Lecture Notes in Mathematics 2179
CIME Foundation Subseries

Bernard Dacorogna · Nicola Fusco
Stefan Müller · Vladimir Sverak

# Vector-Valued Partial Differential Equations and Applications

## Cetraro, Italy 2013

John Ball · Paolo Marcellini *Editors*

Fondazione C.I.M.E.
Springer

# Lecture Notes in Mathematics 2179



More information about this series at http://www.springer.com/series/304

# Fondazione C.I.M.E., Firenze

C.I.M.E. stands for *Centro Internazionale Matematico Estivo*, that is, International Mathematical Summer Centre. Conceived in the early fifties, it was born in 1954 in Florence, Italy, and welcomed by the world mathematical community: it continues successfully, year for year, to this day.

Many mathematicians from all over the world have been involved in a way or another in C.I.M.E.'s activities over the years. The main purpose and mode of functioning of the Centre may be summarised as follows: every year, during the summer, sessions on different themes from pure and applied mathematics are offered by application to mathematicians from all countries. A Session is generally based on three or four main courses given by specialists of international renown, plus a certain number of seminars, and is held in an attractive rural location in Italy.

The aim of a C.I.M.E. session is to bring to the attention of younger researchers the origins, development, and perspectives of some very active branch of mathematical research. The topics of the courses are generally of international resonance. The full immersion atmosphere of the courses and the daily exchange among participants are thus an initiation to international collaboration in mathematical research.

**C.I.M.E. Director (2002 – 2014)**
Pietro Zecca
Dipartimento di Energetica "S. Stecco"
Università di Firenze
Via S. Marta, 3
50139 Florence
Italy
*e-mail: zecca@unifi.it*

**C.I.M.E. Director (2015 – )**
Elvira Mascolo
Dipartimento di Matematica "U. Dini"
Università di Firenze
viale G.B. Morgagni 67/A
50134 Florence
Italy
*e-mail: mascolo@math.unifi.it*

**C.I.M.E. Secretary**
Paolo Salani
Dipartimento di Matematica "U. Dini"
Università di Firenze
viale G.B. Morgagni 67/A
50134 Florence
Italy
*e-mail: salani@math.unifi.it*

CIME activity is carried out with the collaboration and financial support of INdAM (Istituto Nazionale di Alta Matematica)

For more information see CIME's homepage: **http://www.cime.unifi.it**

Bernard Dacorogna • Nicola Fusco •
Stefan Müller • Vladimir Sverak

# Vector-Valued Partial Differential Equations and Applications

## Cetraro, Italy 2013

John Ball • Paolo Marcellini
*Editors*

Springer

FONDAZIONE CIME ROBERTO CONTI
CENTRO INTERNAZIONALE MATEMATICO ESTIVO
INTERNATIONAL MATHEMATICAL SUMMER CENTER

*Authors*

Bernard Dacorogna
Section de Mathématiques, EPFL
Lausanne, Switzerland

Nicola Fusco
Dipartimento di Matematica e Applicazioni
"R. Caccioppoli"
Università degli Studi di Napoli
"Federico II"
Napoli, Italy

Stefan Müller
Hausdorff Center for Mathematics &
Department of Applied Mathematics
University of Bonn
Bonn, Germany

Vladimir Sverak
School of Mathematics
University of Minnesota
Minneapolis, MN, USA

*Editors*

John Ball
Mathematical Institute
University of Oxford
Oxford, United Kingdom

Paolo Marcellini
Dipartimento di Matematica
Università di Firenze
Firenze, Italy

ISSN 0075-8434 ISSN 1617-9692 (electronic)
Lecture Notes in Mathematics
C.I.M.E. Foundation Subseries
ISBN 978-3-319-54513-4 ISBN 978-3-319-54514-1 (eBook)
DOI 10.1007/978-3-319-54514-1

Library of Congress Control Number: 2017940495

Mathematics Subject Classification (2010): 49-XX, 35-XX, 35QXX



This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

We are proud to introduce, as the scientific organisers, the 2013 CIME Course *Vector-valued Partial Differential Equations and Applications*, which took place at Cetraro (Cosenza, Italy) from July 8 to 12, 2013, with the following speakers and courses of lectures:

Bernard Dacorogna (École Polytechnique Fédérale de Lausanne, Switzerland), *The pullback equation.*

Nicola Fusco (Università degli Studi di Napoli Federico II, Italy), *The stability of the isoperimetric inequality*.

Stefan Müller (Universität Bonn, Germany), *Mathematical problems in thin elastic sheets: scaling limits, packing, crumpling and singularities*.

Vladimir Šverák (University of Minnesota, USA), *Aspects of PDEs related to fluid flows*.

The programme included a special session to celebrate the 60th birthday of *Bernard Dacorogna*, with lectures by *Gianni Dal Maso, Carlo Sbordone, Giovanni Cupini, Emanuele Paolini* and *Giovanni Pisante*.

That the meeting was such a success was a consequence of the distinction of the speakers and the high level of their lectures, as evidenced by the quality of the notes in this volume, as well as the participation and active involvement of the participants, who numbered well over 100.

We now briefly describe the course notes included in this set of Lecture Notes, starting with the course of Bernard Dacorogna on the *pullback equation*. A map $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ solves the *pullback equation* $\varphi^* (g) = f$ if it is a diffeomorphism which satisfies the equation with $f, g$ differential $k$-forms with $0 \leq k \leq n$. For instance, in the case $k = n$, the equation takes the form $g(\varphi(x)) \det \nabla\varphi(x) = f(x)$. *Local existence* is analysed, as well as *global existence* in the Hölder space $C^{r,\alpha}$.

In his course, Nicola Fusco considered the stability of the *isoperimetric inequality*. Once we know that, for a given volume, balls are the unique area minimisers,

the next natural question is to understand what happens when a set $E$ has the same volume of a ball $B$ and a slightly bigger surface area. Precisely, one would like to show that in this case $E$ must be close in a proper sense to a translation of $B$. The *stability of the isoperimetric inequality* for general sets of *finite perimeter* is analysed in detail, the proof being based on a suitable symmetrisation argument aimed at reducing a general set of finite perimeter to an axially symmetric bounded set with a centre of symmetry.

Stefan Müller presented in his course an outline of the theory of *thin elastic sheets*; in particular, he considers the *limiting behaviour* of thin elastic objects as the thickness $h$ goes to zero. Mathematically one can distinguish two types of problems: either where the solution has a *well-defined limit* as $h \to 0$, when the natural goal is to characterise the limit, or where the solution develops *increasing complexity*.

The course of Vladimir Šverák concerned two main themes. The first deals with the *long-time behaviour* of solutions of the $2D$ incompressible Euler equations and other Hamiltonian equations. The second theme is related to the *problem of uniqueness* of the Leray–Hopf weak solutions with $L^2$ initial data.

We are pleased to express our appreciation to the speakers for their excellent lectures and to the participants for contributing to the success of the CIME Course. We had in Cetraro an interesting, rich and friendly atmosphere, created by the speakers, by the participants and by the CIME Organisers, in particular *Pietro Zecca (CIME Director)* and *Elvira Mascolo (CIME Secretary)*. At the date of publication, *Elvira* now has the role of CIME Director, while the CIME Secretary is *Paolo Salani*. We thank all of them warmly.

**Acknowledgements** CIME activity is carried out with the collaboration and financial support of: INdAM (Istituto Nazionale di Alta Matematica)—MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca)—Ente Cassa di Risparmio di Firenze.

Oxford, UK — John Ball
Firenze, Italy — Paolo Marcellini

# Contents

# The Pullback Equation

**Bernard Dacorogna**

## 1 Introduction

The aim of this course is the study of the *pullback equation*

$$\varphi^* (g) = f. \tag{1}$$

More precisely we want to find a map $\varphi : \mathbb{R}^n \to \mathbb{R}^n$, preferably we want this map to be a diffeomorphism, that satisfies the above equation, where $f, g$ are differential $k$-forms, $0 \leq k \leq n$. Most of the time we will require these two forms to be closed. Before going further let us examine the exact meaning of (1). We write

$$g(x) = \sum_{1 \leq i_1 < \cdots < i_k \leq n} g_{i_1 \cdots i_k}(x)\, dx^{i_1} \wedge \cdots \wedge dx^{i_k}$$

and similarly for $f$. The meaning of (1) is that

$$\sum_{1 \leq i_1 < \cdots < i_k \leq n} g_{i_1 \cdots i_k}(\varphi)\, d\varphi^{i_1} \wedge \cdots \wedge d\varphi^{i_k} = \sum_{1 \leq i_1 < \cdots < i_k \leq n} f_{i_1 \cdots i_k} dx^{i_1} \wedge \cdots \wedge dx^{i_k}$$

where

$$d\varphi^i = \sum_{j=1}^{n} \frac{\partial \varphi^i}{\partial x_j} dx^j.$$

B. Dacorogna (✉)
Section de Mathématiques, EPFL 1015 Lausanne, Switzerland
e-mail: bernard.dacorogna@epfl.ch



J. Ball, P. Marcellini (eds.), *Vector-Valued Partial Differential Equations and Applications*, Lecture Notes in Mathematics 2179,
DOI 10.1007/978-3-319-54514-1_1

This turns out to be a *non-linear* (if $2 \leq k \leq n$) homogeneous of degree $k$ (in the derivatives) first order system of $\binom{n}{k}$ partial differential equations. Let us see the form that the equation takes when $k = 0, 1, 2, n$.

*Case:* $k = 0$. The Eq. (1) reads as

$$g(\varphi(x)) = f(x)$$

while

$$dg = 0 \quad \Leftrightarrow \quad \operatorname{grad} g = 0.$$

We will be, only marginally, interested in this elementary case, which is trivial for closed forms. In any case (1) is *not*, when $k = 0$, a differential equation.

*Case:* $k = 1$. The form $g$, and analogously for $f$, can be written as

$$g(x) = \sum_{i=1}^{n} g_i(x)\, dx^i.$$

The Eq. (1) becomes then

$$\sum_{i=1}^{n} g_i(\varphi(x))\, d\varphi^i = \sum_{i=1}^{n} f_i(x)\, dx^i$$

while

$$dg = 0 \quad \Leftrightarrow \quad \operatorname{curl} g = 0 \quad \Leftrightarrow \quad \frac{\partial g_i}{\partial x_j} - \frac{\partial g_j}{\partial x_i} = 0,\ 1 \leq i < j \leq n.$$

Writing

$$d\varphi^i = \sum_{j=1}^{n} \frac{\partial \varphi^i}{\partial x_j} dx^j$$

and substituting into the equation, we find that (1) is equivalent to

$$\sum_{j=1}^{n} g_j(\varphi(x)) \frac{\partial \varphi^j}{\partial x_i}(x) = f_i(x), \quad 1 \leq i \leq n.$$

This is a system of $\binom{n}{1} = n$ first order *linear* (in the first derivatives) partial differential equations.

*Case:* $k = 2$. The form $g$, and analogously for $f$, can be written as

$$g = \sum_{1 \le i < j \le n} g_{ij}(x)\, dx^i \wedge dx^j$$

while

$$dg = 0 \quad \Leftrightarrow \quad \frac{\partial g_{ij}}{\partial x_k} - \frac{\partial g_{ik}}{\partial x_j} + \frac{\partial g_{jk}}{\partial x_i} = 0,\ 1 \le i < j < k \le n.$$

The equation $\varphi^*(g) = f$ becomes

$$\sum_{1 \le p < q \le n} g_{pq}(\varphi(x))\, d\varphi^p \wedge d\varphi^q = \sum_{1 \le i < j \le n} f_{ij}(x)\, dx^i \wedge dx^j.$$

We get, as before, that (1) is equivalent, for every $1 \le i < j \le n$, to

$$\sum_{1 \le p < q \le n} g_{pq}(\varphi(x)) \left( \frac{\partial \varphi^p}{\partial x_i} \frac{\partial \varphi^q}{\partial x_j} - \frac{\partial \varphi^p}{\partial x_j} \frac{\partial \varphi^q}{\partial x_i} \right) = f_{ij}(x)$$

which is a *non-linear* homogeneous of degree 2 (in the derivatives) system of $\binom{n}{2} = \frac{n(n-1)}{2}$ first order partial differential equations.

*Case:* $k = n$. In this case we always have $df = dg = 0$. By abuse of notations, if we identify volume forms and functions, we get that the equation $\varphi^*(g) = f$ becomes

$$g(\varphi(x)) \det \nabla \varphi(x) = f(x).$$

It is then a non-linear homogeneous of degree $n$ (in the derivatives) first order partial differential equation.

The main questions that we will discuss are the following.

(1) *Algebraic case*. When the forms are constants, it is natural to seek for solutions $\varphi$ of the form $\varphi(x) = Ax$ where $A$ is an invertible $n \times n$ matrix. Therefore the problem turns out to be of linear algebraic nature. For example when $k = 2$ we can associate, in a unique way, to any 2-form

$$g = \sum_{1 \le i < j \le n} g_{ij} dx^i \wedge dx^j$$

a *skew symmetric matrix* $G \in \mathbb{R}^{n\times n}$ (i.e. $G^t = -G$)

$$G = \begin{pmatrix} 0 & g_{12} & g_{13} & \cdots & g_{1n} \\ -g_{12} & 0 & g_{23} & \cdots & g_{2n} \\ -g_{13} & -g_{23} & 0 & \cdots & g_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -g_{1n} & -g_{2n} & -g_{3n} & \cdots & 0 \end{pmatrix}.$$

We therefore have

$$\varphi^*(g) = f \quad \Leftrightarrow \quad AGA^t = F.$$

Since any skew symmetric matrix has even rank, we have that if

$$\operatorname{rank} G = \operatorname{rank} F = 2m \leq n$$

then it is always possible to find an invertible matrix $A$. The canonical form is then

$$J_m = \begin{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} & \cdots & 0 & 0 \cdots 0 \\ \vdots & \ddots & \vdots & \vdots \cdots \vdots \\ 0 & \cdots & \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} & 0 \cdots 0 \\ 0 & \cdots & 0 & 0 \cdots 0 \\ \vdots & \cdots & \vdots & \vdots \ddots \vdots \\ 0 & \cdots & 0 & 0 \cdots 0 \end{pmatrix}.$$

(2) *Local existence*. This is the easiest question. We will handle fairly completely the case of closed 2-forms, which is the case of Darboux theorem. The cases of 1 and $(n-1)$-forms as well as the case of $n$-forms will also be dealt with. It will turn out that the case $3 \leq k \leq n-2$ is much more difficult and we will be able to handle only closed $k$-forms with special structure.

(3) *Existence of canonical forms*. It will turn out that (for closed forms):

- when $k = 1$, then the canonical form is $g = dx^1$
- when $k = 2$, then the canonical form is, depending of the rank of the form,

$$g = \sum_{i=1}^{m} dx^{2i-1} \wedge dx^{2i}$$

– when $k = n - 1$, then the canonical form is

$$g = dx^1 \wedge \cdots \wedge dx^{n-1}$$

– when $k = n$, then the canonical form is

$$g = dx^1 \wedge \cdots \wedge dx^n.$$

(4) *Global existence*. This is a much more difficult problem. We will obtain results in the case of volume forms and of closed 2-forms.
(5) *Regularity*. A special emphasis will be given on getting sharp regularity results. For this reason we will have to work with Hölder spaces $C^{r,q}$, $0 < q < 1$, and not with spaces $C^r$. We will not deal with Sobolev spaces, apart for some linear problems. In the present context the reason is that Hölder spaces form an algebra contrary to Sobolev spaces (with low exponents).
(6) *Selection principle*. In all cases discussed here, once the existence part has been settled, it turns out that there are, in general, infinitely many solutions. Therefore the problem of selecting a solution with further properties becomes an important one. We will discuss very briefly this difficult problem in the cases $k = 2$ and $k = n$.
(7) *Invariants*. Finally the question of the invariants will be discussed. We will see that the rank and the closedness are two invariants. They are the only one (at least for the local problem) when $k = 1, 2, n - 1, n$, but there are others when $3 \leq k \leq n - 2$.

The course is based on the recent book [16], to which we refer for all missing proofs.

## 2 Algebraic Preliminaries

We now gather some algebraic results about exterior forms that are used throughout the course. Let $1 \leq k \leq n$ be an integer (if $k > n$, we set $f = 0$). An exterior $k$-form will be denoted by

$$f = \sum_{1 \leq i_1 < \cdots < i_k \leq n} f_{i_1 \cdots i_k} e^{i_1} \wedge \cdots \wedge e^{i_k}.$$

The set of exterior $k$-forms over $\mathbb{R}^n$ is a vector space and is denoted $\Lambda^k(\mathbb{R}^n)$ and its dimension is

$$\dim(\Lambda^k(\mathbb{R}^n)) = \binom{n}{k}.$$

If $k = 0$, we set

$$\Lambda^0(\mathbb{R}^n) = \mathbb{R}.$$

By abuse of notations, we will, when convenient and in order not to burden the notations, identify $k$-forms with vectors in $\mathbb{R}^{\binom{n}{k}}$.

(i) The *exterior product* of $f \in \Lambda^k(\mathbb{R}^n)$ with $g \in \Lambda^l(\mathbb{R}^n)$, denoted by $f \wedge g$, is defined as usual and it belongs to $\Lambda^{k+l}(\mathbb{R}^n)$. For example if $k = l = 1$,

$$f = \sum_{1\le i\le n} f_i e^i \quad \text{and} \quad g = \sum_{1\le i\le n} g_j e^j$$

then

$$f \wedge g = \sum_{1\le i,j\le n} f_i g_j e^i \wedge e^j = \sum_{1\le i<j\le n} \left(f_i g_j - f_j g_i\right) e^i \wedge e^j.$$

If $k = 2$ and $l = 1$,

$$f = \sum_{1\le i<j\le n} f_{ij} e^i \wedge e^j \quad \text{and} \quad g = \sum_{1\le l\le n} g_l e^l$$

then

$$f \wedge g = \sum_{1\le i<j<l\le n} \left(f_{ij} g_l - f_{il} g_j + f_{jl} g_i\right) e^i \wedge e^j \wedge e^l.$$

(ii) The *scalar product* between two $k$-forms $f$ and $g$ is denoted by

$$\langle g; f\rangle = \sum_{1\le i_1<\cdots<i_k\le n} g_{i_1\cdots i_k} f_{i_1\cdots i_k}.$$

(iii) The *Hodge star operator* associates to $f \in \Lambda^k(\mathbb{R}^n)$ a form $(*f) \in \Lambda^{n-k}(\mathbb{R}^n)$ defined by

$$f \wedge g = \langle *f; g\rangle\, e^1 \wedge \cdots \wedge e^n$$

for every $g \in \Lambda^{n-k}(\mathbb{R}^n)$. For example if $k = 1,\ n = 3$ and

$$f = \sum_{1\le i\le 3} f_i e^i = f_1 e^1 + f_2 e^2 + f_3 e^3$$

then

$$*f = f_3 e^1 \wedge e^2 - f_2 e^1 \wedge e^3 + f_1 e^2 \wedge e^3.$$

(iv) We define, for $0 \leq l \leq k \leq n$, the *interior product* of $f \in \Lambda^k(\mathbb{R}^n)$ with $g \in \Lambda^l(\mathbb{R}^n)$ by

$$g \lrcorner f = (-1)^{n(k-l)} * (g \wedge (*f)) \in \Lambda^{k-l}(\mathbb{R}^n).$$

For example if $k = l$, then

$$g \lrcorner f = \langle g; f \rangle$$

or if $k = 1$ and $l = 2$, then

$$g \lrcorner f = \sum_{j=1}^{n} \left[ \sum_{i=1}^{n} f_{ij} g_i \right] e^j \in \Lambda^1(\mathbb{R}^n).$$

These definitions are linked through the following elementary facts. For every $f \in \Lambda^k(\mathbb{R}^n)$, $g \in \Lambda^{k+1}(\mathbb{R}^n)$ and $h \in \Lambda^1(\mathbb{R}^n)$

$$|h|^2 f = h \lrcorner (h \wedge f) + h \wedge (h \lrcorner f)$$

$$\langle h \wedge f; g \rangle = \langle f; h \lrcorner g \rangle.$$

(v) Let $A \in \mathbb{R}^{n \times n}$ be a matrix and $f \in \Lambda^k(\mathbb{R}^n)$ be given by

$$f = \sum_{1 \leq i_1 < \cdots < i_k \leq n} f_{i_1 \cdots i_k} e^{i_1} \wedge \cdots \wedge e^{i_k}.$$

We define the *pullback of $f$ by $A$*, denoted $A^*(f)$, by

$$A^*(f) = \sum_{1 \leq i_1 < \cdots < i_k \leq n} f_{i_1 \cdots i_k} A^{i_1} \wedge \cdots \wedge A^{i_k} \in \Lambda^k(\mathbb{R}^n)$$

where $A^j$ is the $j$-th row of $A$ and is identified with

$$A^j = \sum_{k=1}^{n} A^j_k e^k \in \Lambda^1(\mathbb{R}^n).$$

If $k = 0$, we then let

$$A^*(f) = f.$$

The present definition is consistent with the one given in the introduction; just set $\varphi(x) = Ax$ in (1).

(vi) We next define the notion of *rank* of $f \in \Lambda^k(\mathbb{R}^n)$. We first associate to the linear map

$$g \in \Lambda^1(\mathbb{R}^n) \to g \lrcorner f \in \Lambda^{k-1}(\mathbb{R}^n)$$

a matrix $\overline{f} \in \mathbb{R}^{\binom{n}{k-1} \times n}$ such that, by abuse of notations,

$$g \lrcorner f = \overline{f}\, g \quad \text{for every } g \in \Lambda^1(\mathbb{R}^n).$$

In this case, we have

$$g \lrcorner f = \sum_{1 \le j_1 < \cdots < j_{k-1} \le n} \left( \sum_{\gamma=1}^{k} (-1)^{\gamma-1} \sum_{j_{\gamma-1} < i < j_\gamma} f_{j_1 \cdots j_{\gamma-1} i j_\gamma \cdots j_{k-1}}\, g_i \right) e^{j_1} \wedge \cdots \wedge e^{j_{k-1}}.$$

More explicitly, using the lexicographical order for the columns (index below) and the rows (index above) of the matrix $\overline{f}$, we have

$$(\overline{f})_i^{j_1 \cdots j_{k-1}} = f_{i j_1 \cdots j_{k-1}}$$

for $1 \le i \le n$ and $1 \le j_1 < \cdots < j_{k-1} \le n$. The rank of the $k$-form $f$ is then the rank of the $\binom{n}{k-1} \times n$ matrix $\overline{f}$. We then write

$$\operatorname{rank}[f] = \operatorname{rank}\left(\overline{f}\right).$$

*Example 1* For example if $k = 2$, then

$$(\overline{f})_i^j = f_{ij}$$

i.e. when $n = 4$

$$\overline{f} = \begin{pmatrix} 0 & f_{12} & f_{13} & f_{14} \\ f_{21} = -f_{12} & 0 & f_{23} & f_{24} \\ f_{31} = -f_{13} & f_{32} = -f_{23} & 0 & f_{34} \\ f_{41} = -f_{14} & f_{42} = -f_{24} & f_{43} = -f_{34} & 0 \end{pmatrix}.$$

Since $f_{ij} = -f_{ji}$, we have that $\overline{f} \in \mathbb{R}^{n \times n}$ is skew symmetric and therefore can never be invertible if $n$ is odd. The canonical form when $n = 4$ is the standard symplectic

matrix

$$J = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}.$$

*Example 2* When $k = n$, then, identifying the form with its component, we have that $\overline{f} \in \mathbb{R}^{n\times n}$ and, up to a sign,

$$\overline{f} = \begin{pmatrix} f & 0 & \cdots & 0 \\ 0 & -f & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (-1)^{n-1} f \end{pmatrix}.$$

Note that only when $k = 2$ or $k = n$ the matrix $\overline{f}$ is a square matrix. Our best results are obtained precisely in these cases and when the matrix $\overline{f}$ is invertible.

We then have the following elementary result.

**Proposition 3** *Let* $f \in \Lambda^k(\mathbb{R}^n)$, $f \neq 0$.

*(i) If* $k = 1$, *then the rank of* $f$ *is always* 1.

*(ii) If* $k = 2$, *then the rank of* $f$ *is even. The forms*

$$\omega_m = \sum_{i=1}^{m} e^{2i-1} \wedge e^{2i}$$

*are such that* $\operatorname{rank}[\omega_m] = 2m$. *Moreover* $\operatorname{rank}[f] = 2m$ *if and only if*

$$f^m \neq 0 \quad \textit{and} \quad f^{m+1} = 0$$

*where* $f^m = \underbrace{f \wedge \cdots \wedge f}_{m-\textit{times}}$.

*(iii) If* $3 \leq k \leq n$, *then*

$$\operatorname{rank}[f] \in \{k, k+2, \cdots, n\}$$

*and any of the values in* $\{k, k+2, \cdots, n\}$ *can be achieved by the rank of a* $k$*-form. In particular if* $k = n-1$, *then* $\operatorname{rank}[f] = n-1$, *while if* $k = n$, *then* $\operatorname{rank}[f] = n$.

*(iv) If* $\operatorname{rank}[f] = k$, *then there exist* $f_1, \cdots, f_k \in \Lambda^1(\mathbb{R}^n)$ *such that*

$$f = f_1 \wedge \cdots \wedge f_k.$$

*Remark 4* The rank is an invariant for the pullback equation. More precisely if there exists $A \in \mathrm{GL}(n)$, i.e. $A$ is an invertible $n \times n$ matrix, such that

$$A^*(g) = f$$

(or equivalently if $\varphi(x) = Ax$, then the above equation is equivalent to $\varphi^*(g) = f$) then (cf. Theorem 64)

$$\operatorname{rank}[g] = \operatorname{rank}[f].$$

Conversely, when $k = 1, 2, n-1, n$, if $\operatorname{rank}[g] = \operatorname{rank}[f]$, then there exists $A \in \mathrm{GL}(n)$ such that

$$A^*(g) = f.$$

However the converse is not anymore true, in general, if $3 \leq k \leq n-2$ (cf. Examples 60 and 61).

# 3 Harmonic Fields and Poincaré Lemma

## 3.1 Preliminaries

**Definition 5** Let $\Omega \subset \mathbb{R}^n$ be open and $f \in C^1\left(\Omega; \Lambda^k\right)$, namely

$$f = \sum_{1 \leq i_1 < \cdots < i_k \leq n} f_{i_1 \cdots i_k}(x)\, dx^{i_1} \wedge \cdots \wedge dx^{i_k}.$$

(i) The *exterior derivative* of $f$ denoted $df$ belongs to $C^0(\Omega; \Lambda^{k+1})$ and is defined by

$$df = \sum_{1 \leq i_1 < \cdots < i_k \leq n} \sum_{m=1}^{n} \frac{\partial f_{i_1 \cdots i_k}}{\partial x_m}\, dx^m \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}.$$

If $k = n$, then $df = 0$.

