy martingale.*

ii) There exist disjoint Borel subsets $(A_n)_{n=1}^{\infty}$ of U, each separated from 0*, such that the Lévy measure of P_n is $\nu|_{A_n}$ and $\bigcup_{n=1}^{\infty} A_n = U \setminus \{0\}$.*

iii) The covariance operator of $\mathscr{L} := M - W = \sum_{n=1}^{\infty} \widehat{P}_n$ is given by

$$(Q_1 x, y)_U = \int_U (u, x)_U \, (u, y)_U \, \mathrm{d}\nu(u) \tag{42}$$

for all $x, y \in U$.

iv) Let Q_0 denote the covariance operator of W. Then the covariance operator of M is $Q_0 + Q_1$.

Proof By Theorem 2.15 there exists a vector $a \in U$, a U-valued Wiener process W and U-valued compound Poisson processes $(L_n)_{n=0}^{\infty}$, all independent, such that

$$M(t) = at + W(t) + L_0(t) + \sum_{n=1}^{\infty} \widehat{L}_n(t),$$

where the series converges uniformly in t on compact subsets of $[0, \infty)$. Furthermore, the compound Poisson processes can be chosen as $L_n(t) := \sum_{s \in (0,t]} \Delta L(s) \cdot \chi_{A_n}(\Delta L(s))$, where $A_0 := B(0, 1)^c$ and $A_n := B(0, 1/n) \setminus B(0, 1/(n+1))$. In this case, $(L_n)_{n=0}^{\infty}$ are indeed independent (see Lemma 4.24 in [15]) CPPs and L_n has Lévy measure $\nu|_{A_n}$. Since M, W and $(\widehat{L}_n)_{n=1}^{\infty}$ are mean-zero we have

$$a = -\int_{A_0} u \, \mathrm{d}\nu(u) = -\mathbf{E}[L_0(1)].$$

The first equality comes from Theorem 4.47 of [15] and the second from Proposition 2.11. Next, since M is square-integrable we have $\int_U |u|_U^2 \, \mathrm{d}\nu(u) < \infty$ by Theorem 4.47 of [15]. Since the Lévy measure of L_0 is $\nu|_{A_0}$, Proposition 2.11 implies that L_0 is square-integrable. Therefore, the sum $L_0(t) + at = \widehat{L}_0(t)$ is a square-integrable CCPP. Since W and $(\widehat{L}_n)_{n=0}^{\infty}$ are independent processes equation (41) follows by setting $P_n := L_{n-1}$ for each positive integer n. □

Remark 5.2 Since the Wiener process W is continuous a.s., M and $\mathscr{L}$ have the same jumps. That is, they jump at the same times and with the same values. This means that they have the same jump measure, say π, and the same Lévy measure. Indeed, the Lévy measure is determined uniquely by the jump measure via (9). So the Lévy measure of $\mathscr{L}$ is ν.

As we will show, the Lévy-Khinchin decomposition allows one to decompose stochastic integrals with respect to M using the formal rule $\mathrm{d}M = \mathrm{d}W + \mathrm{d}\mathscr{L}$. Suppose that M, W and $\mathscr{L}$ each satisfy Assumption 3.3 with respect to *the same filtration* $(\mathscr{F}_t)_{t \geq 0}$. Then the same space $\mathscr{S}(U, H)$ of simple processes is used for stochastic integration with respect to M, W and $\mathscr{L}$. In this case, it is clear from Definition 3.8 that $\int_0^t \Psi(s)\,\mathrm{d}M(s) = \int_0^t \Psi(s)\,\mathrm{d}W(s) + \int_0^t \Psi(s)\,\mathrm{d}\mathscr{L}(s)$ for every $\Psi \in \mathscr{S}(U,H)$ and every $t \geq 0$. A rigorous interpretation of the statement $\mathrm{d}M = \mathrm{d}W + \mathrm{d}\mathscr{L}$ on the entire space of admissible integrands with respect to M can be achieved in two additional steps. The first step is to map the space of admissible integrands with respect to M continuously into the spaces of admissible integrands with respect to W and $\mathscr{L}$ in a natural way. The second step is an approximation argument. We do both steps in Lemma 5.6 in a more general context. Before proving Lemma 5.6 we show that it is possible for all three processes M, W and $\mathscr{L}$ to satisfy Assumption 3.3 with respect to the same filtration. The given process M satisfies Assumption 3.3 with respect to its natural filtration but also with respect to a complete and right-continuous filtration $(\mathscr{F}_t)_{t\geq 0}$ (see Remark 3.2). The purpose of the next two results is to show that W and $\mathscr{L}$ also satisfy Assumption 3.3 with respect to the same filtration $(\mathscr{F}_t)_{t\geq 0}$. The first of these is similar to Lemma 13.5 in [11]; note that no integrability assumptions are placed on the Lévy process L in Lemma 5.3 below.

Lemma 5.3 *Let L be a U-valued Lévy process. Let $A \in \mathscr{B}(U)$ be separated from* 0 *and define the process*

$$L_A(t) := \sum_{s\in(0,t]} \chi_A(\Delta L(s))\Delta L(s) \tag{43}$$

as in Lemma 2.14. Then for all $t' \geq t \geq 0$ the random variable $L_A(t') - L_A(t)$ is measurable with respect to the σ-field $\mathscr{G}_{t,t'} := \sigma(L(s) - L(t) : s \in [t,t'])$.

Proof Let π denote the jump measure of L (see (13)) and let ν be the Lévy measure of L. We begin with a representation of L_A as a random integral with respect to π. Fix $\rho > 0$ and let B_ρ denote the closed ball of radius ρ centered at 0 in U. For each $t' \geq 0$ the random variable $\pi((0,t'] \times B_\rho^c)$ has a Poisson distribution with mean $t'\nu(B_\rho^c)$, which is finite. Therefore,

$$\int_{(0,t']}\int_{B_\rho^c} |u|_U \,\mathrm{d}\pi(s,u) = \sum_{s\in(0,t']} |\Delta L(s)|_U \chi_{B_\rho^c}(\Delta L(s)) < \infty \qquad \mathbf{P}\text{-a.s.}, \tag{44}$$

because the number of terms in the sum is finite a.s. In particular, for every set $A \in \mathscr{B}(U)$ that is separated from 0 the function which is zero outside of A and agrees with the identity function within A is integrable with respect to the measure π_ω for $\mathbf{P}$-a.e. $\omega \in \Omega$. So, for every $t \geq 0$ we have

$$L_A(t) = \int_{(0,t]}\int_A u\,\mathrm{d}\pi(s,u) \qquad \mathbf{P}\text{-a.s.},$$

and for every $t' \geq t \geq 0$ we have

$$L_A(t') - L_A(t) = \int_{(t,t']} \int_A u \, \mathrm{d}\pi(s,u) \qquad \mathbf{P}\text{-a.s.} \tag{45}$$

We will use an approximation argument to show that the random integral on the right-hand side of (45) is $\mathscr{G}_{t,t'}$-measurable for every $A \in \mathscr{B}(U \setminus B_\rho)$ for each fixed $\rho > 0$.

Let $f : U \to U$ be a bounded continuous function that vanishes in a neighborhood of 0. For every sequence of partitions $t = t_0^{(n)} < t_1^{(n)} < \cdots < t_{m_n}^{(n)} = t'$ with mesh $\max_{1\leq k\leq m_n}(t_k^{(n)} - t_{k-1}^{(n)})$ tending to 0 as $n \to \infty$ we have

$$\sum_{k=1}^{m_n} f\big(L(t_k^{(n)}) - L(t_{k-1}^{(n)})\big) \to \sum_{s\in(t,t']} f(\Delta L(s)) = \int_{(t,t']} \int_U f(u)\, \mathrm{d}\pi(s,u) \qquad \mathbf{P}\text{-a.s.} \tag{46}$$

Each term in the sum on the left-hand side of (46) is $\mathscr{G}_{t,t'}$-measurable, hence so is the limit of these sums, which is $\int_{(t,t']} \int_U f(u)\, \mathrm{d}\pi(s,u)$. Let C be a bounded, closed subset of U that is disjoint from B_ρ. For $\varepsilon > 0$ consider the closed set $F_\varepsilon := \{u \in C^c : d(u, C) \geq \varepsilon\}$. Note that F_ε contains a neighborhood of 0 for all sufficiently small ε. It is clear that the function $f(u) := u\chi_C(u)$ is continuous on the closed set $C \cup F_\varepsilon$. There exists a continuous function $\bar{f}_\varepsilon : U \to U$ such that $\bar{f}_\varepsilon = f$ on $C \cup F_\varepsilon$ and such that its image $\bar{f}_\varepsilon(U)$ is contained in the convex hull of $f(U)$. This is an application of Dugundji's generalization of the Tietze extension theorem to Hilbert space-valued functions, see [9]. Since we assume that C is bounded there exists some $R > 0$ such that $|\bar{f}_\varepsilon(u)|_U \leq R$ for every $u \in U$ and every $\varepsilon > 0$. We have $\bar{f}_\varepsilon \to f$ pointwise on U as $\varepsilon \downarrow 0$. For each fixed $\omega \in \Omega$ we can apply the dominated convergence theorem using the finite measure $\chi_{(t,t']}(s)\, \mathrm{d}\pi_\omega(s,u)$ to conclude that

$$\int_{(t,t']} \int_U \bar{f}_\varepsilon(u)\, \mathrm{d}\pi(s,u) \to \int_{(t,t']} \int_C u\, \mathrm{d}\pi(s,u) \qquad \mathbf{P}\text{-a.s.}$$

This shows that $\int_{(t,t']} \int_C u\, \mathrm{d}\pi(s,u)$ is $\mathscr{G}_{t,t'}$-measurable for every bounded, closed subset of U contained in $U \setminus B_\rho$. The final step is to replace C by any Borel subset of $U \setminus B_\rho$ using a monotone class argument. The class $\mathscr{C}$ consisting of bounded, closed subsets of U that are contained in $U \setminus B_\rho$ is closed under finite intersections. We have just shown that $\mathscr{C}$ is contained in the class

$$\mathcal{M} := \{A \in \mathscr{B}(U \setminus B_\rho) : \textstyle\int_{(t,t']} \int_A u\, \mathrm{d}\pi(s,u) \text{ is } \mathscr{G}_{t,t'}\text{-measurable}\}.$$

Note that the random integral $\int_{(t,t']} \int_A u\, \mathrm{d}\pi(s,u)$ is well-defined for every $A \in \mathscr{B}(U \setminus B_\rho)$ by (44). We need to show that $\mathcal{M}$ is closed under increasing countable unions. Let $A_n \uparrow A$ with $A_n \in \mathcal{M}$ for each $n \in \mathbb{N}$. Since $u\chi_{A_n}(u) \to u\chi_A(u)$ pointwise on U we can apply the dominated convergence theorem using the measure

$\chi_{(t,t']}(s)\,\mathrm{d}\pi_\omega(s,u)$ and dominating function $u \mapsto |u|_U \chi_{B_\rho^c}(u)$ to conclude that

$$\int_{(t,t']}\int_{A_n} u\,\mathrm{d}\pi(s,u) \to \int_{(t,t']}\int_{A} u\,\mathrm{d}\pi(s,u) \qquad \mathbf{P}\text{-a.s.}$$

This shows that $A \in \mathcal{M}$, so $\mathcal{M}$ is a monotone class. The monotone class lemma implies that $\mathscr{B}(U \setminus B_\rho) = \sigma(\mathscr{C})$ is contained in $\mathcal{M}$. Since $\rho > 0$ is arbitrary, the proof is complete. □

An immediate consequence is that if L is an $\mathscr{F}_t$-Lévy process, then so are the processes W and $\left(L_{A_n}\right)_{n=0}^{\infty}$ in any Lévy-Khinchin decomposition of L.

Corollary 5.4 *Let L be a U-valued Lévy process on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$ and assume that L is an $\mathscr{F}_t$-Lévy process. Then for every set $A \in \mathscr{B}(U \setminus \{0\})$ that is separated from 0 the process L_A in (43) is an $\mathscr{F}_t$-compound Poisson process and the Wiener part in the Lévy-Khinchin decomposition of L is an $\mathscr{F}_t$-Wiener process.*

Proof Since $\mathscr{G}_{0,t} \subseteq \mathscr{F}_t$ for each $t \geq 0$ we see that L_A is $\mathscr{F}_t$-adapted. It follows from the independence condition (21) that $\mathscr{G}_{s,t}$ is independent of $\mathscr{F}_s$ for all $t \geq s \geq 0$, so L_A satisfies condition (21). This shows that L_A is an $\mathscr{F}_t$-Lévy process; the fact that L_A is a CPP is a restatement of Lemma 2.14. Let W be the Wiener part in the Lévy-Khinchin decomposition of L in equation (11). The difference $L - W$ is an $\mathscr{F}_t$-Lévy process, so $W = L - (L - W)$ is an $\mathscr{F}_t$-Wiener process. □

Recall that in this section we are given a square-integrable, U-valued Lévy martingale M which we have decomposed in Lemma 5.1 as $M = W + \mathscr{L}$, where W is a Wiener process and $\mathscr{L}$ is a square-integrable Lévy process with covariance operator given by (42). There exists a complete, right-continuous filtration $(\mathscr{F}_t)_{t\geq 0}$ such that M is an $\mathscr{F}_t$-Lévy process, equivalently, so that M satisfies Assumption 3.3. Corollary 5.4 says that W is an $\mathscr{F}_t$-Wiener process and $\mathscr{L}$ is an $\mathscr{F}_t$-Lévy process. Since W and $\mathscr{L}$ are both square-integrable and mean-zero, both processes satisfy Assumption 3.3 with respect to the same filtration $(\mathscr{F}_t)_{t\geq 0}$ as M. Below we show that the space of integrands corresponding to M is naturally mapped into the spaces of integrands corresponding to W and $\mathscr{L}$, respectively, and in such a way that the formal rule $\mathrm{d}M = \mathrm{d}W + \mathrm{d}\mathscr{L}$ holds. We make a more general argument below and show that such a decomposition holds for every sum of independent processes that satisfy Assumption 3.3. We begin with the result that allows us to define a natural continuous map from the space of integrands for stochastic integration with respect to M to the spaces of integrands for W and $\mathscr{L}$.

Lemma 5.5 *Let $Q_1, Q_2 \in L_1^+(U)$ with $Q_1 \leq Q_2$. Then*

i) For every $\Phi \in L(U, H)$ we have $||\Phi||^2_{L_2(Q_1^{1/2}(U),H)} = \mathrm{Tr}(\Phi Q_1 \Phi^*)$, and hence

$$||\Phi||_{L_2(Q_1^{1/2}(U),H)} \leq ||\Phi||_{L_2(Q_2^{1/2}(U),H)}\,.$$

ii) For every $\Psi \in \mathscr{S}(U, H)$ we have

$$\mathbf{E}\int_0^t ||\Psi(s)||^2_{L_2(Q_1^{1/2}(U),H)}\,\mathrm{d}s \le \mathbf{E}\int_0^t ||\Psi(s)||^2_{L_2(Q_2^{1/2}(U),H)}\,\mathrm{d}s,$$

for every $t \in [0, T]$. Thus, the identity map from $\mathscr{S}(U, H)$ endowed with the $\mathbf{L}^2_{Q_2^{1/2}(U),T}(H)$-norm to $\mathscr{S}(U, H)$ endowed with the $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$-norm extends uniquely by continuity to a linear map $\Psi \mapsto \iota(\Psi)$ from $\mathbf{L}^2_{Q_2^{1/2}(U),T}(H) \to \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ with norm at most 1.

Proof *i*) Recall from Definition 3.10 that if $(u_n)_{n=1}^{\infty}$ is an ONB for U, then the nonzero elements of $\{Q_1^{1/2}u_n : n \ge 1\}$ form an ONB for the space $Q_1^{1/2}(U)$. Therefore, for every $\Phi \in L(U, H)$ we have

$$||\Phi||^2_{L_2(Q_1^{1/2}(U),H)} = \mathrm{Tr}(Q_1^{1/2}\Phi^*\Phi Q_1^{1/2}) = \mathrm{Tr}(\Phi Q_1 \Phi^*) \le \mathrm{Tr}(\Phi Q_2 \Phi^*)$$

and the right-hand side of the inequality above is equal to $||\Phi||^2_{L_2(Q_2^{1/2}(U),H)}$.

ii) The statements in part *ii*) follow immediately from part *i*). □

Lemma 5.6 *Let M_1, M_2 be independent Lévy processes on a probability space $(\Omega, \mathscr{F}, \mathbf{P})$ satisfying Assumption 3.3 with respect to the same filtration $(\mathscr{F}_t)_{t\ge0}$. Let Q_j be the covariance operator of M_j for $j = 1, 2$. Then the following statements hold:*

i) The covariance operator of $M := M_1 + M_2$ is $Q := Q_1 + Q_2$.

ii) For $j = 1, 2$ let $\iota_j : \mathbf{L}^2_{Q^{1/2}(U),T}(H) \to \mathbf{L}^2_{Q_j^{1/2}(U),T}(H)$ denote the continuous extension of the identity map on $\mathscr{S}(U, H)$ from Lemma 5.5. Then for all $t \in [0, T]$ and all $\Psi \in \mathbf{L}^2_{Q^{1/2}(U),T}(H)$ we have

$$\int_0^t \Psi(s)\,\mathrm{d}M(s) = \int_0^t \iota_1(\Psi)(s)\,\mathrm{d}M_1(s) + \int_0^t \iota_2(\Psi)(s)\,\mathrm{d}M_2(s) \tag{47}$$

a.s. in H.

Proof *i*) For all $x, y \in U$ we have

$$\begin{aligned}(Qx, y)_U &= \mathbf{E}[(M(1), x)_U\,(M(1), y)_U]\\ &= \mathbf{E}[(M_1(1), x)_U\,(M_1(1), y)_U] + \mathbf{E}[(M_2(1), x)_U\,(M_2(1), y)_U]\\ &\quad + \mathbf{E}[(M_1(1), x)_U\,(M_2(1), y)_U] + \mathbf{E}[(M_2(1), x)_U\,(M_1(1), y)_U].\end{aligned}$$

The top line of the last expression on the right-hand side is $((Q_1 + Q_2)x, y)_U$ and the bottom line is zero because $M_1(1)$ and $M_2(1)$ are independent and have zero mean.

ii) For $j = 1, 2$ we have $Q_j \leq Q$, so the map ι_j exists by Lemma 5.5. Fix $t \in [0, T]$. It is obvious that (47) holds when $\Psi \in \mathscr{S}(U, H)$. For $j = 1, 2$ the map $\Psi \mapsto \int_0^t \iota_j(\Psi)(s)\,\mathrm{d}M_j(s)$ is a continuous linear mapping of $\mathbf{L}^2_{Q^{1/2}(U),T}(H) \to L^2(\Omega; H)$. Indeed, it is simply the composition $I_t^{M_j} \circ \iota_j$. Since $\mathscr{S}(U, H)$ is dense in $\mathbf{L}^2_{Q^{1/2}(U),T}(H)$ it follows that (47) holds for all $\Psi \in \mathbf{L}^2_{Q^{1/2}(U),T}(H)$.

□

Let $M = W + \mathscr{L}$ be as in the Lévy-Khinchin decomposition in Lemma 5.1. Lemma 5.6 shows that for every $\Psi \in \mathbf{L}^2_{Q^{1/2}(U),T}(H)$ and $t \geq 0$ we have

$$\int_0^t \Psi(s)\,\mathrm{d}M(s) = \int_0^t \iota_0(\Psi)(s)\,\mathrm{d}W(s) + \int_0^t \iota_1(\Psi)(s)\,\mathrm{d}\mathscr{L}(s), \tag{48}$$

where ι_0 and ι_1 denote the respective continuous extensions of the identity map on $\mathscr{S}(U, H)$ defined in Lemma 5.5. In order to find the jumps and quadratic variation of the stochastic integral $\big(I_t^M(\Psi)\big)_{t\geq 0}$ we would like to show that the Wiener integral on the right-hand side of (48) is the continuous part of $\big(I_t^M(\Psi)\big)_{t\geq 0}$ and that the stochastic integral with respect to $\mathscr{L}$ on the right-hand side of (48) is the purely discontinuous part of $\big(I_t^M(\Psi)\big)_{t\geq 0}$. To do this it is sufficient to show that the process $\Big(I_t^{\mathscr{L}}(\Psi)\Big)_{t\in[0,T]}$ is purely discontinuous, i.e., that it is a limit in $\mathscr{M}_T^2(H)$ of finite variation processes that start at 0. We do this by relating the stochastic integration map $I_t^{\mathscr{L}}$ to the stochastic integral $I_t^{\widehat{\pi}}$ from Definition 4.8, where π is the jump measure of $\mathscr{L}$. Since $\mathscr{L}$ is a sum of independent CCPPs, it is natural to study stochastic integration with respect to CCPPs first and to then extend the results to $\mathscr{L}$ using an approximation argument. Before turning our attention to CCPPs in Sect. 5.2, we record some properties of processes of the form $\mathscr{L}$ that will be used later on.

Proposition 5.7 *Let $\mathscr{L}$ be a Lévy process on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$ satisfying Assumption 3.3. Let ν be the Lévy measure of $\mathscr{L}$ (it is possible that $\nu(U \setminus \{0\}) = \infty$). Suppose that the covariance operator Q_1 of $\mathscr{L}$ is given by*

$$(Q_1 x, y)_U = \int_U (u, x)_U\,(u, y)_U\;\mathrm{d}\nu(u) \qquad \text{for all } x, y \in U. \tag{49}$$

Then the following statements hold.

i) For every $\Phi \in L(U, H)$ we have

$$||\Phi||^2_{L_2(Q_1^{1/2}(U),H)} = \int_U |\Phi u|_H^2\,\mathrm{d}\nu(u). \tag{50}$$

ii) For every $\Psi \in \mathscr{S}(U, H)$ the function $f_\Psi^{\mathscr{L}}(s, u) := \Psi(s)u$ belongs to $\mathbf{F}^2_{\nu,T}(H)$ and the linear map $f^{\mathscr{L}}$ on $\mathscr{S}(U, H)$ extends to an isometry $f^{\mathscr{L}} \colon \mathbf{L}^2_{Q_1^{1/2}(U),T}(H) \to \mathbf{F}^2_{\nu,T}(H)$. Furthermore, for every $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ one has

$$\mathbf{E}\int_0^t \int_U |f_\Psi^{\mathscr{L}}(s, u)|_H^2 \, \mathrm{d}\nu(u) \, \mathrm{d}s = \mathbf{E}\int_0^t ||\Psi(s)||^2_{L_2(Q_1^{1/2}(U),H)} \, \mathrm{d}s, \tag{51}$$

for every $t \in [0, T]$.

Proof $i)$ Let $\Phi \in L(U, H)$ and let $(e_k)_{k=1}^\infty$ be an ONB for H. By Tonelli's Theorem we have

$$\int_U |\Phi u|_H^2 \, \mathrm{d}\nu(u) = \sum_{k=1}^\infty \int_U (e_k, \Phi u)_H^2 \, \mathrm{d}\nu(u) = \sum_{k=1}^\infty \int_U \left(\Phi^* e_k, u\right)_U^2 \, \mathrm{d}\nu(u)$$

$$= \sum_{k=1}^\infty \left(Q_1 \Phi^* e_k, \Phi^* e_k\right) = \mathrm{Tr}(\Phi Q_1 \Phi^*) = ||\Phi||^2_{L_2(Q_1^{1/2}(U),H)}.$$

$ii)$ Let $\Psi \in \mathscr{S}(U, H)$ be of the form (23). Define $f_\Psi^{\mathscr{L}} : \Omega \times [0, \infty) \times U \to H$ by

$$f_\Psi^{\mathscr{L}}(\omega, s, u) := \Psi(\omega, s)u = \sum_{j=1}^{m-1} \chi_{A_j}(\omega)\chi_{(t_j,t_{j+1}]}(s)\Phi_j u.$$

Each summand on the right-hand side is $\mathscr{P}_{[0,T]} \otimes \mathscr{B}(U)$-measurable. Using part $i)$ we see that

$$\mathbf{E}\int_0^t \int_U |f_\Psi^{\mathscr{L}}(s, u)|_H^2 \, \mathrm{d}\nu(u) \, \mathrm{d}s = \sum_{j=1}^{m-1} \mathbf{P}(A_j)(t \wedge t_{j+1} - t \wedge t_j) \int_U |\Phi_j u|_H^2 \, \mathrm{d}\nu(u)$$

$$= \sum_{j=1}^{m-1} \mathbf{P}(A_j)(t \wedge t_{j+1} - t \wedge t_j) \left|\left|\Phi_j\right|\right|^2_{L_2(Q_1^{1/2}(U),H)}$$

$$= \mathbf{E}\int_0^t ||\Psi(s)||^2_{L_2(Q_1^{1/2}(U),H)} \, \mathrm{d}s.$$

This shows that $f_\Psi^{\mathscr{L}} \in \mathbf{F}^2_{\nu,T}(H)$ and establishes (51) when Ψ is a simple process. As a result, the map $\Psi \mapsto f_\Psi^{\mathscr{L}}$ on $\mathscr{S}(U, H)$ extends uniquely by continuity to a map from $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H) \to \mathbf{F}^2_{\nu,T}(H)$. The extension continues to satisfy (51) for all $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ and all $t \in [0, T]$. □

Remark 5.8 Condition (49) is satisfied when $\mathscr{L}$ is a compensated compound Poisson process (Proposition 4.18 in [15]), in which case $\nu(U \setminus \{0\}) < \infty$. Condition (49) is also satisfied when $\mathscr{L}$ is as in Lemma 5.1, however, it is possible that $\nu(U \setminus \{0\}) = \infty$.

Example 5.9 Let $\mathscr{L}$ be a U-valued Lévy process satisfying Assumption 3.3. Denote its Lévy measure by ν and assume that its covariance operator Q_1 satisfies (49). We will compute $f_\Psi^{\mathscr{L}}$ for a process $\Psi \in L^2(\Omega \times [0,T], \mathscr{P}_{[0,T]}, \mathrm{d}\mathbf{P} \otimes \mathrm{d}t; L(U,H))$ whose values are bounded operators. Note that equation (50) implies that the space $L(U,H)$ of bounded operators is continuously included in the space $L_2(Q_1^{1/2}(U), H)$. Therefore, the space $L^2(\Omega \times [0,T], \mathscr{P}_{[0,T]}, \mathrm{d}\mathbf{P} \otimes \mathrm{d}t; L(U,H))$ is contained in the space of integrands $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ for stochastic integration with respect to $\mathscr{L}$. Let $\Psi \in L^2(\Omega \times [0,T], \mathscr{P}_{[0,T]}, \mathrm{d}\mathbf{P} \otimes \mathrm{d}t; L(U,H))$ and let $(\Phi_n)_{n=1}^\infty \subset \mathscr{S}(U,H)$ with $\Phi_n \to \Psi$ in the space $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$. Using (50) we see that

$$\mathbf{E} \int_0^T \int_U |\Psi(s)u - \Phi_n(s)u|_H^2 \, \mathrm{d}\nu(u) \, \mathrm{d}s = \mathbf{E} \int_0^T ||\Psi(s) - \Phi_n(s)||^2_{L_2(Q_1^{1/2}(U),H)} \, \mathrm{d}s.$$

The right-hand side tends to 0 as $n \to \infty$, so it follows that $\big(f_{\Phi_n}^{\mathscr{L}}\big)_{n=1}^\infty$ converges in the space $\mathbf{F}^2_{\nu,T}(H)$ to the function $(\omega, s, u) \mapsto \Psi(\omega, s)u$. Since $f^{\mathscr{L}}$ is continuous we conclude that

$$f_\Psi^{\mathscr{L}}(\omega, s, u) = \Psi(\omega, s)u, \qquad \text{for all } (\omega, s, u) \in \Omega \times [0,T] \times U.$$

The result below will be used frequently in Sect. 6.

Lemma 5.10 *Fix a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$. Let ν be a Borel measure on U (possibly with $\nu(U) = \infty$), let $E_1 \subseteq E_2 \in \mathscr{B}(U)$ and set $\nu_j := \nu|_{E_j}$. For $j = 1, 2$ define operators $Q_1, Q_2 \in L_1^+(U)$ by*

$$\big(Q_j x, y\big)_U := \int_{E_j} (x,u)_U \, (y,u)_U \, \mathrm{d}\nu(u), \qquad \text{for all } x, y \in U.$$

Let $\iota \colon \mathbf{L}^2_{Q_2^{1/2}(U),T}(H) \to \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ be the continuous extension of the inclusion map on simple processes (see Lemma 5.5). For $j = 1, 2$ let $f^{(j)} \colon \mathbf{L}^2_{Q_j^{1/2}(U),T}(H) \to \mathbf{F}^2_{\nu_j,T}$ be the map defined in Proposition 5.7. Then for every $\Psi \in \mathbf{L}^2_{Q_2^{1/2}(U),T}(H)$ we have

$$f_\Psi^{(2)}|_{\Omega \times [0,T] \times E_1} = f_{\iota(\Psi)}^{(1)} \tag{52}$$

in the space $\mathbf{F}^2_{\nu_1,T}$.

Proof Suppose that $\Psi \in \mathscr{S}(U, H)$. Then both sides of equation (52) are equal to the function $(\omega, s, u) \mapsto \Psi(\omega, s)u$ on $\Omega \times [0, T] \times E_1$. It is clear that both sides of equation (52) are continuous maps of $\mathbf{L}^2_{Q_2^{1/2}(U),T}(H) \to \mathbf{F}^2_{\nu_1,T}$, so (52) holds for all $\Psi \in \mathbf{L}^2_{Q_2^{1/2}(U),T}(H)$. □

5.2 Integration with Respect to Compensated Compound Poisson Processes

Let $\mathscr{L}$ be as in Lemma 5.1. Our goal in this subsection is to show that $\mathrm{d}\mathscr{L} = \mathrm{d}\widehat{\pi}$, formally, where π is the jump measure of $\mathscr{L}$. We do this first in the case where $\mathscr{L} = \widehat{P}$ is a CCPP, then we use a limiting argument to generalize to U-valued Lévy processes satisfying Assumption 3.3 with covariance operator of the form (49), such as the process $\mathscr{L}$ from Lemma 5.1.

Remark 5.11 In this section we work on a filtered probability space $(\Omega,\mathscr{F}, (\mathscr{F}_t)_{t\geq 0},\mathbf{P})$ and first consider a square-integrable $\mathscr{F}_t$-compound Poisson process P taking values in a real, separable Hilbert space U. The Lévy measure of P is denoted ν and the jump measure of P is denoted π. We recall several properties of square-integrable compound Poisson processes that will be used in this section.

i) According to Definition 2.9 we have $\nu(U \setminus \{0\}) < \infty$.
ii) $\mathbf{F}^2_{\nu,T}(H) \subseteq \mathbf{F}^1_{\nu,T}(H)$ by Corollary 4.12.
iii) Since P is integrable, the CCPP $\widehat{P}(t) := P(t) - \mathbf{E}P(t)$ can be defined. Since P is square integrable, so is $\widehat{P}$ and $\widehat{P}$ satisfies Assumption 3.3 with respect to the filtration $(\mathscr{F}_t)_{t\geq 0}$.
iv) The square-integrable, Lévy martingale $\widehat{P}$ has a covariance operator, say $Q_1 \in L_1^+(U)$, by Theorem 3.4. The covariance operator Q_1 satisfies condition (49) by Proposition 4.18 in [15].

As we have seen so far, there are two notions of stochastic integration that come with a square-integrable compound Poisson process P. First, we can integrate processes $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ using the map $I_T^{\widehat{P}} : \mathbf{L}^2_{Q_1^{1/2}(U),T}(H) \to L^2(\Omega; H)$ defined in Theorem 3.12. Second, by Proposition 4.16 the jump measure π has the form $\pi = N_\Xi$, where Ξ is the stationary Poisson process induced by the jumps of P. So, we can integrate functions $f \in \mathbf{F}^2_{\nu,T}(H)$ using the isometry $I_T^{\widehat{\pi}} : \mathbf{F}^2_{\nu,T}(H) \to L^2(\Omega; H)$ from Definition 4.8. In the next result we show that these two notions of stochastic integration coincide via $I_T^{\widehat{P}} = I_T^{\widehat{\pi}} \circ f^{\widehat{P}}$, where $f^{\widehat{P}}$ is the isometry defined in Proposition 5.7.

Proposition 5.12 *Let P be a square-integrable $\mathscr{F}_t$-compound Poisson process on U with Lévy measure ν, covariance operator Q_1 and jump measure π. The the following statements hold.*

i) *For every* $\Psi \in \mathscr{S}(U, H)$ *we have* $f_\Psi^{\widehat{P}} \in \mathbf{F}^1_{\nu,T}(H) \cap \mathbf{F}^2_{\nu,T}(H)$. *Furthermore, for every* $t \in [0, T]$ *we have*

$$\int_0^t \Psi(s)(\mathbf{E}P(1))\,\mathrm{d}s = \int_0^t \int_U \Psi(s)u\,\mathrm{d}\nu(u)\,\mathrm{d}s, \tag{53}$$

$$\int_0^t \Psi(s)\,\mathrm{d}\widehat{P}(s) = \int_0^t \int_U f_\Psi^{\widehat{P}}(s,u)\,\mathrm{d}\widehat{\pi}(s,u) \tag{54}$$

$$= \sum_{s\in(0,t]} \Psi(s)\Delta P(s) - \int_0^t \int_U \Psi(s)u\,\mathrm{d}\nu(u)\,\mathrm{d}s, \tag{55}$$

and there are finitely many terms in the sum a.s.

ii) *For every* $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ *we have* $f_\Psi^{\widehat{P}} \in \mathbf{F}^1_{\nu,T}(H) \cap \mathbf{F}^2_{\nu,T}(H)$. *Furthermore for every* $t \in [0, T]$ *we have*

$$\int_0^t \Psi(s)\,\mathrm{d}\widehat{P}(s) = \int_0^t \int_U f_\Psi^{\widehat{P}}(s,u)\,\mathrm{d}\widehat{\pi}(s,u) \tag{56}$$

$$= \sum_{s\in(0,t]} f_\Psi^{\widehat{P}}(s, \Delta P(s)) - \int_0^t \int_U f_\Psi^{\widehat{P}}(s,u)\,\mathrm{d}\nu(u)\,\mathrm{d}s, \tag{57}$$

and there are finitely many terms in the sum a.s.

Proof *i)* Since $\nu(U \setminus \{0\}) < \infty$ we have $\mathbf{F}^2_{\nu,T}(H) \subseteq \mathbf{F}^1_{\nu,T}(H)$ by Corollary 4.12, so $f_\Psi^{\widehat{P}} \in \mathbf{F}^1_{\nu,T}(H) \cap \mathbf{F}^2_{\nu,T}(H)$ for every $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$. To prove (53) note that $\Phi\mathbf{E}(P(1)) = \Phi \int_U u\,\mathrm{d}\nu(u) = \int_U \Phi u\,\mathrm{d}\nu(u)$ for all $\Phi \in L(U, H)$ by Proposition 2.11. Therefore, if $\Psi \in \mathscr{S}(U, H)$ is of the form (23), then we have

$$\int_0^t \Psi(s)(\mathbf{E}P(1))\,\mathrm{d}s = \sum_{j=1}^{m-1} \mathbf{P}(A_j)(t \wedge t_{j+1} - t \wedge t_j) \int_U \Phi_j u\,\mathrm{d}\nu(u)$$

and the right-hand side of the equation above is equal to $\int_0^t \int_U \Psi(s)u\,\mathrm{d}\nu(u)\,\mathrm{d}s$. To prove (54) and (55) write $P(t) = \sum_{k=1}^{\Pi(t)} Z_k$, as in Theorem 2.10, where $(Z_k)_{k=1}^\infty$ are i.i.d. U-valued random variables with law $\frac{1}{\nu(U)}\nu$ and Π is a Poisson process with rate $\nu(U)$ that is independent of $(Z_k)_{k=1}^{\infty}$. We have

$$\begin{aligned}\int_0^t \Psi(s)\,\mathrm{d}\widehat{P}(s) &= \sum_{j=1}^{m-1} \chi_{A_j}\Phi_j(\widehat{P}(t\wedge t_{j+1}) - \widehat{P}(t\wedge t_j))\\
&= \sum_{j=1}^{m-1} \chi_{A_j}\Phi_j(P(t\wedge t_{j+1}) - P(t\wedge t_j))\\
&\quad - \sum_{j=1}^{m} \chi_{A_j}(t\wedge t_{j+1} - t\wedge t_j)\Phi_j(\mathbf{E}P(1))\\
&= \sum_{j=1}^{m-1}\sum_{k=\Pi(t\wedge t_j)+1}^{\Pi(t\wedge t_{j+1})} \chi_{A_j}\Phi_j Z_k - \int_0^t \Psi(s)(\mathbf{E}P(1))\,\mathrm{d}s\\
&= \sum_{s\in(0,t]} \Psi(s)\Delta P(s) - \int_0^t \Psi(s)(\mathbf{E}P(1))\,\mathrm{d}s,\end{aligned}$$

which is (55) because of (53). Since ν is a finite measure, P has finitely many jumps in $(0,t]$ a.s., so there are finitely many terms in the sum in (55). Since $f_{\Psi}^{\widehat{P}} \in \mathbf{F}^1_{\nu,T}(H) \cap \mathbf{F}^2_{\nu,T}(H)$ we obtain (54) from (55) and (38).

ii) Equation (54) says that $I_t^{\widehat{P}} = I_t^{\widehat{\pi}} \circ f^{\widehat{P}}$ on the dense subspace $\mathscr{S}(U,H)$ of $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$, so (56) follows by continuity. We have observed that $f_{\Psi}^{\widehat{P}} \in \mathbf{F}^1_{\nu,T}(H) \cap \mathbf{F}^2_{\nu,T}(H)$ for all $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$, so (57) follows from (38). There are finitely many terms in the sum because P has finitely many jumps in $(0,t]$ a.s. □

The jumps of a stochastic integral with respect to $\widehat{P}$ can be identified immediately from equation (57).

Corollary 5.13 *Let P be a square-integrable $\mathscr{F}_t$-compound Poisson process on U with covariance operator Q_1. Then for every $\Psi \in \mathbf{L}^2_{Q^{1/2}(U),T}(H)$ and every $t \in [0,T]$ we have*

$$\Delta \int_0^t \Psi(s)\,\mathrm{d}\widehat{P}(s) = \begin{cases} f_{\Psi}^{\widehat{P}}(t,\Delta P(t)) & \text{if } \Delta P(t) \neq 0\\ 0 & \text{if } \Delta P(t) = 0.\end{cases}$$

The final step is to use a limiting argument to extend formula (56) to integration with respect to square-integrable Lévy martingales with covariance operator satisfying (49). In the remainder of Sect. 5.2 we assume that $\mathscr{L}$ is a square-integrable U-valued Lévy martingale with Lévy measure ν and covariance operator satisfying (49). If the filtration is larger than the natural filtration generated by $\mathscr{L}$, then we also assume that $\mathscr{L}$ satisfies the independence condition (21).

Proposition 5.14 *Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$ be a filtered probability space and suppose that $\mathscr{L}$ is a U-valued Lévy martingale satisfying Assumption 3.3 with Lévy measure ν and covariance operator Q_1 satisfying (49). Then for every $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ and every $t \in [0, T]$ we have*

$$\int_0^t \Psi(s)\,\mathrm{d}\mathscr{L}(s) = \int_0^t \int_U f_\Psi^{\mathscr{L}}(s,u)\,\mathrm{d}\widehat{\pi}(s,u), \tag{58}$$

where π is the jump measure of $\mathscr{L}$ and $f^{\mathscr{L}}$ is the map defined in Proposition 5.7.

Proof In preparation for an approximation argument we begin by applying the Lévy-Khinchin decomposition to $\mathscr{L}$. By Lemma 5.1 there exists a U-valued Wiener process and square-integrable compound Poisson processes $(P_n)_{n=1}^\infty$, all independent, such that $\mathscr{L}(t) = W(t) + \sum_{n=1}^\infty \widehat{P}_n(t)$ in U a.s. for all $t \geq 0$. Furthermore, the series converges in U uniformly in time on compact subsets of $[0,\infty)$ and there exist disjoint Borel sets $(A_n)_{n=1}^\infty$, each separated from 0, such that P_n has Lévy measure $\nu|_{A_n}$ and $\bigcup_{n=1}^\infty A_n = U \setminus \{0\} =: E$. Finally, by (42) it follows that $\mathscr{L}$ has the same covariance operator as the sum $\sum_{n=1}^\infty \widehat{P}_n$. Therefore, the covariance operator of the Wiener part, W, of $\mathscr{L}$ is zero by Lemma 5.6. That is, $W \equiv 0$ and therefore $\mathscr{L} = \sum_{n=1}^\infty \widehat{P}_n$.

Since $\mathscr{L}$ is square-integrable with covariance operator satisfying (49) the hypotheses of Proposition 5.7 apply to $\mathscr{L}$, so the map $f^{\mathscr{L}}$ is defined. For each fixed $t \in [0,T]$, both sides of (58) are continuous functions of Ψ sending $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H) \to L^2(\Omega; H)$. Therefore, it is sufficient to establish (58) for $\Psi \in \mathscr{S}(U,H)$.

Fix $\Psi \in \mathscr{S}(U,H)$. For each positive integer N consider the sum $\mathscr{L}_N := \sum_{n=1}^N \widehat{P}_n$ of independent compensated compound Poisson processes. By (56) we have

$$\int_0^t \Psi(s)\,\mathrm{d}\mathscr{L}_N(s) = \sum_{n=1}^N \int_0^t \Psi(s)\,\mathrm{d}\widehat{P}_n(s) = \sum_{n=1}^N \int_0^t \int_U \Psi(s)u\,\mathrm{d}\widehat{\pi}_n(s,u) \tag{59}$$

where π_n is the jump measure of P_n. We claim that π_n is just π restricted to the set $(0,\infty) \times A_n$. Because the convergence $\mathscr{L}_N \to \mathscr{L}$ is a.s. uniform one can see that

$$\Delta\mathscr{L}(t) = \begin{cases} \Delta P_n(t) & \text{if } \Delta P_n(t) \neq 0 \text{ for some } n \\ 0 & \text{otherwise,} \end{cases}$$

for $t > 0$. Since the jumps of P_n belong to A_n with probability one, we see that $\Delta\mathscr{L}(t) = \Delta P_n(t)$ if and only if $\Delta\mathscr{L}(t) \in A_n$. Therefore,

$$\pi_n = \sum_{\substack{s>0\\ \Delta P_n(s)\neq 0}} \delta_{(s,\Delta P_n(s))} = \sum_{\substack{s>0\\ \Delta\mathscr{L}(s)\neq 0}} \chi_{A_n}(\Delta\mathscr{L}(s))\delta_{(s,\Delta\mathscr{L}(s))} = \pi(\cdot\cap((0,\infty)\times A_n)).$$

This shows that π_n is the restriction of π to $(0,\infty)\times A_n$. On the right-hand side of equation (59) the integrand in each term in the sum belongs to $\mathbf{F}^1_{\nu_n,T}(H)\cap\mathbf{F}^2_{\nu_n,T}(H)$, where $\nu_n := \nu|_{A_n}$, because ν_n is a finite measure. With $E_N := \bigcup_{n=1}^N A_n$ we obtain

$$\begin{aligned}\int_0^t \Psi(s)\,\mathrm{d}\mathscr{L}_N(s) &= \sum_{n=1}^N \Big[\int_0^t\int_U \Psi(s)u\,\mathrm{d}\pi_n(s,u) - \int_0^t\int_U \Psi(s)u\,\mathrm{d}\nu_n(u)\,\mathrm{d}s\Big]\\ &= \sum_{n=1}^N \Big[\int_0^t\int_{A_n} \Psi(s)u\,\mathrm{d}\pi(s,u) - \int_0^t\int_{A_n} \Psi(s)u\,\mathrm{d}\nu(u)\,\mathrm{d}s\Big]\\ &= \int_0^t\int_{E_N} \Psi(s)u\,\mathrm{d}\pi(s,u) - \int_0^t\int_{E_N} \Psi(s)u\,\mathrm{d}\nu(u)\,\mathrm{d}s\\ &= \int_0^t\int_U \chi_{E_N}(u)\Psi(s)u\,\mathrm{d}\widehat{\pi}(s,u).\end{aligned}$$

Using the isometric formula (36) it is easy to see that the right-hand side of the equation above converges to $\int_0^t\int_U \Psi(s)u\,\mathrm{d}\widehat{\pi}(s,u)$ in $L^2(\Omega;H)$ as $N\to\infty$. On the other hand, we have

$$\int_0^t \Psi(s)\,\mathrm{d}\mathscr{L}(s) - \int_0^t \Psi(s)\,\mathrm{d}\mathscr{L}_N(s) = \int_0^t \Psi(s)\,\mathrm{d}(\mathscr{L}-\mathscr{L}_N)(s),$$

because $\Psi \in \mathscr{S}(U,H)$ (note that $\mathscr{L}-\mathscr{L}_N$ satisfies Assumption 3.3 by Corollary 5.4). By writing $\mathscr{L} = (\mathscr{L}-\mathscr{L}_N)+\mathscr{L}_N$ and noting that the summands are independent we find that the covariance operator of $\mathscr{L}-\mathscr{L}_N$ is $Q_1 - Q_{\mathscr{L}_N}$ by part $i)$ of Lemma 5.6, where $Q_{\mathscr{L}_N}$ is the covariance operator of $\mathscr{L}_N$. The same result also implies that

$$\big(Q_{\mathscr{L}_N}x,y\big)_U = \int_{E_N} (x,u)_U\,(y,u)_U\,\mathrm{d}\nu(u) \qquad \text{for all } x,y\in U. \tag{60}$$

Therefore, the covariance operator of $\mathscr{L}-\mathscr{L}_N$ satisfies condition (49) with the measure $\nu|_{E\setminus E_N}$. Using the isometric formula (36) and Proposition 5.7 we see that

$$\mathbf{E}\Big|\int_0^t \Psi(s)\,\mathrm{d}(\mathscr{L}-\mathscr{L}_N)(s)\Big|_H^2 = \mathbf{E}\int_0^t\int_{E\setminus E_N} |\Psi(s)u|_H^2\,\mathrm{d}\nu(u)\,\mathrm{d}s,$$

and the right-hand side tends to 0 as $N \to \infty$ by the dominated convergence theorem. This establishes (58) for simple processes and the general case follows by continuity. □

Remark 5.15 Equation (58) generalizes the special case in (56) of a single square-integrable compensated compound Poisson process to the type of process $\mathscr{L}$ that appears as the square-integrable jump part of a Lévy process in the Lévy-Khinchin decomposition (Lemma 5.1). However, one should not expect that (57) holds for $\mathscr{L}$, because the Lévy measure of $\mathscr{L}$ may not be finite.