(ii) The *interior derivative* or *codifferential* of $f$ denoted $\delta f$ belongs to $C^0(\Omega; \Lambda^{k-1})$ and is defined by

$$\delta f = (-1)^{n(k-1)} * (d(*f)).$$

By abuse of notations one can write

$$df = \nabla \wedge f \quad \text{and} \quad \delta f = \nabla \lrcorner f.$$

Note that, for example, when $k = 2$

$$df = \sum_{1 \le i<j<l \le n} \left( \frac{\partial f_{ij}}{\partial x_l} - \frac{\partial f_{il}}{\partial x_j} + \frac{\partial f_{jl}}{\partial x_i} \right) dx^i \wedge dx^j \wedge dx^l.$$

*Remark 6*

(i) If $k = 0$, then the operator $d$ can be identified with the gradient operator, while $\delta f = 0$ for any $f$.
(ii) If $k = 1$, then the operator $d$ can be identified with the curl operator, while the operator $\delta$ is the divergence operator.

We next gather some well known properties of the operators $d$ and $\delta$.

**Theorem 7** *Let* $f \in C^2\left(\Omega; \Lambda^k\right)$ *and* $g \in C^2(\Omega; \Lambda^l)$. *Then*

$$d(f \wedge g) = df \wedge g + (-1)^k f \wedge dg$$
$$\delta(f \lrcorner g) = (-1)^{k+l} df \lrcorner g - f \lrcorner \delta g.$$
$$ddf = 0, \quad \delta\delta f = 0 \quad \textit{and} \quad d\delta f + \delta df = \Delta f.$$

We also need the following definition. In the sequel we will denote the exterior unit normal of $\partial\Omega$ by $\nu$.

**Definition 8** The *tangential component* of a $k$-form $f$ on $\partial\Omega$ is the $(k+1)$-form

$$\nu \wedge f \in \Lambda^{k+1}.$$

The *normal component* of a $k$-form $f$ on $\partial\Omega$ is the $(k-1)$-form

$$\nu \lrcorner f \in \Lambda^{k-1}.$$

*Example 9* If $f$ is a 1-form and

$$\Omega = \{x \in \mathbb{R}^n : x_n > 0\}$$

then $\nu = -e_n$ and

$$\nu \wedge f = 0 \quad \Leftrightarrow \quad f_1 = \cdots = f_{n-1} = 0$$
$$\nu \lrcorner f = 0 \quad \Leftrightarrow \quad f_n = 0$$

We easily deduce the following properties.

**Proposition 10** *Let $0 \leq k \leq n$ and $f \in C^1\left(\overline{\Omega}; \Lambda^k\right)$, then*

$$\begin{aligned} \nu \wedge f = 0 \text{ on } \partial\Omega \quad &\Rightarrow \quad \nu \wedge df = 0 \text{ on } \partial\Omega \\ \nu \lrcorner f = 0 \text{ on } \partial\Omega \quad &\Rightarrow \quad \nu \lrcorner \delta f = 0 \text{ on } \partial\Omega. \end{aligned}$$

We will constantly use the integration by parts formula.

**Theorem 11 (Integration by Parts Formula)** *Let $1 \leq k \leq n$, $f \in C^1(\overline{\Omega}; \Lambda^{k-1})$ and $g \in C^1\left(\overline{\Omega}; \Lambda^k\right)$. Then*

$$\int_\Omega \langle df; g\rangle + \int_\Omega \langle f; \delta g\rangle = \int_{\partial\Omega} \langle \nu \wedge f; g\rangle = \int_{\partial\Omega} \langle f; \nu \lrcorner g\rangle.$$

We will adopt the following notations.

**Notation 12** Let $\Omega \subset \mathbb{R}^n$ be open, $r \geq 0$ be an integer and $0 \leq q \leq 1 \leq p \leq \infty$. Spaces with vanishing tangential or normal component will be denoted in the following way

$$\begin{aligned} C_T^{r,q}\left(\overline{\Omega}; \Lambda^k\right) &= \{f \in C^{r,q}\left(\overline{\Omega}; \Lambda^k\right) : \nu \wedge f = 0 \quad \text{on } \partial\Omega\} \\ C_N^{r,q}\left(\overline{\Omega}; \Lambda^k\right) &= \{f \in C^{r,q}\left(\overline{\Omega}; \Lambda^k\right) : \nu \lrcorner f = 0 \quad \text{on } \partial\Omega\} \\ W_T^{r+1,p}\left(\Omega; \Lambda^k\right) &= \{f \in W^{r+1,p}\left(\Omega; \Lambda^k\right) : \nu \wedge f = 0 \quad \text{on } \partial\Omega\} \\ W_N^{r+1,p}\left(\Omega; \Lambda^k\right) &= \{f \in W^{r+1,p}\left(\Omega; \Lambda^k\right) : \nu \lrcorner f = 0 \quad \text{on } \partial\Omega\}. \end{aligned}$$

The different sets of *harmonic fields* will be denoted by

$$\begin{aligned} \mathcal{H}\left(\Omega; \Lambda^k\right) &= \{f \in W^{1,2}\left(\Omega; \Lambda^k\right) : df = 0 \text{ and } \delta f = 0 \text{ in } \Omega\} \\ \mathcal{H}_T\left(\Omega; \Lambda^k\right) &= \{f \in \mathcal{H}\left(\Omega; \Lambda^k\right) : \nu \wedge f = 0 \quad \text{on } \partial\Omega\} \\ \mathcal{H}_N\left(\Omega; \Lambda^k\right) &= \{f \in \mathcal{H}\left(\Omega; \Lambda^k\right) : \nu \lrcorner f = 0 \quad \text{on } \partial\Omega\}. \end{aligned}$$

We now list some properties of the harmonic fields.

**Theorem 13** *Let $\Omega \subset \mathbb{R}^n$ be an open set. Then*

$$\mathcal{H}\left(\Omega; \Lambda^k\right) \subset C^\infty\left(\Omega; \Lambda^k\right).$$

*Moreover if $\Omega$ is bounded and smooth, then the next statements are valid.*

*(i) The following inclusion holds*

$$\mathcal{H}_T\left(\Omega; \Lambda^k\right) \cup \mathcal{H}_N\left(\Omega; \Lambda^k\right) \subset C^\infty\left(\overline{\Omega}; \Lambda^k\right).$$

*Furthermore if $r \geq 0$ is an integer and $0 \leq q \leq 1$, then there exists $C = C(r, \Omega)$ such that, for every $f \in \mathcal{H}_T\left(\Omega; \Lambda^k\right) \cup \mathcal{H}_N\left(\Omega; \Lambda^k\right)$,*

$$\|f\|_{W^{r,2}} \leq C\|f\|_{L^2} \quad \textit{and} \quad \|f\|_{C^{r,q}} \leq C\|f\|_{C^0}\,.$$

*(ii) The spaces $\mathcal{H}_T\left(\Omega; \Lambda^k\right)$ and $\mathcal{H}_N\left(\Omega; \Lambda^k\right)$ are finite dimensional and closed in $L^2\left(\Omega; \Lambda^k\right)$.*

*(iii) Furthermore if $\Omega$ is contractible, then*

$$\begin{aligned} &\mathcal{H}_T\left(\Omega; \Lambda^k\right) = \{0\} \text{ if } 0 \leq k \leq n-1 \\ &\mathcal{H}_N\left(\Omega; \Lambda^k\right) = \{0\} \text{ if } 1 \leq k \leq n. \end{aligned}$$

*(iv) If $k = 0$ or $k = n$ and $h \in \mathcal{H}\left(\Omega; \Lambda^k\right)$, then $h$ is constant on each connected component of $\Omega$. In particular $\mathcal{H}_T\left(\Omega; \Lambda^0\right) = \{0\}$ and $\mathcal{H}_N\left(\Omega; \Lambda^n\right) = \{0\}$.*

*Remark 14* If $k = 1$ and assuming that $\Omega$ is smooth, then the sets $\mathcal{H}_T$ and $\mathcal{H}_N$ can be rewritten, as usual by abuse of notations, as

$$\mathcal{H}_T\left(\Omega; \Lambda^1\right) = \left\{ f \in C^\infty\left(\overline{\Omega}; \mathbb{R}^n\right) : \begin{bmatrix} \operatorname{curl} f = 0 \text{ and } \operatorname{div} f = 0 \\ f_i \nu_j - f_j \nu_i = 0,\ \forall\, 1 \leq i < j \leq n \end{bmatrix} \right\}$$

$$\mathcal{H}_N\left(\Omega; \Lambda^1\right) = \left\{ f \in C^\infty\left(\overline{\Omega}; \mathbb{R}^n\right) : \begin{bmatrix} \operatorname{curl} f = 0 \text{ and } \operatorname{div} f = 0 \\ \sum_{i=1}^n f_i \nu_i = 0 \end{bmatrix} \right\}.$$

Moreover if $\Omega$ is simply connected, then

$$\mathcal{H}_T\left(\Omega; \Lambda^1\right) = \mathcal{H}_N\left(\Omega; \Lambda^1\right) = \{0\}.$$

In particular if $n = 2$ and $\Omega = B_1 \setminus \{(0,0)\}$, then

$$f = \frac{-x_2}{x_1^2 + x_2^2}\, dx^1 + \frac{x_1}{x_1^2 + x_2^2}\, dx^2 \in \mathcal{H}_N\left(\Omega; \Lambda^1\right).$$

*Proof* We only prove the inclusion

$$\mathcal{H}\left(\Omega; \Lambda^k\right) \subset C^\infty\left(\Omega; \Lambda^k\right)$$

which follows from Weyl Lemma (cf. for example [22]). Indeed let $\phi \in C_0^\infty\left(\Omega; \Lambda^k\right)$

$$\int_\Omega \langle \omega; \Delta\phi \rangle = \int_\Omega \langle \omega; d\delta\phi + \delta d\phi \rangle = \int_\Omega \langle d\omega; d\phi \rangle + \int_\Omega \langle \delta\omega; \delta\phi \rangle = 0$$

Choose $\phi = \varphi\, dx^I$, $\varphi \in C_0^\infty(\Omega)$ and thus $\omega_I \in C^\infty(\Omega)$. ■

### 3.2 The Hodge-Morrey Decomposition

We now turn to the celebrated Hodge-Morrey (see [43–45]) decomposition and an equivalent formulation (see [16] for details).

**Theorem 15 (Hodge-Morrey Decomposition)** *Let $\Omega \subset \mathbb{R}^n$ be a bounded open smooth set and $\nu$ be the exterior unit normal. Let $0 \leq k \leq n$ and $f \in L^2\left(\Omega; \Lambda^k\right)$. Then there exist*

$$\alpha \in W_T^{1,2}\left(\Omega; \Lambda^{k-1}\right), \quad \beta \in W_T^{1,2}\left(\Omega; \Lambda^{k+1}\right),$$
$$h \in \mathcal{H}_T\left(\Omega; \Lambda^k\right) \quad \text{and} \quad \omega \in W_T^{2,2}\left(\Omega; \Lambda^k\right)$$

*such that, in $\Omega$,*

$$f = d\alpha + \delta\beta + h, \quad \alpha = \delta\omega \quad \text{and} \quad \beta = d\omega.$$

*Moreover the decomposition is an orthogonal one, i.e.*

$$\int_\Omega \langle d\alpha; \delta\beta\rangle = \int_\Omega \langle d\alpha; h\rangle = \int_\Omega \langle \delta\beta; h\rangle = 0.$$

*Remark 16*

(i) We have quoted only one of the three decompositions. Another one, completely similar, is by replacing $T$ by $N$ and the other one mixing both $T$ and $N$.
(ii) If $k \leq n-1$ and if the domain $\Omega$ is contractible, then $h = 0$.
(iii) If $k = 0$, then the theorem reads as

$$f = \delta\beta = \delta d\omega = \Delta\omega \text{ in } \Omega \quad \text{with} \quad \omega = 0 \text{ on } \partial\Omega.$$

(iv) When $k = 1$ and $n = 3$, the decomposition reads as follows. For any $f \in L^2(\Omega; \mathbb{R}^3)$, there exist

$$\omega \in W^{2,2}\left(\Omega; \mathbb{R}^3\right) \quad \text{with } \omega_i \nu_j - \omega_j \nu_i = 0 \text{ on } \partial\Omega, \ \forall\, 1 \leq i < j \leq 3$$
$$\alpha \in W_0^{1,2}(\Omega) \quad \text{and} \quad \alpha = \operatorname{div}\omega$$
$$\beta \in W^{1,2}\left(\Omega; \mathbb{R}^3\right) \quad \text{with } \beta = -\operatorname{curl}\omega \quad \text{and} \quad \langle \nu; \beta\rangle = 0 \text{ on } \partial\Omega$$
$$h \in \left\{ h \in C^\infty\left(\overline{\Omega}; \mathbb{R}^3\right) : \begin{bmatrix} \operatorname{curl} h = 0 \text{ and } \operatorname{div} h = 0 \\ h_i \nu_j - h_j \nu_i = 0, \ \forall\, 1 \leq i < j \leq 3 \end{bmatrix} \right\}$$

such that

$$f = \operatorname{grad}\alpha + \operatorname{curl}\beta + h \text{ in } \Omega.$$

Furthermore if $\Omega$ is simply connected, then $h = 0$.

(v) If $f$ is more regular than in $L^2$, then $\alpha, \beta$ and $\omega$ are in the corresponding class of regularity (see [7, 16, 44]). More precisely if, for example, $r \geq 0$ is an integer, $0 < q < 1$ and $f \in C^{r,q}\left(\overline{\Omega}; \Lambda^k\right)$, then

$$\alpha \in C^{r+1,q}\left(\overline{\Omega}; \Lambda^{k-1}\right), \quad \beta \in C^{r+1,q}\left(\overline{\Omega}; \Lambda^{k+1}\right) \quad \text{and} \quad \omega \in C^{r+2,q}\left(\overline{\Omega}; \Lambda^k\right).$$

(vi) The proof of Morrey uses the direct methods of the calculus of variations. One minimizes

$$D_f(\omega) = \int_\Omega \left(\frac{1}{2}|d\omega|^2 + \frac{1}{2}|\delta\omega|^2 + \langle f; \omega\rangle\right)$$

in an appropriate space, Gaffney inequality (see, for example, [14]) giving the coercivity of the integral.

It turns out that the Hodge-Morrey decomposition is in fact equivalent to solving the first order system

$$\begin{cases} d\omega = f \quad \text{and} \quad \delta\omega = g \text{ in } \Omega \\ \nu \wedge \omega = \nu \wedge \omega_0 \qquad \text{on } \partial\Omega \end{cases}$$

or the similar one

$$\begin{cases} d\omega = f \quad \text{and} \quad \delta\omega = g \text{ in } \Omega \\ \nu \lrcorner \omega = \nu \lrcorner \omega_0 \qquad \text{on } \partial\Omega. \end{cases}$$

We here state a simplified version for the first system (see also [38, 54–56]).

**Theorem 17** *Let $r \geq 0$ and $1 \leq k \leq n-2$ be integers, $0 < q < 1$ and $\Omega \subset \mathbb{R}^n$ be a bounded contractible open smooth set and with exterior unit normal $\nu$. The following statements are then equivalent.*

*(i) Let $g \in C^{r,q}\left(\overline{\Omega}; \Lambda^{k-1}\right)$ and $f \in C^{r,q}\left(\overline{\Omega}; \Lambda^{k+1}\right)$ be such that*

$$\delta g = 0 \text{ in } \Omega, \quad df = 0 \text{ in } \Omega \quad \text{and} \quad \nu \wedge f = 0 \text{ on } \partial\Omega.$$

*(ii) There exists $\omega \in C^{r+1,q}\left(\overline{\Omega}; \Lambda^k\right)$, such that*

$$\begin{cases} d\omega = f \quad \text{and} \quad \delta\omega = g \text{ in } \Omega \\ \nu \wedge \omega = 0 \qquad \text{on } \partial\Omega. \end{cases}$$

*Remark 18*

(i) When $k = n-1$, the result is valid provided

$$\int_\Omega f = 0.$$

Note that in this case the conditions $df = 0$ and $\nu \wedge f = 0$ are automatically fulfilled.

(ii) Completely analogous results can be given for Sobolev spaces.

(iii) If the domain $\Omega$ is not contractible or if $\omega_0 \neq 0$, then additional necessary conditions have to be added, namely $\nu \wedge \omega_0 \in C^{r+1,q}\left(\partial\Omega; \Lambda^{k+1}\right)$,

$$\nu \wedge d\omega_0 = \nu \wedge f \text{ on } \partial\Omega$$

and, for every $\chi \in \mathcal{H}_T\left(\Omega; \Lambda^{k+1}\right)$ and $\psi \in \mathcal{H}_T\left(\Omega; \Lambda^{k-1}\right)$,

$$\int_\Omega \langle f; \chi\rangle - \int_{\partial\Omega} \langle \nu \wedge \omega_0; \chi\rangle = 0 \quad \text{and} \quad \int_\Omega \langle g; \psi\rangle = 0.$$

(iv) When $k = 1$ and $n = 3$, the theorem reads as follows. Let $\Omega \subset \mathbb{R}^3$ be a bounded contractible smooth open set, $g \in C^{r,q}\left(\overline{\Omega}\right)$ and $f \in C^{r,q}\left(\overline{\Omega}; \mathbb{R}^3\right)$ be such that

$$\operatorname{div} f = 0 \text{ in } \Omega \quad \text{and} \quad \langle f; \nu\rangle = 0 \text{ on } \partial\Omega.$$

Then there exists $\omega \in C^{r+1,q}\left(\overline{\Omega}; \mathbb{R}^3\right)$, such that

$$\begin{cases} \operatorname{curl} \omega = f \quad \text{and} \quad \operatorname{div} \omega = g \quad \text{in } \Omega \\ \omega_i \nu_j - \omega_j \nu_i = 0 \ \forall\, 1 \leq i < j \leq 3 \text{ on } \partial\Omega. \end{cases}$$

## 3.3 Poincaré Lemma

We now have a global version of Poincaré lemma with optimal regularity.

**Theorem 19** *Let $r \geq 0$ and $0 \leq k \leq n-1$ be integers, $0 < q < 1$ and $\Omega \subset \mathbb{R}^n$ be a bounded open smooth set. The following statements are equivalent.*

*(i) Let $f \in C^{r,q}\left(\overline{\Omega}; \Lambda^{k+1}\right)$, be such that*

$$df = 0 \text{ in } \Omega \quad \text{and} \quad \int_\Omega \langle f; \psi\rangle = 0 \text{ for every } \psi \in \mathcal{H}_N\left(\Omega; \Lambda^{k+1}\right).$$

*(ii) There exists $\omega \in C^{r+1,q}\left(\overline{\Omega}; \Lambda^k\right)$, such that*

$$d\omega = f \quad \text{in } \Omega.$$

*Moreover there exists a constant* $C = C(r, q, \Omega)$ *such that*

$$\|\omega\|_{C^{r+1,q}} \leq C\|f\|_{C^{r,q}}.$$

*Remark 20*

(i) When $k = n - 1$ in Theorem 19, there is no restriction on the solvability of $d\omega = f$ (since $df = 0$ automatically and $\mathcal{H}_N(\Omega; \Lambda^n) = \{0\}$).
(ii) If $r = 0$, then the conditions $df = 0$ have to be understood in the sense of distributions.
(iii) The above results remain valid if $\Omega$ is $C^{r+3,q}$.
(iv) The same result holds true for Sobolev spaces.
(v) The construction is linear and universal.

*Proof* (ii) $\Rightarrow$ (i). Suppose first that there exists $\omega \in C^{r+1,q}\left(\overline{\Omega}; \Lambda^k\right)$ such that $f = d\omega$. Clearly $df = 0$ and the other assertion follows by partial integration, since, for every $\psi \in \mathcal{H}_N$,

$$\int_\Omega \langle f; \psi\rangle = \int_\Omega \langle d\omega; \psi\rangle = -\int_\Omega \langle \omega; \delta\psi\rangle + \int_{\partial\Omega} \langle \omega; \nu \lrcorner \psi\rangle = 0.$$

(i) $\Rightarrow$ (ii). Suppose now that

$$df = 0 \text{ in } \Omega \quad \text{and} \quad \int_\Omega \langle f; \psi\rangle = 0 \text{ for every } \psi \in \mathcal{H}_N(\Omega; \Lambda^{k+1}).$$

We then appeal to the dual version of Theorem 17 to solve the problem

$$\begin{cases} d\omega = f \quad \text{and} \quad \delta\omega = 0 \text{ in } \Omega \\ \qquad \nu \lrcorner \omega = 0 \qquad\quad \text{on } \partial\Omega. \end{cases}$$

This concludes the proof. ■

A much more elementary proof can be obtained in a star shaped domain if we are ready to give up the gain of regularity. The formula is standard and the proof is done by straightforward differentiation (see Dacorogna (unpublished, 2016)).

**Theorem 21** *Let* $0 \leq k \leq n - 1$, $\Omega \subset \mathbb{R}^n$ *be a star shaped (with respect to the origin) open set and* $f \in C^1\left(\Omega; \Lambda^{k+1}\right)$ *be such that* $df = 0$. *Then* $F \in C^1\left(\Omega; \Lambda^k\right)$ *defined by*

$$F(x) = \int_0^1 [x \lrcorner f(tx)]\, t^k dt$$

*verifies* $dF = f$.

*Remark 22*

*(i)* The case $k = 0$ is the most classical and reads as

$$F(x) = \sum_{j=1}^{n} \int_0^1 \left[x_j f_j(tx)\right] dt \quad \Rightarrow \quad F_{x_i} = f_i \text{ (i.e. } \operatorname{grad} F = f).$$

When $k = 1$ the formula becomes

$$F^j(x) = \sum_{i=1}^{n} \int_0^1 \left[x_i f_{ij}(tx)\right] t\, dt \quad \Rightarrow \quad F^j_{x_i} - F^i_{x_j} = f_{ij} \text{ (i.e. } \operatorname{curl} F = f).$$

*(ii)* Using the Hodge $*$ operator, we have the dual version, namely if $1 \le k \le n$ and $\varphi \in C^1\left(\Omega; \Lambda^{k-1}\right)$ is such that $\delta\varphi = 0$, then $\Phi \in C^1\left(\Omega; \Lambda^k\right)$ defined by

$$\Phi(x) = \int_0^1 [x \wedge \varphi(tx)]\, t^{n-k} dt$$

verifies $\delta\Phi = \varphi$. In particular when $k = 1$ the formula reads as

$$\Phi(x) = x \int_0^1 [\varphi(tx)]\, t^{n-1}\, dt \quad \Rightarrow \quad \operatorname{div} \Phi = \varphi.$$

## 3.4 Poincaré Lemma on the Boundary

We start with a slight improvement of a lemma proved in Dacorogna-Moser [30].

**Lemma 23** *Let $r \ge 0$ be an integer, $0 \le q \le 1$ and $\Omega \subset \mathbb{R}^n$ be a bounded open $C^{r+1,q}$ set with exterior unit normal $\nu$. Let $c \in C^{r,q}(\partial\Omega)$. Then there exists*

$$b \in C^{r+1,q}\left(\overline{\Omega}\right)$$

*satisfying all over $\partial\Omega$*

$$\operatorname{grad} b = c\,\nu \quad \text{and} \quad b = 0.$$

*Furthermore there exists a constant $C = C(r, \Omega) > 0$ such that*

$$\|b\|_{C^{r+1,q}\left(\overline{\Omega}\right)} \le C \|c\|_{C^{r,q}(\partial\Omega)}.$$

*Proof* If one is not interested in the sharp regularity result a solution of the problem is given by

$$b(x) = -c(x)\,\zeta(d(x, \partial\Omega))$$

where $c$ has been extended to $\overline{\Omega}$ and $d(x, \partial\Omega)$ stands for the distance from $x$ to the boundary (recalling that the distance function is as regular as the set $\Omega$ near the boundary, see for example [34]) and $\zeta$ is a smooth function so that $\zeta(0) = 0$, $\zeta'(0) = 1$ and $\zeta \equiv 0$ outside a small neighborhood of 0.

We give here a proof that uses elliptic regularity and hence only works whenever $0 < q < 1$ (and also works in $L^p$ for $1 < p < \infty$) in this case the constant obtained depends also on $q$. Another proof exists which is valid also when $q = 0, 1$ (cf. [16]).

The desired solution $b$ is obtained by solving

$$\begin{cases} \Delta^2 b = 0 & \text{in } \Omega \\ b = 0 \text{ and } \frac{\partial b}{\partial \nu} = c \text{ on } \partial\Omega. \end{cases}$$

The solution $b$ is in $C^\infty(\Omega) \cap C^{r+1,q}\left(\overline{\Omega}\right)$ and satisfies the estimate

$$\|b\|_{C^{r+1,q}\left(\overline{\Omega}\right)} \leq C\,\|c\|_{C^{r,q}(\partial\Omega)}\,.$$

Clearly $b$ solves on $\partial\Omega$

$$\operatorname{grad} b = c\,\nu \quad \text{and} \quad b = 0.$$

This concludes the proof. ■

We now need a generalization of the above lemma to differential forms, as achieved in [20].