Now we can identify the jumps of a stochastic integral with respect to $\mathscr{L}$ (cf. Corollary 5.13).

Corollary 5.16 *Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$ be a filtered probability space and suppose that $\mathscr{L}$ is a U-valued Lévy process satisfying Assumption 3.3 with Lévy measure ν and covariance operator Q_1 satisfying (49). Then for every $t \in [0, T]$ and every $\Psi \in \mathbf{L}^2_{Q^{1/2}(U),T}(H)$ we have*

$$\Delta \int_0^t \Psi(s)\, \mathrm{d}\mathscr{L}(s) = \begin{cases} f_\Psi^{\mathscr{L}}(t, \Delta\mathscr{L}(t)) & \text{if } \Delta\mathscr{L}(t) \neq 0 \\ 0 & \text{if } \Delta\mathscr{L}(t) = 0. \end{cases}$$

Proof As in the proof of Proposition 5.14 we use the Lévy-Khinchin decomposition to write $\mathscr{L} = \sum_{n=1}^\infty \widehat{P}_n$. Recall that $(P_n)_{n=1}^\infty$ are independent square-integrable compound Poisson processes and that the series converges uniformly in time on compact subsets of $[0, \infty)$ a.s. For each positive integer n, the Lévy measure of P_n is $\nu_n := \nu|_{A_n}$, where $(A_n)_{n=1}^\infty$ are disjoint Borel sets, each separated from 0, such that $\bigcup_{n=1}^\infty A_n = U \setminus \{0\}$. As before we define $\mathscr{L}_N := \sum_{n=1}^N \widehat{P}_n$ and $E_N := \bigcup_{n=1}^N A_n$. Since $\mathscr{L}_N$ is a finite sum of independent compound Poisson processes, $\mathscr{L}_N$ is also a compound Poisson process by Proposition 2.11 and its Lévy measure is $\nu|_{E_N}$. Corollary 5.13 tells us the jumps of a stochastic integral with respect to $\mathscr{L}_N$. We denote by Q_N the covariance operator of $\mathscr{L}_N$. Recall from the proof of Proposition 5.14 that Q_N is given by (60) for each N, whence $Q_N \leq Q$. Let $\iota_N \colon \mathbf{L}^2_{Q^{1/2}(U),T}(U) \to \mathbf{L}^2_{Q_N^{1/2}(U),T}(U)$ be the unique continuous extension of the identity map on $\mathscr{S}(U, H)$ (see Lemma 5.5). We show below that for every $\Psi \in \mathbf{L}^2_{Q^{1/2}(U),T}(U)$ one has

$$\lim_{N\to\infty} \mathbf{E}\Big(\sup_{t\in[0,T]} \Big| \int_0^t \iota_N(\Psi)(s)\, \mathrm{d}\mathscr{L}_N(s) - \int_0^t \Psi(s)\, \mathrm{d}\mathscr{L}(s) \Big|_H^2 \Big) = 0. \tag{61}$$

Condition (61) is a stronger form of convergence than what we established in the proof of Proposition 5.14. In that proof we showed that $\int_0^t \iota_N(\Psi)(s)\, \mathrm{d}\mathscr{L}_N(s) \to \int_0^t \Psi(s)\, \mathrm{d}\mathscr{L}(s)$ in the space $L^2(\Omega; H)$ for each fixed $t \in [0, T]$ and $\Psi \in \mathscr{S}(U, H)$.

By Propositions 5.12 and 5.14, (61) is equivalent to

$$\lim_{N\to\infty} \mathbf{E}\Big(\sup_{t\in[0,T]} \Big| \int_0^t \int_U f^{\mathscr{L}_N}_{\iota_N(\Psi)}(s,u)\, \mathrm{d}\widehat{\pi}_N(s,u) - \int_0^t \int_U f^{\mathscr{L}}_{\Psi}(s,u)\, \mathrm{d}\widehat{\pi}(s,u) \Big|_H^2 \Big) = 0, \tag{62}$$

where π is the jump measure of $\mathscr{L}$ and π_N is the jump measure of $\mathscr{L}_N$. Recall from the proof of Proposition 5.14 that $\pi_N = \pi|_{(0,\infty)\times E_N}$; thus

$$\int_0^t \int_U f^{\mathscr{L}_N}_{\iota_N(\Psi)}(s,u)\, \mathrm{d}\widehat{\pi}_N(s,u) = \int_0^t \int_U f^{\mathscr{L}_N}_{\iota_N(\Psi)}(s,u) \chi_{E_N}(u)\, \mathrm{d}\widehat{\pi}(s,u),$$

for each positive integer N and $t \in [0,T]$. By the BDG inequality (cf. (40)) and Theorem 4.7 we have

$$\begin{aligned} \mathbf{E}\Big(\sup_{t\in[0,T]} \Big| \int_0^t \int_U \big(f^{\mathscr{L}_N}_{\iota_N(\Psi)}(s,u) \chi_{E_N}(u) - f^{\mathscr{L}}_{\Psi}(s,u) \big)\, \mathrm{d}\widehat{\pi}(s,u) \Big|_H^2 \Big) & \\ \lesssim \mathbf{E} \int_0^t \int_U \big| f^{\mathscr{L}_N}_{\iota_N(\Psi)}(s,u) \chi_{E_N}(u) - f^{\mathscr{L}}_{\Psi}(s,u) \big|_H^2\, \mathrm{d}\pi(s,u) & \\ = \mathbf{E} \int_0^t \int_U \big| f^{\mathscr{L}_N}_{\iota_N(\Psi)}(s,u) \chi_{E_N}(u) - f^{\mathscr{L}}_{\Psi}(s,u) \big|_H^2\, \mathrm{d}\nu(u)\, \mathrm{d}s. & \end{aligned}$$

Here and below we use $\lesssim$ to denote the inequality $\leq$ up to a universal multiplicative constant; in this case, up to a constant that is independent of N. Lemma 5.10 says that $f^{\mathscr{L}_N}_{\iota_N(\Psi)} = f^{\mathscr{L}}_{\Psi}$ on $\Omega \times [0,T] \times E_N$, so the right-hand side of the inequality above is equal to the square-norm of $f^{\mathscr{L}}_{\Psi}(s,u)\chi_{U\setminus E_N}$ in the space $\mathbf{F}^2_{\nu,T}(H)$, which tends to 0 as $N \to \infty$ by the dominated convergence theorem. This proves (62) and the equivalent statement (61). As a result, we see that $\int_0^t \iota_N(\Psi)(s)\, \mathrm{d}\mathscr{L}_N(s) \to \int_0^t \Psi(s)\, \mathrm{d}\mathscr{L}(s)$ in H, uniformly in $t \in [0,T]$, along a subsequence a.s. Since the convergence is uniform and since $\big(\big(\int_0^t \iota_N(\Psi)(s)\, \mathrm{d}\mathscr{L}_N(s) \big)_{t\in[0,T]} \big)_{N=1}^{\infty}$ are purely discontinuous processes whose sets of jumps increase with N it follows that $\big(\int_0^t \Psi(s)\, \mathrm{d}\mathscr{L}(s) \big)_{t\in[0,T]}$ has a jump at time $t \in [0,T]$ if and only if some process $\big(\int_0^t \iota_N(\Psi)(s)\, \mathrm{d}\mathscr{L}_N(s) \big)_{t\in[0,T]}$ has a jump at time t. Furthermore, when this occurs we have

$$\Delta \int_0^t \Psi(s)\, \mathrm{d}\mathscr{L}(s) = \Delta \int_0^t \iota_N(\Psi)(s)\, \mathrm{d}\mathscr{L}_N(s),$$

for every positive integer N such that $\big(\int_0^t \iota_N(\Psi)(s)\, \mathrm{d}\mathscr{L}_N(s) \big)_{t\in[0,T]}$ has a jump at time t. We have used the same reasoning in the proof of Proposition 5.14 to show that $\Delta\mathscr{L}(t) = \Delta\mathscr{L}_N(t)$ for any N such that $\Delta\mathscr{L}_N(t) \neq 0$ and $\Delta\mathscr{L}(t) = 0$

otherwise. By Corollary 5.13 we have

$$\Delta \int_0^t \iota_N(\Psi)(s)\,\mathrm{d}\mathscr{L}_N(s) = \begin{cases} f^{\mathscr{L}_N}_{\iota_N(\Psi)}(t, \Delta\mathscr{L}_N(t)) & \text{if } \Delta\mathscr{L}_N(t) \neq 0 \\ 0 & \text{if } \Delta\mathscr{L}_N(t) = 0 \end{cases}$$

for each positive integer N. Since $\Delta\mathscr{L}_N(t) \in E_N$, Lemma 5.10 implies that

$$\Delta \int_0^t \Psi(s)\,\mathrm{d}\mathscr{L}(s) = \begin{cases} f^{\mathscr{L}_N}_{\iota_N(\Psi)}(t, \Delta\mathscr{L}_N(t)) & \text{if } \Delta\mathscr{L}_N(t) \neq 0 \text{ for some } N \\ 0 & \text{if } \Delta\mathscr{L}(t) = 0. \end{cases}$$

From this we find that

$$\Delta \int_0^t \Psi(s)\,\mathrm{d}\mathscr{L}(s) = \begin{cases} f^{\mathscr{L}}_{\Psi}(t, \Delta\mathscr{L}(t)) & \text{if } \Delta\mathscr{L}(t) \neq 0 \\ 0 & \text{if } \Delta\mathscr{L}(t) = 0. \end{cases}$$

□

We can find the quadratic variation of a stochastic integral with respect to $\mathscr{L}$ by combining Proposition 5.14 with Theorem 4.14.

Corollary 5.7 *Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$ be a filtered probability space and suppose that $\mathscr{L}$ is a U-valued Lévy process satisfying Assumption 3.3 with Lévy measure ν and covariance operator Q_1 satisfying (49). Then for every $t \in [0, T]$ and every $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ we have*

$$\left[\textstyle\int_0^{\cdot} \Psi(s)\,\mathrm{d}\mathscr{L}(s)\right]_t = \int_0^t \int_U |f^{\mathscr{L}}_{\Psi}(s, u)|^2_H\,\mathrm{d}\pi(s, u), \tag{63}$$

where π is the jump measure of $\mathscr{L}$ and $f^{\mathscr{L}}$ is the map defined in Proposition 5.7.

5.3 Summary of the Square-Integrable Case

Now we return to the setting where M is a square-integrable Lévy martingale satisfying Assumption 3.3, W is its Wiener part and π is its jump measure. We combine the results in this section to give a rigorous interpretation and proof of the informal statement $\mathrm{d}M = \mathrm{d}W + \mathrm{d}\widehat{\pi}$. This will show the precise way in which the stochastic integral I^M as presented by Peszat and Zabczyk is a special case of the stochastic integrals $I^W + I^{\widehat{\pi}}$ that appears in the setting presented by Ikeda and Watanabe. We recall the setting.

- We are given a square-integrable, U-valued Lévy martingale M with covariance operator Q.

- A filtration $(\mathscr{F}_t)_{t\geq 0}$ is chosen such that M is an $\mathscr{F}_t$-Lévy process, so M satisfies Assumption 3.3. We may assume that the filtration is complete and right-continuous.
- The process M can be decomposed as $M = W + \mathscr{L}$ as in Lemma 5.1, where W is an $\mathscr{F}_t$-Wiener process on U and $\mathscr{L}$ is an $\mathscr{F}_t$-Lévy process and a sum of independent CCPPs on U.
- Let Q_0 be the covariance operator of W and let Q_1 be the covariance operator of $\mathscr{L}$.
- For $j = 0, 1$ let $\iota_j\colon \mathbf{L}^2_{Q^{1/2}(U),T}(H) \to \mathbf{L}^2_{Q_j^{1/2}(U),T}(H)$ denote the continuous extension of the identity map on $\mathscr{S}(U, H)$ from Lemma 5.5.
- Let ν be the Lévy measure of M and let π be the jump measure of M.
- Let $f^{\mathscr{L}}\colon \mathbf{L}^2_{Q_1^{1/2}(U),T}(H) \to \mathbf{F}^2_{\nu,T}(H)$ be the map defined in Proposition 5.7.

By combining Lemma 5.6 and Proposition 5.14 we obtain the following.

Theorem 5.18 *In the setup above, for every $\Psi \in \mathbf{L}^2_{Q^{1/2}(U),T}(H)$ and for every $t \in [0, T]$ we have*

$$\int_0^t \Psi(s)\,\mathrm{d}M(s) = \int_0^t \iota_0(\Psi)(s)\,\mathrm{d}W(s) + \int_0^t \int_U f^{\mathscr{L}}_{\iota_1(\Psi)}(s, u)\,\mathrm{d}\widehat{\pi}(s, u), \tag{64}$$

i.e., $I^M_t(\Psi) = I^W_t(\iota_0(\Psi)) + I^{\widehat{\pi}}_t(f^{\mathscr{L}}_{\iota_1(\Psi)})$.

Question 1.2 from Sect. 1 can now be answered for the stochastic integral with respect to M using Corollary 5.16. For $\Psi \in \mathbf{L}^2_{Q^{1/2}(U),T}(H)$ the jumps of $I^M(\Psi)$ are

$$\Delta I^M_t(\Psi) = \begin{cases} f^{\mathscr{L}}_{\iota_1(\Psi)}(t, \Delta M(t)) & \text{if } \Delta M(t) \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{65}$$

Question 1.3 from Sect. 1 can be answered as well. We can compute the quadratic variation of the process $I^M(\Psi)$, where $\Psi \in \mathbf{L}^2_{Q^{1/2}(U),T}(H)$, using Theorem 2.29 by decomposing the H-valued L^2-martingale $I^M(\Psi)$ into its continuous and purely discontinuous parts. We have seen in Example 3.13 that $I^W(\iota_0(\Psi))$ is continuous and in Remark 4.13 that $I^{\widehat{\pi}}(f^{\mathscr{L}}_{\iota_1(\Psi)})$ is purely discontinuous. From this, Theorem 2.29, (29) and (39) we obtain an answer to Question 1.3 using Corollary 5.7.

Theorem 5.19 *In the setup above, for every $\Psi \in \mathbf{L}^2_{Q^{1/2}(U),T}(H)$, the continuous part of the stochastic integral $I^M(\Psi)$ is $I^W(\iota_0(\Psi))$ and the purely discontinuous part of $I^M(\Psi)$ is $I^{\widehat{\pi}}(f^{\mathscr{L}}_{\iota_1(\Psi)})$. Therefore, the quadratic variation of $I^M(\Psi)$ is given by*

$$\left[I^M(\Psi)\right]_t = \int_0^t ||\iota_0(\Psi)(s)||^2_{L_2(Q_0^{1/2},H)}\,\mathrm{d}s + \int_0^t \int_U |f^{\mathscr{L}}_{\iota_1(\Psi)}|^2_H(s, u)\,\mathrm{d}\pi(s, u). \tag{66}$$

We can now state the upper bound in the BDG inequality (Theorem 2.23) applied to the stochastic integral $I^M(\Psi)$.

Corollary 5.20 *For every $1 \leq p < \infty$ there exists a constant $C_p > 0$ with the property that for every square-integrable, mean-zero Lévy process M as in the setup above, for every $\mathscr{F}_t$-stopping time τ and for every $\Psi \in \mathbf{L}^2_{Q^{1/2}(U),T}(H)$ one has*

$$\mathbf{E}\Big(\sup_{t\in[0,\tau]} |I_t^M(\Psi)|_H^p\Big) \leq C_p\mathbf{E}\Big[\Big(\int_0^\tau ||\iota_0(\Psi)(s)||^2_{L_2(Q_0^{1/2}(U),H)}\,\mathrm{d}s \\ + \int_{(0,\tau]}\int_U |f^{\mathscr{L}}_{\iota_1(\Psi)}(s,u)|_H^2\,\mathrm{d}\pi(s,u)\Big)^{p/2}\Big]. \tag{67}$$

In particular,

$$\mathbf{E}\Big(\sup_{t\in[0,\tau]} |I_t^M(\Psi)|_H^2\Big) \leq C_2\mathbf{E}\int_0^\tau ||\Psi(s)||^2_{L_2(Q^{1/2}(U),H)}\,\mathrm{d}s. \tag{68}$$

Proof Inequality (67) is a direct application of the BDG inequality (Theorem 2.23) using formula (66) for the quadratic variation of $I^M(\Psi)$. When $p = 2$ inequality (67) becomes

$$\mathbf{E}\Big(\sup_{t\in[0,\tau]} |I_t^M(\Psi)|_H^2\Big) \leq C_2\mathbf{E}\Big[\int_0^\tau ||\iota_0(\Psi)(s)||^2_{L_2(Q_0^{1/2}(U),H)}\,\mathrm{d}s \\ + \int_{(0,\tau]}\int_U |f^{\mathscr{L}}_{\iota_1(\Psi)}(s,u)|_H^2\,\mathrm{d}\nu(u)\,\mathrm{d}s\Big], \tag{69}$$

by Theorem 4.7. We obtain (68) from (69) using the fact that ι_0, ι_1 and $f^{\mathscr{L}}$ all have norm at most 1 on their respective domains. Due to the presence of the stopping time τ in (69), there is also a limiting argument required to deduce (68), wherein Ψ is approximated by simple processes. Alternatively, inequality (68) can be deduced from (69) using the more general observations in Lemma 6.15, Lemma 6.16 and Remark 6.17. □

We turn our attention next to applying the Itô formula to the solution of an SDE with noise from a square-integrable Lévy martingale M. Let M be as above and let $\Psi \in \mathbf{L}^2_{Q^{1/2}(U),T}(H)$. Let A be an H-valued process with paths of finite variation and define

$$X(t) := X_0 + A(t) + \int_0^t \Psi(s)\,\mathrm{d}M(s), \tag{70}$$

where $X_0 \in L^2(\Omega, \mathscr{F}_0, \mathbf{P}; H)$. Solutions to SDEs will have the form of the process X when, for instance, the process A is of the form $A(t) = \int_0^t F(s)\,\mathrm{d}s$ for some

$F\colon \Omega \times [0,T] \to H$ such that $F \in L^1([0,T];H)$ a.s. By (65) the jumps of X are given by

$$\Delta X(t) = \Delta I_t^M(\Psi) = \begin{cases} f_{\iota_1(\Psi)}^{\mathscr{L}}(t, \Delta M(t)) & \text{if } \Delta M(t) \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{71}$$

In order to apply Theorem 1.1 to X we also need to compute the continuous part of the tensor quadratic variation of $I^M(\Psi)$. According to Definition 2.30, this is given by

$$[[I^M(\Psi), I^M(\Psi)]]_t^c = \sum_{j,k=1}^{\infty} [\Big(e_k, I^M(\Psi)\Big)_H^c, \Big(e_j, I^M(\Psi)\Big)_H^c]_t (e_k \otimes e_j) \tag{72}$$

for any orthonormal basis $(e_k)_{k=1}^{\infty}$ of H. By Theorem 3.12 we have $\big(e_k, I^M(\Psi)\big)_H = I^M((e_k,\Psi)_H)$ and by Theorem 5.19 its continuous part is $I^W(\iota_{0,\mathbb{R}}((e_k,\Psi)_H))$, where $\iota_{0,\mathbb{R}}\colon \mathbf{L}^2_{Q^{1/2}(U),T}(\mathbb{R}) \to \mathbf{L}^2_{Q_0^{1/2}(U),T}(\mathbb{R})$ denotes the continuous extension of the identity map on $\mathscr{S}(U,\mathbb{R})$. When Ψ is a simple process we clearly have

$$\iota_{0,\mathbb{R}}((e_k,\Psi)_H) = (e_k, \iota_0(\Psi))_H\,. \tag{73}$$

It follows by continuity that (73) holds for all $\Psi \in \mathbf{L}^2_{Q^{1/2}(U),T}(H)$. This shows that $\big(e_k, I^M(\Psi)\big)_H^c = I^W((e_k,\iota_0(\Psi))_H)$. We use Theorem 3.12 again to conclude that $\big(e_k, I^M(\Psi)\big)_H^c = \big(e_k, I^W(\iota_0(\Psi))\big)_H$. Returning to (72) we see that

$$[[I^M(\Psi), I^M(\Psi)]]_t^c = [[I^W(\iota_0(\Psi)), I^W(\iota_0(\Psi))]]_t. \tag{74}$$

Applying Theorem 1.1 to the solution X to (70) yields the following. Let $\psi\colon H \to \mathbb{R}$ be a C^2 function with the property that ψ, $D\psi$ and $D^2\psi$ are uniformly continuous on bounded subsets of H. Then for each $t \geq 0$ we have

$$\begin{aligned} \psi(X(t)) = \psi(X_0) &+ \int_0^t (D\psi(X(s-)), \mathrm{d}X(s))_H \\ &+ \frac{1}{2}\int_0^t D^2\psi(X(s-))\,\mathrm{d}[[I^W(\iota_0(\Psi)), I^W(\iota_0(\Psi))]]_s \\ &+ \sum_{s\in(0,t]} \big(\Delta(\psi(X(s)) - (D\psi(X(s-)), \Delta X(s))_H\big) \qquad \mathbf{P}\text{-a.s.} \end{aligned}$$

Using (71) and (74) we can write the equation above explicitly in terms of the coefficients F and Ψ that appear in (70):

$$\begin{aligned}\psi(X(t)) =&\psi(X_0)+\int_0^t (D\psi(X(s-)), F(s))_H \,\mathrm{d}s+\int_0^t (D\psi(X(s-)), \Psi(s)\,\mathrm{d}M(s))_H\\ &+\frac{1}{2}\int_0^t \mathrm{Tr}[D^2\psi(X(s-))\iota_0(\Psi)(s)\big(\iota_0(\Psi)(s)\big)^*]\,\mathrm{d}s\\ &+\int_0^t\int_U \big[\psi(X(s-)+f^{\mathscr{L}}_{\iota_1(\Psi)}(s,u))-\psi(X(s-))\\ &\qquad -\big(D\psi(X(s-)), f^{\mathscr{L}}_{\iota_1(\Psi)}(s,u)\big)_H\big]\,\mathrm{d}\pi(s,u) \qquad \mathbf{P}\text{-a.s.} \qquad (75)\end{aligned}$$

5.4 An Application to Markov Properties

We consider an SDE with Lévy noise as in the setting of Peszat and Zabczyk and show how Theorem 5.18 and the Itô formula (75) can be used to analyze the solution as a Markov process. With M as above we consider equation (70) in the following special form:

$$\begin{cases}\mathrm{d}X(t) = F(X(t-))\,\mathrm{d}t + G(X(t-))\,\mathrm{d}M(t),\\ X(0)=x,\end{cases} \qquad (76)$$

where $F\colon H \to H$, $G\colon H \to L_2(Q^{1/2}(U), H)$ and $x \in H$. We assume that F and G are Lipschitz with linear growth. Under these conditions equation (76) possesses a unique solution X belonging to the space $L^2(\Omega; L^\infty([0,T], H))$, see e.g. [5] or [15]. Furthermore, X is predictable, X has a càdlàg version and X is a Markov process, see e.g. Theorem 9.30 in [15]. As an application of the Itô formula (75) we will determine the transition semigroup associated to X on certain test functions $\psi\colon H \to \mathbb{R}$ for which (75) holds. Recall that the transition semigroup $(T_t)_{t\geq 0}$ associated to X is defined on bounded measurable functions $\psi\colon H \to \mathbb{R}$ by

$$(T_t\psi)(x) := \mathbf{E}_x[\psi(X(t))], \qquad \text{for all } x \in H. \qquad (77)$$

A continuous, bounded function $\psi\colon H \to \mathbb{R}$ is said to belong to the domain of the weak generator of $(T_t)_{t\geq 0}$ if the limit

$$\lim_{t\downarrow 0}\frac{(T_t\psi)(x)-\psi(x)}{t} =: (\mathscr{A}\psi)(x)$$

exists for every $x \in H$, the function $\mathscr{A}\psi$ is continuous and bounded, and

$$(T_t\psi)(x) - \psi(x) = \int_0^t (T_s(\mathscr{A}\psi))(x)\,\mathrm{d}s, \qquad \text{for all } x \in H. \tag{78}$$

We will also determine the weak generator $\mathscr{A}$ of $(T_t)_{t\geq 0}$ on certain test functions. Before doing so, we will examine the terms $\iota_0(G(X))(s)$ and $f^{\mathscr{L}}_{\iota_i(G(X))}(s,u)$ that appear when (75) is applied to the solution X of (76). Specifically, we will show that $\iota_0(G(X))(s)$ and $f^{\mathscr{L}}_{\iota_i(G(X))}(s,u)$ can be expressed as deterministic mappings of H to H evaluated at the point $X(s-)$.

Let $Q_1, Q_2 \in L_1^+(U)$ with $Q_1 \leq Q_2$. By part $i)$ of Lemma 5.5 it follows that the identity operator on $L(U,H)$ extends to a continuous linear mapping $\gamma\colon L_2(Q_2^{1/2}(U),H) \to L_2(Q_1^{1/2}(U),H)$ with norm at most 1. Let $\iota\colon \mathbf{L}^2_{Q_2^{1/2}(U),T}(H) \to \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ be the unique continuous extension of the identity operator on the simple processes $\mathscr{S}(U,H)$, as defined in Lemma 5.5. It is clear that for every $\Psi \in \mathscr{S}(U,H)$ we have $\iota(\Psi) = \Psi = \gamma \circ \Psi$. Since γ is continuous it follows that $\Psi \mapsto \gamma \circ \Psi$ is a continuous linear mapping of $\mathbf{L}^2_{Q_2^{1/2}(U),T}(H)$ into $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$. Since ι is unique we conclude that $\iota(\Psi) = \gamma\circ\Psi$ in the space $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ for every $\Psi \in \mathbf{L}^2_{Q_2^{1/2}(U),T}(H)$. Equivalently, for every $\Psi \in \mathbf{L}^2_{Q_2^{1/2}(U),T}(H)$ we have

$$\iota(\Psi)(s) = \gamma(\Psi(s)), \qquad \text{in } L_2(Q_1^{1/2}(U),H),\ \ \mathrm{d}\mathbf{P}\otimes\mathrm{d}t\text{-a.e.} \tag{79}$$

Now we explain how (79) will be used when applying the Itô formula (75) to the solution X of equation (76). Recall that the square-integrable Lévy martingale M has been decomposed as $M = W + \mathscr{L}$, where W is a U-valued $\mathscr{F}_t$-Wiener process and where $\mathscr{L}$ is an $\mathscr{F}_t$-Lévy process that is a sum of independent CCPPs on U. The covariance operator of M is $Q = Q_0 + Q_1$, where Q_0 is the covariance operator of W and Q_1 is the covariance operator of $\mathscr{L}$. For $j = 0,1$ we denote by $\iota_j\colon \mathbf{L}^2_{Q^{1/2}(U),T}(H) \to \mathbf{L}^2_{Q_j^{1/2}(U),T}(H)$ the continuous linear extension of the identity map on $\mathscr{S}(U,H)$, as in Lemma 5.5. We will also denote by $\gamma_j\colon L_2(Q^{1/2}(U),H) \to L_2(Q_j^{1/2}(U),H)$ the continuous extension of the identity operator on $L(U,H)$ that was defined at the beginning of this paragraph. We apply (79) to the process $\Psi \in \mathbf{L}^2_{Q^{1/2}(U),T}(H)$ defined by $\Psi(s) := G(X(s-))$ and find that

$$\iota_j(G(X))(s) = \gamma_j(G(X(s-))), \qquad \text{in } L_2(Q_j^{1/2}(U),H),\ \ \mathrm{d}\mathbf{P}\otimes\mathrm{d}t\text{-a.e.}, \tag{80}$$

for $j = 0,1$. This shows that the process $\iota_j(G(X))$ is nothing but the continuous function $\gamma_j \circ G\colon H \to L_2(Q_j^{1/2}(U),H)$ evaluated along the paths of X (after modifying them to be left-continuous). Our next task is to show that $f^{\mathscr{L}}_{\iota_i(G(X))}(s,\cdot)$ can also be written as a deterministic continuous function evaluated at $X(s-)$. Note that equation (50) says that the map $\Phi \mapsto \Phi u$ is an isometry from the space $L(U,H)$ endowed with the $L_2(Q_1^{1/2}(U),H)$-norm to the space $L^2(U,\mathscr{B}(U),\nu;H)$. Therefore, this map extends uniquely to an isometry $\phi\colon L_2(Q_1^{1/2}(U),H) \to L^2(U,\mathscr{B}(U),\nu;H)$. Note that for each $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ the composition $\phi \circ \Psi$ belongs to the space

$$L^2(\Omega \times [0,T], \mathscr{P}_{[0,T]}, \mathrm{d}\mathbf{P} \otimes \mathrm{d}t; L^2(U,\mathscr{B}(U),\nu;H)),$$

which can be naturally identified with $\mathbf{F}^2_{\nu,T}(H)$. Furthermore, we have

$$\mathbf{E}\int_0^T \int_U |\phi(\Psi(s))(u)|_H^2 \,\mathrm{d}\nu(u)\,\mathrm{d}s = \mathbf{E}\int_0^T ||\Psi(s)||^2_{L_2(Q_1^{1/2}(U),H)}\,\mathrm{d}s,$$

because ϕ is an isometry from $L_2(Q_1^{1/2}(U),H)$ to $L^2(U,\mathscr{B}(U),\nu;H)$. Since we have $\phi(\Psi(s))(u) = \Psi(s)u$ for every $\Psi \in \mathscr{S}(U,H)$ and every $s \in [0,T]$ and $u \in U$, it follows that

$$f^{\mathscr{L}}_{\Psi} = \phi \circ \Psi \quad \text{in the space } \mathbf{F}^2_{\nu,T}(H), \text{ for every } \Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H). \tag{81}$$

For the process $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ defined by $\Psi(s) := \iota_1(G(X))(s)$ equations (81) and (80) show that

$$f^{\mathscr{L}}_{\iota_1(G(X))}(s,\cdot) = \phi(\iota_1(G(X))(s)) = \phi(\gamma_1(G(X(s-)))), \qquad \mathrm{d}\mathbf{P}\otimes\mathrm{d}t\text{-a.e.}, \tag{82}$$

in the space $L^2(U,\mathscr{B}(U),\nu;H)$.

We can now compute the transition semigroup of the solution X to equation (76). Let $\psi\colon H \to \mathbb{R}$ be of class C^2 such that ψ, $D\psi$ and $D^2\psi$ are uniformly continuous on bounded subsets of H. By applying the Itô formula (75) to equation (76) and taking expectations we obtain

$$\begin{aligned}\mathbf{E}[\psi(X(t))] = \psi(x) &+ \int_0^t \mathbf{E}[(D\psi(X(s-)), F(X(s-)))_H]\,\mathrm{d}s \\ &+ \frac{1}{2}\int_0^t \mathbf{E}\big[\mathrm{Tr}[D^2\psi(X(s-))\iota_0(G(X))(s)\big(\iota_0(G(X))(s)\big)^*]\big]\,\mathrm{d}s \\ &+ \mathbf{E}\int_0^t \int_U \Big[\psi(X(s-) + f^{\mathscr{L}}_{\iota_1(G(X))}(s,u)) - \psi(X(s-)) \\ &\qquad - (D\psi(X(s-)), f^{\mathscr{L}}_{\iota_1(G(X))}(s,u))_H\Big]\,\mathrm{d}\nu(u)\,\mathrm{d}s.\end{aligned}$$

Using equations (81) and (80) we find that

$$\begin{aligned}(T_t\psi)(x) = \psi(x) &+ \int_0^t \mathbf{E}[(D\psi(X(s-)),\, F(X(s-)))_H]\,\mathrm{d}s \\ &+ \frac{1}{2}\int_0^t \mathbf{E}\big[\mathrm{Tr}[D^2\psi(X(s-))\gamma_0(G(X(s-)))\big(\gamma_0(G(X(s-)))\big)^*]\big]\,\mathrm{d}s \\ &+ \int_0^t \mathbf{E}\int_U \Big[\psi(X(s-) + \phi(\gamma_1(G(X(s-))))(u)) - \psi(X(s-)) \\ &\qquad - (D\psi(X(s-)), \phi(\gamma_1(G(X(s-))))(u))_H\Big]\,\mathrm{d}\nu(u)\,\mathrm{d}s. \end{aligned} \tag{83}$$

If we define $\mathscr{A}\psi\colon H \to \mathbb{R}$ by

$$\begin{aligned}(\mathscr{A}\psi)(x) :=\ &(D\psi(x),\, F(x))_H + \frac{1}{2}\mathrm{Tr}[D^2\psi(x)\gamma_0(G(x))\big(\gamma_0(G(x))\big)^*] \\ &+ \int_U \big[\psi(x + \phi(\gamma_1(G(x)))(u)) - \psi(x) \\ &\qquad - (D\psi(x), \phi(\gamma_1(G(x)))(u))_H\big]\,\mathrm{d}\nu(u), \end{aligned} \tag{84}$$

then (83) can be expressed succinctly as

$$(T_t\psi)(x) - \psi(x) = \int_0^t (T_s(\mathscr{A}\phi))(X(s))\,\mathrm{d}s. \tag{85}$$

Using the fundamental theorem of calculus we find that

$$\lim_{t\downarrow 0} \frac{\mathbf{E}[\psi(X(t))] - \psi(x)}{t} = (\mathscr{A}\psi)(x), \qquad \text{for all } x \in H. \tag{86}$$

In the language of Markov processes (see e.g. [15]), we have shown that the transition semigroup of X is given by (83) for test functions ψ that are C^2 with ψ, $D\psi$ and $D^2\psi$ uniformly continuous on bounded subsets of H. Equations (85) and (86) show that such functions ψ for which $\mathscr{A}\psi$ is bounded belong to the domain of the weak generator $\mathscr{A}$ of $(T_t)_{t\geq 0}$.

6 Comparing Lévy Noise: The Non-Square-Integrable Case

In order to define stochastic integration with noise from a general Lévy process in the setting presented by Peszat and Zabczyk, one must consider stochastic integration with respect to a compound Poisson process P (see Definition 2.9) along with stochastic integration with respect to a square-integrable Lévy martingale M (cf. Theorem 2.15). A compound Poisson process P is not a martingale and is not necessarily square-integrable, so the results of Sect. 5 do not apply to P directly. If P is integrable but not square-integrable, then the results of Sect. 5 do not apply to the compensated compound Poisson process $\widehat{P}$ either. In fact, $\widehat{P}$ does not have a trace-class covariance operator, so one cannot even define the space of integrands $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ in the same way as before. For these reasons, stochastic integration with respect to a compound Poisson process P must be defined in a different way from Sect. 5. In this section we summarize Peszat and Zabczyk's presentation of the construction of stochastic integration with respect to a compound Poisson process by localization. However, we adopt the abstract framework of projective limits of Hilbert spaces as the setting for the localization procedure. We choose to use this abstract setting in order to explain subtle points that are not mentioned by Peszat and Zabczyk. Specifically, in Peszat and Zabczyk's book [15]

- it is not clear on page 123 which processes satisfy hypothesis (H3) and make up the space of integrands for stochastic integration with respect to P,
- it is not immediately clear how to interpret Ψu when $\Psi \in L_2(U_0, H)$ and $u \in U$, which is required to define the term $\chi_{[0,\tau_m]}(s)\Psi_2(s)u_m$ on page 125 and
- it is not clear how to define the integral $\int_0^t \chi_{[0,\tau_m]}(s)\Psi_2(s)u_m \, \mathrm{d}s$.

Our abstract setting is also used to address the additional questions below that are not treated by Peszat and Zabczyk.

Question 6.1 What is the appropriate space of integrands for stochastic integration with respect to P?

Question 6.2 Are simple processes dense in the space of integrands?

In order to state the remaining questions, suppose that Ψ belongs to the space of integrands for stochastic integration with respect to P.

Question 6.3 Can the stochastic integral of Ψ with respect to P be expressed as a stochastic integral with respect to the jump measure of P?

Question 6.4 Is the stochastic integral of Ψ with respect to P a random sum of finitely many vectors in H?

Question 6.5 What are the jumps of the stochastic integral of Ψ with respect to P?

Question 6.6 If Ψ takes values in the space $L(U, H)$ of bounded linear operators, then does the stochastic integral of Ψ with respect to P agree with the pathwise Riemann-Stieltjes integral $\sum\limits_{s\in(0,t]} \Psi(s)\Delta P(s)$?

In preparation for constructing the stochastic integral with respect to a compound Poisson process P we begin by considering the case where P is square-integrable in Sect. 6.1. In Sect. 6.2 we construct the stochastic integral with respect to a general, not necessarily integrable, compound Poisson process P. We define the space of integrands for stochastic integration with respect to a CPP in (105), answering Question 6.1. We answer Question 6.2 in Proposition 6.14. The stochastic integral with respect to P is defined in Definition 6.20. We answer Questions 6.3 and 6.4 in Proposition 6.24 and use this to answer Question 6.5 in Corollary 6.26. We answer Question 6.6 in Proposition 6.27. In Sect. 6.3 we compare two notions of stochastic integration with respect to a square-integrable compound Poisson process, the first being Definition 6.7 in Sect. 6.1, the second being Definition 6.20 in Sect. 6.2. In Sect. 6.4 we show how the notion of stochastic integration with respect to a general, non-square-integrable Lévy process presented by Peszat and Zabczyk can be converted into the setting of stochastic integration with respect to a Lévy process presented by Ikeda and Watanabe.

6.1 Integration with Respect to a Square-Integrable Compound Poisson Process

In this section we define stochastic integration with respect to a *square-integrable* compound Poisson process P. We work on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$ such that P is an $\mathscr{F}_t$-compound Poisson process. Since we assume that P is square-integrable, the compensated compound Poisson process $\widehat{P}$ satisfies Assumption 3.3. Several times below we will use the fact that the Lévy measure of a compound Poisson process is a finite measure (see Definition 2.9). We will also refer to the covariance operator of P, by which we mean the covariance operator of $\widehat{P}$ (see Definition 3.5).

Definition 6.7 Let P be a square-integrable compound Poisson process on U with covariance operator Q_1 and jump measure π. For every $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ and every $t \in [0, T]$ we define the H-valued process

$$\int_0^t \Psi(s)\,\mathrm{d}P(s) := \int_0^t \int_U f_\Psi^{\widehat{P}}(s,u)\,\mathrm{d}\pi(s,u) = \sum_{s\in(0,t]} f_\Psi^{\widehat{P}}(s, \Delta P(s)), \tag{87}$$

where $f^{\widehat{P}} : \mathbf{L}^2_{Q_1^{1/2}(U),T}(H) \to \mathbf{F}^2_{\nu,T}(H)$ is the map defined in Proposition 5.7. Note that the right-hand side of (87) is well-defined for every $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ because $\nu(U) < \infty$ implies that $f_\Psi^{\widehat{P}} \in \mathbf{F}^2_{\nu,T}(H) \subseteq \mathbf{F}^1_{\nu,T}(H)$.

Proposition 6.8 *Let P be a square-integrable compound Poisson process on U with covariance operator Q_1 and Lévy measure ν. Then for every $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ we have*

$$\int_0^t \Psi(s)\,\mathrm{d}P(s) = \int_0^t \Psi(s)\,\mathrm{d}\widehat{P}(s) + \int_0^t \int_U f_\Psi^{\widehat{P}}(s,u)\,\mathrm{d}\nu(u)\,\mathrm{d}s \tag{88}$$

for all $t \in [0, T]$ *and*

$$\mathbf{E}\Big|\int_0^t \Psi(s)\,\mathrm{d}P(s)\Big|_H^2 \le 2(1+t\nu(U))\cdot\mathbf{E}\int_0^t ||\Psi(s)||^2_{L_2(Q_1^{1/2}(U),H)}\,\mathrm{d}s. \tag{89}$$

Thus, for every $t \in [0, T]$*, the map* $\Psi \mapsto \int_0^t \Psi(s)\,\mathrm{d}P(s)$ *is linear and continuous from* $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H) \to L^2(\Omega; H)$.

Proof Equation (88) follows by combining the definition (87) with (57). To obtain the estimate (89) we use equation (88), the triangle inequality and the Cauchy-Schwarz inequality to obtain

$$\begin{aligned}\mathbf{E}\Big|\int_0^t \Psi(s)\,\mathrm{d}P(s)\Big|_H^2 &\le 2\mathbf{E}\Big(\Big|\int_0^t \Psi(s)\,\mathrm{d}\widehat{P}(s)\Big|_H^2 + \Big|\int_0^t\int_U f_\Psi^{\widehat{P}}(s,u)\,\mathrm{d}\nu(u)\,\mathrm{d}s\Big|_H^2\Big)\\ &\le 2\mathbf{E}\Big|\int_0^t \Psi(s)\,\mathrm{d}\widehat{P}(s)\Big|_H^2 + 2t\nu(U)\mathbf{E}\int_0^t\int_U |f_\Psi^{\widehat{P}}(s,u)|_H^2\,\mathrm{d}\nu(u)\,\mathrm{d}s.\end{aligned}$$

Using the Itô isometry (see Theorem 3.12) and the isometric property (51) of the map $f^{\widehat{P}}$ we obtain inequality (89). Since P is a compound Poisson process we have $\nu(U) < \infty$, so (88) and (89) show that the map $\Psi \mapsto \int_0^t \Psi(s)\,\mathrm{d}P(s)$ is linear and continuous from $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H) \to L^2(\Omega; H)$. □

Remark 6.9 Recall from Proposition 5.7 that $f^{\widehat{P}}$ sends a simple process $\Psi \in \mathscr{S}(U, H)$ to the function $f_\Psi^{\widehat{P}}(s, u) := \Psi(s)u$ in $\mathbf{F}^2_{\nu,T}(H)$. Therefore, when the integrand Ψ is a simple process Definition 6.7 says that

$$\int_0^t \Psi(s)\,\mathrm{d}P(s) = \sum_{s\in(0,t]} \Psi(s)\Delta P(s).$$

So (87) is a natural definition because it agrees, a.s., with the Riemann-Stieltjes integral of $\Psi \in \mathscr{S}(U, H)$ with respect to the process P.

Our Definition 6.7 is an alternative to the definition that is used by Peszat and Zabczyk on page 125 of [15]. In a similar way to (88), they define

$$\int_0^t \Psi(s)\,\mathrm{d}P(s) := \int_0^t \Psi(s)\,\mathrm{d}\widehat{P}(s) + \int_0^t \Psi(s)(\mathbf{E}P(1))\,\mathrm{d}s, \tag{90}$$

when P is a square-integrable compound Poisson process and $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$. However, since $\Psi(s) \in L_2(U_0, H)$ may be an unbounded operator on U, it is not immediately clear how the term $\Psi(s)(\mathbf{E}P(1))$ on the right-hand side of (90) should be interpreted and it is even less clear that this can be done in such a way so that the H-valued process $\Psi(s)(\mathbf{E}P(1))$ is integrable on $[0, T]$, $\mathbf{P}$-a.s. Our aim below is to show that the natural interpretation is

$$\Psi(s)(\mathbf{E}P(1)) := \int_U f_\Psi^{\widehat{P}}(s,u)\,\mathrm{d}\nu(u), \tag{91}$$

making our definition of $\int_0^t \Psi(s)\,\mathrm{d}P(s)$ in (87) coincide, via (88), with Peszat and Zabczyk's definition (90). If $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$, then $f_\Psi^{\widehat{P}} \in \mathbf{F}^2_{\nu,T}(H) \subseteq \mathbf{F}^1_{\nu,T}(H)$. Therefore, the right-hand side of (91) is well-defined and belongs to $L^1([0,T];H)$, $\mathbf{P}$-a.s., for every $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$. The purpose of the results below is to show that (91) is a natural interpretation of $\Psi(s)(\mathbf{E}P(1))$, in the sense that equality holds in (91) when the left-hand side of (91) is well-defined in H, e.g., when $\Psi(s) \in L(U,H)$, $\mathbf{P}$-a.s.

Lemma 6.10 *Let $Q \in L_1^+(U)$ and let $U_0 := Q^{1/2}(U)$. Then the subspace $\{\Phi|_{U_0} : \Phi \in L(U,H)\}$ is dense in $L_2(U_0,H)$.*

Proof Let $(u_k)_{k=1}^\infty$ be an ONB of $\mathscr{N}(Q)^\perp$ consisting of eigenvectors of Q with corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots > 0$. Then $\big(Q^{1/2}u_k\big)_{k=1}^\infty = (\lambda_k^{1/2}u_k)_{k=1}^\infty$ is an orthonormal basis for U_0. For each positive integer n let P_n denote the orthogonal projection onto the linear span of $\{u_1,\ldots,u_n\}$ in U. Since each u_k is an eigenvector for $Q^{1/2}$ we see that the range of P_n is contained in the range of $Q^{1/2}$. Thus, for any $\Phi \in L_2(U_0,H)$ the composition ΦP_n is a well-defined linear mapping from $U \to H$. The proof will be complete if we show that $\Phi P_n \in L(U,H)$ and that $\Phi P_n|_{U_0} \to \Phi$ in $L_2(U_0,H)$.

For boundedness we begin with the inequality

$$|\Phi P_n u|_H \leq ||\Phi||_{L_2(U_0,H)}\,|P_n u|_{U_0},$$

which holds for all $u \in U$. This follows because the Hilbert-Schmidt norm dominates the operator norm (extend a unit vector in U_0 to an ONB). Next, we have

$$|P_n u|_{U_0}^2 = (Q^{-1/2}P_n u, Q^{1/2}P_n u)_U = |v|_U^2,$$

where v is the unique vector in $\mathscr{N}(Q^{1/2})^\perp$ such that $Q^{1/2}v = P_n u$ (recall the definition of $(\cdot,\cdot)_{U_0}$ in (25)). Since

$$P_n u = \sum_{k=1}^n (u,u_k)_U\,u_k = Q^{1/2}\Big(\sum_{k=1}^n (u,u_k)_U\,\lambda_k^{-1/2}u_k\Big),$$

and since $\operatorname{span}\{u_1,\ldots,u_n\} \subseteq \mathscr{N}(Q^{1/2})^\perp = \overline{\operatorname{span}\{u_1,u_2,\ldots\}}$, we see that

$$v = \sum_{k=1}^n (u,u_k)_U\,\lambda_k^{-1/2}u_k.$$

This means that

$$|P_n u|_{U_0}^2 = \Big|\sum_{k=1}^{n} (u, u_k)_U\, \lambda_k^{-1/2} u_k\Big|_U^2 = \sum_{k=1}^{n} (u, u_k)_U^2\, \lambda_k^{-1} \leq \lambda_n^{-1} |u|_U^2.$$

This shows that $\Phi P_n \in L(U, H)$ with norm at most $||\Phi||_{L_2(U,H)}\, \lambda_n^{-1/2}$.