**Lemma 24** *Let $r \geq 0$ and $1 \leq k \leq n-1$ be integers, $0 \leq q \leq 1$ and $\Omega \subset \mathbb{R}^n$ be a bounded open $C^{r+1,q}$ set with exterior unit normal $\nu$.*

*(i) If $c \in C^{r,q}\left(\partial\Omega; \Lambda^k\right)$ is such that*

$$\nu \wedge c = 0 \quad \text{on } \partial\Omega,$$

*then there exists $b \in C^{r+1,q}\left(\overline{\Omega}; \Lambda^{k-1}\right)$ satisfying all over $\partial\Omega$*

$$db = c, \quad \delta b = 0 \quad \text{and} \quad b = 0.$$

*Moreover there exists a constant $C = C(r, \Omega) > 0$ such that*

$$\|b\|_{C^{r+1,q}\left(\overline{\Omega}\right)} \leq C\,\|c\|_{C^{r,q}(\partial\Omega)}\,.$$

*(ii) If $c \in C^{r,q}(\partial\Omega; \Lambda^k)$ is such that*

$$\nu \lrcorner c = 0 \quad \text{on } \partial\Omega,$$

*then there exists $b \in C^{r+1,q}\left(\overline{\Omega}; \Lambda^{k+1}\right)$ satisfying all over $\partial\Omega$*

$$\delta b = c, \quad db = 0 \quad \text{and} \quad b = 0.$$

*Furthermore there exists a constant $C = C(r, \Omega) > 0$ such that*

$$\|b\|_{C^{r+1,q}(\overline{\Omega})} \leq C \ \|c\|_{C^{r,q}(\partial\Omega)} .$$

*Remark 25*

(i) If $k = 0$ in Statement (ii) (and analogously if $k = n$ in Statement (i)) and $0 < q < 1$, then it is easy to find $b$ such that (and without any restriction on $c$)

$$\delta b = c \quad \text{and} \quad db = 0 \quad \text{in } \overline{\Omega}$$

where $c$ has been extended to $\overline{\Omega}$ with the appropriate regularity. Indeed choose $b = \operatorname{grad} B$ where $B$ solves

$$\begin{cases} \Delta B = c \text{ in } \Omega \\ B = 0 \ \text{ on } \partial\Omega. \end{cases}$$

(ii) The above result remains valid, with the same proof, in the Sobolev setting. More precisely Statement (i) (and similarly for Statement (ii)) reads as follows. Let $r \geq 1$ be an integer, $1 < p < \infty$ and $\Omega \subset \mathbb{R}^n$ be a bounded open $C^{r+1}$ set with exterior unit normal $\nu$. Let $c \in W^{r,p}\left(\Omega; \Lambda^k\right)$, then there exists

$$b \in W^{r+1,p}\left(\Omega; \Lambda^{k-1}\right)$$

satisfying all over $\partial\Omega$

$$db = c, \quad \delta b = 0 \quad \text{and} \quad b = 0.$$

Moreover there exists a constant $C = C(r, p, \Omega) > 0$ such that

$$\|b\|_{W^{r+1,p}(\Omega)} \leq C \ \|c\|_{W^{r-1/p,p}(\partial\Omega)} .$$

*Proof* *Step 1.* We start with the case (i). First solve with Lemma 23 the problem, on $\partial\Omega$,

$$\operatorname{grad} b_{i_1 \cdots i_{k-1}} = (\nu \lrcorner c)_{i_1 \cdots i_{k-1}} \nu \quad \text{and} \quad b_{i_1 \cdots i_{k-1}} = 0$$

for every multiindex $1 \le i_1 < \ldots < i_{k-1} \le n$ and set

$$b = \sum_{1 \le i_1 < \ldots < i_{k-1} \le n} b_{i_1 \cdots i_{k-1}} dx^{i_1} \wedge \cdots \wedge dx^{i_{k-1}}.$$

The classical formulas immediately imply that, on $\partial\Omega$,

$$db = \nu \wedge (\nu \lrcorner c) \quad \text{and} \quad \delta b = \nu \lrcorner (\nu \lrcorner c) = 0.$$

We combine the first equation with the hypothesis $\nu \wedge c = 0$ and use the fact that

$$c = \nu \lrcorner (\nu \wedge c) + \nu \wedge (\nu \lrcorner c) = \nu \wedge (\nu \lrcorner c)$$

to get

$$db = \nu \wedge (\nu \lrcorner c) = \nu \wedge (\nu \lrcorner c) + \nu \lrcorner (\nu \wedge c) = c \quad \text{on } \partial\Omega$$

We have therefore proved the assertion.
*Step 2.* For (ii) we first solve, on $\partial\Omega$,

$$\operatorname{grad} b_{i_1 \ldots i_{k+1}} = (\nu \wedge c)_{i_1 \ldots i_{k+1}}\, \nu \quad \text{and} \quad b_{i_1 \cdots i_{k+1}} = 0$$

and then proceed exactly as in Step 1. This concludes the proof of the lemma. ■

## 3.5 *Poincaré Lemma with Dirichlet Boundary Data*

We now consider the boundary value problems

$$\begin{cases} d\omega = f \text{ in } \Omega \\ \omega = \omega_0 \text{ on } \partial\Omega \end{cases} \quad \text{and} \quad \begin{cases} \delta\omega = g \text{ in } \Omega \\ \omega = \omega_0 \text{ on } \partial\Omega. \end{cases}$$

In contrast to the problems studied in the previous sections $\delta\omega$ (respectively $d\omega$) is not prescribed, but however both the tangential and normal components of $\omega$ are given on the boundary. It turns out that the problems can be solved under exactly the same hypotheses on $f$, $g$ and $\omega_0$ as above. We follow exactly the construction in Dacorogna [20] for Hölder spaces; a very similar method is used in Schwarz [51] for Sobolev spaces.

**Theorem 26** *Let $r \ge 0$ and $0 \le k \le n-1$ be integers, $0 < q < 1$ and $\Omega \subset \mathbb{R}^n$ be a bounded connected open smooth set with exterior unit normal $\nu$. Let $f : \overline{\Omega} \to \Lambda^{k+1}$. Then the following statements are equivalent.*

*(i) Let* $f \in C^{r,q}\left(\overline{\Omega}; \Lambda^{k+1}\right)$ *satisfy*

$$df = 0 \text{ in } \Omega, \quad \nu \wedge f = 0 \text{ on } \partial\Omega \tag{A1}$$

*and, for every* $\chi \in \mathcal{H}_T\left(\Omega; \Lambda^{k+1}\right)$,

$$\int_\Omega \langle f; \chi \rangle = 0. \tag{A2}$$

*(ii) There exists* $\omega \in C^{r+1,q}\left(\overline{\Omega}; \Lambda^k\right)$ *such that*

$$\begin{cases} d\omega = f \text{ in } \Omega \\ \omega = 0 \text{ on } \partial\Omega \end{cases}$$

*and there exists a constant* $C = C(r, q, \Omega)$ *such that*

$$\|\omega\|_{C^{r+1,q}(\overline{\Omega})} \leq C \|f\|_{C^{r,q}(\overline{\Omega})}.$$

*Remark 27*

(i) Instead of imposing the boundary data to be 0 we can have any boundary data $\omega_0$ satisfying, in addition to $df = 0$,

$$\nu \wedge d\omega_0 = \nu \wedge f \text{ on } \partial\Omega$$

and, for every $\chi \in \mathcal{H}_T\left(\Omega; \Lambda^{k+1}\right)$,

$$\int_\Omega \langle f; \chi \rangle = \int_{\partial\Omega} \langle \nu \wedge \omega_0; \chi \rangle.$$

(ii) In the case $k = n - 1$, the conditions (A1) are trivially satisfied, while (A2) reads as

$$\int_\Omega f = 0.$$

(iii) When $k = 0$, then the result is still valid for $q = 0, 1$.
(iv) If $r = 0$, then the condition $df = 0$ is understood in the sense of distributions.
(v) The above results remain valid if the set $\Omega$ is $C^{r+3,q}$.
(vi) If $\Omega$ is contractible, then $\mathcal{H}_T\left(\Omega; \Lambda^k\right) = \{0\}$.
(vii) The construction is linear and universal.
(viii) Analogous results hold in Sobolev spaces.
(ix) If $f$ has compact support, one can find (see Takahashi [52], following the earlier work of Bogovski [6]) a solution $\omega$ with compact support and optimal regularity.

*Proof* The implication (ii) $\Rightarrow$ (i) is straightforward using partial integration, cf. Theorem 11. To show the other implication, we first use Theorem 17 to find a solution $u \in C^{r+1,q}\left(\overline{\Omega}; \Lambda^k\right)$ of the problem

$$\begin{cases} du = f \quad \text{and} \quad \delta u = 0 \text{ in } \Omega \\ \qquad \nu \wedge u = 0 \qquad \text{on } \partial\Omega. \end{cases}$$

Since $\nu \wedge u = 0$, we can apply Lemma 24 Part (i) to find $v \in C^{r+2,q}(\overline{\Omega}; \Lambda^{k-1})$ such that

$$dv = -u \quad \text{on } \partial\Omega.$$

We finally set $\omega = u + dv$ to obtain the result. ■

We have now the dual version.

**Theorem 28** *Let $r \geq 0$ and $1 \leq k \leq n$ be integers, $0 < q < 1$ and $\Omega \subset \mathbb{R}^n$ be a bounded connected open smooth set with exterior unit normal $\nu$. Let $g : \overline{\Omega} \to \Lambda^{k-1}$. Then the following claims are equivalent.*

*(i) Let $g \in C^{r,q}\left(\overline{\Omega}; \Lambda^{k-1}\right)$ satisfy*

$$\delta g = 0 \text{ in } \Omega, \quad \nu \lrcorner g = 0 \text{ on } \partial\Omega \tag{C1}$$

*and, for every $\chi \in \mathcal{H}_N\left(\Omega; \Lambda^{k-1}\right)$,*

$$\int_\Omega \langle g; \chi \rangle = 0. \tag{C2}$$

*(ii) There exists $\omega \in C^{r+1,q}\left(\overline{\Omega}; \Lambda^k\right)$ such that*

$$\begin{cases} \delta\omega = g \text{ in } \Omega \\ \omega = 0 \text{ on } \partial\Omega \end{cases}$$

*and there exists a constant $C = C(r, q, \Omega)$ such that*

$$\|\omega\|_{C^{r+1,q}(\overline{\Omega})} \leq C \|g\|_{C^{r,q}(\overline{\Omega})}.$$

*Remark 29*

(i) Instead of imposing the boundary data to be 0 we can have any boundary data $\omega_0$ satisfying, in addition to $\delta g = 0$,

$$\nu \lrcorner \delta\omega_0 = \nu \lrcorner g \text{ on } \partial\Omega$$

and, for every $\chi \in \mathcal{H}_N\left(\Omega; \Lambda^{k-1}\right)$,

$$\int_\Omega \langle g; \chi\rangle = \int_{\partial\Omega} \langle \nu \lrcorner \omega_0; \chi\rangle .$$

(ii) In the case $k = 1$ (cf. Theorem 30 below), then $\delta\omega = \operatorname{div}\omega$ and the conditions (C1) are trivially satisfied, while (C2) reads as

$$\int_\Omega g = 0.$$

(iii) When $k = n$, then the result is still valid for $q = 0, 1$ with a simple argument.
(iv) If $r = 0$, then the condition $\delta g = 0$ is understood in the sense of distributions.
(v) The above results remains valid if $\Omega$ is $C^{r+3,q}$.
(vi) If $\Omega$ is contractible, then $\mathcal{H}_N\left(\Omega; \Lambda^k\right) = \{0\}$.
(vii) The construction is linear and universal.
(viii) Analogous results hold in Sobolev spaces.

As a corollary, but it can be proved in a more direct way, we have the following.

**Theorem 30** *Let $r \geq 0$ be an integer and $0 < q < 1$. Let $\Omega \subset \mathbb{R}^n$ be a bounded connected open $C^{r+2,q}$ set. The following conditions are then equivalent.*

*(i) $f \in C^{r,q}\left(\overline{\Omega}\right)$ satisfies*

$$\int_\Omega f = 0.$$

*(ii) There exists $u \in C^{r+1,q}\left(\overline{\Omega}; \mathbb{R}^n\right)$ verifying*

$$\begin{cases} \operatorname{div} u = f \text{ in } \Omega \\ u = 0 \quad \text{on } \partial\Omega . \end{cases} \tag{2}$$

*Furthermore the correspondence $f \to u$ can be chosen linear and there exists $C = C(r, q, \Omega) > 0$ such that*

$$\|u\|_{C^{r+1,q}} \leq C \|f\|_{C^{r,q}} .$$

*Remark 31*

(i) The above theorem has been proved by several authors independently of the general Poincaré lemma, see [6, 8, 20, 21, 30, 32, 33, 35, 36, 39, 40, 47, 53].
(ii) The result is false when $r = q = 0$, as established in [9, 24, 42] and [48].
(iii) Similar results (see [15]) can be obtained for the inhomogeneous problem

$$\begin{cases} \operatorname{div} u + \langle a; u\rangle = f \text{ in } \Omega \\ u = 0 \qquad \text{on } \partial\Omega \end{cases}$$

where $a$ is a vector field and $\langle .;. \rangle$ stands for the scalar product in $\mathbb{R}^n$. As a matter of curiosity this last result has applications to nonlinear problems (cf. [23]) of the form

$$\begin{cases} \det \nabla \varphi(x) = f(x, \varphi(x)) & x \in \Omega \\ \varphi(x) = x & x \in \partial\Omega. \end{cases}$$

## 4 The Flow Method

We start by recalling a well known result of differential geometry. In the sequel we will write

$$u = u(t,x) = u_t(x) \quad \text{and} \quad \varphi = \varphi(t,x) = \varphi_t(x).$$

**Theorem 32** *Let $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ be open sets, $T > 0$ and $1 \leq k \leq n$ be an integer. Let*

$$u \in C^1([0,T] \times \Omega_2; \mathbb{R}^n) \quad \text{and} \quad \varphi \in C^1([0,T] \times \Omega_1; \Omega_2)$$

*be such that, in $\Omega_1$,*

$$\frac{d}{dt}\varphi_t = u_t \circ \varphi_t = u_t(t, \varphi_t(t,x)), \quad \text{for every } 0 \leq t \leq T. \tag{3}$$

*Then, for every $f \in C^1\left([0,T] \times \Omega_2; \Lambda^k\right)$, the following equality holds in $\Omega_1$ and for $0 \leq t \leq T$*

$$\frac{d}{dt}[\varphi_t^*(f_t)] = \varphi_t^*\left(\frac{d}{dt}f_t + d(u_t \lrcorner f_t) + u_t \lrcorner (df_t)\right) \tag{4}$$

*where $u_t$ has been identified with a 1-form.*

*Remark 33*

(i) The right-hand side of (4) is related to the so called Lie derivative which we define now. Let $u \in C^1(U; \mathbb{R}^n)$ and $f \in C^1(U; \Lambda^k)$. The *Lie derivative* $\mathcal{L}_u(f)$ is defined by

$$\mathcal{L}_u(f) = \left.\frac{d}{dt}\right|_{t=0} \varphi_t^*(f)$$

where $\varphi = \varphi(t,x) = \varphi_t(x)$ is the flow associated to the vector field $u$, that is

$$\begin{cases} \frac{d}{dt}\varphi_t = u \circ \varphi_t \\ \varphi_0 = \mathrm{id} \end{cases}$$

for $t$ small enough. *Cartan formula* gives that

$$\mathcal{L}_u(f) = d(u \lrcorner f) + u \lrcorner df.$$

(ii) When $k = n$ the formula reads as

$$\frac{d}{dt}[f_t(\varphi_t) \det \nabla \varphi_t] = \left[ (\frac{d}{dt} f_t)(\varphi_t) + \mathrm{div}(f_t u_t)(\varphi_t) \right] \det \nabla \varphi_t .$$

*Proof* We prove this result only for $k = n$. Note that when $k = n$, then necessarily $df_t = 0$ and (identifying as usual the $(n-1)$-form $u_t \lrcorner f_t$ with a vector field whose components have the appropriate sign)

$$d(u_t \lrcorner f_t) = \mathrm{div}(f_t u_t).$$

*Step 1.* First recall the well known Abel formula. Any solution of

$$F'(t) = A(t) F(t)$$

satisfies

$$\det F(t) = [\det F(0)] \exp\left[ \int_0^t \mathrm{trace}\, A(s)\, ds \right].$$

It follows that in the nonlinear case (cf., for example, Theorem 7.2 in Chap. 1 of Coddington-Levinson [13]) the solution of (3) satisfies

$$\det \nabla \varphi_t(x) = [\det \nabla \varphi_0(x)] \exp\left[ \int_0^t (\mathrm{div}\, u_s)(\varphi_s(x))\, ds \right].$$

Since the right hand side of the above identity is $C^1$ in $t$, we get

$$\frac{d}{dt}[\det \nabla \varphi_t(x)] = \det \nabla \varphi_t(x) \cdot (\mathrm{div}\, u_t)(\varphi_t(x)). \tag{5}$$

*Step 2.* Using (3), we obtain

$$\begin{aligned}\frac{d}{dt}[\varphi_t^*(f_t)] &= \frac{d}{dt}[\det \nabla\varphi_t \cdot f_t(\varphi_t)] \\ &= \frac{d}{dt}[\det \nabla\varphi_t] f_t(\varphi_t) + \det \nabla\varphi_t \left[ (\frac{d}{dt} f_t)(\varphi_t) + \langle \nabla f_t(\varphi_t); \frac{d}{dt}\varphi_t \rangle \right]\end{aligned}$$

and thus, appealing to (5), we find

$$\begin{aligned}\frac{d}{dt}[\varphi_t^*(f_t)] &= \det \nabla\varphi_t \left[ (\operatorname{div} u_t)(\varphi_t) \cdot f_t(\varphi_t) + (\frac{d}{dt} f_t)(\varphi_t) + \langle \nabla f_t(\varphi_t); u_t(\varphi_t) \rangle \right] \\ &= \det \nabla\varphi_t \left[ (\frac{d}{dt} f_t)(\varphi_t) + \operatorname{div}(f_t u_t)(\varphi_t) \right] = \varphi_t^* \left( \frac{d}{dt} f_t + \operatorname{div}(f_t u_t) \right)\end{aligned}$$

which concludes the proof. ■

As a consequence we have the following result essentially established by Moser [46].

**Theorem 34** *Let $r \geq 1$ and $1 \leq k \leq n$ be integers, $0 \leq q \leq 1$, $T > 0$ and $\Omega \subset \mathbb{R}^n$ be a bounded open Lipschitz set. Let*

$$u \in C^{r,q}\left([0,T] \times \overline{\Omega}; \mathbb{R}^n\right) \quad \text{and} \quad f \in C^{r,q}\left([0,T] \times \overline{\Omega}; \Lambda^k\right)$$

*be such that, for every $t \in [0,T]$,*

$$u_t = 0 \quad \text{on } \partial\Omega, \quad df_t = 0 \text{ in } \Omega$$

$$d(u_t \lrcorner f_t) = -\frac{d}{dt} f_t \quad \text{in} \quad \Omega.$$

*Then, for every $t \in [0,T]$, the solution $\varphi_t$ of*

$$\begin{cases} \frac{d}{dt}\varphi_t = u_t \circ \varphi_t \; 0 \leq t \leq T \\ \varphi_0 = \mathrm{id} \end{cases} \tag{6}$$

*belongs to* $\mathrm{Diff}^{r,q}\left(\overline{\Omega}; \overline{\Omega}\right)$, *satisfies $\varphi_t = \mathrm{id}$ on $\partial\Omega$ and*

$$\varphi_t^*(f_t) = f_0 \quad \text{in} \quad \Omega.$$

*Proof* Standard results show that, for every $0 \leq t \leq T$, the solution $\varphi_t$ of (6) belongs to $\mathrm{Diff}^{r,q}(\overline{\Omega}; \overline{\Omega})$ and verifies $\varphi_t = \mathrm{id}$ on $\partial\Omega$. Moreover, defining $\varphi : [0,T] \times \overline{\Omega} \to \overline{\Omega}$ by $\varphi(t,x) = \varphi_t(x)$, we have

$$\varphi \in C^{r,q}([0,T] \times \overline{\Omega}; \overline{\Omega}).$$

Using Theorem 32 and the hypotheses on $u_t$ and $f_t$, we find that, in $\Omega$,

$$\frac{d}{dt}[\varphi_t^*(f_t)] = \varphi_t^*\left(\frac{d}{dt}f_t + d(u_t \lrcorner f_t) + u_t \lrcorner (df_t)\right) = 0,$$

which implies the result since $\varphi_0 = \mathrm{id}$. ■

# 5 The Case of Volume Forms

## *5.1 Statement of the Problem*

We start with the case $k = n$. We first observe that the local problem is here elementary. Indeed if we want to solve (if $g \equiv 1$, the general case being treated similarly)

$$\varphi^*(1) = f \quad \Leftrightarrow \quad \det \nabla \varphi = f$$

we just choose $\varphi(x) = \left(\varphi^1(x), x_2, \cdots x_n\right)$ where

$$\varphi^1(x_1, x_2, \cdots x_n) = \int^{x_1} f(t, x_2, \cdots x_n)\, dt.$$

Note however that this does not give the optimal regularity for $\varphi$. A less trivial problem, discussed in Theorem 39, concerns the case where several equations are considered, namely for $1 \leq i \leq n$

$$\varphi^*(g_i) = f_i \quad \Leftrightarrow \quad g_i(\varphi) \det \nabla \varphi = f_i.$$

In the case $k = n$ we will therefore (apart from Theorem 39) discuss only the global case which takes the following form. Given $\Omega$ a bounded open set in $\mathbb{R}^n$ and $f, g : \mathbb{R}^n \to \mathbb{R}$, we want to find $\varphi : \overline{\Omega} \to \mathbb{R}^n$ verifying

$$\begin{cases} g(\varphi(x)) \det \nabla \varphi(x) = f(x) & x \in \Omega \\ \varphi(x) = x & x \in \partial\Omega. \end{cases} \tag{7}$$

Writing the functions $f$ and $g$ as volume forms through the straightforward identification

$$g = g(x)\, dx^1 \wedge \cdots \wedge dx^n \quad \text{and} \quad f = f(x)\, dx^1 \wedge \cdots \wedge dx^n,$$

the problem (7) can be written as

$$\begin{cases} \varphi^*(g) = f \text{ in } \Omega \\ \varphi = \mathrm{id} \quad \text{on } \partial\Omega \end{cases}$$

where $\varphi^*(g)$ is the pullback of $g$ by $\varphi$.

The following preliminary remarks are in order.

(i) The case $n = 1$ is completely elementary and is discussed in the section below.

(ii) When $n \geq 2$, the equation in (7) is a non-linear first order *partial differential equation* homogeneous of degree $n$ in the derivatives. It is *underdetermined*, in the sense that we have $n$ unknowns (the components of $\varphi$) and only one equation. Related to this observation, we have that if there exists a solution to our problem then there are infinitely many ones. Indeed, for example, if $n = 2$, $\Omega$ is the unit ball and $f = g = 1$, the maps $\varphi_m$ (written in polar and in Cartesian coordinates) defined by

$$\begin{aligned} \varphi_m(x) = \varphi_m(x_1, x_2) &= \begin{pmatrix} r\cos\left(\theta + 2m\pi r^2\right) \\ r\sin\left(\theta + 2m\pi r^2\right) \end{pmatrix} \\ &= \begin{pmatrix} x_1\cos\left(2m\pi\left(x_1^2 + x_2^2\right)\right) - x_2\sin\left(2m\pi\left(x_1^2 + x_2^2\right)\right) \\ x_2\cos\left(2m\pi\left(x_1^2 + x_2^2\right)\right) + x_1\sin\left(2m\pi\left(x_1^2 + x_2^2\right)\right) \end{pmatrix} \end{aligned}$$

satisfy (7) for every $m \in \mathbb{Z}$.

(iii) An integration by parts, or, what amounts to the same thing, an elementary topological degree argument immediately gives the *necessary condition* (independently of the fact that $\varphi$ is a diffeomorphism or not and of the fact that $\varphi(\Omega)$ contains strictly or not the domain $\Omega$)

$$\int_\Omega f = \int_\Omega g. \tag{8}$$

In most of our analysis, it will turn out that this condition is also sufficient.

(iv) We will always assume that $g > 0$. If $g$ is not strictly positive, then other hypotheses than (8) are necessary; for example $f$ cannot be strictly positive. Indeed if for example $f \equiv 1$ and $g$ is allowed to vanish even at a single point, then no $C^1$ solution of our problem exists. However in some very special cases, one can deal with functions $f$ and $g$ that *both* change sign.

(v) We will however allow $f$ to change sign, but the analysis is very different if $f > 0$ or if $f$ vanishes, even at a single point, let alone if it becomes negative. The first problem will be discussed below, while the second is discussed in [16] and [18]. One of the main differences is that in the first case any solution of (7) is necessarily a diffeomorphism, while this is never true in the second case.

(vi) It is easy to see that any solution of (7) satisfies

$$\varphi(\Omega) \supset \Omega \quad \text{and} \quad \varphi(\overline{\Omega}) \supset \overline{\Omega}. \tag{9}$$

If $f > 0$, we have, since $\varphi$ is a diffeomorphism, that

$$\varphi(\Omega) = \Omega \quad \text{and} \quad \varphi(\overline{\Omega}) = \overline{\Omega}.$$

If this is not the case, then, in general the inclusions can be strict. This matter is discussed in details in [16].

(vii) Problem (7) admits a *weak formulation*. Indeed if $\varphi$ is a diffeomorphism, we can write the equation $g(\varphi) \det \nabla\varphi = f$ as

$$\int_{\varphi(E)} g = \int_E f \quad \text{for every open set } E \subset \Omega$$

or equivalently

$$\int_\Omega g\,\zeta\left(\varphi^{-1}\right) = \int_\Omega f\,\zeta \quad \text{for every } \zeta \in C_0^\infty(\Omega).$$

We observe that both new writings make sense if $\varphi$ is only a homeomorphism.

(viii) The problem can be seen as a question of *mass transportation*. Indeed we want to transport the mass distribution $g$ to the mass distribution $f$, without moving the points of the boundary of $\Omega$. In this context the equation is usually written as

$$\int_E g = \int_{\varphi^{-1}(E)} f \quad \text{for every open set } E \subset \Omega.$$

The problem of *optimal* mass transportation has received considerable attention. We should point out that our analysis is not in this framework (except in an indirect way in Sect. 5.6.4). The two main strong points of our analysis are that we are able to find smooth solutions, sometimes with the optimal regularity, and to deal with fixed boundary data.

## 5.2 The One Dimensional Case

As already said the case $n = 1$ is completely elementary, but it exhibits some striking differences with the case $n \geq 2$. However it may shed some light on some issues that we will discuss in the higher dimensional case. Let $\Omega = (a, b)$,

$$F(x) = \int_a^x f(t)\,dt \quad \text{and} \quad G(x) = \int_a^x g(t)\,dt.$$

Then Problem (7) becomes

$$\begin{cases} G(\varphi(x)) = F(x) & \text{if } x \in (a,b) \\ \varphi(a) = a \quad \text{and} \quad \varphi(b) = b. \end{cases}$$

If $G$ is invertible, and this happens if, for example, $g > 0$ and if

$$F([a,b]) \subset G(\mathbb{R}), \tag{10}$$

and this happens if, for example, $g \geq g_0 > 0$, then the problem has the solution

$$\varphi(x) = G^{-1}(F(x)).$$

The necessary condition (8)

$$\int_a^b f = \int_a^b g$$

ensures that

$$\varphi(a) = a \quad \text{and} \quad \varphi(b) = b.$$

This very elementary analysis leads to the following conclusions.

(1) Contrary to the case $n \geq 2$, the necessary condition (8) is not sufficient. We need the extra condition (10).
(2) The problem has a *unique* solution, contrary to the case $n \geq 2$.
(3) If $f$ and $g$ are in the space $C^r$, then the solution $\varphi$ is in $C^{r+1}$.
(4) If $f > 0$, then $\varphi$ is a diffeomorphism from $[a,b]$ onto itself.
(5) If $f$ is allowed to change sign, then, in general,

$$[a,b] \underset{\neq}{\subset} \varphi([a,b]).$$

For example, this always happens if $f(a) < 0$ or $f(b) < 0$.

## 5.3 The Case $f \cdot g > 0$

We now discuss the problem (7) when $f \cdot g > 0$. It will be seen that (8) is sufficient to solve (7) and that any solution is in fact a diffeomorphism from $\overline{\Omega}$ to $\overline{\Omega}$. This last observation implies, in particular, a symmetry in $f$ and $g$ and allows us to restrict ourselves, without loss of generality, to the case $g \equiv 1$. Our main result will be the following.

**Theorem 35 (Dacorogna-Moser Theorem)** *Let $r \geq 0$ be an integer and $0 < q < 1$. Let $\Omega \subset \mathbb{R}^n$ be a bounded connected open $C^{r+2,q}$ set. Then the two following statements are equivalent.*

*(i) The function $f \in C^{r,q}\left(\overline{\Omega}\right)$, $f > 0$ in $\overline{\Omega}$ and satisfies*

$$\int_\Omega f = \operatorname{meas} \Omega .$$

*(ii) There exists $\varphi \in \operatorname{Diff}^{r+1,q}\left(\overline{\Omega}; \overline{\Omega}\right)$ satisfying*

$$\begin{cases} \det \nabla \varphi (x) = f(x) & x \in \Omega \\ \varphi (x) = x & x \in \partial\Omega . \end{cases}$$

*Furthermore if $c > 0$ is such that*

$$\left\| \frac{1}{f} \right\|_{C^0}, \ \|f\|_{C^{0,q}} \leq c,$$

*then there exists a constant $C = C(c, r, q, \Omega) > 0$ such that*

$$\|\varphi - \mathrm{id}\|_{C^{r+1,q}} \leq C \|f - 1\|_{C^{r,q}} .$$

The study of this problem originated in the seminal work of Moser [46]. This result has generated a considerable amount of work, notably by Banyaga [3], Dacorogna [19], Reimann [49], Tartar (unpublished, 1978), Zehnder [58]. The above optimal theorem was obtained by Dacorogna-Moser [30], the estimate is however in [16]. Posterior contributions can be found in Carlier-Dacorogna [12], Rivière-Ye [50] and Ye [57]. Burago-Kleiner [10] and Mc Mullen [42], independently, proved that the result is false if $r = q = 0$, suggesting that the gain of regularity is to be expected only when $0 < q < 1$.

**Corollary 36** *Let $r \geq 0$ be an integer and $0 < q < 1$. Let $\Omega \subset \mathbb{R}^n$ be a bounded connected open $C^{r+2,q}$ set. Let $f, g \in C^{r,q}\left(\overline{\Omega}\right)$ be such that $f \cdot g > 0$ in $\overline{\Omega}$ and*

$$\int_\Omega f = \int_\Omega g. \tag{11}$$

*Then there exists $\varphi \in \operatorname{Diff}^{r+1,q}\left(\overline{\Omega}; \overline{\Omega}\right)$ satisfying*

$$\begin{cases} g(\varphi (x)) \det \nabla \varphi (x) = f(x) & x \in \Omega \\ \varphi (x) = x & x \in \partial\Omega . \end{cases} \tag{12}$$

*Remark 37*

(i) Recall that $\mathrm{Diff}^{r,q}\left(\overline{\Omega};\overline{\Omega}\right)$ denotes the set of diffeomorphisms $\varphi$ so that $\varphi\left(\overline{\Omega}\right)=\overline{\Omega}, \varphi\in C^{r,q}\left(\overline{\Omega};\mathbb{R}^n\right)$ and $\varphi^{-1}\in C^{r,q}\left(\overline{\Omega};\mathbb{R}^n\right)$.

(ii) If the domain is not connected, then the condition (11) has to hold on each connected component.

(iii) The sufficient conditions are also necessary. More precisely if $\varphi$ satisfies (12), then necessarily, for non-vanishing $f$ and $g$, we have $f\cdot g>0$ in $\overline{\Omega}$ and (11) holds. Moreover the function

$$\frac{f}{g\circ\varphi}\in C^{r,q}\left(\overline{\Omega}\right),$$

hence, if one of the functions $f$ or $g$ is in $C^{r,q}$, then so is the other one.

*Proof (Corollary 36)* First find, by Theorem 35,

$$\psi_1,\psi_2\in \mathrm{Diff}^{r+1,q}(\overline{\Omega};\overline{\Omega})$$

satisfying

$$\begin{cases}\det\nabla\psi_2(x)=\dfrac{f(x)\,\mathrm{meas}\,\Omega}{\displaystyle\int_\Omega f(x)\,dx} & x\in\Omega\\ \det\nabla\psi_1(x)=\dfrac{g(x)\,\mathrm{meas}\,\Omega}{\displaystyle\int_\Omega g(x)\,dx} & x\in\Omega\\ \psi_1(x)=\psi_2(x)=x & x\in\partial\Omega.\end{cases}$$

It is then easy to see that $\varphi=\psi_1^{-1}\circ\psi_2$ satisfies (12). ■

There are other more constructive methods to solve the equation, cf. for example Dacorogna-Moser [30]. These methods do not use the regularity of elliptic differential operators; in this sense they are more elementary. The drawback is that they do not provide any gain of regularity, which is the strong point of the above theorem. However the advantage is that they are much more flexible. For example, if we assume in (7) that

$$\mathrm{supp}(f-g)\subset\Omega$$

then we are able to find $\varphi$ such that

$$\mathrm{supp}(\varphi-\mathrm{id})\subset\Omega.$$

In this last case see also [37].

## 5.4 The Case with no Sign Hypothesis on $f$

We start by observing that if $f$ vanishes even at a single point, then the solution $\varphi$ cannot be anymore a diffeomorphism, though it can be a homeomorphism. In any case if $f$ is negative somewhere, it can never be a homeomorphism. Furthermore if $f$ is negative in some parts of the boundary, then any solution $\varphi$ must go out of the domain, more precisely

$$\overline{\Omega} \underset{\neq}{\subset} \varphi(\overline{\Omega}).$$

A special case of the theorem proved by Cupini-Dacorogna-Kneuss [18] is the following.

**Theorem 38** *Let $n \geq 2$ and $r \geq 1$ be integers. Let $B_1 \subset \mathbb{R}^n$ be the open unit ball. Let $f \in C^r(\overline{B}_1)$ be such that*

$$\int_{B_1} f = \operatorname{meas} B_1 .$$

*Then there exists $\varphi \in C^r(\overline{B}_1; \mathbb{R}^n)$ satisfying*

$$\begin{cases} \det \nabla \varphi (x) = f(x) & x \in B_1 \\ \varphi (x) = x & x \in \partial B_1 . \end{cases}$$

*Furthermore the following conclusions also hold.*

*(i) If either $f > 0$ on $\partial B_1$ or $f \geq 0$ in $\overline{B}_1$, then $\varphi$ can be chosen so that*

$$\varphi(\overline{B}_1) = \overline{B}_1 .$$

*(ii) If $f \geq 0$ in $\overline{B}_1$ and $f^{-1}(0) \cap B_1$ is countable, then $\varphi$ can be chosen as a homeomorphism from $\overline{B}_1$ onto $\overline{B}_1$.*

## 5.5 Multiple Jacobian Equations

When dealing with several equations we have the following local result established in [28].

**Theorem 39** *Let $n, r \geq 2$ be integers, $x_0 \in \mathbb{R}^n$ and $g_i, f_i \in C^r(\mathbb{R}^n)$, $1 \leq i \leq n$, be such that $g_i(x_0), f_i(x_0) \neq 0$ for every $1 \leq i \leq n$,*

$$\operatorname{rank} \left[ \begin{pmatrix} \nabla (g_2/g_1) \\ \vdots \\ \nabla (g_n/g_1) \end{pmatrix} (x_0) \right] = \operatorname{rank} \left[ \begin{pmatrix} \nabla (f_2/f_1) \\ \vdots \\ \nabla (f_n/f_1) \end{pmatrix} (x_0) \right] = n - 1 \tag{13}$$

*and*

$$\frac{g_i}{g_1}(x_0) = \frac{f_i}{f_1}(x_0) \quad \text{for every } 2 \le i \le n. \tag{14}$$

Part 1 (Existence and regularity). *There exist a neighborhood $U$ of $x_0$ and $\varphi \in \mathrm{Diff}^{r-1}(U; \varphi(U))$ such that $\varphi(x_0) = x_0$ and*

$$g_i(\varphi) \det \nabla \varphi = f_i \quad \text{in } U \text{ for every } 1 \le i \le n. \tag{15}$$

*The regularity is, in general, optimal.*

Part 2 (Uniqueness). *Let $h \in C^{r-1}(\mathbb{R}^n)$ be such that $h(x_0) = 0$ and*

$$\det \left[ \begin{pmatrix} \nabla h \\ \nabla(f_2/f_1) \\ \vdots \\ \nabla(f_n/f_1) \end{pmatrix} (x_0) \right] \neq 0.$$

*Let $U$ be a neighborhood of $x_0$, $\varphi \in \mathrm{Diff}^{r-1}(U; \varphi(U))$ and $\psi \in \mathrm{Diff}^{r-1}(U; \psi(U))$ be two solutions of (15) verifying*

$$\varphi = \psi \quad \text{on} \quad \{x \in U : h(x) = 0\}.$$

*Then, up to further restricting $U$,*

$$\varphi \equiv \psi \quad \text{on } U.$$

*Remark 40*

(i) The fact that the regularity that we obtain is optimal (even in Hölder spaces), is, at first glance, surprising.
(ii) The hypothesis (14) is obviously necessary to have $\varphi(x_0) = x_0$. The hypothesis (13) (although not necessary in general) is very reasonable: for example if $n = 2$ and $g_1 = g_2$ near $x_0$ (and thus $\nabla(g_2/g_1) = 0$), then obviously $f_2$ has to be equal to $f_1$ near $x_0$ to be able to solve (15) and vice versa.
(iii) The solution in Part 1 of the previous theorem is easily seen not to be unique. However (cf. Part 2) it becomes unique as soon as the value of the solution is prescribed not only at the point $x_0$ but on a $(n-1)$ surface (near $x_0$) compatible with the data.
(iv) It is to be noted that the equivalent problem for 1 and 2 forms has already been considered in Chap. 15 of [16] (for a particular case see Theorem 63).

*Proof* We discuss here only the existence part, for the other statements we refer to [28]. With no loss of generality we can assume throughout the proof that $x_0 = 0$.

Obviously (15) is equivalent to $\varphi(0) = 0$,

$$g_1(\varphi) \det \nabla\varphi = f_1 \quad \text{and} \quad \frac{g_i}{g_1}(\varphi) = \frac{f_i}{f_1} \text{ if } 2 \le i \le n.$$

For $a \in \mathbb{R}^n$ we set $v_a(x) = \langle a; x\rangle$. We claim that $\varphi = G^{-1} \circ F$ has all the desired properties where

$$G = \left(v_a, \frac{g_2}{g_1}, \cdots, \frac{g_n}{g_1}\right) \quad \text{and} \quad F = \left(u, \frac{f_2}{f_1}, \cdots, \frac{f_n}{f_1}\right)$$

where $a \in \mathbb{R}^n$ and $u : \mathbb{R}^n \to \mathbb{R}$ are determined as follows. Using (13) we can select $a \in \mathbb{R}^n$ such that $\det \nabla G(0) \neq 0$ which immediately implies that $G \in \mathrm{Diff}^r(B_\epsilon; G(B_\epsilon))$ for $\epsilon > 0$ small enough ($B_\epsilon$ being the ball centered at 0 and of radius $\epsilon$). Then, for any $u \in C^{r-1}$ with $u(0) = 0$, we have that $\varphi = G^{-1} \circ F$ satisfies, using (14), $\varphi(0) = 0$ and

$$\frac{g_i}{g_1}(\varphi) = \frac{f_i}{f_1} \quad \text{near } 0, \text{for every } 2 \le i \le n.$$

In view of the previous considerations, it only remains to find $u \in C^{r-1}$ such that $u(0) = 0$ and

$$g_1(\varphi) \det \nabla\varphi = f_1 \quad \text{near } 0 \tag{16}$$

(note that the above equation implies, in particular, that $\varphi$ is a local diffeomorphism) or equivalently

$$g_1\left(G^{-1} \circ F\right) \cdot \left(\det \nabla\left(G^{-1}\right)\right)(F) \cdot \det \nabla F = f_1.$$

Let us investigate the terms in the left hand side of the last equation. Note that

$$g_1(G^{-1} \circ F)(x) \quad \text{and} \quad \left(\det \nabla\left(G^{-1}\right)\right)(F(x))$$

have, respectively, the form

$$\alpha(x, u(x)) \quad \text{and} \quad \beta(x, u(x))$$

where $\alpha \in C^r$ with $\alpha(0,0) \neq 0$ and $\beta \in C^{r-1}$ with $\beta(0,0) \neq 0$. Finally, using (13), we obtain that

$$\det \nabla F = \langle \nabla u; H\rangle$$

for some $H \in C^{r-1}(\mathbb{R}^n;\mathbb{R}^n)$ with $H(0) \neq 0$. We hence deduce that (16) can be written, near 0, as

$$\langle \nabla u(x); H(x)\rangle = \gamma(x, u(x)) = \frac{f_1(x)}{\alpha(x, u(x))\,\beta(x, u(x))}$$

where $\gamma \in C^{r-1}$. Using the method of characteristics, we can find, near 0, $u \in C^{r-1}$ verifying the last equation as well as $u(0) = 0$. This concludes the proof of the existence part. ■

## 5.6 *Proof of the Main Theorem*

Before proving Theorem 35 (without the estimates) we prove two intermediate results.

### 5.6.1 The Flow Method

We first present the flow method introduced by Moser [46] (cf. also Sect. 4), who did not however consider the boundary condition. Note that the theorem does not provide the optimal regularity.

**Theorem 41** *Let $r \geq 1$ be a integer, $0 \leq q \leq 1$ and $\Omega \subset \mathbb{R}^n$ be a bounded connected open $C^{r+2,q}$ set. Let also $f \in C^{r,q}(\overline{\Omega})$ be such that $f > 0$ in $\overline{\Omega}$ and*

$$\int_\Omega f = \operatorname{meas}\Omega.$$

*Then there exists $\varphi \in \mathrm{Diff}^{r,q}(\overline{\Omega};\overline{\Omega})$ satisfying*

$$\begin{cases} \det \nabla\varphi(x) = f(x) & x \in \Omega \\ \varphi(x) = x & x \in \partial\Omega. \end{cases} \tag{17}$$

*Proof* Define, for $0 \leq t \leq 1$, $x \in \overline{\Omega}$,

$$f_t(x) = (1-t)f(x) + t$$

and

$$u_t(x) = \frac{u(x)}{f_t(x)} \tag{18}$$

where $u \in C^{r,q}\left(\overline{\Omega};\mathbb{R}^n\right)$ (if $0 < q < 1$, then $u \in C^{r+1,q}\left(\overline{\Omega};\mathbb{R}^n\right)$) satisfies

$$\begin{cases} \operatorname{div} u = f - 1 \text{ in } \Omega \\ u = 0 \qquad \text{on } \partial\Omega. \end{cases} \tag{19}$$

Such a $u$ exists by Theorem 30 (cf. in particular the remark following the theorem) or Theorem 30. Note however that $u_t$ (see (18)) is only in $C^{r,q}$ (even if $0 < q < 1$), since $f$ is only in $C^{r,q}$. Since (18) and (19) hold, we have

$$\begin{cases} \operatorname{div}(u_t f_t) = -\frac{d}{dt} f_t = f - 1 \text{ in } \Omega \\ u_t = 0 \qquad\qquad \text{on } \partial\Omega. \end{cases} \tag{20}$$

We can then apply Theorem 34 and have, defining $\phi_t : \overline{\Omega} \to \mathbb{R}^n$ for every $t \in [0,1]$ as the solution of

$$\begin{cases} \dfrac{d}{dt}\phi_t = u_t \circ \phi_t \ 0 \le t \le 1 \\ \phi_0 = \mathrm{id}, \end{cases}$$

that

$$\varphi = \phi_1$$

has all the desired properties. ■

### 5.6.2 The Fixed Point Method

We now prove Theorem 36 when $g \equiv 1$ and under a smallness assumption on the $C^{0,s}$ norm of $f - 1$. The following result is in Dacorogna-Moser [30] and follows earlier considerations by Zehnder [58].

**Theorem 42** *Let $r \ge 0$ be an integer and $0 < s \le q < 1$. Let $\Omega \subset \mathbb{R}^n$ be a bounded connected open $C^{r+2,q}$ set. Let $f \in C^{r,q}\left(\overline{\Omega}\right)$, $f > 0$ in $\overline{\Omega}$ and*

$$\int_\Omega f = \operatorname{meas}\Omega.$$

*Then there exists $\epsilon = \epsilon(r,q,s,\Omega) > 0$ such that if $\|f - 1\|_{C^{0,s}} \le \epsilon$, then there exists $\varphi \in \mathrm{Diff}^{r+1,q}\left(\overline{\Omega};\overline{\Omega}\right)$ satisfying*

$$\begin{cases} \det \nabla\varphi(x) = f(x) \ x \in \Omega \\ \varphi(x) = x \qquad x \in \partial\Omega. \end{cases} \tag{21}$$

*Moreover there exists a constant* $c = c(r, q, s, \Omega) > 0$ *such that if* $\|f - 1\|_{C^{0,s}} \leq \epsilon$, *then* $\varphi$ *satisfies*

$$\|\varphi - \mathrm{id}\|_{C^{r+1,q}} \leq c \|f - 1\|_{C^{r,q}} \quad \textit{and} \quad \|\varphi - \mathrm{id}\|_{C^{1,s}} \leq c \|f - 1\|_{C^{0,s}} .$$

*Proof* For the convenience of the reader we will not use the abstract fixed point theorem (cf. Theorem 81) but we will redo the proof. We divide the proof into two steps.

*Step1*. We start by introducing some notations.

(i) Let

$$X = \left\{a \in C^{r+1,q}\left(\overline{\Omega}; \mathbb{R}^n\right) : \ a = 0 \text{ on } \partial\Omega\right\}$$

$$Y = \left\{b \in C^{r,q}\left(\overline{\Omega}\right) : \ \int_\Omega b = 0\right\}.$$

Define $L : X \to Y$ by $La = \operatorname{div} a$. Note that $L$ is well defined by the divergence theorem. As seen in Theorem 30, there exist a bounded linear operator $L^{-1} : Y \to X$ and a constant $K_1 > 0$, such that

$$LL^{-1} = \mathrm{id}, \quad \text{in } Y$$

$$\left\|L^{-1}b\right\|_{C^{1,s}} \leq K_1 \|b\|_{C^{0,s}} \tag{22}$$

$$\left\|L^{-1}b\right\|_{C^{r+1,q}} \leq K_1 \|b\|_{C^{r,q}} . \tag{23}$$

(ii) Let for $\xi$, any $n \times n$ matrix,

$$Q(\xi) = \det(I + \xi) - 1 - \operatorname{trace}(\xi) \tag{24}$$

where $I$ stands for the identity matrix. Note that $Q$ is a sum of monomials of degree $t$, $2 \leq t \leq n$. Hence there exists a constant $k > 0$ such that, for every $\xi, \eta \in \mathbb{R}^{n\times n}$,

$$|Q(\xi) - Q(\eta)| \leq k\left(|\xi| + |\eta| + |\xi|^{n-1} + |\eta|^{n-1}\right)|\xi - \eta| .$$

With the same method, we can find (cf. Theorem 75) a constant $K_2 > 0$ such that if $v, w \in C^{r+1,q}$ with $\|v\|_{C^{1,s}}$ , $\|w\|_{C^{1,s}} \leq 1$, then

$$\begin{aligned} &\|Q(\nabla v) - Q(\nabla w)\|_{C^{0,s}} \leq K_2 \left(\|v\|_{C^{1,s}} + \|w\|_{C^{1,s}}\right) \|v - w\|_{C^{1,s}} \\ &\|Q(\nabla v)\|_{C^{r,q}} \leq K_2 \|v\|_{C^1} \|v\|_{C^{r+1,q}} . \end{aligned} \tag{25}$$

*Step 2*. In order to solve (21) we set $v(x) = \varphi(x) - x$ and we rewrite it as

$$\begin{cases} \operatorname{div} v = f - 1 - Q(\nabla v) & \text{in } \Omega \\ v = 0 & \text{on } \partial\Omega. \end{cases} \tag{26}$$

If we set

$$N(v) = f - 1 - Q(\nabla v)$$

then (26) is satisfied for any $v \in X$ with

$$v = L^{-1} N(v). \tag{27}$$

Note first that the equation is well defined (i.e. $N : X \to Y$), since if $v = 0$ on $\partial\Omega$ then $\int_\Omega N(v(x))\, dx = 0$. Indeed from (24) we have that

$$\begin{aligned} \int_\Omega N(v(x))\, dx &= \int_\Omega [f(x) - 1 - Q(\nabla v(x))]\, dx \\ &= \int_\Omega [f(x) + \operatorname{div} v(x) - \det(I + \nabla v(x))]\, dx; \end{aligned}$$

since $v = 0$ on $\partial\Omega$ and $\int_\Omega f = \operatorname{meas}\Omega$, it follows immediately that the right hand side of the above identity is 0.
We now solve (27) by the contraction principle. We first let

$$B = \left\{ u \in C^{r+1,q}\left(\overline{\Omega}; \mathbb{R}^n\right) : \begin{array}{c} u = 0 \text{ on } \partial\Omega \\ \|u\|_{C^{1,s}} \le 2K_1 \|f - 1\|_{C^{0,s}} \\ \|u\|_{C^{r+1,q}} \le 2K_1 \|f - 1\|_{C^{r,q}} \end{array} \right\}.$$

We endow $B$ with the $C^{1,s}$ norm. We observe that $B$ is complete and we will show that by choosing $\|f - 1\|_{C^{0,s}}$ small enough, then $L^{-1}N : B \to B$ is a contraction mapping. The contraction principle will then immediately lead to a solution $v \in B$ and hence in $C^{r+1,q}$ of (27). Indeed let

$$\|f - 1\|_{C^{0,s}} \le \min\left\{ \frac{1}{8K_1^2 K_2}, \frac{1}{2K_1} \right\}. \tag{28}$$

If $v, w \in B$ (note that by construction $2K_1 \|f - 1\|_{C^{0,s}} \le 1$), we will show that

$$\left\| L^{-1}N(v) - L^{-1}N(w) \right\|_{C^{1,s}} \le \frac{1}{2} \|v - w\|_{C^{1,s}} \tag{29}$$

$$\left\| L^{-1}N(v) \right\|_{C^{1,s}} \le 2K_1 \|f - 1\|_{C^{0,s}}, \quad \left\| L^{-1}N(v) \right\|_{C^{r+1,q}} \le 2K_1 \|f - 1\|_{C^{r,q}}. \tag{30}$$

The inequality (29) follows from (22), (25) and (28) through

$$\begin{aligned}\left\|L^{-1}N(v)-L^{-1}N(w)\right\|_{C^{1,s}} &\leq K_1\left\|N(v)-N(w)\right\|_{C^{0,s}}\\ &= K_1\left\|Q(\nabla v)-Q(\nabla w)\right\|_{C^{0,s}}\\ &\leq K_1K_2\left(\|v\|_{C^{1,s}}+\|w\|_{C^{1,s}}\right)\|v-w\|_{C^{1,s}}\\ &\leq 4K_1^2K_2\left\|f-1\right\|_{C^{0,s}}\|v-w\|_{C^{1,s}}\\ &\leq \frac{1}{2}\|v-w\|_{C^{1,s}}\ .\end{aligned}$$

To obtain the first inequality in (30) we observe that

$$\left\|L^{-1}N(0)\right\|_{C^{1,s}} \leq K_1\left\|N(0)\right\|_{C^{0,s}} = K_1\left\|f-1\right\|_{C^{0,s}}$$

and hence combining (29) with the above inequality we have immediately the first inequality in (30). To obtain the second one we just have to observe that

$$\left\|L^{-1}N(v)\right\|_{C^{r+1,q}} \leq K_1\left\|N(v)\right\|_{C^{r,q}} \leq K_1\left\|f-1\right\|_{C^{r,q}}+K_1\left\|Q(\nabla v)\right\|_{C^{r,q}} \tag{31}$$

and use the second inequality in (25) to get, recalling that $v\in B$,

$$\begin{aligned}\left\|Q(\nabla v)\right\|_{C^{r,q}} &\leq K_2\|v\|_{C^1}\|v\|_{C^{r+1,q}} \leq K_2\|v\|_{C^{1,s}}\|v\|_{C^{r+1,q}}\\ &\leq 2K_1K_2\left\|f-1\right\|_{C^{0,s}}\|v\|_{C^{r+1,q}}\ .\end{aligned}$$

The above inequality combined with (28) gives

$$\left\|Q(\nabla v)\right\|_{C^{r,q}} \leq \frac{1}{4K_1}\|v\|_{C^{r+1,q}}\ .$$

Combining this last inequality, (31) and the fact that $v\in B$ we deduce that

$$\left\|L^{-1}N(v)\right\|_{C^{r+1,q}} \leq 2K_1\left\|f-1\right\|_{C^{r,q}}\ .$$

Thus the contraction principle gives immediately the existence of a $C^{r+1,q}$ solution.

It now remains to show that $\varphi(x)=v(x)+x$ is a diffeomorphism. This is a consequence of the fact that $\det\nabla\varphi = f>0$ and $\varphi(x)=x$ on $\partial\Omega$. The estimate in the statement of the theorem follows by construction, since $v\in B$. ■

### 5.6.3 Proof of the Main Theorem

We prove Theorem 35, following the original proof of Dacorogna-Moser [30].

*Proof* We divide the proof into three steps. The first step is to prove that $(ii) \Rightarrow (i)$ and the two others to prove the reverse implication.

*Step 1.* Assume that $\varphi \in \text{Diff}^{r+1,q}\left(\overline{\Omega};\overline{\Omega}\right)$ satisfies

$$\begin{cases} \det \nabla\varphi\,(x) = f\,(x) & x \in \Omega \\ \varphi\,(x) = x & x \in \partial\Omega\,. \end{cases}$$

Then clearly $f \in C^{r,q}\left(\overline{\Omega}\right)$. We easily have that $f > 0$ in $\overline{\Omega}$ and

$$\int_\Omega f = \text{meas}\,\Omega\,.$$

*Step 2 (approximation).* We first approximate $f \in C^{r,q}$ by a function $h \in C^\infty\left(\overline{\Omega}\right)$ with $h > 0$ in $\overline{\Omega}$ so that

$$\int_\Omega \frac{f}{h} = \text{meas}\,\Omega$$

$$\left\| \frac{f}{h} - 1 \right\|_{C^{0,q/2}} \leq \epsilon \tag{32}$$

where $\epsilon$ is as in the statement of Theorem 42.
*Step 3 (conclusion).* Using (32) and Theorem 42 we can find $\varphi_1 \in \text{Diff}^{r+1,q}\left(\overline{\Omega};\overline{\Omega}\right)$ a solution of

$$\begin{cases} \det \nabla\varphi_1\,(x) = \dfrac{f\,(x)}{h\,(x)} & x \in \Omega \\ \varphi_1\,(x) = x & x \in \partial\Omega\,. \end{cases}$$

We further let $\varphi_2 \in \text{Diff}^{r+1,q}\left(\overline{\Omega};\overline{\Omega}\right)$ to be a solution of

$$\begin{cases} \det \nabla\varphi_2\,(y) = h\left(\varphi_1^{-1}\,(y)\right) & y \in \Omega \\ \varphi_2\,(y) = y & y \in \partial\Omega\,. \end{cases}$$

Such a solution exists by Theorem 41, since $h \circ \varphi_1^{-1} \in C^{r+1,q}\left(\overline{\Omega}\right)$ and

$$\int_\Omega h\left(\varphi_1^{-1}\,(y)\right) dy = \int_\Omega h\,(x) \det \nabla\varphi_1\,(x)\, dx = \int_\Omega f\,(x)\, dx = \text{meas}\,\Omega\,.$$

Finally observe that the function $\varphi = \varphi_2 \circ \varphi_1$ has all the claimed properties. ■

### 5.6.4 An Alternative Proof

We here outline the approach of Carlier-Dacorogna [12] using known results on Monge-Ampère equation. The proof that we briefly discuss below is the nonlinear analogue of the solution (for zero mean $f$) of

$$\operatorname{div}\varphi = f \text{ in } \Omega \quad \text{and} \quad \varphi = 0 \text{ on } \partial\Omega\,.$$

Indeed, the "optimal" way to solve this linear problem (cf. Theorem 30) is to look for solutions of the form $\varphi = \nabla\Phi + s$.