For convergence we begin by finding $\Phi P_n|_{U_0}$. Observe that for each $h \in U_0$ we have $h = \sum_{k=1}^{\infty} (h, \lambda_k^{1/2} u_k)_{U_0} \lambda_k^{1/2} u_k$ and the series converges not only in U_0 but also in U. Indeed, since $Q^{-1/2} h \perp \mathscr{N}(Q)$ we have

$$Q^{1/2}(Q^{-1/2}h) = Q^{1/2} \sum_{k=1}^{\infty} (Q^{-1/2}h, u_k)_U u_k = \sum_{k=1}^{\infty} (Q^{-1/2}h, u_k)_U \lambda_k^{1/2} u_k,$$

where we use the fact that $Q^{1/2}$ is bounded to conclude in the last step that the sum converges in U. The right-hand side of the equation above is equal to $\sum_{k=1}^{\infty} (h, \lambda_k^{1/2} u_k)_{U_0} \lambda_k^{1/2} u_k$. Since this sum converges in U we have

$$\Phi P_n h = \Phi\Big(\sum_{k=1}^{n} (h, \lambda_k^{1/2} u_k)_{U_0} \lambda_k^{1/2} u_k\Big) = \Phi \widetilde{P}_n h,$$

where $\widetilde{P}_n$ is the orthogonal projection onto span$\{u_1, \cdots, u_n\}$ in U_0. Write $\widetilde{P}_n^{\perp} := I - \widetilde{P}_n$. We need to show that $\Phi \widetilde{P}_n^{\perp} \to 0$ in $L_2(U_0, H)$. We have

$$||\Phi \widetilde{P}_n^{\perp}||_{L_2(U_0,H)}^2 = \sum_{k=1}^{\infty} |\Phi \widetilde{P}_n^{\perp} (Q^{1/2} u_k)|_H^2 = \sum_{k=n+1}^{\infty} |\Phi (Q^{1/2} u_k)|_H^2. \tag{92}$$

On the right-hand side of (92) we have tail sums of the series $\sum_{k=1}^{\infty} |\Phi(Q^{1/2} u_k)|_H^2$ and this series converges to the finite number $||\Phi||_{L_2(U_0,H)}^2$. So the right-hand side of (92) tends to 0 as $n \to \infty$. □

Proposition 6.11 *Let P be a square-integrable $\mathscr{F}_t$-compound Poisson process on U with covariance operator Q_1 and Lévy measure ν. Define a map $h^P : L(U, H) \to H$ by*

$$h^P(\Phi) := \Phi(\mathbf{E} P(1)) = \int_U \Phi u \, \mathrm{d}\nu(u). \tag{93}$$

Then the following statements hold.

i) h^P is continuous in the $L_2(Q_1^{1/2}(U), H)$-norm and therefore has a unique continuous, linear extension to a map $h^P : L_2(Q_1^{1/2}(U), H) \to H$.

ii) For every simple process $\Psi \in \mathscr{S}(U,H)$ we have

$$h^P(\Psi(s)) = \Psi(s)(\mathbf{E}P(1)) = \int_U \Psi(s)u \,\mathrm{d}\nu(u), \tag{94}$$

in H, $\mathbf{P}$-a.s., for every $s \in [0,T]$.

iii) For every $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ we have $h^P \circ \Psi \in L^2(\Omega \times [0,T]; H)$ and

$$h^P(\Psi(s)) = \int_U f_\Psi^{\widehat{P}}(s,u)\,\mathrm{d}\nu(u), \tag{95}$$

in H, $\mathrm{d}\mathbf{P} \otimes \mathrm{d}s$-a.e.

iv) For every $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ and every sequence $(\Psi_n)_{n=1}^\infty \subset \mathscr{S}(U,H)$ such that $\Psi_n \to \Psi$ in the space $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ we have

$$\int_U f_\Psi^{\widehat{P}}(s,u)\,\mathrm{d}\nu(u) = \lim_{n\to\infty}\int_U \Psi_n(s)u\,\mathrm{d}\nu(u) = \lim_{n\to\infty}\Psi_n(s)(\mathbf{E}P(1)), \tag{96}$$

in the space $L^2(\Omega \times [0,T]; H)$ and, for every $t \in [0,T]$, we have

$$\int_0^t\int_U f_\Psi^{\widehat{P}}(s,u)\,\mathrm{d}\nu(u)\,\mathrm{d}s = \lim_{n\to\infty}\int_0^t\int_U \Psi_n(s)u\,\mathrm{d}\nu(u)\,\mathrm{d}s = \lim_{n\to\infty}\int_0^t \Psi_n(s)(\mathbf{E}P(1))\,\mathrm{d}s, \tag{97}$$

in the space $L^2(\Omega; H)$. Furthermore, if we also have $\Psi_n \to \Psi$ in $L_2(Q_1^{1/2}(U),H)$, $\mathrm{d}\mathbf{P}\otimes\mathrm{d}t$-a.e., then the convergence (96) holds in H, $\mathrm{d}\mathbf{P}\otimes\mathrm{d}t$-a.e.

Proof *i*) Let $\Phi \in L(U,H)$. Note that the equality on the right of (93) follows from Proposition 2.11. By the Cauchy-Schwarz inequality and part *i*) of Proposition 5.7 we have

$$|h^P(\Phi)|_H^2 \le \nu(U)\int_U |\Phi u|_H^2\,\mathrm{d}\nu(u) = \nu(U)\,||\Phi||^2_{L_2(Q_1^{1/2}(U),H)}\,. \tag{98}$$

This shows that h^P is continuous in the $L_2(Q_1^{1/2}(U),H)$-norm. In Lemma 6.10 we showed that $\{\Phi|_{Q_1^{1/2}(U)} : \Phi \in L(U,H)\}$ is dense in $L_2(Q_1^{1/2}(U),H)$, so h^P extends uniquely by continuity to a linear map $h^P : L_2(Q_1^{1/2}(U),H) \to H$ for which inequality (98) continues to hold for all $\Phi \in L_2(Q_1^{1/2}(U),H)$.

ii) Since simple processes take values in $L(U,H)$, $\mathbf{P}$-a.s. for all $s \in [0,T]$, equation (94) follows from the definition of h^P in (93).

iii) According to the definition of $f^{\widehat{P}}$ in Proposition 5.7, equation (94) says that (95) holds whenever $\Psi \in \mathscr{S}(U,H)$ is a simple process. Therefore, in order to establish (95) for a general integrand $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$, it suffices to show that both sides of (95) are continuous linear functions of Ψ from $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H) \to L^2(\Omega \times [0,T]; H)$. This is clearly true of the right-hand side of (95) because

$$\mathbf{E}\int_0^T \left|\int_U f_\Psi^{\widehat{P}}(s,u)\,\mathrm{d}\nu(u)\right|_H^2 \mathrm{d}s \le \nu(U)\mathbf{E}\int_0^T \int_U |f_\Psi^{\widehat{P}}(s,u)|_H^2\,\mathrm{d}\nu(u)\,\mathrm{d}s$$

$$= \nu(U)\mathbf{E}\int_0^T ||\Psi(s)||^2_{L_2(Q_1^{1/2}(U),H)}\,\mathrm{d}s,$$

by the Cauchy-Schwarz inequality and (51). For the left-hand side of (95), let $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ and use (98) to obtain

$$\mathbf{E}\int_0^T |h^P(\Psi(s))|_H^2\,\mathrm{d}s \le \nu(U)\mathbf{E}\int_0^T ||\Psi(s)||^2_{L_2(Q_1^{1/2}(U),H)}.$$

This shows that both sides of (95) are continuous mappings of Ψ from the space $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ to the space $L^2(\Omega \times [0,T]; H)$, so (95) follows from (94) by continuity.

iv) We have observed in the proof of part *iii*) that the left-hand side of (96) is a continuous linear map of Ψ from $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H) \to L^2(\Omega \times [0,T]; H)$. Equation (96) is just a restatement of this fact combined with (94). We have already observed in the proof of Proposition 6.8 that $\Psi \mapsto \int_0^t \int_U f_\Psi^{\widehat{P}}(s,u)\,\mathrm{d}\nu(u)\,\mathrm{d}s$ is a continuous linear map from $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ to $L^2(\Omega; H)$. Equation (97) is just a restatement of this fact combined with (94). Finally, if $\Psi_n \to \Psi$ in $L_2(Q_1^{1/2}(U), H)$, $\mathrm{d}\mathbf{P} \otimes \mathrm{d}t$-a.e., then we have

$$h^P(\Psi(s)) = \lim_{n\to\infty} h^P(\Psi_n(s)),$$

in H, $\mathrm{d}\mathbf{P} \otimes \mathrm{d}t$-a.e., because h^P is continuous. Because of (94) and (95), the equation above says exactly that the convergence (96) holds in H, $\mathrm{d}\mathbf{P} \otimes \mathrm{d}t$-a.e. □

According to (94) and (95), it is natural to interpret $\Psi(s)(\mathbf{E}P(1))$, which appears on the right-hand side of (90), as

$$\Psi(s)(\mathbf{E}P(1)) := h^P(\Psi(s)) = \int_U f_\Psi^{\widehat{P}}(s,u)\,\mathrm{d}\nu(u), \tag{99}$$

for each $s \in [0,T]$ and $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$. Since $h^P \circ \Psi \in L^2(\Omega \times [0,T]; H)$ for $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ by part iii) of Proposition 6.11, we see that $h^P \circ \Psi \in L^1([0,T]; H)$, $\mathbf{P}$-a.s. Therefore, the natural interpretation (99) of the expression $\Psi(s)(\mathbf{E}P(1))$ also gives a well-defined meaning to the integral $\int_0^t \Psi(s)(\mathbf{E}P(1))\,\mathrm{d}s$ as

$$\int_0^t \Psi(s)(\mathbf{E}P(1))\,\mathrm{d}s := \int_0^t h^P(\Psi(s))\,\mathrm{d}s = \int_0^t \int_U f_\Psi^{\widehat{P}}(s,u)\,\mathrm{d}\nu(u)\,\mathrm{d}s.$$

When we adopt (99) to interpret the right-hand side of (90), we see that the definition (90) agrees with our definition (87) because of (88).

6.2 Integration with Respect to a Compound Poisson Process

In this section we develop a notion of stochastic integration with respect to a general (not necessarily square-integrable) compound Poisson process P. When P is not square-integrable, new difficulties arise when attempting to define stochastic integration with respect to P, which we alluded to above. First, because the compensated compound Poisson process $\widehat{P}$ cannot be defined, neither can the space $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ nor the map $f^{\widehat{P}}$. Thus, stochastic integration with respect to a non-square-integrable compound Poisson process P cannot be defined in such a simple way as (87) or using the related formulas (88) or (90). Second, since the space $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ cannot be defined, one must look elsewhere for a natural space of integrands for stochastic integration with respect to P. Peszat and Zabczyk handle these difficulties using localization. We sketch the argument next, then give a detailed and rigorous treatment in the remainder of this section.

Up until the time that the jumps of P leave the ball $B(0,m) \subset U$, denoted τ_m, P agrees with the square-integrable compound Poisson process P_m formed by taking from P only the jumps that lie in $B(0,m)$. For this reason it makes sense to use the stochastic integral $\int_0^t \Psi(s)\,\mathrm{d}P_m(s)$, as defined in (87), as the definition of $\int_0^t \Psi(s)\,\mathrm{d}P(s)$ on the event $\{t < \tau_m\}$. The expression $\int_0^t \Psi(s)\,\mathrm{d}P_m(s)$ makes sense when $\Psi \in \mathbf{L}^2_{U_0^m,T}(H)$, where $U_0^m := Q_m^{1/2}(U)$ and Q_m denotes the covariance operator of P_m. So, given a function Ψ from $\Omega \times [0,T]$ to the space of (possibly unbounded) linear operators from U to H, we have a way to define the stochastic integral $\int_0^t \Psi(s)\,\mathrm{d}P(s)$ on the event $\{t < \tau_m\}$ for every m such that Ψ can be viewed as an element of $\mathbf{L}^2_{U_0^m,T}(H)$. This is the condition that Peszat and Zabczyk use on page 125 in hypothesis (H3) to describe integrands for stochastic integration with respect to a compound Poisson process. While Peszat and Zabczyk stop the discussion of integrands here, we would like to capture this description of integrands in the definition of a topological vector space that will play the role of the space of integrands for stochastic integration with respect to P. Our aim in this section is to expound on the notion of stochastic integration with respect to P by localization as presented by Peszat and Zabczyk and to answer Questions 6.1 and 6.2. We define an explicit Fréchet space of integrands in which simple processes are naturally represented as a dense subspace, define the stochastic integral with respect to P and show that our notion agrees with the stochastic integral constructed by localization by Peszat and Zabczyk. We go on to answer Question 6.5 and the additional Questions 6.6, 6.4 and 6.3.

We begin by laying out the notation in the localization argument explicitly. Let U be a real, separable Hilbert space, let μ be a finite Borel measure on U with $\mu(\{0\}) = 0$ and suppose that P is a U-valued compound Poisson process with Lévy measure μ (see Definition 2.9) defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \geq 0}, \mathbf{P})$. In the remainder of this section we assume that the filtration is complete and right-continuous and that P is an $\mathscr{F}_t$-compound Poisson process. It is always possible to construct such a filtration (see Remark 3.2). We do not necessarily assume that $\int_U |y|_U^2 \, \mathrm{d}\mu(y) < \infty$ or that $\int_U |y|_U \, \mathrm{d}\mu(y) < \infty$, so that P may not be square-integrable or even integrable. By Theorem 2.10 there exists a Poisson process Π with rate $\mu(U)$ (see Definition 2.6) and i.i.d. U-valued random variables $\left(Z_j\right)_{j=1}^{\infty}$ with law $\frac{1}{\mu(U)}\mu$, which are independent of Π, such that

$$P(t) = \sum_{j=1}^{\Pi(t)} Z_j. \tag{100}$$

For each positive integer m define the random variable

$$\tau_m := \inf\{t > 0 : |\Delta P(t)|_U \geq m\} \wedge T, \tag{101}$$

and the U-valued process

$$P_m(t) := \sum_{j=1}^{\Pi(t)} Z_j \chi_{B_m}(Z_j), \tag{102}$$

where $B_m := B(0, m)$ denotes the open ball of radius m centered at 0 in U. We gather facts about τ_m and P_m below. Part $ii)$ of the following Lemma 6.12 is stated as Lemma 8.18 in [15] but the proof there does not seem best, so we include here an alternative one.

Lemma 6.12 *Suppose that P is a U-valued $\mathscr{F}_t$-compound Poisson process. Then*

i) each τ_m is an $\mathscr{F}_t$-stopping time,

ii) $\mathbf{P}$-a.s., there exists an $M = M(\omega) \in \mathbb{N}$ such that $\tau_m = T$ for all $m \geq M$.

iii) P_m is an $\mathscr{F}_t$-compound Poisson process on U with Lévy measure $\mu|_{B_m}$. In particular, P_m is square-integrable and its covariance operator, denoted Q_m, *is given by*

$$(Q_m x, y)_u = \int_{B_m} (u, x)_U \, (u, y)_U \, \mathrm{d}\mu(u). \tag{103}$$

Proof *i*) Since $\{\tau_m \le t\} = \bigcap_{\substack{s>t \\ s\in\mathbb{Q}}} \{\tau_m < s\}$ and the filtration is right-continuous it suffices to show that $\{\tau_m < s\} \in \mathscr{F}_s$ for every $s > 0$. For $s \le T$ we have

$$\begin{aligned}\{\tau_m < s\} &= \Big\{ \inf\{s' > 0 : |\Delta P(s')|_U \ge m\} < s \Big\} \\ &= \bigcap_{\substack{\varepsilon>0 \\ \varepsilon\in\mathbb{Q}}} \bigcup_{\substack{0\le q_1<q_2<s \\ q_2-q_1<\varepsilon \\ q_1,q_2\in\mathbb{Q}}} \{|P(q_2) - P(q_1)|_U \ge m - \varepsilon\}. \end{aligned} \tag{104}$$

The union in (104) belongs to $\mathscr{F}_s$ for each $\varepsilon > 0$, so the intersection belongs to $\mathscr{F}_s$. This shows that τ_m is an $\mathscr{F}_t$-stopping time.

ii) Let $\left(T_j\right)_{j=1}^{\infty}$ be the jump times of P. Since μ is a finite measure, P has finitely many jumps in each compact interval, **P**-a.s. Therefore, we may assume that $T_j < T_{j+1}$ for all j and we have $T_j \uparrow \infty$, **P**-a.s. It is clear that the Poisson process Π and jumps $\left(Z_j\right)_{j=1}^{\infty}$ satisfy $\Pi(t) = \sum_{j=1}^{\infty} \chi_{[0,t]}(T_j)$ and $Z_j = \Delta P(T_j)$ (the value of the j^{th} jump of P). Since $\tau_m \le \tau_{m+1}$ a.s. it suffices to show that

$$\mathbf{P}\Big(\bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \{\tau_m \ge T_k \wedge T\} \Big) = 1.$$

That is, we must show that $\mathbf{P}\big(\bigcup_{m=1}^{\infty} \{\tau_m \ge T_k \wedge T\} \big) = 1$ for each fixed positive integer k. Since $\tau_m \le \tau_{m+1}$ we have

$$\mathbf{P}\Big(\bigcup_{m=1}^{\infty} \{\tau_m \ge T_k \wedge T\} \Big) = \lim_{m\to\infty} \mathbf{P}[\tau_m \ge T_k \wedge T]$$

by continuity from below. To compute the limit observe that

$$\mathbf{P}[Z_1 \in B_m, \cdots, Z_k \in B_m] \le \mathbf{P}[\tau_m \ge T_k] \le \mathbf{P}[\tau_m \ge T_k \wedge T].$$

Since $\left(Z_j\right)_{j=1}^{\infty}$ are i.i.d. with law $\frac{1}{\mu(U)}\mu$ we get

$$\left(\frac{\mu(B_m)}{\mu(U)} \right)^k \le \mathbf{P}[\tau_m \ge T_k \wedge T].$$

The left-hand side of the equation above tends to 1 as $m \to \infty$ because $B_m \uparrow U$.

iii) We cannot apply Theorem 2.10 directly to the sum $P_m(t) = \sum_{j=1}^{\Pi(t)} Z_j \chi_{B_m}(Z_j)$ because although the random variables $\left(Z_j \chi_{B_m}(Z_j)\right)_{j=1}^{\infty}$ are i.i.d., they are 0 with probability $\mu(B_m^c)/\mu(U) > 0$. In Theorem 2.10 the summands must be nonzero a.s. For this reason we consider a related process with the same distribution as $P_m(t)$. We have

$$P_m(t) \overset{\mathscr{D}}{=} \sum_{j=1}^{\Pi_m(t)} Y_j^m,$$

where the $\left(Y_j^m\right)_{j=1}^{\infty}$ are independent, Y_j^m has the distribution of Z_j conditioned on $Z_j \in B_m$, Π_m is a Poisson process with intensity $\mu(B_m)$ and is independent of $\left(Y_j^m\right)_{j=1}^{\infty}$. This is intuitively clear because the nonzero terms in $P_m(t)$ occur at the arrival times of Π for which $\Delta P \in B_m$. These arrival times occur with probability $\mathbf{P}[Z_1 \in B_m] = \mu(B_m)/\mu(U)$. Hence, selecting these arrival times forms a Poisson process with intensity $\mu(B_m)$. In addition, at these times, the jumps of P have the distribution of Z_j conditioned on $Z_j \in B_m$. A rigorous argument can be made by conditioning $P_m(t)$ on the number of nonzero terms in the sum on the right-hand side of (102). It is clear that the law of Y_j^m is $\frac{1}{\mu(B_m)}\mu|_{B_m}$. By Theorem 2.10, $\sum_{j=1}^{\Pi_m(t)} Y_j^m$ is a compound Poisson process with Lévy measure $\mu|_{B_m}$, so the same is true for P_m. The fact that P_m is square-integrable follows from Proposition 2.11 because $\int_{B_m} |u|^2 \, \mathrm{d}\mu(u) \leq m^2 \mu(B_m) < \infty$. The fact that the covariance operator of P_m is given by (103) follows from Proposition 4.18 of [15], as mentioned in Proposition 5.12. Finally, the fact that P_m is an $\mathscr{F}_t$-Lévy process follows from Corollary 5.4, which says that the difference $P - P_m$ is an $\mathscr{F}_t$-compound Poisson process. □

Below we list general notations that will be used to define the space of integrands for stochastic integration with respect to a compound Poisson process P. Note that for each positive integer m the compensated compound Poisson process $\widehat{P}_m$ satisfies Assumption 3.3 with respect to *the same filtration* $(\mathscr{F}_t)_{t \geq 0}$.

Notation For each positive integer m let $U_0^m := Q_m^{1/2}(U)$, so that the space of integrands for stochastic integration with respect to the square-integrable compensated compound Poisson process $\widehat{P}_m$ is $\mathbf{L}^2_{U_0^m,T}(H)$. We will use the same filtration $(\mathscr{F}_t)_{t \geq 0}$ to define the space of integrands $\mathbf{L}^2_{U_0^m,T}(H)$ for stochastic integration with respect to $\widehat{P}_m$ for each m. Recall the notation $I_t^{\widehat{P}_m} : \mathbf{L}^2_{U_0^m,T}(H) \to L^2(\Omega; H)$ for the stochastic integration map. From equation (103) we see that

$$((Q_{m+1} - Q_m)x, x)_U = \int_{B_{m+1} \setminus B_m} |\,(u, x)_U\,|^2 \, \mathrm{d}\mu(u) \geq 0,$$

for all $x \in U$, and hence $Q_m \leq Q_{m+1}$. We denote by $\iota_m \colon \mathbf{L}^2_{U_0^{m+1},T}(H) \to \mathbf{L}^2_{U_0^m,T}(H)$ the unique continuous, linear extension of the identity map on $\mathscr{S}(U, H)$ that exists by Lemma 5.5.

As mentioned above, we would like to define the space of integrands for stochastic integration with respect to P to capture the property that an integrand Ψ can be viewed as an element of $\mathbf{L}^2_{U_0^m,T}(H)$ for every m. In view of the structure

$$\mathbf{L}^2_{U_0^1,T}(H) \xleftarrow{\iota_1} \mathbf{L}^2_{U_0^2,T}(H) \xleftarrow{\iota_2} \mathbf{L}^2_{U_0^3,T}(H) \xleftarrow{\iota_3} \cdots,$$

it is natural to define the space of integrands for stochastic integration with respect P as a certain projective limit of the Hilbert spaces $\mathbf{L}^2_{U_0^m,T}(H)$. We recall the notion of projective limit below and give basic properties.

Definition 6.13 Let $(X_m)_{m=1}^{\infty}$ be a sequence of real Banach spaces equipped with continuous linear maps $\phi_m \colon X_{m+1} \to X_m$. The *projective limit* (or inverse limit) of the sequence $(X_m, \phi_m)_{m=1}^{\infty}$ is the subspace

$$\varprojlim \phi_m X_m := \left\{ x = (x_m)_{m=1}^{\infty} \in \prod_{m=1}^{\infty} X_m : x_m = \phi_m(x_{m+1}) \text{ for all } m \right\}$$

of the Cartesian product $\prod_{m=1}^{\infty} X_m$ (see, e.g., [8] or [18]). The projective limit $\varprojlim \phi_m X_m$ is a Fréchet space under the product topology, which is clearly generated by the seminorms $p_m(x) := ||x_m||_{X_m}$.