(i) We first solve the Neumann problem

$$\Delta\Phi = f \text{ in } \Omega \quad \text{and} \quad \frac{\partial\Phi}{\partial\nu} = 0 \text{ on } \partial\Omega\,.$$

(ii) We next seek $s$ so that $\operatorname{div} s = 0$ (cf. Lemma 24) and

$$s = -\nabla\Phi \quad \text{on } \partial\Omega\,.$$

We proceed similarly for the nonlinear problem. We look for solutions of the form $\varphi = s \circ \nabla\Phi$.

(i) We first solve, using Caffarelli result [11],

$$\det\nabla^2\Phi = f \text{ in } \Omega\,, \quad \Phi \quad \text{convex} \quad \text{and} \quad \nabla\Phi(\Omega) = \Omega\,.$$

(ii) We then find $s$ such that

$$\det\nabla s = 1 \text{ in } \Omega \quad \text{and} \quad s = \nabla\Phi^{-1} = \nabla\Phi^* \text{ on } \partial\Omega \tag{33}$$

where $\Phi^*$ is the Legendre transform of $\Phi$. This last construction goes as follows.

– We first solve, using again [11],

$$\det\nabla^2\Psi_t = (1-t) + t\det\nabla\Phi^* \text{ in } \Omega\,, \quad \Psi_t \text{ convex} \quad \text{and} \quad \nabla\Psi_t(\Omega) = \Omega\,.$$

Note that $\nabla\Psi_0 = I$ while $\nabla\Psi_1 = \nabla\Phi^*$.

– We next find $w_t$ through

$$w_t\,(\nabla\Psi_t) = \frac{d}{dt}\nabla\Psi_t\,.$$

It is easy to see that $\langle w_t; \nu\rangle = 0$ on $\partial\Omega$.

– We then solve (cf. Theorem 30)

$$\operatorname{div} v_t = -\operatorname{div} w_t \text{ in } \Omega \quad \text{and} \quad v_t = 0 \text{ on } \partial\Omega.$$

This is possible since $\langle w_t; \nu\rangle = 0$ on $\partial\Omega$.
– Letting $u_t = v_t + w_t$, we define $s_t$ by the flow method (cf. Theorem 41)

$$\frac{d}{dt}s_t = u_t(s_t) \quad \text{and} \quad s_0 = \mathrm{id}.$$

By construction $\det \nabla s_t \equiv 1$ and, by uniqueness, $s_t = \nabla\Psi_t$ on $\partial\Omega$. The map $s = s_1$ therefore solves (33).

## 6 The Case $k = 2$

Our best results besides the ones for volume forms are in the case $k = 2$.

### *6.1 The Case of Constant Forms*

Recall what we have seen in the introduction. To any 2-form $g$

$$g = \sum_{1\le i<j\le n} g_{ij}\, dx^i \wedge dx^j$$

we can associate, in a unique way, a *skew symmetric matrix* $G = \left(g_{ij}\right) \in \mathbb{R}^{n\times n}$ (i.e. $G^t = -G$). We therefore have, if we choose $\varphi(x) = Ax$,

$$\varphi^*(g) = f \quad \Leftrightarrow \quad AGA^t = F.$$

The following theorem is standard in linear algebra (cf. Theorem 65 for a sharper version).

**Theorem 43** *Let $F, G \in \mathbb{R}^{n\times n}$ be two skew symmetric matrices with*

$$\operatorname{rank} G = \operatorname{rank} F = 2m \le n.$$

*Then there exists an invertible matrix $A \in \mathbb{R}^{n\times n}$ such that $AGA^t = F$.*

The theorem says, in particular, that any skew symmetric matrix of rank $2m$ is equivalent to the canonical matrix $J_m$ given by

$$J_m = \begin{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} & \cdots & 0 & 0 \cdots 0 \\ \vdots & \ddots & \vdots & \vdots \cdots \vdots \\ 0 & \cdots & \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} & 0 \cdots 0 \\ 0 & \cdots & 0 & 0 \cdots 0 \\ \vdots & \cdots & \vdots & \vdots \ddots \vdots \\ 0 & \cdots & 0 & 0 \cdots 0 \end{pmatrix}.$$

## 6.2 *Darboux Theorem with Optimal Regularity*

The following result is the classical Darboux theorem [31] (for the importance of this problem in symplectic geometry see, for example, [41]) for closed 2-forms but with optimal regularity. This is a delicate point and it has been obtained by Bandyopadhyay-Dacorogna [1]. The other existing results provide solutions that are, at best, only in $C^{r,q}$, while in the theorem below we find a solution which belongs to $C^{r+1,q}$.

**Theorem 44 (Darboux Theorem with Optimal Regularity)** *Let $r \geq 0$ and $n = 2m \geq 4$ be integers. Let $0 < q < 1$ and $x_0 \in \mathbb{R}^n$. Let $\omega_m$ be the standard symplectic form*

$$\omega_m = \sum_{i=1}^{m} dx^{2i-1} \wedge dx^{2i}.$$

*Let $\omega$ be a 2-form. The two following statements are then equivalent.*

*(i) The 2-form $\omega$ is closed, is in $C^{r,q}$ in a neighborhood of $x_0$ and verifies*

$$\operatorname{rank}\left[\omega\left(x_0\right)\right] = n.$$

*(ii) There exist a neighborhood $U$ of $x_0$ and $\varphi \in \operatorname{Diff}^{r+1,q}(U; \varphi(U))$ such that*

$$\varphi^*\left(\omega_m\right) = \omega \text{ in } U \quad \text{and} \quad \varphi\left(x_0\right) = x_0\,.$$

*Remark 45*

(i) When $r = 0$, the hypothesis $d\omega = 0$ is to be understood in the sense of distributions.
(ii) The theorem is still valid when $n = 2$, but it is then the result of Dacorogna-Moser [30] (cf. Theorem 36).
(iii) One possible proof of the theorem could be to use Theorem 56 with $n = 2m$, i.e. the result for 1-forms. We however will go the other way around and prove Theorem 56 using Theorem 44.

*Proof* The necessary part is obvious and we discuss only the sufficient part. We divide the proof into four steps.

*Step 1*. Without loss of generality we can always assume (cf. Theorem 43) that

$$x_0 = 0 \quad \text{and} \quad \omega(0) = \omega_m .$$

*Step 2*. Our theorem will follow from Theorem 81. So we need to define the spaces and the operators and then check all the hypotheses.

1) We choose $V$ a sufficiently small ball centered at 0 and we define the sets

$$X_1 = C^{1,q}(\overline{V}; \mathbb{R}^n) \quad \text{and} \quad Y_1 = C^{0,q}(\overline{V}; \Lambda^2)$$
$$X_2 = C^{r+1,q}(\overline{V}; \mathbb{R}^n) \quad \text{and} \quad Y_2 = \{b \in C^{r,q}(\overline{V}; \Lambda^2) : db = 0 \text{ in } V\}.$$

It is easy to see that $(H_{XY})$ of Theorem 81 is fulfilled.
2) Define $L : X_2 \to Y_2$ by

$$La = d[a \llcorner \omega_m] = b.$$

We will show that there exists $L^{-1} : Y_2 \to X_2$ a linear right inverse of $L$ and a constant $C_1 = C_1(r, q, V)$ such that

$$\|L^{-1}b\|_{X_i} \leq C_1 \|b\|_{Y_i} \text{ for every } b \in Y_2 \text{ and } i = 1, 2.$$

Once shown this, $(H_L)$ of Theorem 81 will be satisfied. First, using Theorem 19, find $w \in C^{r+1,q}(\overline{V}; \Lambda^1)$ and $C_1 = C_1(r, q, V) > 0$ such that

$$dw = b \quad \text{in } V$$
$$\|w\|_{C^{r+1,q}} \leq C_1 \|b\|_{C^{r,q}} \quad \text{and} \quad \|w\|_{C^{1,q}} \leq C_1 \|b\|_{C^{0,q}} .$$

Moreover the correspondence $b \to w$ can be chosen to be linear. Next, define $a \in C^{r+1,q}(\overline{V}; \mathbb{R}^n)$ by

$$a_{2i-1} = w_{2i} \quad \text{and} \quad a_{2i} = -w_{2i-1}, \quad 1 \leq i \leq m$$

and note that

$$a \lrcorner \omega_m = w.$$

Finally, defining $L^{-1} : Y_2 \to X_2$ by $L^{-1}(b) = a$, we easily check that $L^{-1}$ is linear,

$$L L^{-1} = \mathrm{id} \quad \text{on } Y_2$$

and

$$\|L^{-1} b\|_{X_i} \leq C_1 \|b\|_{Y_i} \text{ for every } b \in Y_2 \text{ and } i = 1, 2.$$

So that $(H_L)$ of Theorem 81 is satisfied.

3) We then let $Q$ be defined by

$$Q(u) = \omega_m - (\mathrm{id} + u)^* \omega_m + d\left[u \lrcorner \omega_m\right].$$

Since

$$d\left[u \lrcorner \omega_m\right] = \sum_{i=1}^{m} \left[du^{2i-1} \wedge dx^{2i} + dx^{2i-1} \wedge du^{2i}\right]$$

$$\omega_m - (\mathrm{id} + u)^* \omega_m = \sum_{i=1}^{m} \left[dx^{2i-1} \wedge dx^{2i} - \left(dx^{2i-1} + du^{2i-1}\right) \wedge \left(dx^{2i} + du^{2i}\right)\right]$$

we get

$$Q(u) = -\sum_{i=1}^{m} du^{2i-1} \wedge du^{2i}.$$

4) Note that $Q(0) = 0$ and $dQ(u) = 0$ in $V$. Appealing to Theorem 75, there exists a constant $C_2 = C_2(r, V)$ such that, for every $u, v \in C^{r+1,q}(\overline{\Omega}; \mathbb{R}^n)$, the following estimates hold

$$\begin{aligned}
\|Q(u) - Q(v)\|_{C^{0,q}} &\leq \sum_{i=1}^{m} \|du^{2i-1} \wedge du^{2i} - dv^{2i-1} \wedge dv^{2i}\|_{C^{0,q}} \\
&\leq \sum_{i=1}^{m} \|du^{2i-1} \wedge \left(du^{2i} - dv^{2i}\right)\|_{C^{0,q}} \\
&\quad + \sum_{i=1}^{m} \|\left(dv^{2i-1} - du^{2i-1}\right) \wedge dv^{2i}\|_{C^{0,q}} \\
&\leq C_2(\|u\|_{C^{1,q}} + \|v\|_{C^{1,q}})\|u - v\|_{C^{1,q}}
\end{aligned}$$

and

$$\begin{aligned}\|Q(u)\|_{C^{r,q}} &\leq \sum_{i=1}^{m} \|du^{2i-1} \wedge du^{2i}\|_{C^{r,q}} \\ &\leq C \sum_{i=1}^{m} \left[\|du^{2i-1}\|_{C^{r,q}} \|du^{2i}\|_{C^0} + \|du^{2i}\|_{C^{r,q}} \|du^{2i-1}\|_{C^0}\right] \\ &\leq C_2 \|u\|_{C^{1,q}} \|u\|_{C^{r+1,q}} .\end{aligned}$$

We therefore see that property $\left(H_Q\right)$ is valid for every $\rho$ and we choose $\rho = 1/(2n)$, where $c = C_2$.

5) Setting $\varphi = \mathrm{id} + u$, we can rewrite the equation $\varphi^*(\omega_m) = \omega$ as

$$\begin{aligned}Lu = d[u \lrcorner \omega_m] &= \omega - (\mathrm{id} + u)^* \omega_m + d[u \lrcorner \omega_m] \\ &= \omega - \omega_m + [\omega_m - (\mathrm{id} + u)^* \omega_m + d[u \lrcorner \omega_m]] \\ &= \omega - \omega_m + Q(u).\end{aligned}$$

*Step 3*. We may now apply Theorem 81 and get that there exists $\psi \in \mathrm{Diff}^{r+1,q}(\overline{V}; \psi(\overline{V}))$ such that

$$\psi^*(\omega_m) = \omega \text{ in } V \quad \text{and} \quad \|\nabla\psi - I\|_{C^0} \leq 1/(2n),$$

provided

$$\|\omega - \omega_m\|_{C^{0,q}} \leq \frac{1}{2C_1 \max\{4C_1C_2, 2n\}}. \tag{34}$$

Setting $\varphi(x) = \psi(x) - \psi(0)$, we have indeed proved that there exists $\varphi \in \mathrm{Diff}^{r+1,q}(\overline{V}; \varphi(\overline{V}))$ satisfying

$$\varphi^*(\omega_m) = \omega \text{ in } V, \quad \|\nabla\varphi - I\|_{C^0} \leq 1/(2n) \quad \text{and} \quad \varphi(0) = 0.$$

*Step 4*. We may now conclude the proof of the theorem.
*Step 4.1*. Let $0 < \epsilon < 1$ and define

$$\omega^\epsilon(x) = \omega(\epsilon x).$$

Observe that $\omega^\epsilon \in C^{r,q}(\overline{V}; \Lambda^2)$, $d\omega^\epsilon = 0$, $\omega^\epsilon(0) = \omega_m$ and

$$\|\omega^\epsilon - \omega_m\|_{C^{0,q}(\overline{V})} \to 0 \quad \text{as } \epsilon \to 0.$$

Choose $\epsilon$ sufficiently small so that

$$\|\omega^\epsilon - \omega_m\|_{C^{0,q}(\overline{V})} \leq \frac{1}{2C_1 \max\{4C_1C_2\,,\,2n\}}\,.$$

Apply Step 3 to find $\psi_\epsilon \in \mathrm{Diff}^{r+1,q}(\overline{V}; \psi_\epsilon\,(\overline{V}))$ satisfying

$$\psi_\epsilon^*\,(\omega_m) = \omega^\epsilon \text{ in } V, \quad \|\nabla\psi_\epsilon - I\|_{C^0} \leq 1/\,(2n) \quad \text{and} \quad \psi_\epsilon\,(0) = 0\,.$$

*Step 4.2*. Let

$$\chi_\epsilon\,(x) = \frac{x}{\epsilon}$$

and define

$$\varphi = \epsilon\,\psi_\epsilon \circ \chi_\epsilon\,.$$

Define $U = \epsilon V$. It is easily seen that $\varphi \in C^{r+1,q}(\overline{U}; \varphi\,(U))$

$$\varphi^*\,(\omega_m) = \omega \text{ in } U \quad \text{and} \quad \varphi\,(0) = 0.$$

Note in particular that

$$\|\nabla\varphi - I\|_{C^0(\overline{U})} = \|\nabla\psi_\epsilon - I\|_{C^0(\overline{V})} \leq 1/\,(2n)$$

and therefore $\det\nabla\varphi > 0$ in $\overline{U}$. Hence, restricting $U$, if necessary, we can assume that $\varphi \in \mathrm{Diff}^{r+1,q}(U; \varphi\,(U))$. This concludes the proof of the theorem. ■

## 6.3 *Darboux Theorem for Lower Rank Forms*

We next discuss the case of forms of lower rank. This is also well known in the literature. However our theorem (proved in [2] by Bandyopadhyay-Dacorogna-Kneuss) provides, as the previous theorem, one class higher degree of regularity than the other results. Indeed in all other theorems it is proved that if $\omega \in C^{r,q}$, then, at best, $\varphi \in C^{r-1,q}$. It may appear that the theorem below is still not optimal, since it only shows that $\varphi \in C^{r,q}$ when $\omega \in C^{r,q}$. But since there are some missing variables, it is probably the best possible regularity (in [17], we can allow in the theorem below $q = 0, 1$).

**Theorem 46** *Let $n \geq 3,\ r, m \geq 1$ be integers and $0 < q < 1$. Let $x_0 \in \mathbb{R}^n$ and $\omega_m$ be the standard symplectic form with* $\operatorname{rank}[\omega_m] = 2m < n$, *namely*

$$\omega_m = \sum_{i=1}^{m} dx^{2i-1} \wedge dx^{2i}.$$

*Let $\omega$ be a $C^{r,q}$ closed 2-form such that*

$$\operatorname{rank}[\omega] = 2m \quad \text{in a neighborhood of } x_0\,.$$

*Then there exist a neighborhood $U$ of $x_0$ and $\varphi \in \operatorname{Diff}^{r,q}(U; \varphi(U))$ such that*

$$\varphi^*(\omega_m) = \omega \text{ in } U \quad \text{and} \quad \varphi(x_0) = x_0\,.$$

*Proof* *Step 1.* Without loss of generality, we can assume $x_0 = 0$. We first find, appealing to Theorem 47 below, a neighborhood $V \subset \mathbb{R}^n$ of 0 and $\psi \in \operatorname{Diff}^{r,q}(V; \psi(V))$ with $\psi(0) = 0$ and

$$\psi^*(\omega)(x_1, \cdots, x_n) = \widetilde{\omega}(x_1, \cdots, x_{2m}) = \sum_{1 \leq i < j \leq 2m} \widetilde{\omega}_{ij}(x_1, \cdots, x_{2m})\, dx^i \wedge dx^j.$$

Therefore $\psi^*(\omega) = \widetilde{\omega} \in C^{r-1,q}$ in a neighborhood of 0 in $\mathbb{R}^{2m}$ and $\operatorname{rank}[\widetilde{\omega}] = 2m$ in a neighborhood of 0.
*Step 2.* We then apply Theorem 44 to $\widetilde{\omega}$ and find a neighborhood $W \subset \mathbb{R}^{2m}$ of 0 and $\chi \in \operatorname{Diff}^{r,q}(W; \chi(W))$, with $\chi(0) = 0$, such that

$$\chi^*(\omega_m) = \widetilde{\omega} \quad \text{in } W.$$

We set

$$\widetilde{\chi}(x) = \widetilde{\chi}(x_1, \cdots, x_{2m}, x_{2m+1}, \cdots, x_n) = (\chi(x_1, \cdots, x_{2m}), x_{2m+1}, \cdots, x_n)$$

We then choose $V$ smaller, if necessary, so that

$$V \subset W \times \mathbb{R}^{n-2m}.$$

We finally have that $U = \psi(V)$ and $\varphi = \widetilde{\chi} \circ \psi^{-1}$ have all the desired properties. ■

In the above theorem we used a very useful result (cf. Theorem 4.5 in [16]).

**Theorem 47 (Reduction of Dimension)** *Let $r \geq 1,\ 1 \leq k \leq l \leq n-1$ be integers and $x_0 \in \mathbb{R}^n$. Let $g$ be a $C^r$ $k$-form verifying*

$$dg = 0 \quad \text{and} \quad \operatorname{rank}[g] = l \quad \text{in a neighborhood of } x_0\,.$$

*Then, there exist a neighborhood $U$ of $x_0$ and $\varphi \in \mathrm{Diff}^r(U;\varphi(U))$ with $\varphi(x_0) = x_0$ and such that, for every $x = (x_1, \cdots, x_n) \in U$,*

$$\begin{aligned}\varphi^*(g)(x_1,\cdots,x_n) &= f(x_1,\cdots,x_l)\\ &= \sum_{1\le i_1<\cdots<i_k\le l} f_{i_1\cdots i_k}(x_1,\cdots,x_l)\, dx^{i_1}\wedge\cdots\wedge dx^{i_k}.\end{aligned}$$

*Thus $f = \varphi^*(g)$ can be seen as a k-form with maximal rank (i.e.* $\mathrm{rank}\,[f] = l$*) on* $\mathbb{R}^l$.

*Remark 48*

(i) The result is still valid in Hölder spaces.
(ii) Note that $\varphi^*(g)$ is only in $C^{r-1}$ although $g \in C^r$.

## 6.4 A Global Result

### 6.4.1 The Main Result

We now state our main theorem. It has been obtained under slightly more restrictive hypotheses by Bandyopadhyay-Dacorogna [1]; as stated it is due to Dacorogna-Kneuss [27]. We will only outline the main steps of the proof (for complete details see [16]).

**Theorem 49** *Let $n > 2$ be even and $\Omega \subset \mathbb{R}^n$ be a bounded open smooth set with exterior unit normal $\nu$. Let $0 < q < 1$ and $r \ge 1$ be an integer. Let $f, g \in C^{r,q}(\overline{\Omega};\Lambda^2)$ satisfying $df = dg = 0$ in $\Omega$,*

$$\nu\wedge f, \nu\wedge g \in C^{r+1,q}\left(\partial\Omega;\Lambda^3\right) \quad \text{and} \quad \nu\wedge f = \nu\wedge g \text{ on } \partial\Omega$$

$$\int_\Omega \langle f;\psi\rangle\, dx = \int_\Omega \langle g;\psi\rangle\, dx, \quad \text{for every } \psi \in \mathcal{H}_T(\Omega;\Lambda^2) \tag{35}$$

*and, for every $t \in [0,1]$,*

$$\mathrm{rank}\,[tg + (1-t)f] = n, \quad \text{in } \overline{\Omega}.$$

*Then there exists $\varphi \in \mathrm{Diff}^{r+1,q}\left(\overline{\Omega};\overline{\Omega}\right)$ such that*

$$\varphi^*(g) = f \text{ in } \Omega \quad \text{and} \quad \varphi = \mathrm{id} \text{ on } \partial\Omega.$$

*Remark 50*

(i) We can consider, in a similar way, a general homotopy $f_t$ with $f_0 = f$, $f_1 = g$,

$$df_t = 0, \quad \nu \wedge f_t = \nu \wedge f_0 \text{ on } \partial\Omega \quad \text{and} \quad \operatorname{rank}[f_t] = n \text{ in } \overline{\Omega}$$

$$\int_\Omega \langle f_t; \psi \rangle \, dx = \int_\Omega \langle f_0; \psi \rangle \, dx, \quad \text{for every } \psi \in \mathcal{H}_T(\Omega; \Lambda^2).$$

Note that the non-degeneracy condition $\operatorname{rank}[f_t] = n$ implies (identifying, as usual, volume forms with functions)

$$f^{n/2} \cdot g^{n/2} > 0 \text{ in } \overline{\Omega}.$$

(ii) The non-degeneracy condition

$$\operatorname{rank}[tg + (1-t)f] = n, \quad \text{for every } t \in [0,1]$$

is equivalent to the condition that the matrix $(\overline{g})\left(\overline{f}\right)^{-1}$ has no negative eigenvalues.

(iii) If $\Omega$ is contractible, then $\mathcal{H}_T(\Omega; \Lambda^2) = \{0\}$ and therefore (35) is automatically satisfied.

(iv) Note that the extra regularity on $f$ and $g$ holds only on the boundary and only for their tangential parts. More precisely recall that, for $x \in \partial\Omega$, we denote by $\nu = \nu(x)$ the exterior unit normal to $\Omega$. By

$$\nu \wedge f \in C^{r+1,q}(\partial\Omega; \Lambda^3)$$

we mean that the tangential part of $f$ is in $C^{r+1,q}$, namely the 3-form $F$ defined by

$$F(x) = \nu(x) \wedge f(x)$$

is such that

$$F \in C^{r+1,q}(\partial\Omega; \Lambda^3).$$

(v) If the support of $(f - g)$ is compact in $\Omega$, then one can find a diffeomorphism $\varphi$ such that the support of $(\varphi - \mathrm{id})$ is also compact in $\Omega$, see [37].

The proof of Theorem 49 follows the same pattern as that of the case $k = n$ (cf. Theorem 35). In the first step (cf. Theorem 51) we establish the result, through the flow method, but without the optimal regularity. We next (cf. Theorem 52) prove the result under a smallness assumption. We finally combine the two intermediate

theorems to get the claim. The technical details are however more delicate than those for the case $k = n$ and we prove only the first step and refer for details to [16].

### 6.4.2 The Flow Method

We now state and prove a weaker version, from the point of view of regularity, of Theorem 49. It has, however, the advantage of having a simple proof. It has been obtained by Bandyopadhyay-Dacorogna [1].

**Theorem 51** *Let $n > 2$ be even and $\Omega \subset \mathbb{R}^n$ be a bounded smooth set with exterior unit normal $\nu$. Let $r \geq 1$ be an integer, $0 < q < 1$ and let $f, g \in C^{r,q}(\overline{\Omega}; \Lambda^2)$ satisfy*

$$df = dg = 0 \text{ in } \Omega, \quad \nu \wedge f = \nu \wedge g \text{ on } \partial\Omega$$

$$\int_\Omega \langle f; \psi \rangle \, dx = \int_\Omega \langle g; \psi \rangle \, dx, \quad \text{for every } \psi \in \mathcal{H}_T(\Omega; \Lambda^2)$$

$$\operatorname{rank} [tg + (1-t)f] = n, \quad \text{in } \overline{\Omega} \text{ and for every } t \in [0,1].$$

*Then there exists $\varphi \in \operatorname{Diff}^{r,q}\left(\overline{\Omega}; \overline{\Omega}\right)$ such that*

$$\varphi^*(g) = f \text{ in } \Omega \quad \text{and} \quad \varphi = \mathrm{id} \text{ on } \partial\Omega.$$

*Proof* We first find $w \in C^{r+1,q}(\overline{\Omega}; \Lambda^1)$ (cf. Theorem 26) such that

$$\begin{cases} dw = f - g \text{ in } \Omega \\ w = 0 \quad \text{on } \partial\Omega. \end{cases}$$

Since $\operatorname{rank} [tg + (1-t)f] = n$, we can find $u_t \in C^{r,q}(\overline{\Omega}; \mathbb{R}^n)$ so that

$$u_t \lrcorner [tg + (1-t)f] = w \quad \Leftrightarrow \quad u_t = \left[ t\overline{g} + (1-t)\overline{f} \right]^{-1} w.$$

We then apply Theorem 34 to $u_t$ and $f_t = tg + (1-t)f$ to find $\varphi$ satisfying

$$\varphi^*(g) = f \text{ in } \Omega \quad \text{and} \quad \varphi = \mathrm{id} \text{ on } \partial\Omega.$$

The proof is therefore complete. ■

### 6.4.3 The Fixed Point Method

As for the case of volume forms $k = n$, the second step in the proof is to obtain the theorem under a smallness condition on $g$ and $f$. The case $k = 2$ is however more delicate and we refer to [16] for details of the proof of next theorem.

**Theorem 52** *Let $n > 2$ be even and $\Omega \subset \mathbb{R}^n$ be a bounded open smooth set. Let $r \geq 1$ be an integer and $0 < s \leq q < 1$. Let $g \in C^{r+1,q}(\overline{\Omega}; \Lambda^2)$ and $f \in C^{r,q}(\overline{\Omega}; \Lambda^2)$ be such that*

$$df = dg = 0 \text{ in } \Omega, \quad \nu \wedge f = \nu \wedge g \text{ on } \partial\Omega$$

$$\int_\Omega \langle f; \psi \rangle \, dx = \int_\Omega \langle g; \psi \rangle \, dx \quad \text{for every } \psi \in \mathcal{H}_T(\Omega; \Lambda^2)$$

$$\operatorname{rank} [g] = n \quad \text{in } \overline{\Omega}.$$

*Let $c > 0$ be such that*

$$\|g\|_{C^0}, \left\| \frac{1}{[g]^{n/2}} \right\|_{C^0} \leq c$$

*and define*

$$\theta(g) = \frac{1}{\|g\|^2_{C^{1,s}}} \min \left\{ \|g\|_{C^{1,s}}, \frac{1}{\|g\|_{C^{2,s}}}, \frac{1}{\|g\|_{C^{r+1,q}}} \right\}.$$

*There exists $C = C(c, r, q, s, \Omega) > 0$ such that if*

$$\|f - g\|_{C^{0,s}} \leq C\theta(g) \quad \text{and} \quad \|f - g\|_{C^{0,s}} \leq C \frac{\|f - g\|_{C^{r,q}}}{\|g\|_{C^{1,s}} \|g\|_{C^{r+1,q}}} \tag{36}$$

*then there exists $\varphi \in \operatorname{Diff}^{r+1,q}(\overline{\Omega}; \overline{\Omega})$ verifying*

$$\varphi^*(g) = f \text{ in } \Omega \quad \text{and} \quad \varphi = \mathrm{id} \text{ on } \partial\Omega. \tag{37}$$

*Furthermore there exists $\widetilde{C} = \widetilde{C}(c, r, q, s, \Omega) > 0$ such that*

$$\|\varphi - \mathrm{id}\|_{C^{r+1,q}} \leq \widetilde{C} \|g\|_{C^{r+1,q}} \|f - g\|_{C^{r,q}}.$$

*Remark 53* Note that since $g \in C^{r+1,q}(\overline{\Omega}; \Lambda^2)$ and $\nu \wedge f = \nu \wedge g$ on $\partial\Omega$, then $\nu \wedge f \in C^{r+1,q}(\partial\Omega; \Lambda^3)$.

# 7 The Other Cases and Necessary Conditions

## 7.1 The Cases $k = 0$ and $k = 1$

We start with the case $k = 0$ which is particularly elementary. The first theorem is of a local nature.

**Theorem 54** *Let $r \geq 1$ be an integer, $x_0 \in \mathbb{R}^n$ and $f, g \in C^r$ in a neighborhood of $x_0$ and such that $f(x_0) = g(x_0)$,*

$$\nabla f(x_0) \neq 0 \quad \text{and} \quad \nabla g(x_0) \neq 0.$$

*Then there exists a neighborhood $U$ of $x_0$ and $\varphi \in \mathrm{Diff}^r(U; \varphi(U))$ such that*

$$\varphi^*(g) = f \text{ in } U \quad \text{and} \quad \varphi(x_0) = x_0.$$

We now have the following global result.

**Theorem 55** *Let $\Omega \subset \mathbb{R}^n$ be a bounded open Lipschitz set. Let $r \geq 1$ be an integer and $f, g \in C^r(\overline{\Omega})$ with $f = g$ on $\partial\Omega$ and*

$$\frac{\partial f}{\partial x_i} \cdot \frac{\partial g}{\partial x_i} > 0 \quad \text{in } \overline{\Omega}$$

*for a certain $1 \leq i \leq n$. Then there exists a diffeomorphism $\varphi \in \mathrm{Diff}^r(\overline{\Omega}; \overline{\Omega})$ satisfying*

$$\varphi^*(g) = f \text{ in } \Omega \quad \text{and} \quad \varphi = \mathrm{id} \text{ on } \partial\Omega.$$

Both results extend in a straightforward way to the case of closed 1-forms.

We now give a theorem for non-closed 1-forms. It can be considered as the 1-form version of Darboux theorem. It is easy to see that it leads to Darboux theorem for closed 2-forms.

**Theorem 56** *Let $2 \leq 2m \leq n$ be integers, $x_0 \in \mathbb{R}^n$ and $\omega$ be a $C^\infty$ 1-form such that $\omega(x_0) \neq 0$ and*

$$\mathrm{rank}[d\omega] = 2m \quad \text{in a neighborhood of } x_0,$$

*Then there exist an open set $U$ and $\varphi \in \mathrm{Diff}^\infty(U; \varphi(U))$ such that $\varphi(U)$ is a neighborhood of $x_0$ and*

$$\varphi^*(\omega) = \begin{cases} \Omega_m(x) & \text{if } \omega \wedge (d\omega)^m = 0 \text{ in a neighborhood of } x_0 \\ \Omega_m(x) + dx^{2m+1} & \text{if } \omega \wedge (d\omega)^m \neq 0 \text{ in a neighborhood of } x_0 \end{cases}$$

*where*

$$\Omega_m(x) = \sum_{i=1}^{m} x_{2i-1} dx^{2i}.$$

*Remark 57*

(i) We recall that

$$(d\omega)^m = \underbrace{d\omega \wedge \cdots \wedge d\omega}_{m \text{ times}} .$$

(ii) Note that if $n = 2m$, then $\omega \wedge (d\omega)^m \equiv 0$.
(iii) Without further hypothesis it is, in general, impossible to guarantee that $\varphi(x_0) = x_0$.
(iv) Of particular interest is the case of *contact forms* where $n = 2m + 1$ and $\omega \wedge (d\omega)^m \neq 0$, see [29] for some improvements on the result.

## 7.2 The Case $k = n - 1$

We have the following result (cf. Bandyopadhyay-Dacorogna-Kneuss [2]).

**Theorem 58** *Let $x_0 \in \mathbb{R}^n$ and $f$ be a $(n-1)$-form such that $f \in C^\infty$ in a neighborhood of $x_0$ and $f(x_0) \neq 0$. Then there exist an open set $U$ and*

$$\varphi \in \mathrm{Diff}^\infty(U; \varphi(U))$$

*such that $\varphi(U)$ is a neighborhood of $x_0$ and*

$$f = \begin{cases} \nabla\varphi^1 \wedge \cdots \wedge \nabla\varphi^{n-1} & \text{if } df = 0 \text{ in a neighborhood of } x_0 \\ \varphi^n \left(\nabla\varphi^1 \wedge \cdots \wedge \nabla\varphi^{n-1}\right) & \text{if } df \neq 0 \text{ in a neighborhood of } x_0 . \end{cases}$$

*Remark 59*

(i) The statement can be rewritten as follows

$$f = \begin{cases} \varphi^* \left(dx^1 \wedge \cdots \wedge dx^{n-1}\right) & \text{if } df = 0 \\ \varphi^* \left(x_n\, dx^1 \wedge \cdots \wedge dx^{n-1}\right) & \text{if } df \neq 0. \end{cases}$$

The present theorem, when $df = 0$, is a consequence of a theorem which is valid for $k$-forms of rank $k$.
(ii) With our usual abuse of notations, identifying a $(n-1)$-form with a vector field and observing that the $d$ operator can then be essentially identified with the divergence operator, we can rewrite the theorem as follows (compare with Barbarosie [4]). For any $C^\infty$ vector field $f$ such that $f(x_0) \neq 0$, there exist an open set $U$ and

$$\varphi \in \mathrm{Diff}^\infty(U; \varphi(U))$$

such that $\varphi(U)$ is a neighborhood of $x_0$ and

$$f = \begin{cases} *\left(\nabla\varphi^1 \wedge \cdots \wedge \nabla\varphi^{n-1}\right) & \text{if } \operatorname{div} f = 0 \\ *\left(\varphi^n\left(\nabla\varphi^1 \wedge \cdots \wedge \nabla\varphi^{n-1}\right)\right) & \text{if } \operatorname{div} f \neq 0 \end{cases}$$

where $*$ denotes the Hodge $*$ operator.

(iii) When $df = 0$ we can also ensure that $\varphi(x_0) = x_0$; but not, in general, when $df \neq 0$.

## 7.3 The Case $3 \leq k \leq n-2$

We now turn to the case $3 \leq k \leq n-2$ which is, as previously said, much more difficult. This is so already at the algebraic level, since there are no known canonical forms. In particular the rank is not the only invariant (cf. Remark 4). And even when the algebraic setting is simple, the analytical situation is more complicated than in the cases $k = 0, 1, 2, n-1, n$. We give three examples; the first two are purely algebraic and the third one is more analytic.

*Example 60 (Example 2.36 in [16])* When $k = 3$, the forms

$$f = e^1 \wedge e^2 \wedge e^3 + e^4 \wedge e^5 \wedge e^6$$
$$g = e^1 \wedge e^2 \wedge e^3 + e^1 \wedge e^4 \wedge e^5 + e^2 \wedge e^4 \wedge e^6 + e^3 \wedge e^5 \wedge e^6$$

have both rank $= 6$, but there is no $A \in \mathrm{GL}(6)$ so that

$$A^*(g) = f.$$

*Example 61 (Example 2.35 in [16])* Similarly and more strikingly, when $k = 4$ and

$$f = e^1 \wedge e^2 \wedge e^3 \wedge e^4 + e^1 \wedge e^2 \wedge e^5 \wedge e^6 + e^3 \wedge e^4 \wedge e^5 \wedge e^6$$

there is no $A \in \mathrm{GL}(6)$ such that

$$A^*(f) = -f$$

although

$$\operatorname{rank}[f] = \operatorname{rank}[-f] = 6.$$

*Example 62 (Proposition 15.14 in [16])* Although every constant 3-form of rank $=$ 5 is a linear pullback of

$$f = dx^1 \wedge dx^2 \wedge dx^5 + dx^3 \wedge dx^4 \wedge dx^5$$

we have the following result. There exists $g \in C^\infty(\mathbb{R}^5; \Lambda^3)$ with

$$dg = 0 \quad \text{and} \quad \text{rank}[g] = 5 \text{ in } \mathbb{R}^5$$

namely

$$g = -dx^1 \wedge dx^2 \wedge dx^5 + \left((x_3)^2 + 1\right) dx^1 \wedge dx^3 \wedge dx^4 + \left((x_3)^4 + 1\right) dx^2 \wedge dx^3 \wedge dx^4$$

which cannot be pulled back locally by a diffeomorphism to $f$.

The only cases that we are able to study are those that are combinations of 1 and 2-forms that we can handle separately. For 1-forms, we easily obtain local as well as global results. We now give a simple theorem (more general statements can be found in [16]) that deals with 3-forms obtained by product of a 1-form and a 2-form (in the same spirit, we can deal with some $k$-forms that are product of 1 and 2-forms).

**Theorem 63** *Let $n = 2m \geq 4$ be integers, $x_0 \in \mathbb{R}^n$ and $f$ be a $C^\infty$ symplectic (i.e. closed and with* rank$[f] = n$*) 2-form and $a$ be a non-zero closed $C^\infty$ 1-form. Then there exist a neighborhood $U$ of $x_0$ and $\varphi \in \text{Diff}^\infty(U; \varphi(U))$ such that $\varphi(x_0) = x_0$ and*

$$\varphi^*(\omega_m) = f \quad \text{and} \quad \varphi^*(dx^n) = a, \quad \text{in } U$$

*where*

$$\omega_m = \sum_{i=1}^{m} dx^{2i-1} \wedge dx^{2i}.$$

*In particular if*

$$G = \left[\sum_{i=1}^{m-1} dx^{2i-1} \wedge dx^{2i}\right] \wedge dx^n = \omega_m \wedge dx^n$$

*then*

$$\varphi^*(G) = f \wedge a, \quad \text{in } U.$$

### 7.4 Necessary Conditions

We point out the following necessary conditions.

**Theorem 64** *Let $\Omega \subset \mathbb{R}^n$ be a bounded open smooth set and $\varphi \in \mathrm{Diff}^1\left(\overline{\Omega}; \varphi\left(\overline{\Omega}\right)\right)$. Let $1 \leq k \leq n$, $f \in C^1\left(\overline{\Omega}; \Lambda^k\right)$ and $g \in C^1\left(\varphi\left(\overline{\Omega}\right); \Lambda^k\right)$ be such that*

$$\varphi^*(g) = f \quad \text{in } \Omega.$$

*(i) For every $x \in \Omega$, then*

$$\operatorname{rank}[g(\varphi(x))] = \operatorname{rank}[f(x)] \quad \text{and} \quad \operatorname{rank}[dg(\varphi(x))] = \operatorname{rank}[df(x)].$$

*In particular*

$$dg = 0 \text{ in } \varphi(\Omega) \quad \Leftrightarrow \quad df = 0 \text{ in } \Omega.$$

*(ii) If $\varphi(x) = x$ for $x \in \partial\Omega$, then*

$$\nu \wedge f = \nu \wedge g \quad \text{on } \partial\Omega$$

*where $\nu$ is the exterior unit normal to $\Omega$.*

## 8 Selection Principle Via Ellipticity

We now very briefly discuss the question of selecting one specific solution of the pullback equation. The problem for $k = n$ has been intensively studied in the context of optimal mass transportation, but very little for general forms (see [25]). We here discuss (following [26]) the choice that can be made by considering appropriate elliptic systems in the cases $k = n$ and $k = 2$.

### 8.1 The Case $k = n$

When $k = n$ a natural choice, which is the one obtained via optimal mass transportation, is $\varphi = \nabla\Phi$ transforming the pullback equation $\varphi^*(g) = f$ to the celebrated *Monge-Ampère equation*

$$g(\nabla\Phi(x)) \det \nabla^2\Phi(x) = f(x).$$

This equation, when coupled with appropriate boundary conditions, leads to uniqueness as well as regularity. In order to understand better the case $k = 2$, we rephrase the earlier considerations in the following way. For the sake of simplicity we assume that $g \equiv 1$, transforming the pullback equation to the single equation $\det \nabla\varphi = f$. We then couple this equation to the $[n(n-1)/2]$ equations $\operatorname{curl} \varphi = 0$. It turns out that this system, when restricted to maps such that $\nabla\varphi + (\nabla\varphi)^t > 0$, is elliptic (see [26] for a general definition of ellipticity).

## 8.2 The Case $k = 2$

It was essentially equivalent, in the case $k = n$, to choose $\varphi = \nabla\Phi$ or to add the equations $\operatorname{curl} \varphi = 0$ in order to find an appropriate elliptic system. This is not any more true when $k = 2$.

### 8.2.1 The Gradient Case

It has been proved in [17] that for constant forms the choice $\varphi = \nabla\Phi$ (here requiring that $\varphi(x) = \nabla\Phi(x) = Ax$ amounts to say that $A$ is symmetric) is appropriate.

**Theorem 65** *Let $n$ be even and $g, f \in \Lambda^2(\mathbb{R}^n)$ be such that* $\operatorname{rank}[g] = \operatorname{rank}[f] = n$. *Then there exists $A \in \mathrm{GL}(n)$ such that*

$$A^*(g) = f \quad \text{and} \quad A^t = A.$$

However as soon as the forms are non-constant, the above theorem is, in general, false.

**Proposition 66** *Let $f \in C^\infty(\mathbb{R}^4; \Lambda^2)$ be defined by*

$$f = (1 + x_3)\, dx^1 \wedge dx^2 + x_2\, dx^1 \wedge dx^3 + 2\, dx^3 \wedge dx^4.$$

*Then there exists no $\Phi \in C^3(\mathbb{R}^4)$ such that near 0*

$$(\nabla\Phi)^* \left(dx^1 \wedge dx^2 + dx^3 \wedge dx^4\right) = f$$

*although there exists a local $C^\infty$ diffeomorphism $\varphi$ such that*

$$\varphi^* \left(dx^1 \wedge dx^2 + dx^3 \wedge dx^4\right) = f.$$

### 8.2.2 The Ellipticity Criterion

A better choice is as follows (cf. [26] for details and proofs). We couple the pullback equation $\varphi^*(g) = f$, which is a first order system of $[n(n-1)/2]$ equations, to the single equation $\langle d\varphi; g\rangle = 0$, i.e.

$$\sum_{1\le i<j\le n} \left(\varphi^j_{x_i} - \varphi^i_{x_j}\right) g^{ij} = 0.$$

It turns out that the new system, when restricted to maps such that $\nabla\varphi + (\nabla\varphi)^t > 0$, is elliptic. In fact one can prove the following theorem. We let $r \ge 0$ and $n = 2m$ be integers, $0 < q < 1$ and $\omega_m$ be the standard symplectic form, namely

$$\omega_m = \sum_{i=1}^{m} dx^{2i-1} \wedge dx^{2i}.$$

We also let $\Omega \subset \mathbb{R}^n$ be a bounded contractible open smooth set with exterior unit normal $\nu$.

**Theorem 67** *Let $f \in C^{r,q}\left(\overline{\Omega}; \Lambda^2\right)$ be closed. Then there exist $\epsilon, \gamma, c > 0$ depending only on $(r, q, \Omega)$ such that if*

$$\|f - \omega_m\|_{C^{0,q/2}} \le \epsilon,$$

*then there exists a unique $\varphi \in \mathrm{Diff}^{r+1,q}\left(\overline{\Omega}; \varphi\left(\overline{\Omega}\right)\right)$ satisfying*

$$\begin{cases} \varphi^*(\omega_m) = f \quad \text{and} \quad d\varphi \lrcorner \omega_m = \langle d\varphi; \omega_m\rangle = 0 \text{ in } \Omega \\ \qquad\qquad \nu \lrcorner ((\varphi - \mathrm{id}) \lrcorner \omega_m) = 0 \qquad\qquad \text{on } \partial\Omega \end{cases} \tag{38}$$

*and such that*

$$\begin{cases} \|\varphi - \mathrm{id}\|_{C^{r+1,q}} \le c \|f - \omega_m\|_{C^{r,q}} \\ |\langle [\nabla\varphi(x)]\xi; \xi\rangle| \ge \gamma |\xi|^2, \quad \forall\, \xi \in \mathbb{R}^n \text{ and } \forall\, x \in \overline{\Omega}. \end{cases} \tag{39}$$

*Furthermore the system (38) is elliptic when restricted to maps satisfying the second inequality in (39).*

*Remark 68* The comparison between the volume form case and the symplectic case becomes now striking. In the first one we have

$$\begin{cases} \det \nabla\varphi = f \ \text{1 equation} \\ \mathrm{curl}\, \varphi = 0 \ \ [n(n-1)/2] \ \text{equations} \end{cases}$$

while in the second one we get

$$\begin{cases} \varphi^*(\omega_m) = f & [n(n-1)/2] \text{ equations} \\ d\varphi \lrcorner \omega_m = \sum_{j=1}^{m} \left( \varphi^{2j}_{x_{2j-1}} - \varphi^{2j-1}_{x_{2j}} \right) = 0 & 1 \text{ equation.} \end{cases}$$

The above theorem can be written as a second order system, which is the counterpart of Monge-Ampère equation when $n = 2$ and therefore $f$ is a volume form.

**Corollary 69 (Second Order Darboux Theorem)** *Let $f \in C^{r,q}\left(\overline{\Omega}; \Lambda^2\right)$ be closed. Then there exists $\epsilon = \epsilon(r, q, \Omega)$ such that if*

$$\|f - \omega_m\|_{C^{0,q/2}} \leq \epsilon,$$

*then there exists a unique $\Phi \in C^{r+2,q}\left(\overline{\Omega}; \Lambda^2\right)$ satisfying the elliptic system*

$$\begin{cases} (\delta\Phi \lrcorner \omega_m)^*(\omega_m) = f \quad \text{and} \quad d\Phi = 0 & \text{in } \Omega \\ \nabla(\delta\Phi \lrcorner \omega_m) + (\nabla(\delta\Phi \lrcorner \omega_m))^t > 0 & \\ \nu \lrcorner \Phi = -\nu \lrcorner H & \text{on } \partial\Omega. \end{cases}$$

*Here $H$ is such that $\delta H = \mathrm{id} \lrcorner \omega_m$.*

*Remark 70*

(i) Writing

$$\Phi = \sum_{i<j} \Phi^{ij} dx^i \wedge dx^j$$

and similarly for $f$ we have that $(\delta\Phi \lrcorner \omega_m)^*(\omega_m) = f$ reads as (recalling that $\Phi^{ij} = -\Phi^{ji}$)

$$\sum_{l=1}^{m} \sum_{s,t=1}^{2m} \left[ \Phi^{s(2l-1)}_{x_s x_i} \Phi^{t(2l)}_{x_t x_j} - \Phi^{s(2l-1)}_{x_s x_j} \Phi^{t(2l)}_{x_t x_i} \right] = f_{ij}, \quad 1 \leq i < j \leq n \tag{40}$$

while $d\Phi = 0$ means that

$$\Phi^{ij}_{x_k} - \Phi^{ik}_{x_j} + \Phi^{jk}_{x_i} = 0, \quad 1 \leq i < j < k \leq n.$$

Note that when $n = 2$ the equation $d\Phi = 0$ is trivially fulfilled, while (40) is exactly Monge-Ampère equation.

(ii) The form $H$ can be taken, for example, as

$$H = \sum_{i=1}^{m} \frac{(x_{2i-1})^2 + (x_{2i})^2}{2} dx^{2i-1} \wedge dx^{2i}.$$

# 9 Hölder Spaces

We now present fine properties of Hölder continuous functions. Most of the results are "standard", but they are scattered in the literature. There does not exist such a huge literature as the one for Sobolev spaces. The most complete reference for this chapter is [16]

## *9.1 Definition and Extension of Hölder Functions*

We give here the definition of Hölder continuous functions.

**Definition 71** Let $\Omega \subset \mathbb{R}^n$ be a bounded open set, $f : \overline{\Omega} \to \mathbb{R}$ and $0 < \alpha \leq 1$. Let

$$[f]_{C^{0,\alpha}(\overline{\Omega})} = \sup_{\substack{x,y \in \overline{\Omega} \\ x \neq y}} \left\{ \frac{\|f(x) - f(y)|}{|x-y|^{\alpha}} \right\}.$$

(i) The set $C^{0,\alpha}\left(\overline{\Omega}\right)$ is the set of $f \in C^0\left(\overline{\Omega}\right)$ so that

$$\|f\|_{C^{0,\alpha}(\overline{\Omega})} = \|f\|_{C^0(\overline{\Omega})} + [f]_{C^{0,\alpha}(\overline{\Omega})} < \infty$$

where

$$\|f\|_{C^0(\overline{\Omega})} = \sup_{x \in \Omega} \{|f(x)|\}.$$

If there is no ambiguity we drop the dependence on the set $\overline{\Omega}$ and write simply

$$\|f\|_{C^{0,\alpha}} = \|f\|_{C^0} + [f]_{C^{0,\alpha}} .$$

(ii) If $r \geq 1$ is an integer, then the set $C^{r,\alpha}\left(\overline{\Omega}\right)$ is the set of functions $f \in C^r\left(\overline{\Omega}\right)$ so that

$$[\nabla^r f]_{C^{0,\alpha}(\overline{\Omega})} < \infty.$$

We equip $C^{r,\alpha}\left(\overline{\Omega}\right)$ with the following norm

$$\|f\|_{C^{r,\alpha}(\overline{\Omega})} = \|f\|_{C^r(\overline{\Omega})} + [\nabla^r f]_{C^{0,\alpha}(\overline{\Omega})}$$

where

$$\|f\|_{C^r(\overline{\Omega})} = \sum_{m=0}^{r} \|\nabla^m f\|_{C^0(\overline{\Omega})} .$$

*Remark 72*

(i) $C^{r,\alpha}\left(\overline{\Omega}\right)$ with its norm $\|\cdot\|_{C^{r,\alpha}}$ is a Banach space.
(ii) If $\alpha = 0$, we set

$$\|f\|_{C^{r,0}} = \|f\|_{C^r} .$$

(iii) If we assume that the domain is Lipschitz, then the following norms

$$\|f\|_{C^{r,\alpha}} = \sum_{m=0}^{r} \|\nabla^m f\|_{C^{0,\alpha}}$$

and

$$\|f\|_{C^{r,\alpha}} = \begin{cases} \|f\|_{C^0} + [\nabla^r f]_{C^{0,\alpha}} \text{ if } 0 < \alpha \leq 1 \\ \|f\|_{C^0} + \|\nabla^r f\|_{C^0} \text{ if } \alpha = 0. \end{cases}$$

are equivalent to the one defined above. We should, however, point out that these norms are, in general, not equivalent for very wild domains.
(iv) When $\alpha = 1$, we note that $C^{0,1}\left(\overline{\Omega}\right)$ is in fact the set of *Lipschitz continuous* functions.

The following extension result is due to Calderon and Stein.

**Theorem 73** *Let $\Omega \subset \mathbb{R}^n$ be a bounded open Lipschitz set. Then there exists a continuous linear extension operator*

$$E : C^{r,\alpha}\left(\overline{\Omega}\right) \to C_0^{r,\alpha}\left(\mathbb{R}^n\right)$$

*for any integer $r \geq 0$ and any $0 \leq \alpha \leq 1$. More precisely there exists a constant $C = C(r, \Omega) > 0$ such that, for every $f \in C^{r,\alpha}\left(\overline{\Omega}\right)$,*

$$E(f)|_{\overline{\Omega}} = f, \quad \text{supp}\,[E(f)] \text{ is compact,}$$
$$\|E(f)\|_{C^{r,\alpha}(\mathbb{R}^n)} \leq C \|f\|_{C^{r,\alpha}(\overline{\Omega})} .$$

*Remark 74* The extension is universal, in the sense that the same extension also leads to

$$\|E(f)\|_{C^{s,\beta}(\mathbb{R}^n)} \leq C \|f\|_{C^{s,\beta}(\overline{\Omega})}$$

for any integer $s$ and any $0 \leq \beta \leq 1$, with, of course, $C = C(s, \Omega)$ as far as $f \in C^{s,\beta}\left(\overline{\Omega}\right)$. The same extension is also valid for Sobolev spaces.

## 9.2 Product, Composition and Inverse

We start with a result on products of Hölder continuous functions.

**Theorem 75** *Let $\Omega \subset \mathbb{R}^n$ be a bounded open Lipschitz set, $r \geq 0$ an integer and $0 \leq \alpha \leq 1$. Then there exists a constant $C = C(r, \Omega) > 0$ such that*

$$\|fg\|_{C^{r,\alpha}} \leq C \left(\|f\|_{C^{r,\alpha}} \|g\|_{C^0} + \|f\|_{C^0} \|g\|_{C^{r,\alpha}}\right).$$

The next theorem has also been intensively used.

**Theorem 76** *Let $\Omega \subset \mathbb{R}^n$, $O \subset \mathbb{R}^m$ be bounded open Lipschitz sets, $r \geq 0$ an integer and $0 \leq \alpha \leq 1$. Let $g \in C^{r,\alpha}\left(\overline{O}\right)$ and $f \in C^{r,\alpha}\left(\overline{\Omega}; \overline{O}\right) \cap C^1\left(\overline{\Omega}; \overline{O}\right)$. Then*

$$\|g \circ f\|_{C^{0,\alpha}(\overline{\Omega})} \leq \|g\|_{C^{0,\alpha}(\overline{o})} \|f\|^{\alpha}_{C^1(\overline{\Omega})} + \|g\|_{C^0(\overline{o})}$$

*while if $r \geq 1$, there exists a constant $C = C(r, \Omega, O) > 0$ such that*

$$\|g \circ f\|_{C^{r,\alpha}(\overline{\Omega})} \leq C \left[\|g\|_{C^{r,\alpha}(\overline{o})} \|f\|^{r+\alpha}_{C^1(\overline{\Omega})} + \|g\|_{C^1(\overline{o})} \|f\|_{C^{r,\alpha}(\overline{\Omega})} + \|g\|_{C^0(\overline{o})}\right].$$

We easily deduce, from the previous results, an estimate on the inverse.