With this notion in hand we are prepared to answer Question 6.1. We define the space of integrands for stochastic integration with respect to a compound Poisson process P as

$$\mathbf{L}_{P,T}(H) := \left\{ \Psi = (\Psi_m)_{m=1}^{\infty} \in \prod_{m=1}^{\infty} \mathbf{L}^2_{U_0^m,T}(H) : \Psi_m = \chi_{[0,\tau_m]} \iota_m(\Psi_{m+1}) \text{ for all } m \right\}, \tag{105}$$

i.e., $\mathbf{L}_{P,T}(H)$ is the projective limit of the sequence $\left(\mathbf{L}^2_{U_0^m,T}(H), \phi_m\right)_{m=1}^{\infty}$, where $\phi_m \colon \mathbf{L}^2_{U_0^{m+1},T}(H) \to \mathbf{L}^2_{U_0^m,T}(H)$ is defined by

$$\phi_m(\Phi) := \chi_{[0,\tau_m]} \iota_m(\Phi), \qquad \text{for all } \Phi \in \mathbf{L}^2_{U_0^{m+1},T}(H). \tag{106}$$

Lemmas 3.15 and 5.5 show that each ϕ_m is a continuous linear map with norm less than or equal to 1. In Proposition 6.14 below we answer Question 6.2 affirmatively by representing simple processes as a dense subspace of $\mathbf{L}_{P,T}(H)$. After that we define the stochastic integral as a continuous linear map on $\mathbf{L}_{P,T}(H)$.

Proposition 6.14 *Define a map* $\vartheta : \mathscr{S}(U,H) \to \prod_{m=1}^{\infty} \mathbf{L}^2_{U_0^m,T}(H)$ *by*

$$\vartheta(\Phi) := \left(\chi_{[0,\tau_m]}\Phi\right)_{m=1}^{\infty}. \tag{107}$$

Then

i) ϑ is linear and one-to-one,
ii) the range of ϑ is a dense subspace of $\mathbf{L}_{P,T}(H)$.

Proof *i*) It is clear that ϑ is linear. Suppose that $\vartheta(\Phi) = 0$ for some $\Phi \in \mathscr{S}(U,H)$, i.e., that $\chi_{[0,\tau_m]}\Phi = 0$ in $\mathbf{L}^2_{U_0^m,T}(H)$ for every positive integer m. By Lemma 6.12 this means that, a.s., $\Phi = 0$ for all time. So ϑ is injective when we identify elements of $\mathscr{S}(U,H)$ that agree $\mathrm{d}\mathbf{P} \otimes \mathrm{d}t$-a.e.

ii) First, we must show that $\vartheta(\Phi) \in \mathbf{L}_{P,T}(H)$ for every $\Phi \in \mathscr{S}(U,H)$. For every positive integer m we have

$$\chi_{[0,\tau_m]}\iota_m(\vartheta(\Phi)_{m+1}) = \chi_{[0,\tau_m]}\iota_m(\chi_{[0,\tau_{m+1}]}\Phi) = \chi_{[0,\tau_m]} \cdot \chi_{[0,\tau_{m+1}]}\iota_m(\Phi) = \chi_{[0,\tau_m]}\Phi.$$

The second equality above is obtained using the fact that ι_m commutes with multiplication by $\chi_{[0,\tau_{m+1}]}$. This is an application of Lemma 6.15 below. Next, we must show that the range of ϑ is dense in the space $\mathbf{L}_{P,T}(H)$. Since the map ϕ_k in (106) has norm at most 1 we see that for every $\Psi \in \mathbf{L}_{P,T}(H)$ one has

$$||\Psi_k||_{\mathbf{L}^2_{U_0^k,T}(H)} = ||\phi_k(\Psi_{k+1})||_{\mathbf{L}^2_{U_0^k,T}(H)} \le ||\Psi_{k+1}||_{\mathbf{L}^2_{U_0^{k+1},T}(H)}. \tag{108}$$

By induction we have

$$||\Psi_k||_{\mathbf{L}^2_{U_0^k,T}(H)} \le ||\Psi_m||_{\mathbf{L}^2_{U_0^m,T}(H)},$$

for all $k \le m$. Now let $\Psi \in \mathbf{L}_{P,T}(H)$ and $\Phi \in \mathscr{S}(U,H)$ and apply this inequality to the difference $\vartheta(\Phi) - \Psi$ to see that

$$\left|\left|\chi_{[0,\tau_k]}\Phi - \Psi_k\right|\right|_{\mathbf{L}^2_{U_0^k,T}(H)} \le \left|\left|\chi_{[0,\tau_m]}\Phi - \Psi_m\right|\right|_{\mathbf{L}^2_{U_0^m,T}(H)}, \tag{109}$$

for all $k \le m$. Fix $\Psi \in \mathbf{L}_{P,T}(H)$. For each positive integer m we can use Lemma 3.11 to select a simple process $\Phi_m \in \mathscr{S}(U,H)$ with the property that

$$||\Phi_m - \iota_m(\Psi_{m+1})||^2_{\mathbf{L}^2_{U_0^m,T}(H)} < \frac{1}{m}$$

Since $\Psi_m = \chi_{[0,\tau_m]}\iota_m(\Psi_{m+1})$ we see that

$$\left|\left|\chi_{[0,\tau_m]}\Phi_m - \Psi_m\right|\right|^2_{\mathbf{L}^2_{U_0^m,T}(H)} \le ||\Phi_m - \iota_m(\Psi_{m+1})||^2_{\mathbf{L}^2_{U_0^m,T}(H)} < \frac{1}{m}. \tag{110}$$

We claim that $(\vartheta(\Phi_m))_{m=1}^{\infty}$ converges to Ψ in the space $\mathbf{L}_{P,T}(H)$. This is equivalent to showing that for each fixed positive integer k we have $\vartheta(\Phi_m)_k \to \Psi_k$ in $\mathbf{L}^2_{U_0^k,T}(H)$ as $m \to \infty$. By (109) and (110) we have

$$\begin{aligned}||\vartheta(\Phi_m)_k - \Psi_k||_{\mathbf{L}^2_{U_0^k,T}(H)} &= \left|\left|\chi_{[0,\tau_k]}\Phi_m - \Psi_k\right|\right|_{\mathbf{L}^2_{U_0^k,T}(H)} \\ &\leq \left|\left|\chi_{[0,\tau_m]}\Phi_m - \Psi_m\right|\right|_{\mathbf{L}^2_{U_0^m,T}(H)} \\ &< \frac{1}{m}\end{aligned}$$

for all $m \geq k$, so $\vartheta(\Phi_m)_k \to \Psi_k$ in $\mathbf{L}^2_{U_0^k,T}(H)$. □

Lemma 6.15 *Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$ be a filtered probability space, let U and H be separable Hilbert spaces, let $Q_1, Q_2 \in L_1^+(U)$ with $Q_1 \leq Q_2$ and, as in Lemma 5.5, let $\iota\colon \mathbf{L}^2_{Q_2^{1/2}(U),T}(H) \to \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ denote the continuous extension of the identity map on $\mathscr{S}(U, H)$. Suppose that τ is an $\mathscr{F}_t$-stopping time with $\mathbf{P}[\tau \leq T] = 1$. Then for every $\Psi \in \mathbf{L}^2_{Q_2^{1/2}(U),T}(H)$ we have*

$$\iota(\chi_{[0,\tau]}\Psi) = \chi_{[0,\tau]}\iota(\Psi) \tag{111}$$

in the space $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$.

Proof Since both sides of (111) are continuous functions of Ψ on $\mathbf{L}^2_{Q_2^{1/2}(U),T}(H)$ it suffices to assume that $\Psi \in \mathscr{S}(U, H)$. First, assume that τ takes finitely many values a.s. In this case, we have $\chi_{[0,\tau]}\Psi \in \mathscr{S}(U, H)$ by Lemma 3.15, so (111) holds because ι is the identity operator on $\mathscr{S}(U, H)$. Now we allow the possibility that τ takes infinitely many values with positive probability. As in the proof of Lemma 3.15 there exists a sequence of stopping times $(\tau_n)_{n=1}^{\infty}$ such that $\tau_n \leq T$ a.s., $\tau_n \downarrow \tau$, and each τ_n has finitely many values a.s. By Lemma 3.15 we have

$$\chi_{[0,\tau_n]}\Psi \to \chi_{[0,\tau]}\Psi \quad \text{in the space } \mathbf{L}^2_{Q_2^{1/2}(U),T}(H)$$

and

$$\chi_{[0,\tau_n]}\iota(\Psi) \to \chi_{[0,\tau]}\iota(\Psi) \quad \text{in the space } \mathbf{L}^2_{Q_1^{1/2}(U),T}(H).$$

Since ι is continuous we find that

$$\iota(\chi_{[0,\tau]}\Psi) = \lim_{n\to\infty} \iota(\chi_{[0,\tau_n]}\Psi) = \lim_{n\to\infty} \chi_{[0,\tau_n]}\iota(\Psi) = \chi_{[0,\tau]}\iota(\Psi)$$

where each limit is taken in the space $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$. □

We introduce and recall more notations in preparation for defining stochastic integration on the space $\mathbf{L}_{P,T}(H)$. We will denote the Lévy measure of P_m by $\mu_m := \mu|_{B_m}$. Recall from Proposition 5.7 that we have the isometry $f^{\widehat{P}_m} : \mathbf{L}^2_{U_0^m,T}(H) \to \mathbf{F}^2_{\mu_m,T}(H)$. For fixed $t \in [0, T)$ define the space

$$Y^{(t)} := \{\psi \in \prod_{m=1}^{\infty} L^2(\Omega, \mathscr{F}_t, \mathbf{P}; H) : \psi_m = \chi_{\{t<\tau_m\}} \cdot \psi_{m+1}\},$$

i.e., $Y^{(t)}$ is the projective limit of $\left(L^2(\Omega, \mathscr{F}_t, \mathbf{P}; H), \eta_m^{(t)}\right)_{m=1}^{\infty}$, where for each positive integer m, $\eta_m^{(t)} : L^2(\Omega, \mathscr{F}_t, \mathbf{P}; H) \to L^2(\Omega, \mathscr{F}_t, \mathbf{P}; H)$ is the continuous linear map defined by

$$\eta_m^{(t)}(\psi) := \chi_{\{t<\tau_m\}} \cdot \psi, \qquad \text{for all } \psi \in L^2(\Omega, \mathscr{F}_t, \mathbf{P}; H).$$

Define a map $I_t^P : \mathbf{L}_{P,T}(H) \to \prod_{m=1}^{\infty} L^2(\Omega, \mathscr{F}_t, \mathbf{P}; H)$ by sending each $\Psi \in \mathbf{L}_{P,T}(H)$ to the sequence

$$I_t^P(\Psi) := \left(\chi_{\{t<\tau_m\}} \int_0^t \Psi_m(s)\,\mathrm{d}P_m(s)\right)_{m=1}^{\infty},$$

where in each coordinate $\int_0^t \Psi_m(s)\,\mathrm{d}P_m(s)$ is the stochastic integral of the process $\Psi_m \in \mathbf{L}^2_{U_0^m,T}(H)$ with respect to the square-integrable compound Poisson process P_m as defined in (87) and given by the equivalent expression (88). We will establish some lemmas in preparation for showing that I_t^P maps into the subspace $Y^{(t)}$.

Lemma 6.16 *Let P be a square-integrable $\mathscr{F}_t$-compound Poisson process on U with Lévy measure ν, covariance operator Q_1 and jump measure π. Let τ be an $\mathscr{F}_t$-stopping time such that* $\mathbf{P}[\tau \le T] = 1$. *Then for every $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ we have*

i)

$$f^{\widehat{P}}_{\chi_{[0,\tau]}\Psi} = \chi_{[0,\tau]} f^{\widehat{P}}_{\Psi} \tag{112}$$

in the space $\mathbf{F}^2_{\nu,T}(H)$ and

ii) for all $t \in [0, T]$ we have

$$\sum_{s\in(0,t]} f^{\widehat{P}}_{\chi_{[0,\tau]}\Psi}(s, \Delta P(s)) = \sum_{s\in(0,t\wedge\tau]} f^{\widehat{P}}_{\Psi}(s, \Delta P(s)), \tag{113}$$

$$\int_0^t \chi_{[0,\tau]}\Psi(s)\,\mathrm{d}P(s) = \int_0^{t\wedge\tau} \Psi(s)\,\mathrm{d}P(s). \tag{114}$$

Proof The proof requires some care due to the fact that $\chi_{[0,\tau]}\Psi$ is not necessarily a simple process for every $\Psi \in \mathscr{S}(U, H)$. As usual, this difficulty is handled by approximating τ with stopping times that take finitely many values a.s.

i) First, assume that τ has finitely many values a.s. In this case, for every $\Psi \in \mathscr{S}(U, H)$ the process $\chi_{(\tau,T]}\Psi$ is also simple by part *ii*) of Lemma 3.15 and it is clear that

$$f^{\widehat{P}}_{\chi_{(\tau,T]}\Psi}(s,u) = \chi_{(\tau,T]}(s)\Psi(s)u = f^{\widehat{P}}_{\Psi}(s,u) - \chi_{[0,\tau]} f^{\widehat{P}}_{\Psi}(s,u).$$

Rearranging this equality gives *i*) for simple processes. Next, we allow τ to attain infinitely many values with positive probability. As in the proof of Lemma 3.15, there exists a sequence of stopping times $(\tau_n)_{n=1}^{\infty}$ such that $\tau_n \le T$ a.s., $\tau_n \downarrow \tau$, and each τ_n has finitely many values a.s. By Lemma 3.15 we have $\chi_{[0,\tau_n]}\Psi \to \chi_{[0,\tau]}\Psi$ in the space $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$. This means that

$$f^{\widehat{P}}_{\chi_{[0,\tau_n]}\Psi} \to f^{\widehat{P}}_{\chi_{[0,\tau]}\Psi}$$

in $\mathbf{F}^2_{\nu,T}(H)$ because $f^{\widehat{P}}$ is continuous. On the other hand, since $f^{\widehat{P}}_{\Psi} \in \mathbf{F}^2_{\nu,T}(H)$ the dominated convergence theorem implies that $\chi_{[0,\tau_n]} f^{\widehat{P}}_{\Psi} \to \chi_{[0,\tau]} f^{\widehat{P}}_{\Psi}$ in $\mathbf{F}^2_{\nu,T}(H)$. We conclude that *i*) holds for all $\Psi \in \mathscr{S}(U, H)$. Finally, both sides of equation *i*) are continuous functions of Ψ from $\mathbf{L}^2_{Q_1^{1/2}(U),T}(H) \to \mathbf{F}^2_{\nu,T}(H)$, so *i*) holds for all $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$.

ii) Fix $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$ and $t \in [0, T]$. Since $\mathbf{F}^2_{\nu,T}(H) \subseteq \mathbf{F}^1_{\nu,T}(H)$ (on account of $\nu(U) < \infty$), equation (112) continues to hold in the space $\mathbf{F}^1_{\nu,T}(H)$. As a result,

$$\begin{aligned}\sum_{s\in(0,t]} f^{\widehat{P}}_{\chi_{[0,\tau]}\Psi}(s, \Delta P(s)) &= \int_{(0,t]}\int_U f^{\widehat{P}}_{\chi_{[0,\tau]}\Psi}(s,u)\,\mathrm{d}\pi(s,u) \\ &= \int_{(0,t]}\int_U \chi_{[0,\tau]}(s) f^{\widehat{P}}_{\Psi}(s,u)\,\mathrm{d}\pi(s,u) \\ &= \int_{(0,t\wedge\tau]}\int_U f^{\widehat{P}}_{\Psi}(s,u)\,\mathrm{d}\pi(s,u) \\ &= \sum_{s\in(0,t\wedge\tau]} f^{\widehat{P}}_{\Psi}(s, \Delta P(s)),\end{aligned}$$

which is (113). By definition (87), equation (114) is equivalent to (113). □

Remark 6.17 We did not use the fact that compound Poisson processes have finite Lévy measures in the proof of part $i)$ in Lemma 6.16. The same argument shows that equation (112) remains true when the square-integrable compensated compound Poisson process $\widehat{P}$ is replaced by a square-integrable $\mathscr{F}_t$-Lévy process $\mathscr{L}$ whose covariance operator satisfies condition (49). That is, equation (112) holds whenever the map $f^{\mathscr{L}}$ can be defined.

We are now ready to show that the range of I_t^P is contained in the subspace $Y^{(t)}$ (cf. Lemma 8.19 in [15]).

Proposition 6.18 *Let P be a U-valued $\mathscr{F}_t$-compound Poisson process (not necessarily square-integrable). For each $t \in [0, T)$, I_t^P is linear, continuous and its range is contained in $Y^{(t)}$, i.e., for every $\Psi \in \mathbf{L}_{P,T}(H)$ and all positive integers $m \leq n$ we have*

$$\int_0^t \Psi_m(s)\,\mathrm{d}P_m(s) = \int_0^t \Psi_n(s)\,\mathrm{d}P_n(s), \tag{115}$$

in H, $\mathbf{P}$-a.s. on the event $\{t < \tau_m\}$.

Proof Linearity and continuity of I_t^P are clear because each coordinate function of I_t^P is linear and continuous from $\mathbf{L}_{P,T}(H) \to L^2(\Omega, \mathscr{F}_t, \mathbf{P}; H)$. Fix $\Psi = (\Psi_m)_{m=1}^{\infty} \in \mathbf{L}_{P,T}(H)$. In order to show that $I_t^P(\Psi) \in Y^{(t)}$, it is necessary and sufficient to show that

$$\chi_{\{t<\tau_m\}} \int_0^t \Psi_m(s)\,\mathrm{d}P_m(s) = \chi_{\{t<\tau_m\}} \Big(\chi_{\{t<\tau_{m+1}\}} \int_0^t \Psi_{m+1}(s)\,\mathrm{d}P_{m+1}(s) \Big), \tag{116}$$

in the space $L^2(\Omega; H)$ for every positive integer m. It is easy to see that this is equivalent to (115). In order to compute the left-hand side (116) we recall that

$$\int_0^t \Psi_m(s)\,\mathrm{d}P_m(s) = \sum_{s\in(0,t]} f_{\Psi_m}^{\widehat{P}_m}(s, \Delta P_m(s)),$$

from (87) and $\Psi_m = \chi_{[0,\tau_m]}\iota_m(\Psi_{m+1})$ because $\Psi \in \mathbf{L}_{P,T}(H)$. Using Lemmas 5.10 and 6.16 we see that

$$\begin{aligned} \int_0^t \Psi_m(s)\,\mathrm{d}P_m(s) &= \sum_{s\in(0,t\wedge\tau_m]} f_{\iota_m(\Psi_{m+1})}^{\widehat{P}_m}(s, \Delta P_m(s)) \\ &= \sum_{s\in(0,t\wedge\tau_m]} f_{\Psi_{m+1}}^{\widehat{P}_{m+1}}(s, \Delta P_m(s)) \end{aligned}$$

in H, $\mathbf{P}$-a.s. It is clear that $\mathbf{P}[P_m(s) = P_{m+1}(s) \text{ for all } s \in [0, t] \mid t < \tau_m] = 1$, in particular, when $t < \tau_m$, all of the jumps of P_{m+1} that occur in $(0, t]$ lie in the ball

B_m. Therefore,

$$\int_0^t \Psi_m(s)\,\mathrm{d}P_m(s) = \sum_{s\in(0,t]} f_{\Psi_{m+1}}^{\widehat{P}_{m+1}}(s, \Delta P_{m+1}(s)) = \int_0^t \Psi_{m+1}(s)\,\mathrm{d}P_{m+1}(s)$$

on the event $\{t < \tau_m\}$. This proves (116) and (115) follows by induction. □

Corollary 6.19 *Let P be a U-valued $\mathscr{F}_t$-compound Poisson process (not necessarily square-integrable). Then for every $\Psi \in \mathbf{L}_{P,T}(H)$ and every $t \in [0, T)$, the sequence $I_t^P(\Psi) \in Y^{(t)}$ converges* **P***-a.s. to an $\mathscr{F}_t$-measurable, H-valued random variable.*

Proof Fix $t \in [0, T)$ and $\Psi \in \mathbf{L}_{P,T}(H)$. With probability one, there exists a positive integer m such that $t < \tau_m$ by Lemma 6.12. By Proposition 6.18 we have $(I_t^P(\Psi))_n = \int_0^t \Psi_m(s)\,\mathrm{d}P_m(s)$, **P**-a.s., for all $n \geq m$. So $\lim_{m\to\infty}(I_t^P(\Psi))_m$ exists in H, **P**-a.s., and the limit is $\mathscr{F}_t$-measurable because each $(I_t^P(\Psi))_m$ is $\mathscr{F}_t$-measurable. □

We can now use the map $I_t^P : \mathbf{L}_{P,T}(H) \to Y^{(t)}$ and Corollary 6.19 to define the stochastic integral of each $\Psi \in \mathbf{L}_{P,T}(H)$ with respect to a compound Poisson process P as an adapted H-valued stochastic process.

Definition 6.20 Let P be a U-valued $\mathscr{F}_t$-compound Poisson process (not necessarily square-integrable) and let $\Psi \in \mathbf{L}_{P,T}(H)$. We define the stochastic integral of Ψ with respect to P as

$$\int_0^t \Psi(s)\,\mathrm{d}P(s) := \lim_{m\to\infty}(I_t^P(\Psi))_m = \lim_{m\to\infty}\int_0^t \Psi_m(s)\,\mathrm{d}P_m(s). \tag{117}$$

By Corollary 6.19, the limit exists and is $\mathscr{F}_t$-measurable. Furthermore, the limit stabilizes, **P**-a.s., at the value $\int_0^t \Psi_m(s)\,\mathrm{d}P_m(s)$ for any m such that $t < \tau_m$. Thus, $\int_0^t \Psi(s)\,\mathrm{d}\mathbf{P}(s)$ is a sum of finitely many vectors in H **P**-a.s., which answers Question 6.4 affirmatively, and $\int_0^t \Psi(s)\,\mathrm{d}\mathbf{P}(s)$ is a càdlàg pure-jump process as a function of t.

Remark 6.21 Our definition of stochastic integration with respect to a compound Poisson process P in (117) agrees with the process constructed by localization by Peszat and Zabczyk for stochastic integration with respect to P. Indeed, the processes considered by Peszat and Zabczyk that satisfy hypothesis (H3) on page 125 of [15] belong to the space $\mathbf{L}_{P,T}(H)$. The stochastic integral presented by Peszat and Zabczyk that is constructed by localization is defined, as is ours, to agree with $\int_0^t \Psi_m(s)\,\mathrm{d}P_m(s)$ for every m such that $t < \tau_m$.

The next result is a partial affirmative answer to Question 6.6. We show that the stochastic integral of (the natural image of) a simple process Ψ with respect to a compound Poisson process P in the sense of Definition 6.20 agrees with the pathwise Riemann-Stieltjes integral of Ψ with respect to P.