**Theorem 77** *Let $\Omega, O \subset \mathbb{R}^n$ be bounded open Lipschitz sets, $r \geq 1$ an integer and $0 \leq \alpha \leq 1$. Let $c > 0$. Let $f \in C^{r,\alpha}\left(\overline{\Omega}; \overline{O}\right)$ and $g \in C^{r,\alpha}\left(\overline{O}; \overline{\Omega}\right)$ be such that*

$$g \circ f = \mathrm{id} \quad \textit{and} \quad \|g\|_{C^1(\overline{o})}, \|f\|_{C^1(\overline{\Omega})} \leq c.$$

*Then there exists a constant $C = C(c, r, \Omega, O) > 0$ such that*

$$\|f\|_{C^{r,\alpha}(\overline{o})} \leq C \|g\|_{C^{r,\alpha}(\overline{\Omega})}.$$

## 9.3 Smoothing Operator

The next theorem is about smoothing $C^r$ or $C^{r,\alpha}$ functions. We should draw the attention that, in order to get the conclusions of the theorem, one proceeds, as usual, by convolution. However the kernel has to be chosen carefully.

**Theorem 78** *Let $\Omega \subset \mathbb{R}^n$ be a bounded open Lipschitz set. Let $s \geq r \geq t \geq 0$ be integers and $0 \leq \alpha, \beta, \gamma \leq 1$ be such that*

$$t + \gamma \leq r + \alpha \leq s + \beta.$$

*Let $f \in C^{r,\alpha}\left(\overline{\Omega}\right)$. Then, for every $0 < \epsilon \leq 1$, there exist a constant $C = C(s, \Omega) > 0$ and $f_\epsilon \in C^\infty\left(\overline{\Omega}\right)$ such that*

$$\|f_\epsilon\|_{C^{s,\beta}} \leq \frac{C}{\epsilon^{(s+\beta)-(r+\alpha)}} \|f\|_{C^{r,\alpha}}$$

$$\|f - f_\epsilon\|_{C^{t,\gamma}} \leq C\epsilon^{(r+\alpha)-(t+\gamma)} \|f\|_{C^{r,\alpha}} .$$

We also need to approximate closed forms in $C^{r,\alpha}\left(\overline{\Omega}; \Lambda^k\right)$ by smooth closed forms in a precise way.

**Theorem 79** *Let $\Omega \subset \mathbb{R}^n$ be a bounded open smooth set and $\nu$ be the exterior unit normal. Let $s \geq r \geq t \geq 0$ with $s \geq 1$ and $1 \leq k \leq n-1$ be integers. Let $0 < \alpha, \beta, \gamma < 1$ be such that*

$$t + \gamma \leq r + \alpha \leq s + \beta.$$

*Let $g \in C^{r,\alpha}\left(\overline{\Omega}; \Lambda^k\right)$ with*

$$dg = 0 \text{ in } \Omega \quad \text{and} \quad \nu \wedge g \in C^{s,\beta}\left(\partial\Omega; \Lambda^{k+1}\right).$$

*Then for every $\epsilon \in (0, 1]$, there exist $g_\epsilon \in C^\infty\left(\Omega; \Lambda^k\right) \cap C^{s,\beta}\left(\overline{\Omega}; \Lambda^k\right)$ and a constant $C = C(s, \alpha, \beta, \gamma, \Omega) > 0$ such that*

$$dg_\epsilon = 0 \text{ in } \Omega, \quad \nu \wedge g_\epsilon = \nu \wedge g \text{ on } \partial\Omega$$

$$\int_\Omega \langle g_\epsilon; \psi \rangle = \int_\Omega \langle g; \psi \rangle, \quad \text{for every } \psi \in \mathcal{H}_T\left(\Omega; \Lambda^k\right)$$

$$\|g_\epsilon\|_{C^{s,\beta}\left(\overline{\Omega}\right)} \leq \frac{C}{\epsilon^{(s+\beta)-(r+\alpha)}} \|g\|_{C^{r,\alpha}\left(\overline{\Omega}\right)} + C \|\nu \wedge g\|_{C^{s,\beta}(\partial\Omega)}$$

$$\|g_\epsilon - g\|_{C^{t,\gamma}\left(\overline{\Omega}\right)} \leq C\epsilon^{(r+\alpha)-(t+\gamma)} \|g\|_{C^{r,\alpha}\left(\overline{\Omega}\right)} .$$

*Remark 80* We recall that if $\Omega$ is contractible and since $1 \leq k \leq n-1$, then

$$\mathcal{H}_T\left(\Omega; \Lambda^k\right) = \{0\}.$$

## 10 An Abstract Fixed Point Theorem

The following theorem is particularly useful when dealing with non-linear problems, once good estimates are known for the linearized problem (see [5] for some applications). We give it under a general form (still a more sophisticated version can be found in [16]), because we have used it this way in Theorems 44 and 52. However, in many instances, Corollary 82 is amply sufficient. Our theorem will lean on the following hypotheses.

$(H_{XY})$ Let $X_1 \supset X_2$ be Banach spaces and $Y_1 \supset Y_2$ be normed spaces such that the following property holds: if

$$u_\nu \xrightarrow{X_1} u \quad \text{and} \quad \|u_\nu\|_{X_2} \leq r$$

then $u \in X_2$ and

$$\|u\|_{X_2} \leq r.$$

$(H_L)$ Let $L : X_2 \to Y_2$ be such that there exists a linear right inverse operator $L^{-1} : Y_2 \to X_2$ (namely $L L^{-1} = \mathrm{id}$ on $Y_2$). Moreover there exist $k > 0$ such that for every $f \in Y_2$

$$\|L^{-1} f\|_{X_i} \leq k \|f\|_{Y_i} \quad i = 1, 2.$$

$\left(H_Q\right)$ There exists $\rho > 0$ such that

$$Q : B_\rho = \{u \in X_2 : \|u\|_{X_1} \leq \rho\} \to Y_2$$

$Q(0) = 0$ and for every $u, v \in B_\rho$, the following two inequalities hold

$$\|Q(u) - Q(v)\|_{Y_1} \leq c(\|u\|_{X_1} + \|v\|_{X_1})\|u - v\|_{X_1} \tag{41}$$

$$\|Q(v)\|_{Y_2} \leq c\|v\|_{X_1}\|v\|_{X_2} \tag{42}$$

where $c > 0$ is a constant.

**Theorem 81 (Fixed Point Theorem)** *Let $X_1, X_2, Y_1, Y_2, L, Q$ satisfy the hypotheses $(H_{XY})$, $(H_L)$ and $(H_Q)$. Then, for every $f \in Y_2$ verifying*

$$\|f\|_{Y_1} \leq \min\left\{\frac{\rho}{2k}, \frac{1}{8k^2c}\right\} \tag{43}$$

*there exists $u \in B_\rho \subset X_2$ such that*

$$Lu = Q(u) + f \quad \text{and} \quad \|u\|_{X_i} \leq 2k\|f\|_{Y_i}\,,\ i = 1, 2. \tag{44}$$

We have as an immediate consequence of the theorem the following result.

**Corollary 82** *Let $X$ be a Banach space and $Y$ a normed space. Let $L : X \to Y$ be such that there exists a linear right inverse operator $L^{-1} : Y \to X$ (namely $LL^{-1} = \mathrm{id}$ on $Y$) and there exists $k > 0$ such that*

$$\|L^{-1}f\|_X \leq k\|f\|_Y\,.$$

*Let $\rho > 0$ and*

$$Q : B_\rho = \{u \in X : \|u\|_X \leq \rho\} \to Y$$

*with $Q(0) = 0$ and, for every $u, v \in B_\rho$,*

$$\|Q(u) - Q(v)\|_Y \leq c(\|u\|_X + \|v\|_X)\|u - v\|_X$$

*and where $c > 0$. If*

$$\|f\|_{Y_1} \leq \min\left\{\frac{\rho}{2k}, \frac{1}{8k^2c}\right\}$$

*then there exists $u \in B_\rho \subset X$ such that*

$$Lu = Q(u) + f \quad \text{and} \quad \|u\|_X \leq 2k\|f\|_Y\,.$$

We now turn to the proof of Theorem 81.

*Proof* We set

$$N(u) = Q(u) + f.$$

We next define

$$B = \{u \in X_2 : \|u\|_{X_i} \leq 2k\|f\|_{Y_i}\ i = 1, 2\}\,.$$

We endow $B$ with $\|\cdot\|_{X_1}$ norm; the property $(H_{XY})$ ensures that $B$ is closed. We now want to show that $L^{-1}N : B \to B$ is a contraction mapping (cf. Claims 1 and 2 below). Applying Banach fixed point theorem we will have indeed found a solution verifying (44), since $LL^{-1} = \mathrm{id}$.

*Claim 1* Let us first show that $L^{-1}N$ is a contraction on $B$. To show this, let $u, v \in B$ and use (41), (43) to get that

$$\begin{aligned}\|L^{-1}N(u) - L^{-1}N(v)\|_{X_1} &\le k\|N(u) - N(v)\|_{Y_1} = k\|Q(u) - Q(v)\|_{Y_1}\\ &\le kc(\|u\|_{X_1} + \|v\|_{X_1})\|u - v\|_{X_1}\\ &\le kc(2k\|f\|_{Y_1} + 2k\|f\|_{Y_1})\|u - v\|_{X_1}\\ &\le \frac{1}{2}\|u - v\|_{X_1}.\end{aligned}$$

*Claim 2* We next show $L^{-1}N : B \to B$ is well-defined. First, note that

$$\|L^{-1}N(0)\|_{X_1} \le k\|N(0)\|_{Y_1} = k\|f\|_{Y_1}.$$

Therefore, using Claim 1, we obtain

$$\begin{aligned}\|L^{-1}N(u)\|_{X_1} &\le \|L^{-1}N(u) - L^{-1}N(0)\|_{X_1} + \|L^{-1}N(0)\|_{X_1}\\ &\le \frac{1}{2}\|u\|_{X_1} + k\|f\|_{Y_1} \le 2k\|f\|_{Y_1}.\end{aligned}$$

It remains to show that

$$\|L^{-1}N(u)\|_{X_2} \le 2k\|f\|_{Y_2}.$$

Using (42), we have

$$\begin{aligned}\|L^{-1}N(u)\|_{X_2} &\le k\|N(u)\|_{Y_2} \le k\|Q(u)\|_{Y_2} + k\|f\|_{Y_2}\\ &\le kc\|u\|_{X_1}\|u\|_{X_2} + k\|f\|_{Y_2}\\ &\le k\left[c(2k\|f\|_{Y_1} \cdot 2k\|f\|_{Y_2}) + \|f\|_{Y_2}\right]\\ &\le k\left[4k^2c\|f\|_{Y_1} + 1\right]\|f\|_{Y_2}\end{aligned}$$

and hence, appealing once more to (43),

$$\|L^{-1}N(u)\|_{X_2} \le 2k\|f\|_{Y_2}.$$

This concludes the proof of Claim 2 and thus of the theorem. ■

## References


1. S. Bandyopadhyay, B. Dacorogna, On the pullback equation $\varphi^*(g) = f$. Ann. Inst. H. Poincaré Anal. Non Linéaire **26**, 1717–1741 (2009)
2. S. Bandyopadhyay, B. Dacorogna, O. Kneuss, The pullback equation for degenerate forms. Discrete Continuous Dyn. Syst. Ser. A **27**, 657–691 (2010)
3. A. Banyaga, Formes-volume sur les variétés à bord. Enseignement Math. **20**, 127–131 (1974)
4. C. Barbarosie, Representation of divergence-free vector fields. Q. Appl. Math. **69**, 309–316 (2011)
5. S. Basterrechea, B. Dacorogna, Existence of solutions for Jacobian and Hessian equations under smallness assumptions. Numer. Funct. Anal. Optim. **35**, 868–892 (2014)
6. M.E. Bogovski, Solution of the first boundary value problem for the equation of continuity of an incompressible medium. Sov. Math. Dokl. **20**, 1094–1098 (1979)
7. J. Bolik, H Weyl's boundary value problems for differential forms. Differ. Integr. Equ. **14**, 937–952 (2001)
8. W. Borchers, H. Sohr, On the equations rot $v = g$ and div $u = f$ with zero boundary conditions. Hokkaido Math. J. **19**, 67–87 (1990)
9. J. Bourgain, H. Brézis, Sur l'équation div $u = f$. C. R. Acad. Sci. Paris Sér. I Math. **334**, 973–976 (2002)
10. D. Burago, B. Kleiner, Separated nets in Euclidean space and Jacobian of biLipschitz maps. Geom. Funct. Anal. **8**, 273–282 (1998)
11. L. Caffarelli, Boundary regularity of maps with convex potentials II. Ann. Math. **144**, 453–496 (1996)
12. G. Carlier, B. Dacorogna, Résolution du problème de Dirichlet pour l'équation du Jacobien prescrit via l'équation de Monge-Ampère. C. R. Acad. Sci. Paris, Ser. I **350**, 371–374 (2012)
13. E.A. Coddington, N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill Book Company Inc., New York/Toronto/London, 1955)
14. G. Csato, B. Dacorogna, An identity involving exterior derivatives and applications to Gaffney inequality. Discrete Continuous Dyn. Syst. Ser. S **5**, 531–544 (2012)
15. G. Csato, B. Dacorogna, A Dirichlet problem involving the divergence operator. Ann. Inst. H. Poincaré Anal. Non Linéaire **33**, 829–848 (2016)
16. G. Csato, B. Dacorogna, O. Kneuss, *The Pullback Equation for Differential Forms*. PNLDE Series, vol. 83 (Birkhaüser, New York, 2012)
17. G. Csato, B. Dacorogna, O. Kneuss, The second order pullback equation. Calc. Var. Partial Differential Equations **49**, 538–611 (2014)
18. G. Cupini, B. Dacorogna, O. Kneuss, On the equation det $\nabla u = f$ with no sign hypothesis. Calc. Var. Partial Differential Equations **36**, 251–283 (2009)
19. B. Dacorogna, A relaxation theorem and its applications to the equilibrium of gases. Arch. Ration. Mech. Anal. **77**, 359–386 (1981)
20. B. Dacorogna, Existence and regularity of solutions of $dw = f$ with Dirichlet boundary conditions. *Nonlinear Problems in Mathematical Physics and Related Topics*. International Mathematical Series (N. Y.), vol. 1 (Kluwer/Plenum, New York, 2002), pp. 67–82
21. B. Dacorogna, *Direct Methods in the Calculus of Variations*, 2nd edn. (Springer, New York, 2007)
22. B. Dacorogna, *Introduction to the Calculus of Variations*, 3rd edn. (Imperial College Press, London, 2014)
23. B. Dacorogna, Sur un problème non linéaire pour la divergence et le déterminant. Confluentes Mathematici **7**, 49–55 (2015)
24. B. Dacorogna, N. Fusco, L. Tartar, On the solvability of the equation div $u = f$ in $L^1$ and in $C^0$. Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl. **14**, 239–245 (2003)

25. B. Dacorogna, W. Gangbo, O. Kneuss, Optimal transport of closed differential forms for convex costs. C. R. Math. Acad. Sci. Paris Ser. I **353**, 1099–1104 (2015)
26. B. Dacorogna, W. Gangbo, O. Kneuss, Symplectic factorization, Darboux theorem and ellipticity (2017, to appear)
27. B. Dacorogna, O. Kneuss, A global version of Darboux theorem with optimal regularity and Dirichlet condition. Adv. Differ. Equ. **16**, 325–360 (2011)
28. B. Dacorogna, O. Kneuss, Multiple Jacobian equations. Commun. Pure Appl. Anal. **13**, 1779–1787 (2014)
29. B. Dacorogna, O. Kneuss, W. Neves, Some remarks on the Lie derivative and the pullback equation for contact forms. To appear in Advanced Nonlinear Studies (2017)
30. B. Dacorogna, J. Moser, On a partial differential equation involving the Jacobian determinant. Ann. Inst. H. Poincaré Anal. Non Linéaire **7**, 1–26 (1990)
31. G. Darboux, Sur le problème de Pfaff. Bull. Sci. Math. **6**, 14–36, 49–68 (1882)
32. R. Dautray, J.L. Lions, *Analyse Mathématique et Calcul Numérique* (Masson, Paris, 1988)
33. G.P. Galdi, *An Introduction to the Mathematical Theory of the Navier-Stokes Equations* (Springer, New York, 1994)
34. D. Gilbarg, N.S. Trudinger, *Elliptic Partial Differential Equations of Second Order* (Springer, Berlin, 1977)
35. V. Girault, P.A. Raviart, *Finite Element Approximation of the Navier-Stokes Equations*. Lecture Notes in Mathematics, vol. 749 (Springer, Berlin, 1979)
36. L.V. Kapitanskii, K. Pileckas, Certain problems of vector analysis. J. Sov. Math. **32**, 469–483 (1986)
37. O. Kneuss, Optimal regularity and control of the support for the pullback equation (2017, to appear)
38. R. Kress, Potentialtheoretische Randwertprobleme bei Tensorfelderrn beliebiger Dimensions und beliebigen Ranges. Arch. Ration. Mech. Anal. **47**, 59–80 (1972)
39. O.A. Ladyzhenskaya, *The Mathematical Theory of Viscous Incompressible Flow* (Gordon and Breach, New York, 1969)
40. O.A. Ladyzhenskaya, V.A. Solonnikov, Some problems of vector analysis and generalized formulations of boundary value problems for the Navier-Stokes equations. J. Sov. Math. **10**, 257–286 (1978)
41. D. Mc Duff, D. Salamon, *Introduction to Symplectic Topology*, 2nd edn. (Oxford Science Publications, Oxford, 1998)
42. C.T. Mc Mullen, Lipschitz maps and nets in Euclidean space. Geom. Funct. Anal. **8**, 304–314 (1998)
43. C.B. Morrey, A Variational method in the theory of harmonic integrals II. Am. J. Math. **78**, 137–170 (1956)
44. C.B. Morrey, *Multiple Integrals in the Calculus of Variations* (Springer, Berlin, 1966)
45. C.B. Morrey, J. Eells, A variational method in the theory of harmonic integrals. Ann. Math. **63**, 91–128 (1956)
46. J. Moser, On the volume elements on a manifold. Trans. Am. Math. Soc. **120**, 286–294 (1965)
47. J. Necas, *Les méthodes Directes en Théorie des Équations Elliptiques* (Masson, Paris, 1967)
48. D. Preiss, Additional regularity for Lipschitz solutions of pde. J. Reine Angew. Math. **485**, 197–207 (1997)
49. H.M. Reimann, Harmonische funktionen und jacobi-determinanten von diffeomorphismen. Comment. Math. Helv. **47**, 397–408 (1972)
50. T. Rivière, D. Ye, Resolutions of the prescribed volume form equation. Nonlinear Differ. Equ. Appl. **3**, 323–369 (1996)
51. G. Schwarz, *Hodge Decomposition - a Method for Solving Boundary Value Problems*. Lecture Notes in Mathematics, vol. 1607 (Springer, Berlin, 1995)

52. S. Takahashi, On the Poincaré-Bogovski lemma on differential forms. Proc. Jpn. Acad. Ser. A Math. Sci. **68**, 1–6 (1992)
53. L. Tartar, *Topics in Nonlinear Analysis* (University of Wisconsin, Madison, 1975); Preprint
54. W. Von Wahl, *Vorlesung über das Aussenraumproblem für die instationären Gleichungen von Navier-Stokes*; Rudolph-Lipschitz-Vorlesung. Sonderforschungsbereich 256 Nichtlineare Partielle Differentialgleichungen, Bonn, 1989
55. W. Von Wahl, *On Necessary and Sufficient Conditions for the Solvability of the Equations* $\operatorname{rot} u = \gamma$ *and* $\operatorname{div} u = \epsilon$ *with u Vanishing on the Boundary*. Lecture Notes in Mathematics, vol. 1431 (Springer, Berlin, 1990), pp. 152–157
56. W. Von Wahl, Estimating $\nabla u$ by $\operatorname{div} u$ and $\operatorname{curl} u$. Math. Methods Appl. Sci. **15**, 123–143 (1992)
57. D. Ye, Prescribing the Jacobian determinant in Sobolev spaces. Ann. Inst. H. Poincaré Anal. Non Linéaire **11**, 275–296 (1994)
58. E. Zehnder, *Note on Smoothing Symplectic and Volume Preserving Diffeomorphisms*. Lecture Notes in Mathematics, vol. 597 (Springer, Berlin, 1976), pp. 828–855

# The Stability of the Isoperimetric Inequality

**Nicola Fusco**

## 1 Introduction

These lecture notes contain the material that I presented in two summer courses in 2013, one at the Carnegie Mellon University and the other one in a CIME school at Cetraro. The aim of both courses was to give a quick but comprehensive introduction to some recent results on the stability of the isoperimetric inequality.

The starting point is the De Giorgi's proof of the *isoperimetric inequality*. Many other proofs of this inequality are now available. Some of them are classical, like the one based on the *Brunn-Minkowski inequality*, see for instance [15, Theorem 8.1.1], or the one based on the Alexandrov rigidity theorem [2]. More recent proofs are the one based on mass transportation due to Gromov, see Sect. 6, and the PDE proof due to Cabré [16]. Among all these proofs the one by De Giorgi still stands as the most intuitive from a geometric point of view and at the same time the most general one since his isoperimetric inequality (17) applies to any measurable set of finite measure.

In order to explain this proof a few basic properties of sets of finite perimeter are required. They are presented in Sect. 2, while Sect. 3 contains a slightly modified version of the original proof of De Giorgi.

The remaining part of these notes are devoted to the *stability of the isoperimetric inequality*. In fact, once we know that for a given volume balls are the unique area minimizers the next natural question is to understand what happens when a set $E$ has the same volume of a ball $B$ and a slightly bigger surface area. Precisely, one would like to show that in this case $E$ must be close in a proper sense to a translation of $B$.

N. Fusco (✉)
Dipartimento di Matematica e Applicazioni "R. Caccioppoli", Università degli Studi di Napoli "Federico II", Napoli, Italy
e-mail: n.fusco@unina.it



J. Ball, P. Marcellini (eds.), *Vector-Valued Partial Differential Equations and Applications*, Lecture Notes in Mathematics 2179,
DOI 10.1007/978-3-319-54514-1_2

Already a few years after the Hurwitz proof [49] of the isoperimetric inequality in the plane, this problem was studied by Bernstein [8] and later on by Bonnesen [11] for planar convex sets. The case of convex sets in any dimension was settled much later by Fuglede in [39]. Section 4 contains the complete proof of the Fuglede's Theorem 26.

The stability of the isoperimetric inequality for general sets of finite perimeter is a different story, see the discussion at the beginning of Sect. 5. The first result in this direction was proved by Hall [47] in 1992 with a not optimal estimate of the distance between $E$ and the closest ball, while the estimate with the sharp exponent was obtained by Maggi, Pratelli and myself in [44], see Theorem 34. Section 5 contains a fairly detailed discussion of this result, whose proof is based on a suitable symmetrization argument aimed to reduce from a general set of finite perimeter to an axially symmetric bounded set with a center of symmetry.

Other proofs and generalizations of the quantitative isoperimetric inequality (35) were later on obtained by Figalli, Maggi and Pratelli in [34] and by Cicalese and Leonardi in [23], see also [42] and [1]. These alternative proofs are presented in Sect. 6.

The aforementioned papers were the starting point for an intensive study of the stability of other geometric and functional inequalities such as other inequalities of isoperimetric type [3, 5, 6, 9, 10, 22, 24, 25, 31, 41, 46, 56, 58], the *Sobolev inequality* [21, 35, 36, 43], the *Brunn-Minkowski inequality* [33], the *Faber-Krahn inequality* [12, 45] and several others [7, 13, 18, 20, 32, 52]. We shall not discuss here these further developments. The interested reader may have a look at a the survey paper [40] which contains a detailed account of all the recent results, updated to Spring 2015.

Finally, I would like to thank Ryan Murray who typed the notes of the course I gave in Pittsburgh, Matteo Rinaldi who added some extra material from some hand written notes of mine and Laura Bufford and Andrea Fusco for all the pictures.

## 2 A Quick Review of Sets of Finite Perimeter

We start by reviewing the definition and the main properties of sets of finite perimeter which are the objects for which the isoperimetric inequality will be proved in the next section. A good reference for the results stated here are the books [4, 29, 51] and the original papers of De Giorgi collected in [28]. Note, however that the definition below is equivalent, but different from the one originally proposed by De Giorgi.

In the following we denote by $B_r(x)$ the ball with radius $r > 0$ and center $x$ and we use the following simplified notation

$$B_r := B_r(0), \quad B(x) := B_1(x) \quad B := B_1(0)\,.$$

The measure of the unit ball $B$ will be denoted by $\omega_n$. As a starting point we consider the classical divergence theorem stating that if $E$ is a smooth bounded open set in $\mathbb{R}^n$, and $\varphi$ is a smooth vector field in $\mathbb{R}^n$ with compact support, then

$$\int_E \mathrm{div}\varphi \, dx = \int_{\partial E} \varphi \cdot \nu d\mathcal{H}^{n-1} . \tag{1}$$

Here, if $k$ is a nonnegative integer, by $\mathcal{H}^k$ we denote the $k$-dimensional *Hausdorff measure* in $\mathbb{R}^n$. Observe that from the previous formula, by taking the supremum over all vector fields $\varphi \in C_c^1(\mathbb{R}^n;\mathbb{R}^n)$, with $||\varphi||_\infty \leq 1$, we get

$$\mathcal{H}^{n-1}(\partial E) = \sup \left\{ \int_E \mathrm{div}\varphi \, dx : \ \varphi \in C_c^1(\mathbb{R}^n;\mathbb{R}^n), ||\varphi||_\infty \leq 1 \right\} . \tag{2}$$

Since the first integral in (1) makes sense for any measurable set, equality (2) suggests how to extend the notion of boundary measure to any measurable set $E \subset \mathbb{R}^n$.

**Definition 1** Let $\Omega$ be an open set in $\mathbb{R}^n$. The *perimeter of $E$ in $\Omega$* is defined as

$$P(E;\Omega) := \sup \left\{ \int_E \mathrm{div}\varphi \, dx : \ \varphi \in C_c^1(\Omega;\mathbb{R}^n), ||\varphi||_\infty \leq 1 \right\} .$$

An important feature of this definition is that the perimeter is not affected by modifications on sets of measure zero. Thus the two sets shown in Fig. 1 have the same perimeter. Note also that $P(E;\Omega) = P(\mathbb{R}^n \setminus E;\Omega)$.