Lemma 6.22 *Let P be a U-valued $\mathscr{F}_t$-compound Poisson process (not necessarily square-integrable). For every $\Psi \in \mathscr{S}(U, H)$ and $t \in [0, T)$, we have*

$$\int_0^t \vartheta(\Psi)(s)\,\mathrm{d}P(s) = \sum_{s\in(0,t]} \Psi(s)\Delta P(s) \qquad \mathbf{P}\text{-a.s.} \tag{118}$$

Furthermore, the right-hand side of (118) *is a sum of finitely many vectors in H a.s.*

Proof Let $\Psi \in \mathscr{S}(U, H)$ and let $t \in [0, T)$. For each positive integer m we have

$$\int_0^t \chi_{[0,\tau_m]}(s)\Psi(s)\,\mathrm{d}P_m(s) = \int_0^{t\wedge\tau_m} \Psi(s)\,\mathrm{d}P_m(s)$$

by (88) and part iv) of Lemma 3.15. On the event $\{t < \tau_m\}$ we have $P(s) = P_m(s)$ for all $s \in [0, t]$, so

$$\int_0^t \chi_{[0,\tau_m]}(s)\Psi(s)\,\mathrm{d}P_m(s) = \int_0^t \Psi(s)\,\mathrm{d}P_m(s) = \sum_{s\in(0,t]} \Psi(s)\Delta P_m(s) = \sum_{s\in(0,t]} \Psi(s)\Delta P(s),$$

by (87). This shows that the limit in (117) stabilizes at $\sum_{s\in(0,t]} \Psi(s)\Delta P(s)$, so (118) holds. Since P has finitely many jumps in $[0, t]$, $\mathbf{P}$-a.s., the right-hand side of (118) is a sum of finitely many vectors in H a.s. □

Corollary 6.23 *Let P be a U-valued $\mathscr{F}_t$-compound Poisson process (not necessarily square-integrable). For every $\Psi \in \mathscr{S}(U, H)$ and $t \in [0, T)$, we have*

$$\int_0^t \vartheta(\Psi)(s)\,\mathrm{d}P(s) = \int_{s\in(0,t]} \int_U \Psi(s)u\,\mathrm{d}\pi(u, s), \tag{119}$$

where π is the jump measure of P.

Proof Combine Lemma 6.22 with the definition of the jump measure $\pi := \sum_{\substack{s\in(0,t]\\ \Delta P(s)\neq 0}} \delta_{(s,\Delta P(s))}$. □

The use of projective limits to define the space of integrands $\mathbf{L}_{P,T}(H)$ and the space $Y^{(t)}$ where the stochastic integration map I_t^P takes values is merely a way to organize the localization construction of the stochastic integral with respect to P presented by Peszat and Zabczyk. We feel that the localization construction fits naturally into the setting of projective limits. The use of projective limits provides the additional desirable features of an explicit space of integrands, $\mathbf{L}_{P,T}(H)$, that naturally includes $\mathscr{S}(U, H)$ as the dense subspace $\vartheta(\mathscr{S}(U, H))$ and an explicit continuous linear operator I_t^P that serves as stochastic integration with respect to P.

Our next goal in this section addresses Question 6.3 by relating the notion of stochastic integration with respect to P as defined in Definition 6.20 with the notion of stochastic integration with respect to the jump measure π of P. We will show that for every $\Psi \in \mathbf{L}_{P,T}(H)$ there exists a $\mathscr{P}_{[0,T]} \otimes \mathscr{B}(U)$-measurable function $f_\Psi^P : \Omega \times [0,T] \times U \to H$ such that

$$\int_0^t \Psi(s)\,\mathrm{d}P(s) = \int_0^t \int_U f_\Psi^P(s,u)\,\mathrm{d}\pi(s,u) = \sum_{s\in(0,t]} f_\Psi^P(s,\Delta P(s)), \tag{120}$$

P-a.s., for each $t \in [0,T]$ (note that the right-hand side of (120) is a sum of finitely many vectors in H a.s.). This endeavor is motivated by related results in the case where P is square-integrable, namely (56) in Proposition 5.12, (58) in Proposition 5.14 and the initial definition (87) of stochastic integration with respect to a square-integrable compound Poisson process. Note that (120) has already been established for simple processes in (118) of Lemma 6.22 and (119) in Corollary 6.23. In this case the integrand on the left-hand side of (120) has the form $\Psi = \vartheta(\Phi)$ for some $\Phi \in \mathscr{S}(U,H)$ and the function f_Ψ^P on the right-hand side of (120) is given by $f_\Psi^P(s,u) = \Phi(s)u$. This is extended in the next result, which gives a complete and affirmative answer to Question 6.3.

Proposition 6.24 *Let P be a U-valued $\mathscr{F}_t$-compound Poisson process (not necessarily square-integrable). Assume P has Lévy measure ν and jump measure π. Then for every $\Psi = (\Psi_m)_{m=1}^\infty \in \mathbf{L}_{P,T}(H)$ the sequence $\big(f_{\Psi_m}^{\widehat{P}_m}\big)_{m=1}^\infty$ converges pointwise in H, $\mathrm{d}\mathbf{P}\otimes\mathrm{d}t\otimes\mathrm{d}\nu$-a.e. on $\Omega\times[0,T]\times U$. Denote the limit by $f_\Psi^P := \lim_{m\to\infty} f_{\Psi_m}^{\widehat{P}_m}$. For every $t \in [0,T)$ we have*

$$\int_0^t \Psi(s)\,\mathrm{d}P(s) = \int_0^t \int_U f_\Psi^P(s,u)\,\mathrm{d}\pi(s,u) = \sum_{s\in(0,t]} f_\Psi^P(s,\Delta P(s)) \tag{121}$$

and the right-hand side of (121) is a sum of finitely many vectors in H a.s.

Proof Let $\Psi = (\Psi_m)_{m=1}^\infty \in \mathbf{L}_{P,T}(H)$. By Lemmas 6.16 and 5.10 we see that

$$f_{\Psi_m}^{\widehat{P}_m} = \chi_{[0,\tau_m]} f_{\iota_m(\Psi_{m+1})}^{\widehat{P}_m} = \chi_{[0,\tau_m]} f_{\Psi_{m+1}}^{\widehat{P}_{m+1}},$$

$\mathrm{d}\mathbf{P}\otimes\mathrm{d}t\otimes\mathrm{d}\nu$-a.e. on $\Omega\times[0,T]\times B_m$. Since $\tau_m \uparrow T$ a.s. by Lemma 6.12 it follows that the limit $f_\Psi^P := \lim_{m\to\infty} f_{\Psi_m}^{\widehat{P}_m}$ exists in H, $\mathrm{d}\mathbf{P}\otimes\mathrm{d}t\otimes\mathrm{d}\nu$-a.e. Furthermore, for $\mathrm{d}t\otimes\mathrm{d}\nu$-a.e. $(s,u)\in[0,T)\times U$, the limit stabilizes a.s. for all m large enough so that $s < \tau_m$ and $u \in B_m$. Using (117) in Definition 6.20 and (87) in Definition 6.7 we see that

$$\int_0^t \Psi(s)\,\mathrm{d}P(s) = \lim_{m\to\infty}\int_0^t \Psi_m(s)\,\mathrm{d}P_m(s) = \lim_{m\to\infty}\sum_{s\in(0,t]} f_{\Psi_m}^{\widehat{P}_m}(s,\Delta P_m(s)).$$

Since $P_m(s) = P(s)$ and $\Delta P(s) \in B_m$ a.s. for all $s \in [0, t]$ on the event $\{t < \tau_m\}$ we find that

$$\int_0^t \Psi(s)\,\mathrm{d}P(s) = \lim_{m\to\infty} \sum_{s\in(0,t]} f_{\Psi_m}^{\widehat{P}_m}(s, \Delta P(s)) = \sum_{s\in(0,t]} f_{\Psi}^{P}(s, \Delta P(s)).$$

We can pass to the limit inside the sum in the last step because P has finitely many jumps in $[0, T]$ a.s. □

Remark 6.25 Proposition 6.24 shows that the construction of the stochastic integral with respect to an $\mathscr{F}_t$-compound Poisson process P by localization as presented by Peszat and Zabczyk is a special case of stochastic integration with respect to the Poisson random measure of a stationary $\mathscr{F}_t$-Poisson point process, namely the jump measure π of P. As seen in Remark 6.21, the left-hand side of (121) agrees with Peszat and Zabczyk's construction of integration with respect to P by localization and the right-hand side of (121) is an integral with respect to the jump measure of P.

Using Proposition 6.24 we can immediately identify the jumps of a stochastic integral with respect to a compound Poisson process P (cf. Corollaries 5.13 and 5.16). This answers Question 6.5.

Corollary 6.26 *Let P be a U-valued $\mathscr{F}_t$-compound Poisson process (not necessarily square-integrable). For every $t \in [0, T]$ and $\Psi \in \mathbf{L}_{P,T}(H)$ we have*

$$\Delta \int_0^t \Psi(s)\,\mathrm{d}P(s) = \begin{cases} f_{\Psi}^{P}(t, \Delta P(t)) & \text{if } \Delta P(t) \neq 0 \\ 0 & \text{if } \Delta P(t) = 0. \end{cases}$$

Below we show that predictable processes with values in $L(U, H)$ belong to $\mathbf{L}_{P,T}(H)$ and that the stochastic integral of such a process with respect to P agrees with the pathwise Riemann-Stieltjes integral. This answers Question 6.6 completely and affirmatively (cf. Proposition 6.14 and Corollary 6.23 for simple processes).

Proposition 6.27 *Let P be a U-valued $\mathscr{F}_t$-compound Poisson process (not necessarily square-integrable). Let P have Lévy measure ν, jump measure π and covariance operator Q. For every process $\Psi \in L^2(\Omega \times [0, T], \mathscr{P}_{[0,T]}, \mathrm{d}\mathbf{P} \otimes \mathrm{d}t; L(U, H))$ the sequence $\alpha(\Psi) := \big(\chi_{[0,\tau_m]}\Psi\big)_{m=1}^{\infty}$ belongs to $\mathbf{L}_{P,T}(H)$ and*

$$\int_0^t \alpha(\Psi)\,\mathrm{d}P = \sum_{s\in(0,t]} \Psi(s)\Delta P(s) = \int_{(0,t]}\int_U \Psi(s)u\,\mathrm{d}\pi(s, u), \tag{122}$$

a.s. in H for every $t \in [0, T)$.

Proof Let $\Psi \in L^2(\Omega \times [0, T], \mathscr{P}_{[0,T]}, \mathrm{d}\mathbf{P} \otimes \mathrm{d}t; L(U, H))$. Recall from Example 5.9 that $\Psi \in \mathbf{L}^2_{U_0^m,T}(H)$ for every positive integer m, so $\alpha(\Psi) \in$

$\prod_{m=1}^{\infty} \mathbf{L}^2_{U_0^m,T}(H)$. Next, since the map ι_m has norm less than or equal to 1 we see that every sequence in $\mathscr{S}(U,H)$ that converges to Ψ in the space $\mathbf{L}^2_{U_0^{m+1},T}(H)$ also converges to Ψ in the space $\mathbf{L}^2_{U_0^m,T}(H)$. This means that $\iota_m(\Psi) = \Psi$ for every positive integer m. It is now easy to see that $\alpha(\Psi) \in \mathbf{L}_{P,T}(H)$. To compute the stochastic integral $\int_0^t \alpha(\Psi)\,\mathrm{d}P$ we recall from Example 5.9 that $f_{\Psi}^{\widehat{P}_m}(s,u) = \Psi(s)u$ for every positive integer m. Using this and Lemma 6.16 we see that

$$\begin{aligned}
\int_0^t \alpha(\Psi)\,\mathrm{d}P &= \lim_{m\to\infty} \int_0^t \chi_{[0,\tau_m]}(s)\Psi(s)\,\mathrm{d}P_m(s) \\
&= \lim_{m\to\infty} \int_0^t \int_U f^{\widehat{P}_m}_{\chi_{[0,\tau_m]}\Psi}(s,u)\chi_{B_m}(u)\,\mathrm{d}\pi(s,u) \\
&= \lim_{m\to\infty} \int_0^{t\wedge\tau_m} \int_{B_m} \Psi(s)u\,\mathrm{d}\pi(s,u) \\
&= \sum_{s\in(0,t]} \Psi(s)\Delta P(s),
\end{aligned}$$

a.s. in H for every $t \in [0,T)$. This proves the first equality in (122) and the second equality follows from the definition of the jump measure π. □

6.3 *Comparing Two Integrals with Respect to a Compound Poisson Process*

We now have two notions of stochastic integration with respect to *square-integrable* compound Poisson processes, namely (87) from Definition 6.7 and (117) from Definition 6.20. We show below that the new definition in (117) extends the old definition in (87), so there is no ambiguity about the meaning of stochastic integration with respect to a square-integrable compound Poisson process. We introduce some notation to make the statement of this result precise. Let P be a square-integrable U-valued $\mathscr{F}_t$-compound Poisson with Lévy measure ν. We continue to assume that the filtration $(\mathscr{F}_t)_{t\geq 0}$ is complete and right-continuous. Recall that the covariance operator, say Q, of P is given by equation (49). For each positive integer m we continue to denote by B_m the open unit ball of radius m in U, by P_m the compound Poisson process $P_m(t) := \sum_{s\in(0,t]} \chi_{B_m}(\Delta P(s))\Delta P(s)$, by $\nu_m := \nu|_{B_m}$ its Lévy measure, by Q_m its covariance operator (given by (103)) and $U_0^m := Q_m^{1/2}(U)$. We will also set $U_0 := Q^{1/2}(U)$. For the notion of stochastic integration in Definition 6.20 we use integrands in the space $\mathbf{L}_{P,T}(H)$, while for the notion in Definition 6.7 use integrands in the space $\mathbf{L}^2_{U_0,T}(H)$. In order to show that the former extends the latter we must show that $\mathbf{L}^2_{U_0,T}(H)$ can be viewed naturally as a subspace of $\mathbf{L}_{P,T}(H)$. We begin by defining the inclusion. For each positive integer m we have $Q_m \leq Q$, so by Lemma 5.5 the identity map on $\mathscr{S}(U,H)$ extends uniquely to a continuous linear map $\beta_m \colon \mathbf{L}^2_{U_0,T}(H) \to \mathbf{L}^2_{U_0^m,T}(H)$ with norm less than or equal to 1. Now we define a map

$$\beta \colon \mathbf{L}^2_{U_0,T}(H) \to \prod_{m=1}^{\infty} \mathbf{L}^2_{U_0^m,T}(H) \qquad \text{by}$$

$$\beta(\Psi) = \left(\chi_{[0,\tau_m]}\beta_m(\Psi)\right)_{m=1}^{\infty}, \qquad \text{for all } \Psi \in \mathbf{L}^2_{U_0,T}(H). \tag{123}$$

Basic properties of the map β are given below.

Lemma 6.28 *In the setup above, the range of β is contained in the subspace $\mathbf{L}_{P,T}(H)$ and β is the unique continuous extension of the map $\vartheta : \mathscr{S}(U,H) \to \mathbf{L}_{P,T}(H)$ defined in Proposition 6.14, where $\mathscr{S}(U,H)$ is endowed with the norm it inherits as a subspace of $\mathbf{L}^2_{U_0,T}(H)$.*

Proof Fix $\Psi \in \mathbf{L}^2_{U_0,T}(H)$ and let m be a positive integer. Using Lemma 6.15 we see that

$$\chi_{[0,\tau_m]}\iota_m((\beta(\Psi))_{m+1}) = \chi_{[0,\tau_m]}\iota_m(\chi_{[0,\tau_{m+1}]}\beta_{m+1}(\Psi)) = \chi_{[0,\tau_m]}\iota_m(\beta_{m+1}(\Psi)).$$

Now observe that the composition $\iota_m \circ \beta_{m+1} \colon \mathbf{L}^2_{U_0,T}(H) \to \mathbf{L}^2_{U_0^m,T}(H)$ is a continuous linear extension of the identity map on $\mathscr{S}(U,H)$, whence $\iota_m \circ \beta_{m+1} = \beta_m$ by uniqueness. This shows that

$$\chi_{[0,\tau_m]}\iota_m((\beta(\Psi))_{m+1}) = \chi_{[0,\tau_m]}\beta_m(\Psi) = (\beta(\Psi))_m,$$

for every positive integer m, which is to say that $\beta(\Psi) \in \mathbf{L}_{P,T}(H)$. Continuity of β is clear because $\Psi \mapsto (\beta(\Psi))_m$ is a continuous map from $\mathbf{L}^2_{U_0,T}(H) \to \mathbf{L}^2_{U_0^m,T}(H)$ for each positive integer m. To show that β extends ϑ we just need to observe that

$$\beta(\Psi) = \left(\chi_{[0,\tau_m]}\Psi\right)_{m=1}^{\infty} = \vartheta(\Psi) \qquad \text{for every } \Psi \in \mathscr{S}(U,H).$$

Since $\mathscr{S}(U,H)$ is dense in $\mathbf{L}^2_{U_0,T}(H)$ it follows that β is the unique continuous extension of ϑ. □

We are now ready to show that stochastic integration with respect to a square-integrable $\mathscr{F}_t$-compound Poisson process P, as originally defined in (87) from Definition 6.7, coincides with the notion of stochastic integration in (117) from Definition 6.20 on the image of the map β. This is the precise sense in which (117) extends (87) when P is square-integrable. To avoid confusion we will use the notation $\int_0^t \beta(\Psi)\,\mathrm{d}P$, for $\Psi \in \mathbf{L}^2_{U_0,T}(H)$, to denote the notion of stochastic integration from (117) in Definition 6.20 on the image of β. We will denote by

$\int_0^t \int_U f_\Psi^{\widehat{P}}(s,u)\,\mathrm{d}\pi(s,u)$, for $\Psi \in \mathbf{L}^2_{U_0,T}(H)$, the notion of stochastic integration from (87) in Definition 6.7.

Proposition 6.29 *Let P be a square-integrable U-valued $\mathscr{F}_t$-compound Poisson process with Lévy measure ν, covariance operator Q and $U_0 := Q^{1/2}(U)$. For every $t \in [0,T]$ and $\Psi \in \mathbf{L}^2_{U_0,T}(H)$ we have $\int_0^t \beta(\Psi)\,\mathrm{d}P \in L^2(\Omega; H)$ and*

$$\int_0^t \beta(\Psi)\,\mathrm{d}P = \int_0^t \int_U f_\Psi^{\widehat{P}}(s,u)\,\mathrm{d}\pi(s,u). \tag{124}$$

Proof Fix $t \in [0,T]$ and $\Psi \in \mathbf{L}^2_{U_0,T}(H)$. We begin by showing that $\int_0^t \beta(\Psi)\,\mathrm{d}P \in L^2(\Omega; H)$. By Fatou's lemma and inequality (89) we have

$$\begin{aligned}\mathbf{E}\Big|\int_0^t \beta(\Psi)\,\mathrm{d}P\Big|_H^2 &\leq \liminf_{m\to\infty}\mathbf{E}\Big|\int_0^t \chi_{[0,\tau_m]}\beta_m(\Psi)(s)\,\mathrm{d}P_m(s)\Big|_H^2\\ &\leq 2\liminf_{m\to\infty}(1+t\nu_m(U))\mathbf{E}\int_0^{t\wedge\tau_m} ||\beta_m(\Psi)(s)||^2_{L_2(U_0^m,H)}\;\mathrm{d}s\\ &\leq 2(1+t\nu(U))\liminf_{m\to\infty}\mathbf{E}\int_0^T ||\beta_m(\Psi)(s)||^2_{L_2(U_0^m,H)}\;\mathrm{d}s\\ &\leq 2(1+t\nu(U))\mathbf{E}\int_0^T ||\Psi(s)||^2_{L_2(U_0,H)}\;\mathrm{d}s.\end{aligned}$$

The last line follows because $\beta_m\colon \mathbf{L}^2_{U_0,T}(H) \to \mathbf{L}^2_{U_0^m,T}(H)$ has norm at most 1. Since P is a compound Poisson process we have $\nu(U) < \infty$, so the estimate above shows that $\int_0^t \beta(\Psi)\,\mathrm{d}P \in L^2(\Omega; H)$ and that the map $\Psi \mapsto \int_0^t \beta(\Psi)\,\mathrm{d}P$ is linear and continuous from $\mathbf{L}^2_{U_0,T}(H) \to L^2(\Omega; H)$. Since the right-hand side of (124) is continuous from $\mathbf{L}^2_{U_0,T}(H) \to L^2(\Omega; H)$ (which was shown in the proof of inequality (89)) it suffices to establish (124) for simple processes. But this has already been done in Lemma 6.22. Indeed, since β extends ϑ and $f_\Psi^{\widehat{P}}(s,u) = \Psi(s)u$ for $\Psi \in \mathscr{S}(U,H)$ we have

$$\int_0^t \beta(\Psi)\,\mathrm{d}P = \int_0^t \vartheta(\Psi)\,\mathrm{d}P = \sum_{s\in(0,t]} \Psi(s)\Delta P(s) = \int_0^t \int_U f_\Psi^{\widehat{P}}(s,u)\,\mathrm{d}\pi(s,u),$$

by (118). It follows by continuity that (124) holds for all $\Psi \in \mathbf{L}^2_{U_0,T}(H)$. □

Given a square-integrable compound Poisson process P, it is natural to ask whether the space of integrands $\mathbf{L}_{P,T}(H)$ for the notion of stochastic integration in Definition 6.20 is strictly larger than the space of integrands $\mathbf{L}^2_{Q^{1/2}(U),T}(H)$ for the notion of stochastic integration in Definition 6.7. More precisely, it is natural to ask when β maps $\mathbf{L}^2_{Q^{1/2}(U),T}(H)$ onto $\mathbf{L}_{P,T}(H)$. The more interesting case is, of course, when β is not surjective. In that case, $\mathbf{L}_{P,T}(H)$ is strictly larger than the space of integrands $\mathbf{L}^2_{Q^{1/2}(U),T}(H)$ and Definition 6.20 strictly extends the notion of stochastic integration in Definition 6.7. To avoid trivialities in this discussion we assume that $H \neq \{0\}$.

Proposition 6.30 *Assume that $H \neq \{0\}$. Suppose that P is a square-integrable $\mathscr{F}_t$-compound Poisson process on U with Lévy measure ν and covariance operator Q. Set $U_0 := Q^{1/2}(U)$ and let $\beta\colon \mathbf{L}^2_{U_0,T}(H) \to \mathbf{L}_{P,T}(H)$ be the map defined in (123). Then the following statements are equivalent:*

i) β maps $\mathbf{L}^2_{Q^{1/2}(U),T}(H)$ onto $\mathbf{L}_{P,T}(H)$,
ii) $\sup_{m\geq 1} ||\Psi_m||_{\mathbf{L}^2_{U_0^m,T}(H)} < \infty$ for every $\Psi = (\Psi_m)_{m=1}^\infty \in \mathbf{L}_{P,T}(H)$,
iii) ν is supported on a bounded subset of U.

Proof $i) \implies ii)$ Suppose that β is onto and let $\Psi = (\Psi_m)_{m=1}^\infty \in \mathbf{L}_{P,T}(H)$. By hypothesis there exists some $\Phi \in \mathbf{L}^2_{U_0,T}(H)$ such that $\beta(\Phi) = \Psi$. Using the definition of β in (123) we see that

$$\sup_{m\geq 1} ||\Psi_m||_{\mathbf{L}^2_{U_0^m,T}(H)} = \sup_{m\geq 1} \left|\left|\chi_{[0,\tau_m]}\beta_m(\Phi)\right|\right|_{\mathbf{L}^2_{U_0^m,T}(H)}.$$

The right-hand side above is less than or equal to $||\Phi||_{\mathbf{L}^2_{U_0,T}(H)}$ because β_m has norm at most 1. This shows that $ii)$ holds.

not $iii) \implies$ not $ii)$ Assume that the support of ν is unbounded. We will construct a process $\Psi = (\Psi_m)_{m=1}^\infty$ in the space $\mathbf{L}_{P,T}(H)$ with the property that $||\Psi_m||_{\mathbf{L}^2_{U_0^m,T}(H)} \uparrow \infty$ as $m \to \infty$. For each positive integer m we take $\Psi_m \in L^2(\Omega \times [0,T], \mathscr{P}_{[0,T]}, \mathrm{d}\mathbf{P} \otimes \mathrm{d}t; L(U,H))$ to be of the form

$$\Psi_m(s) := \Big(\sum_{k=1}^m g_k \chi_{(\tau_{k-1},\tau_k]}(s)\Big) S,$$

where $S \in L(U,H)$ is nonzero on the range of $Q^{1/2}$, $(g_k)_{k=1}^\infty$ is a sequence of positive numbers to be chosen later and where we set $\tau_0 := 0$. Recall from the proof of Lemma 3.15 that the set $\{(\omega,s) \in \Omega \times [0,T] : s \leq \tau(\omega)\}$ is predictable for every $\mathscr{F}_t$-stopping time τ. Therefore, the process

$$\chi_{(\tau_{k-1},\tau_k]}(t) = (1 - \chi_{\{t\leq\tau_{k-1}\}}) \cdot \chi_{\{\tau_k\leq t\}},$$

is predictable for each $k \in \{1,\ldots,m\}$, so Ψ_m is predictable. It follows from Example 5.9 that the sequence $\Psi := (\Psi_m)_{m=1}^\infty$ belongs to the Cartesian product $\prod_{m=1}^\infty \mathbf{L}^2_{U_0^m,T}(H)$. We have $\Psi \in \mathbf{L}_{P,T}(H)$ because

$$\chi_{[0,\tau_m]}\iota_m(\Psi_{m+1}) = \chi_{[0,\tau_m]}\Psi_{m+1} = \Psi_m,$$

for every positive integer m. Above we used the fact that $\iota_m(\Phi) = \Phi$ for every $\Phi \in L^2(\Omega \times [0,T], \mathscr{P}_{[0,T]}, \mathrm{d}\mathbf{P} \otimes \mathrm{d}t; L(U,H))$, which was observed during the proof of Proposition 6.27. We point out that the assumption that ν has unbounded support has not been used yet. We will use it to choose a sequence $(g_k)_{k=1}^{\infty}$ so that $ii)$ is violated. For each positive integer m we use (50) to compute

$$\begin{aligned} ||\Psi_m||^2_{\mathbf{L}^2_{U_0^m,T}(H)} &= \mathbf{E}\int_0^t ||\Psi_m(s)||^2_{L_2(U_0^m,H)}\,\mathrm{d}s \\ &= \sum_{k=1}^{m} g_k^2\,\mathbf{E}\int_{\tau_{k-1}}^{\tau_k} ||S||^2_{L_2(U_0^m,H)}\,\mathrm{d}s \\ &= \Big(\sum_{k=1}^{m} g_k^2\,\mathbf{E}[\tau_k - \tau_{k-1}]\Big)\int_{B_m} |Su|_H^2\,\mathrm{d}\nu(u). \end{aligned}$$

Since S does not vanish on the range of $Q^{1/2}$ we have $||S||_{L_2(U_0,H)} > 0$. By the monotone convergence theorem we have $\int_{B_m} |Su|_H^2\,\mathrm{d}\nu(u) \uparrow ||S||^2_{L_2(U_0,H)}$. This means that $\int_{B_m} |Su|_H^2\,\mathrm{d}\nu(u) > \frac{1}{2}\,||S||^2_{L_2(U_0,H)} > 0$ for all sufficiently large m. Next, we claim that $\mathbf{E}[\tau_k - \tau_{k-1}] > 0$ for all k. This is clearly true for $k = 1$. We have $\tau_k > \tau_{k-1}$ a.s. on the event $\{\tau_{k-1} < T\}$ and

$$\mathbf{P}[\tau_{k-1} < T] \geq \mathbf{P}[\Pi(T) = 1, Z_1 \notin B_{k-1}] = e^{-\nu(U)T}\cdot\nu(B_{k-1}^c),$$

where we write $P(t) = \sum_{j=1}^{\Pi(t)} Z_j$ as in Theorem 2.10. Since the support of ν is unbounded we have $\nu(B_{k-1}^c) > 0$ for every $k \geq 2$. Now set $g_k := (\mathbf{E}[\tau_k - \tau_{k-1}])^{-1/2}$ so that

$$||\Psi_m||^2_{\mathbf{L}^2_{U_0^m,T}(H)} = m\int_{B_m} |Su|_H^2\,\mathrm{d}\nu(u) \geq \frac{m}{2}\,||S||^2_{L_2(U_0,H)}\,,$$

for all sufficiently large m. This shows that Ψ is an element of $\mathbf{L}_{P,T}(H)$ for which $ii)$ does not hold.

$iii) \implies i)$ Assume that ν has bounded support, then $\nu(B_{m_0}^c) = 0$ for some positive integer m_0. As a result, the following statements hold whenever $m \geq m_0$:

- $\tau_m = T$ a.s.,
- $Q_m = Q$,
- $U_0^m = U_0$,
- ι_m is the identity map on $\mathbf{L}^2_{U_0,T}(H)$.

For every $\Psi \in \mathbf{L}_{P,T}(H)$ we have $\Psi_{m+1} = \Psi_m$ in the space $\mathbf{L}^2_{U_0,T}(H)$ for every $m \geq m_0$ and therefore $\Psi_m = \Psi_{m_0}$ for every $m \geq m_0$. We claim that $\Psi = \beta(\Psi_{m_0})$ in the space $\mathbf{L}_{P,T}(H)$. We need to show that $\Psi_m = \chi_{[0,\tau_m]}\beta_m(\Psi_{m_0})$ in the space

$\mathbf{L}^2_{U_0^m,T}(H)$ for every positive integer m. We have already observed that this is true for all $m \geq m_0$. For $m \leq m_0$ we use induction on the ordered set $(m_0, m_0-1, \ldots, 2, 1)$. Suppose that $\Psi_{m+1} = \chi_{[0,\tau_{m+1}]}\beta_{m+1}(\Psi_{m_0})$ for some $m \in \{1, \ldots, m_0 - 1\}$. Using Lemma 6.15 we see that

$$\Psi_m = \chi_{[0,\tau_m]}\iota_m(\Psi_{m+1}) = \chi_{[0,\tau_m]}\iota_m(\chi_{[0,\tau_{m+1}]}\beta_{m+1}(\Psi_{m_0})) = \chi_{[0,\tau_m]}\iota_m(\beta_{m+1}(\Psi_{m_0})).$$

We have seen in the proof of Lemma 6.28 that $\iota_m \circ \beta_{m+1} = \beta_m$ for every positive integer m, so $\Psi_m = \chi_{[0,\tau_m]}\beta_m(\Psi_{m_0})$. This shows that $\Psi = \beta(\Psi_{m_0})$ in the space $\mathbf{L}_{P,T}(H)$, so β is surjective. □

Remark 6.31 The assumption in Proposition 6.30 that P be a square-integrable compound Poisson process is only required to define the map β from (123). We show here that $ii)$ and $iii)$ fail when P is a non-square-integrable $\mathscr{F}_t$-compound Poisson process. Since the Lévy measure of a compound Poisson process is finite it is clear that the Lévy measure of a non-square-integrable compound Poisson process cannot be supported on a bounded set. To show that $ii)$ fails we can use a similar construction as in the proof of not $iii) \implies$ not $ii)$ in Proposition 6.30. In that construction we chose $S \in L(U, H)$ so that the quantity $\int_U |Su|_H^2 \, \mathrm{d}\nu(u) = ||S||^2_{L_2(Q^{1/2}(U),H)}$ was strictly positive. The operator Q is no longer available when P is not square-integrable but it will still be possible to choose $S \in L(U, H)$ such that $\int_U |Su|_H^2 \, \mathrm{d}\nu(u) > 0$, which is sufficient to repeat the reasoning in the construction used to prove not $iii) \implies$ not $ii)$. Let V be the closed subspace of U generated by the support of ν and let $\{v_n\}_n$ be an orthonormal basis for V (which could be finite dimensional). Let h be a unit vector in H and define $S \in L(U, H)$ by $Su := (u, v_1)_U \, h$. We claim that $\int_U |Su|_H^2 \, \mathrm{d}\nu(u) > 0$, or equivalently, that $\nu(\{S \neq 0\}) > 0$. Since V is the minimal closed subspace containing the support of ν there exists a vector u in the support of ν such that $(u, v_1)_U \neq 0$. Since u belongs to the support of ν we have $\nu(u + B_r(0)) > 0$ for every $r > 0$. For $r \in (0, |\,(u, v_1)_U\,|)$ we have $u + B_r(0) \subseteq \{S \neq 0\}$, so $\nu(\{S \neq 0\}) > 0$. With this operator S we can use the same construction as in the proof of not $iii) \implies$ not $ii)$ to construct a sequence $(\Psi_m)_{m=1}^\infty \in \mathbf{L}_{P,T}(H)$ that violates $ii)$.

Corollary 6.32 *Assume that $H \neq \{0\}$. Let P be an $\mathscr{F}_t$-compound Poisson process on U with Lévy measure ν and covariance operator Q and set $U_0 := Q^{1/2}(U)$. If ν is supported on a bounded subset of U, then $\mathbf{L}_{P,T}(H)$ is a Banach space. Under the norm*

$$\left|\left|(\Psi_m)_{m=1}^\infty\right|\right|_{\mathbf{L}_{P,T}(H)} := \sup_m ||\Psi_m||_{\mathbf{L}^2_{U_0^m,T}(H)} = \lim_{m\to\infty} ||\Psi_m||_{\mathbf{L}^2_{U_0^m,T}(H)}, \tag{125}$$

the map $\beta\colon \mathbf{L}^2_{U_0,T}(H) \to \mathbf{L}_{P,T}(H)$ defined in (123) is an isometric isomorphism.

Proof Since the support of ν is bounded we have $\sup_m ||\Psi_m||_{\mathbf{L}^2_{U_0^m,T}(H)} < \infty$ for every $\Psi = (\Psi_m)_{m=1}^\infty \in \mathbf{L}_{P,T}(H)$ by Proposition 6.30. The fact that the limit equals the supremum follows from the estimate (108)[2]. Equation (125) obviously defines a norm on $\mathbf{L}_{P,T}(H)$ that induces the product topology. Since $\mathbf{L}_{P,T}(H)$ is complete it follows that $\mathbf{L}_{P,T}(H)$ is a Banach space. We know from Proposition 6.30 that β is onto. We have also seen during the proof of that proposition that β is isometric. Indeed, there exists a positive integer m_0 such that $\nu(B^c_{m_0}) = 0$ and for every $\Phi \in \mathbf{L}^2_{U_0,T}(H)$ we have $\beta(\Phi)_m = \Phi$ for all $m \geq m_0$, so $||\beta(\Phi)||_{\mathbf{L}_{P,T}(H)} = ||\Phi||_{\mathbf{L}^2_{U_0,T}(H)}$. □

We have shown that the two notions of stochastic integration with respect to a square-integrable U-valued $\mathscr{F}_t$-compound Poisson process P in Definition 6.7 and Definition 6.20 coincide when the Lévy measure of P is supported on a bounded subset of U. Corollary 6.32 shows that the spaces of integrands used in Definition 6.7 and Definition 6.20 are isometrically isomorphic via the map β defined in (123). Proposition 6.29 shows that the two notions of stochastic integration coincide. On the other hand, Proposition 6.30 shows that the space $\mathbf{L}_{P,T}(H)$ is strictly larger than the range of β when P is square-integrable but has a Lévy measure with unbounded support. We still know from Proposition 6.29 that the notion of stochastic integration in Definition 6.20 coincides with the notion in Definition 6.7 after composing with β, but we see from Proposition 6.30 that the notion in Definition 6.20 strictly extends the notion in Definition 6.7.

6.4 Summary of the Non-Square-Integrable Case

For every $\mathscr{F}_t$-compound Poisson process P, we have rigorously constructed a space of integrands $\mathbf{L}_{P,T}(H)$ for stochastic integration with respect to P as a projective limit of Hilbert spaces. We have constructed the stochastic integral with respect to P on $\mathbf{L}_{P,T}(H)$ in two steps, by applying the continuous map $I^P_t : \mathbf{L}_{P,T}(H) \to Y^{(t)}$ to $\Psi \in \mathbf{L}_{P,T}(H)$ and then taking the limit of the resulting sequence, which stabilizes as a sum of finitely many vectors in H, $\mathbf{P}$-a.s. We have shown that this notion of stochastic integration with respect to P agrees with the stochastic integral constructed by localization as presented by Peszat and Zabczyk. We believe that the use of projective limits helps to organize the localization procedure and offers additional benefits. First, we can define the space of integrands explicitly as the Fréchet space $\mathbf{L}_{P,T}(H)$. Second, $\mathbf{L}_{P,T}(H)$ contains the natural image of the simple processes $\mathscr{S}(U, H)$ as a dense subspace by Proposition 6.14. Third, stochastic integration with respect to P is defined via a continuous map on the space of integrands, namely I^P_t. Fourth, when P is square-integrable, we are able to show in Proposition 6.30 that the new notion of stochastic integration defined using projective limits in Definition 6.20 is exactly the same as the original notion

[2] This is true in general, even if P is not square-integrable, but the supremum can be ∞ for some $\Psi \in \mathbf{L}_{P,T}(H)$ when the support of ν is unbounded.

defined in Definition 6.7 when the Lévy measure of P has bounded support and that the new notion is a strict extension of the old notion when the Lévy measure of P has unbounded support. We have also shown in Proposition 6.24 precisely how the stochastic integral $\int_0^t \Psi(s)\,\mathrm{d}P(s)$, for $\Psi \in \mathbf{L}_{P,T}(H)$, can be written in the form $\int_0^t \int_U f_\Psi^P(s,u)\,\mathrm{d}\pi(u,s)$, where $f_\Psi^P : \Omega \times [0,T] \times U \to H$ and π is the jump measure of P. We have used this in Corollary 6.26 to identify the jumps of the stochastic integral with respect to P. We have also shown in Proposition 6.27 that the stochastic integral of a predicable process Ψ with values in $L(U,H)$ with respect to a compound Poisson process P agrees a.s. with the Riemann-Stieltjes integral of Ψ with respect to P.

We close this section by showing how to write a stochastic integral with respect to a Lévy process in the framework presented by Peszat and Zabczyk in the form used by Ikeda and Watanabe. To be slightly more precise, let L be a U-valued Lévy process with Wiener part W and jump measure π; we show how to rigorously interpret the formal decomposition $\mathrm{d}L = a\,\mathrm{d}t + \mathrm{d}W + \mathrm{d}\widehat{\pi} + \mathrm{d}\pi$ suggested by the Lévy-Khinchin decomposition in (11). Let L be a U-valued $\mathscr{F}_t$-Lévy process with Lévy measure ν and jump measure π. We assume that the filtration $(\mathscr{F}_t)_{t\geq 0}$ is complete and right-continuous. By Theorem 2.15 there exists a vector $a \in U$, a U-valued Wiener process W and independent compound Poisson processes $(P_n)_{n=0}^\infty$ on U (also independent of W) such that

$$L(t) = at + W(t) + P_0(t) + \sum_{n=1}^{\infty} \widehat{P}_n(t), \tag{126}$$

and, with probability 1, the series converges uniformly in U on compact subsets of $[0,\infty)$. Furthermore, the compound Poisson processes can be chosen so that P_0 has Lévy measure $\nu|_{B_1^c}$ and P_n has Lévy measure $\nu|_{B_{1/n}\setminus B_{1/(n+1)}}$ for every $n \geq 1$. In particular, P_n is square-integrable for $n \geq 1$ but P_0 is not necessarily square-integrable. Corollary 5.4 shows that W is an $\mathscr{F}_t$-Wiener process and P_n is an $\mathscr{F}_t$-compound Poisson process for each nonnegative integer n. Therefore, the processes W and $\mathscr{L} := \sum_{n=1}^\infty \widehat{P_n}$ satisfy Assumption 3.3 with respect to the filtration $(\mathscr{F}_t)_{t\geq 0}$. Let $Q_0 \in L_1^+(U)$ be the covariance operator of W and let $Q_1 \in L_1^+(U)$ be the covariance operator of $\mathscr{L}$. We have separate notions of stochastic integration with respect to each term on the right-hand side of (126). If we want to integrate a single process Ψ with respect to L, then Ψ should belong to, or at least have a natural image in, the spaces of integrands for stochastic integration with respect to each term on the right-hand side of (126). If $\Psi \in L^2(\Omega \times [0,T], \mathscr{P}_{[0,T]}, \mathrm{d}\mathbf{P} \otimes \mathrm{d}t; L(U,H))$, then we have $\Psi \in \mathbf{L}^2_{Q_0^{1/2}(U),T}(H)$, $\Psi \in \mathbf{L}^2_{Q_1^{1/2}(U),T}(H)$, $\alpha(\Psi) \in \mathbf{L}_{P_0,T}(H)$ and the integral $\int_0^t \Psi(s)a\,\mathrm{d}s$ is well-defined pathwise as a Riemann integral. Therefore, it makes sense to define the stochastic integral of $\Psi \in L^2(\Omega \times [0,T], \mathscr{P}_{[0,T]}, \mathrm{d}\mathbf{P} \otimes$

dt; $L(U, H)$) with respect to L by

$$\int_0^t \Psi(s)\,\mathrm{d}L(s) := \int_0^t \Psi(s)a\,\mathrm{d}s + \int_0^t \Psi(s)\,\mathrm{d}W(s) \\ + \int_0^t \Psi(s)\,\mathrm{d}\mathscr{L}(s) + \int_0^t \alpha(\Psi)(s)\,\mathrm{d}P_0(s). \tag{127}$$

Using Proposition 5.14, Example 5.9 and Proposition 6.27 we see that

$$\begin{aligned} \int_0^t \Psi(s)\,\mathrm{d}L(s) &= \int_0^t \Psi(s)a\,\mathrm{d}s + \int_0^t \Psi(s)\,\mathrm{d}W(s) \\ &\quad + \int_0^t \int_{B_1} f_\Psi^{\mathscr{L}}(s,u)\,\mathrm{d}\widehat{\pi}(s,u) + \int_0^t \int_{B_1^c} f_{\alpha(\Psi)}^{P_0}(s,u)\,\mathrm{d}\pi(s,u) \\ &= \int_0^t \Psi(s)a\,\mathrm{d}s + \int_0^t \Psi(s)\,\mathrm{d}W(s) \\ &\quad + \int_0^t \int_{B_1} \Psi(s)u\,\mathrm{d}\widehat{\pi}(s,u) + \sum_{s\in(0,t]} \Psi(s)\Delta P_0(s). \end{aligned} \tag{128}$$

Next, we discuss more general integrands. For this, let $Q := Q_0 + Q_1$ and $M := W + \mathscr{L}$. Recall from Lemma 5.6 that Q is the covariance operator of the sum M. Given processes $\Psi_1 \in L^2(\Omega \times [0,T], \mathscr{P}_{[0,T]}, \mathrm{d}\mathbf{P} \otimes \mathrm{d}t; L(U,H))$, $\Psi_2 \in \mathbf{L}^2_{Q^{1/2}(U),T}(H)$, and $\Psi_3 \in \mathbf{L}_{P_0,T}(H)$ it is reasonable to define the stochastic integral of the tuple $\Psi := (\Psi_1, \Psi_2, \Psi_3)$ with respect to L as the sum

$$\int_0^t \Psi(s)\,\mathrm{d}L(s) := \int_0^t \Psi_1(s)a\,\mathrm{d}s + \int_0^t \Psi_2(s)\,\mathrm{d}M(s) + \int_0^t \Psi_3(s)\,\mathrm{d}P_0(s), \tag{129}$$

as done by Peszat and Zabczyk. As we have seen, the two stochastic integral terms on the right-hand side of (129) can be expressed as a sum of stochastic integrals with respect to W, $\widehat{\pi}$ and π. Specifically, using Theorem 5.18 and Proposition 6.24 we see that (129) can be stated equivalently as

$$\int_0^t \Psi(s)\,\mathrm{d}L(s) = \int_0^t \Psi_1(s)a\,\mathrm{d}s + \int_0^t \iota_0(\Psi_2)(s)\,\mathrm{d}W(s) \\ + \int_0^t \int_{B_1} f^{\mathscr{L}}_{\iota_1(\Psi_2)}(s,u)\,\mathrm{d}\widehat{\pi}(s,u) + \int_0^t \int_{B_1^c} f^{P_0}_{\Psi_3}(s,u)\,\mathrm{d}\pi(s,u). \tag{130}$$

Equation (130) gives a rigorous interpretation to the heuristic $\mathrm{d}L = a\,\mathrm{d}t + \mathrm{d}W + \mathrm{d}\widehat{\pi} + \mathrm{d}\pi$ suggested by the Lévy-Khinchin decomposition and shows how to express stochastic integration with respect to a general Lévy process as defined in the setting of Peszat and Zabczyk in the framework presented by Ikeda and Watanabe. In this way stochastic integration with respect to a general Lévy process as defined in the setting of Peszat and Zabczyk can be viewed as a special case of the theory of stochastic integration presented by Ikeda and Watanabe.

Acknowledgments This work was partially supported by the National Science Foundation (NSF) under the grant DMS-1510249 and by the Research Fund of Indiana University. The authors gratefully acknowledge the use of the book [15] during a semester-long seminar, and some private exchanges of emails with the authors.

References

1. Albeverio, S., Brzeźniak, Z., Wu, J.: Existence of global solutions and invariant measures for stochastic differential equations driven by Poisson type noise with non-Lipschitz coefficients. J. Math. Anal. Appl. **371**, 309–322 (2010)
2. Applebaum, D.: Lévy Processes and Stochastic Calculus. Cambridge University Press, Cambridge (2009)
3. Bensoussan, A., Lions, J.L.: Contrôle Impulsionnel et in Équations Quasi Variationnelles. Gauthier-Villars, Paris (1982)
4. Brzeźniak, Z., Liu, W., Zhu, J.: Strong solutions for SPDE with locally monotone coefficients driven by Lévy noise. Nonlinear Anal. Real World Appl. **17**, 283–310 (2014)
5. Cyr, J., Tang, S., Temam, R.: Review of local and global existence results for stochastic PDEs with Lévy noise. To appear
6. Da Prato, G., Zabczyk, J.: Stochastic Equations in Infinite Dimensions. Cambridge University Press, Cambridge (2014)
7. Debussche, A., Högele, M., Imkeller, P.: The Dynamics of Nonlinear Reaction-Diffusion Equations with Small Lévy Noise. Springer, Cham (2013)
8. Dubinsky, E.: Projective and inductive limits of Banach spaces. Studia Math. **42**, 259–263 (1972)
9. Dugundji, J.: An extension of Tietze's theorem. Pac. J. Math. **1**, 353–367 (1951)
10. Ikeda, N., Watanabe, S.: Stochastic Differential Equations and Diffusion Processes. North-Holland, Amsterdam; Kodansha, Tokyo (1989)
11. Kallenberg O.: Foundations of Modern Probability. Springer, New York (1997)
12. Marinelli, C., Röckner, M.: On the maximal inequalities of Burkholder, Davis and Gundy. Expo. Math. **34**, 1–26 (2016)
13. Métivier, M.: Semimartingales: A Course on Stochastic Processes. Walter de Gruyter & Co., Berlin-New York (1982)
14. Motyl, E.: Stochastic Navier-Stokes equations driven by Lévy noise in unbounded 3D domains. Potential Anal. **38**, 863–912 (2013)
15. Peszat, S., Zabczyk, J.: Stochastic Partial Differential Equations with Lévy Noise: An Evolution Equation Approach. Cambridge University Press, Cambridge (2007)
16. Prévôt, C., Röckner, M.: A Concise Course on Stochastic Partial Differential Equations. Springer, Berlin (2007)
17. Rüdiger, B.: Stochastic integration with respect to compensated Poisson random measures on separable Banach spaces. Stoch. Stoch. Rep. **76**, 213–242 (2004)
18. Schaefer, H.H., Wolff, M.P.: Topological Vector Spaces. Springer, New York (1999)