Observe that if $P(E;\Omega) < \infty$, then the map

$$\varphi \in C_c^1(\Omega;\mathbb{R}^n) \mapsto \int_E \mathrm{div}\varphi \, dx$$

is linear and continuous with respect to the uniform convergence on $C_c^1(\Omega;\mathbb{R}^n)$. Therefore Riesz's theorem yields that there exists a vector valued Radon measure

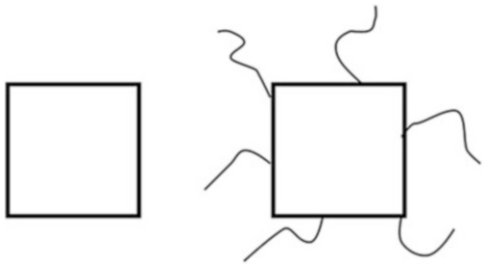

**Fig. 1** Two sets with the same perimeter

$\mu = (\mu_1, \dots, \mu_n)$ in $\Omega$ such that

$$\int_\Omega \chi_E \mathrm{div}\varphi \, dx = \int_E \mathrm{div}\varphi \, dx = \int_\Omega \varphi \cdot d\mu = \sum_{i=1}^n \int_\Omega \varphi_i \, d\mu_i$$

for all $\varphi \in C_c^1(\Omega, \mathbb{R}^n)$. Thus $\mu = -D\chi_E$, where $D\chi_E$ is the distributional derivative of $\chi_E$ and the above formula can be rewritten as

$$\int_E \mathrm{div}\varphi \, dx = -\int_\Omega \varphi \cdot dD\chi_E. \tag{3}$$

In conclusion, $E$ has finite perimeter in $\Omega$ if and only if $D\chi_E$ is a Radon measure with values in $\mathbb{R}^n$ and finite total variation. In fact, from Definition 1 we immediately get that

$$P(E; \Omega) = |D\chi_E|(\Omega).$$

If $\Omega = \mathbb{R}^n$ we simply write $P(E)$ in place of $P(E; \mathbb{R}^n)$ and if $P(E) < \infty$ we say that $E$ is a *set of finite perimeter*. If $P(E; \Omega) < \infty$ for every bounded open set, then we say that $E$ has *locally finite perimeter*. The following properties are immediate consequences of Definition 1. For any measurable set $E$

$$P(\lambda E) = \lambda^{n-1} P(E) \qquad \text{for all } \lambda > 0\,; \tag{4}$$

moreover, for any open set $\Omega$,

$$P(E; \mathring{E} \cap \Omega) = P(E; \Omega \setminus \overline{E}) = 0$$

Therefore the measure $D\chi_E$ is concentrated on $\partial E \cap \Omega$ and (3) can be rewritten as

$$\int_E \mathrm{div}\varphi \, dx = -\int_{\partial E \cap \Omega} \varphi \cdot D\chi_E, \qquad \text{for all } \varphi \in C_c^1(\Omega; (\mathbb{R}^n). \tag{5}$$

Observe also that from Besicovitch derivation theorem [4, Theorem 2.22] we have that for $|D\chi_E|$–a.e. $x \in \mathrm{supp}|D\chi_E|$ there exists the derivative of $D\chi_E$ with respect to its total variation $|D\chi_E|$ and that it is a vector of length 1. For such points we have

$$\frac{D\chi_E}{|D\chi_E|}(x) = \lim_{r \to 0} \frac{D\chi_E(B_r(x))}{|D\chi_E|(B_r(x))} =: -\nu^E(x) \quad \text{and} \quad |\nu^E(x)| = 1. \tag{6}$$

**Definition 2** We shall denote by $\partial^* E$ the set of all points in $\mathrm{supp}|D\chi_E|$ where (6) holds. The set $\partial^* E$ is called the *reduced boundary* of $E$, while the vector $\nu^E(x)$ is the *generalized exterior normal at x*.

From (6) it follows that the measure $D\chi_E$ can be represented by integrating $-\nu^E$ with respect to $|D\chi_E|$, i.e.,

$$D\chi_E = -\nu^E |D\chi_E|.$$

Thus (5) can be rewritten as

$$\int_E \mathrm{div}\varphi dx = \int_{\partial^* E \cap \Omega} \varphi \cdot \nu^E d|D\chi_E|, \qquad \forall\, \varphi \in C_c^1(\Omega, \mathbb{R}^n). \tag{7}$$

Since $\partial^* E \subset \mathrm{supp}|D\chi_E| \subset \partial E$, the reduced boundary of $E$ is a subset of the topological boundary. Moreover, as a consequence of De Giorgi structure Theorem 6, if $E$ has finite perimeter, then $\mathcal{H}^{n-1}(\partial^* E) = P(E) < \infty$. Next example shows that in general $\partial^* E$ can be much smaller than $\partial E$.

*Example 3* Let us take a sequence $\{q_i\}$ dense in $\mathbb{R}^n$ and set $E := \bigcup_{i=1}^{\infty} B_{2^{-i}}(q_i)$. Observe that $|\partial E| = \infty$. Nevertheless $E$ is a set of finite perimeter. To see this take $\varphi \in C_c^1(\mathbb{R}^n, \mathbb{R}^n)$, $||\varphi||_\infty \leq 1$, and note that

$$\begin{aligned}
\int_E \mathrm{div}\varphi \, dx &= \lim_{N\to\infty} \int_{\bigcup_{i=1}^N B_{2^{-i}}(q_i)} \mathrm{div}\varphi \, dx = \lim_{N\to\infty} \int_{\partial(\bigcup_{i=1}^N B_{2^{-i}}(q_i))} \varphi \cdot \nu \, d\mathcal{H}^{n-1} \\
&\leq \lim_{N\to\infty} \mathcal{H}^{n-1}\Big(\partial\Big(\bigcup_{i=1}^N B_{2^{-i}}(q_i)\Big)\Big) \leq \lim_{N\to\infty} \sum_{i=1}^N \mathcal{H}^{n-1}\big(\partial B_{2^{-i}}(q_i)\big) \\
&= n\omega_n \sum_{i=1}^{\infty} 2^{-i(n-1)} < \infty\,.
\end{aligned}$$

In dimension 1, sets of finite perimeter are easily characterized (see [4, Proposition 3.52]).

**Theorem 4** *Let $E \subset \mathbb{R}$ be a measurable set. Then $E$ has finite perimeter in $\mathbb{R}$ if and only if there exist $-\infty \leq a_1 < b_1 < a_2 < b_2 < \ldots < b_n \leq +\infty$ such that*

$$E = \bigcup_{i=1}^{n} (a_i, b_i)$$

*up to a set of zero Lebesgue measure. Moreover, if $\Omega \subset \mathbb{R}$ is an open set,*

$$P(E; \Omega) = \#(\{a_i, b_i \in \Omega\}).$$

*Remark 5* Thus, if for instance $E = (0,1) \cup (1,2)$, then $P(E) = 2$ and $\partial^* E = \{0, 2\}$. In fact, as we already observed, the measure $D\chi_E$ does not change if we modify $E$ by a set of measure zero and thus $E$ and $(0, 2)$ have the same reduced boundary.

The characterization of sets of finite perimeter in $\mathbb{R}^n$ is more complicate and is contained in the next theorem due to De Giorgi. For a proof see [29, Sect. 5.7.3] or [4, Theorem 3.59].

**Theorem 6 (De Giorgi)** *Let $E \subset \mathbb{R}^n$ be a measurable set of finite perimeter. Then the following hold:*

(i) *$\partial^* E$ is $(n-1)$-countably rectifiable, i.e., $\partial^* E = \bigcup_{i=1}^{\infty} K_i \cup N_0$, where $\mathcal{H}^{n-1}(N_0) = 0$ and $K_i$ are compact subsets of $C^1$ manifolds $M_i$ of dimension $n-1$;*

(ii) *$|D\chi_E| = \mathcal{H}^{n-1} \llcorner \partial^* E$;*

(iii) *for $\mathcal{H}^{n-1}$-a.e. $x \in K_i$, the generalized exterior normal $\nu^E(x)$ is orthogonal to the tangent plane $T_x M_i$ to the manifold $M_i$ at $x$;*

(iv) *for all $x \in \partial^* E$, $\dfrac{|E \cap B_r(x)|}{B_r(x)} \to \dfrac{1}{2}$ as $r \to 0$;*

(v) *for all $x \in \partial^* E$, $\lim\limits_{r \to 0} \dfrac{\mathcal{H}^{n-1}(\partial^* E \cap B_r(x))}{\omega_{n-1} r^{n-1}} = 1$.*

As a consequence of the equality (ii) above we have that (7) can we rewritten as

$$\int_E \mathrm{div}\varphi \, dx = \int_{\partial E} \varphi \cdot \nu^E d\mathcal{H}^{n-1}, \qquad \forall\, \varphi \in C_c^1(\mathbb{R}^n, \mathbb{R}^n).$$

*Example 7* Let $Q$ be a square in $\mathbb{R}^2$. The reduced boundary is given by $\partial^* Q = \partial Q \setminus \bigcup_{i=1}^4 \{v_i\}$, where $v_i$ are the vertices of $Q$. In fact, for any sufficiently small ball $B_r(v_i)$ we have that $|Q \cap B_r(v_i)|/|B_r| = \frac{1}{4}$. Therefore from the property $(iv)$ in Theorem 6 it follows that the $v_i$ do not belong to the reduced boundary $\partial^* Q$, see Fig. 2.

Property $(v)$ tells us that if $x \in \partial^* E$ then the reduced boundary $\partial^* E$ looks flatter and flatter at small scales. Observe in fact that if we rescale $\partial^* E$ around $x$, we have, see Fig. 3,

$$\mathcal{H}^{n-1}\left(\frac{\partial^* E - x}{r} \cap B\right) = \frac{\mathcal{H}^{n-1}(\partial^* E \cap B_r(x))}{r^{n-1}} \to \omega_{n-1}.$$

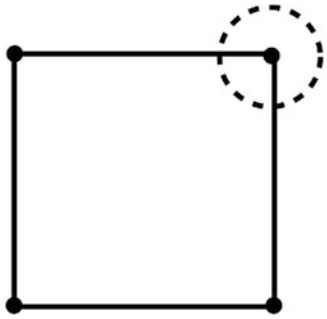

**Fig. 2** The density of the vertices is $1/4$

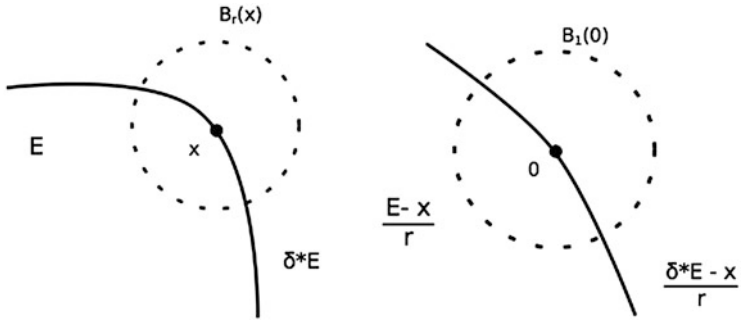


**Fig. 3** Rescaling around $x$

**Definition 8** Given a measurable set $E$ and $x \in \mathbb{R}^n$, the *density of $E$ at $x$*, $D(x;E)$, is defined as

$$D(x;E) := \lim_{r\to 0} \frac{|E \cap B_r(x)|}{\omega_n r^n}.$$

If $0 \le a \le 1$ we denote by $E^{(a)}$ the set of all points where the density of $E$ is equal to $a$.

Observe that from the above definition it follows immediately that

$$x \in E^{(1)} \text{ if and only if } \lim_{r\to 0} \frac{|E \cap Q_r(x)|}{2^n r^n} = 1, \tag{8}$$

where $Q_r(x)$ is the cube with center at $x$ with edge length equal to $2r$ and faces parallel to the coordinate planes. A similar characterization holds also for the points in $E^{(0)}$.

Using densities, part (iv) of De Giorgi's Theorem 6 can be written as $\partial^* E \subset E^{(1/2)}$. We recall also that if $E$ is a measurable set in $\mathbb{R}^n$ its *measure theoretic boundary* $\partial^M E$ is defined by setting

$$\partial^M E := \mathbb{R}^n \setminus \left(E^{(0)} \cup E^{(1)}\right). \tag{9}$$

The next result gives a precise description of what is going on with sets of finite perimeter. For the proof see for instance [4, Theorem 3.61].

**Theorem 9 (Federer)** *Let $E$ be a set of finite perimeter in $\mathbb{R}^n$. Then*

$$\partial^* E \subset E^{(1/2)} \subset \partial^M E \qquad \text{and} \qquad \mathcal{H}^{n-1}\left(\partial^M E \setminus \partial^* E\right) = 0\,.$$

Note that if $E$ is a set of finite perimeter in $\Omega$ Theorems 6 and 9 hold in local form.

*Example 10* Let $U \subset \mathbb{R}^{n-1}$ be a bounded open set and $\Omega = U \times \mathbb{R}$. Let $f : U \to \mathbb{R}$ be a Lipschitz function. Let us denote by $\mathcal{S}_f := \{(x,t) \in \Omega : t < f(x)\}$ the *subgraph* of $f$. Then it may be easily checked that $\mathcal{S}_f$ has finite perimeter in $\Omega$ and that $\partial^* \mathcal{S}_f$ coincides with $\Gamma_f := \{(x, f(x)) : x \in U\}$ up to a set of zero $\mathcal{H}^{n-1}$ measure. Moreover, the generalized normal $\nu^{\mathcal{S}_f}(x)$ coincides $\mathcal{H}^{n-1}$-a.e. on $\Gamma_f$ with the usual exterior normal $\dfrac{(-\nabla f, 1)}{\sqrt{1 + |\nabla f|^2}}$.

*Example 11* Let $f : \mathbb{R} \to \mathbb{R}$ be the function $f(x) = x^2 \sin \frac{1}{x}$ and let $E := \mathcal{S}_f$ be the subgraph of $f$. Using the fact that $f'(0) = 0$ we get easily that

$$\frac{|E \cap B_r|}{|B_r|} \to \frac{1}{2}.$$

However $(0,0) \notin \partial^* E$ since it can be checked that

$$\limsup_{r \to 0} \frac{\mathcal{H}^1(\partial^* E \cap B_r)}{2r} > 1 .$$

Thus property (v) stated in Theorem 6 does not hold.

Approximating sets of finite perimeter with nicer sets is very useful to deduce various properties from the corresponding ones of smooth sets. To this aim we introduce the following notion of convergence.

**Definition 12** Given a sequence of measurable sets $E_j$ and a measurable set $E$, we say that $E_j \to E$ *in measure* in $\Omega$ if $\chi_{E_j} \to \chi_E$ in $L^1(\Omega)$, i.e., $|(E_j \Delta E) \cap \Omega| \to 0$, as $j \to \infty$.

An important property of the perimeters is the lower semicontinuity with respect to the convergence in measure. This is a straightforward consequence of Definition 1. Precisely, if $E_j$ is a sequence of measurable sets converging in measure in $\Omega$ to $E$, then

$$P(E; \Omega) \le \liminf_{j \to \infty} P(E_j; \Omega).$$

For the proof of the next approximation result see for instance [4, Theorem 3.42].

**Theorem 13** *Let $E$ be a set of finite perimeter. Then there exists a sequence of smooth, bounded open sets $E_j$ such that $E_j \to E$ in measure in $\mathbb{R}^n$ and $P(E_j) \to P(E)$.*

In view of this theorem and of the lower semicontinuity of the perimeter we have that $E$ is a set of finite perimeter in $\mathbb{R}^n$ if and only if there exists a sequence of smooth open sets $E_j \subset \mathbb{R}^n$, such that

$$E_j \to E \text{ in measure in } \mathbb{R}^n \quad \text{and} \quad \sup_j P(E_j) < \infty.$$

Note also that in Theorem 13 one may replace the smooth sets $E_j$ with polyhedra, i.e., bounded open sets obtained as the intersection of finitely many half–spaces. A local version of Theorem 13 is also true (see [4, Remark 3.43]). As a consequence of Theorem 13 observe that if $E$ and $F$ are sets of finite perimeter, the same is true for $E \cup F$, $E \cap F$ and $E \setminus F$ and that

$$P(E \cap F) + P(E \cup F) \leq P(E) + P(F)\,.$$

Simple examples show that the above inequality may be strict. In general the precise expression of the reduced boundaries of $E \cap F$ or $E \cup F$ in terms of the reduced boundaries of $E$ and $F$ is a little involved. The next statement provides the precise picture. For a proof see for instance [34, (2.8), (2.9) and Lemma 2.2].

**Proposition 14** *Let $E, F \subset \mathbb{R}^n$ be sets of finite perimeter. Then, up to a set of zero $\mathcal{H}^{n-1}$ measure*

$$\partial^*(E \cap F) = \{y \in \partial^* E \cap \partial^* F :\ \nu^E(y) = \nu^F(y)\} \cup [\partial^* E \cap F^{(1)}] \cup [\partial^* F \cap E^{(1)}]$$

*and for $\mathcal{H}^{n-1}$-a.e. $x \in \partial^*(E \cap F)$*

$$\nu^{E\cap F}(x) = \begin{cases} \nu^E(x) = \nu^F(x) & \text{if } x \in \{y \in \partial^* E \cap \partial^* F :\ \nu^E(y) = \nu^F(y)\}, \\ \nu^E(x) & \text{if } x \in \partial^* E \cap F^{(1)}, \\ \nu^F(x) & \text{if } x \in \partial^* F \cap E^{(1)}\,. \end{cases}$$

*Moreover, if $|E \cap F| = 0$, then, up to a set of zero $\mathcal{H}^{n-1}$-measure, $\partial^*(E \cup F) = \partial^* E \Delta \partial^* F$ and*

$$\nu^{E\cup F}(x) = \begin{cases} \nu^E(x) & \text{if } x \in \partial^* E \setminus \partial^* F, \\ \nu^F(x) & \text{if } x \in \partial^* F \setminus \partial^* E\,. \end{cases}$$

The next result, is just the Rellich-Kondrachov compactness theorem stated in the framework of sets of finite perimeter (see [4, Theorem 3.39]).

**Theorem 15** *Given a bounded open set $\Omega \subset \mathbb{R}^n$ and a sequence of measurable sets $E_j$ such that $\sup_j P(E_j; \Omega) < \infty$, there exists a set $E$ of finite perimeter in $\Omega$ such that, up to a subsequence, $E_j \to E$ in measure in $\Omega$.*

The theory of sets of finite perimeter can be viewed as a special part of the theory of *functions of bounded variation*. Recall that if $\Omega$ is an open set a function $u \in L^1(\Omega)$ is said to be of bounded variation if the distributional gradient $Du$ is a vector-valued measure in $\Omega$ with finite total variation. Observe that by definition of distributional gradient

$$\int_\Omega u \,\mathrm{div}\varphi \, dx = -\int_\Omega \varphi \, dDu,$$

for all $C^1$ vector fields $\varphi$ with compact support in $\Omega$. From this formula it follows immediately that the total variation $|Du|(\Omega)$ of $Du$ in $\Omega$ is given by

$$|Du|(\Omega) = \sup\left\{ \int_\Omega u \,\mathrm{div}\varphi \, dx : \ \varphi \in C^1_c(\Omega;\mathbb{R}^n), ||\varphi||_\infty \le 1 \right\}.$$

We shall denote by $BV(\Omega)$ the space of all functions of bounded variation in $\Omega$.

We conclude by recalling the *coarea formula* for sets of finite perimeter. For our purposes it will be enough to consider only $C^1$ maps, though these formulas may easily generalized to Lipschitz and even less regular maps, see [4, Chap. 2] and [30, Sect. 3.2]. Thus, let $f : \mathbb{R}^n \to \mathbb{R}^k$ be a $C^1$ map, $1 \le k \le n-1$, and $E$ a set of finite perimeter. By Definition 2 at every point $x$ of the reduced boundary $\partial^* E$ we have a generalized exterior normal $\nu^E(x)$, hence a generalized tangent plane, that we denote by $T_x\partial^* E$. Therefore, we can consider the *tangential differential* of $f$ at $x$, that is the map $df(x) : T_x\partial^* E \to \mathbb{R}^k$ given by

$$df(x)(\tau) = \nabla f(x)(\tau), \qquad \text{for all } \tau \in T_x\partial^* E\,. \tag{10}$$

Furthermore, we define the *coarea factor at x* as

$$\mathbf{C}_k df(x) = \sqrt{\det(df(x) \circ (df(x))^T)},$$

where $(df(x))^T$ is the transpose of the matrix $df(x)$. It can be shown that $\mathbf{C}_k df(x)$ is the square root of the sum of the squares of the $k$–order minors of the matrix representing $df(x)$ with respect to a base in $T_x\partial^* E$ and a base in $\mathbb{R}^k$ (see [4, (2.71)]).

**Theorem 16 (Coarea Formula for Sets of Finite Perimeter)** *Let $E \subset \mathbb{R}^n$ be a set of finite perimeter and $f : \mathbb{R}^n \to \mathbb{R}^k$ a $C^1$ map, $1 \le k \le n-1$. If $g : \mathbb{R}^n \to [0, +\infty]$ is a Borel function, then*

$$\int_{\partial^* E} g(x)\mathbf{C}_k df(x)\, d\mathcal{H}^{n-1}(x) = \int_{\mathbb{R}^k} dz \int_{f^{-1}(z)\cap\partial^* E} g(x)\, d\mathcal{H}^{n-1-k}(x).$$

Observe that if $\pi : \mathbb{R}^n \to \mathbb{R}^k$ is the projection over the first $k$ components, i.e. $\pi(x,y) = x$ for all $(x,y) \in \mathbb{R}^k \times \mathbb{R}^{n-k}$, then $\mathbf{C}_k d\pi(x,y) = |\nu^E_y(x,y)|$, for all $(x,y) \in \partial^* E$, where $\nu^E = (\nu^E_x, \nu^E_y) \in \mathbb{R}^k \times \mathbb{R}^{n-k}$. To prove this consider an orthonormal base $\{\tau_1, \dots, \tau_{n-1}\}$ for $T_{(x,y)}\partial^* E$, such that the frame $\{\tau_1, \dots, \tau_{n-1}, \nu^E(x,y)\}$ is positively oriented. Then the matrix representing $d\pi(x,y)$ with respect to the given orthonormal base of $T_{(x,y)}\partial^* E$ and the standard base $\{e_1, \dots, e_k\}$ in $\mathbb{R}^k$ has coefficients $e_i \cdot \tau_\ell$, for $i = 1, \dots, k$, $\ell = 1, \dots, n-1$. Therefore, the matrix representing $\det(d\pi(x) \circ (d\pi(x))^T)$ has coefficients

$$a_{ij} := \sum_{\ell=1}^{n-1} (e_i \cdot \tau_\ell)(e_j \cdot \tau_\ell) = \delta_{ij} - \nu^E_i \nu^E_j \qquad \text{for } i,j = 1, \dots, k\,.$$

Recall that if $a, b \in \mathbb{R}^k$ and $I$ denotes the identity matrix, then

$$\det(I + a \otimes b) = 1 + a \cdot b\,. \tag{11}$$

Thus $\mathbf{C}_k d\pi(x,y) = \sqrt{\det(a_{ij})} = \sqrt{1 - |\nu_x^E|^2} = |\nu_y^E|$ and the coarea formula reduces to

$$\int_{\partial^* E} g(x,y)|\nu_y^E(x,y)|d\mathcal{H}^{n-1}(x,y) = \int_{\mathbb{R}^k} dx \int_{(\partial^* E)_x} g(x,y)d\mathcal{H}^{n-1-k}(y), \tag{12}$$

where

$$(\partial^* E)_x = \{y \in \mathbb{R}^{n-k} : (x,y) \in \partial^* E\},$$

see Fig. 4. If we apply (12) to the particular case of the projection over the first $n-1$ components, recalling that $\mathcal{H}^0$ is the counting measure, we have that for every Borel function $g : \mathbb{R}^n \to [0, +\infty]$.

$$\int_{\partial^* E} g(x,y)|\nu_y^E(x,y)|d\mathcal{H}^{n-1}(x,y) = \int_{\mathbb{R}^{n-1}} \Big( \sum_{y \in (\partial^* E)_x} g(x,y) \Big) dx\,. \tag{13}$$

From this formula we deduce that the vertical part of the reduced boundary $\partial^* E$ "is not seen from below".

To be precise, let us define the *vertical part* of the reduced boundary by setting $V := \{(x,y) \in \partial^* E : \nu_y^E(x,y) = 0\}$. If we apply (13) with $g = \chi_V$ we get $\int_{\partial^* E} \chi_V(x,y)|\nu_y^E(x,y)|d\mathcal{H}^{n-1}(x,y) = 0$. Therefore, the right hand side of (13) is

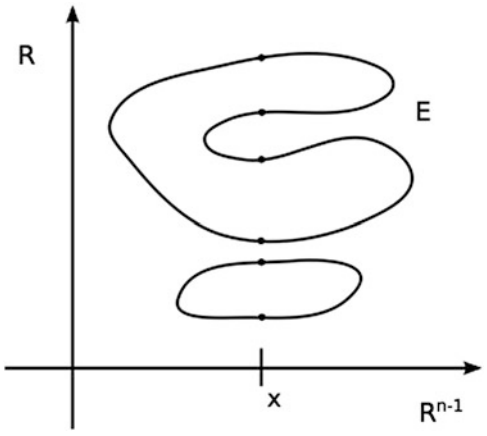


**Fig. 4** Section of $\partial^* E$

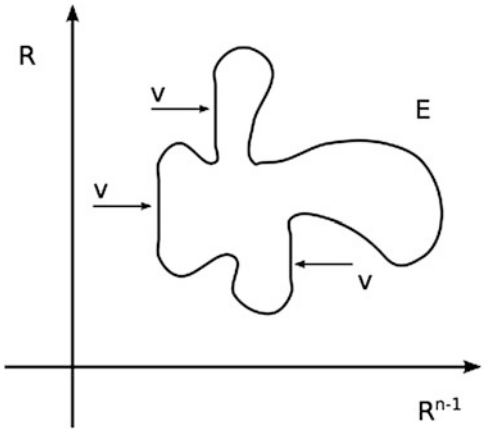


**Fig. 5** The set $V$

also zero, i.e.,

$$\int_{\mathbb{R}^{n-1}} \#\left(\{y \in \mathbb{R} : (x, y) \in V\}\right) dx = 0.$$

This implies that for $\mathcal{H}^{n-1}$-a.e. $x \in \mathbb{R}^{n-1}$, the section $V_x$ is empty, see Fig. 5.

## 3 De Giorgi's Proof of the Isoperimetric Inequality

In the framework of sets of finite perimeter the isoperimetric inequality takes the following very general form.

**Theorem 17** *Let $E \subset \mathbb{R}^n$ be a measurable set with $|E| = |B_r|$. Then*

$$P(B_r) \leq P(E) \tag{14}$$

*with the equality holding if and only if $E$ is a ball.*

De Giorgi's proof follows an idea that Steiner had one century before [59]. Actually, the proof of the isoperimetric property of the ball was the original motivation for Steiner to introduce the symmetrization that nowadays bears his name.

**Definition 18** Let $E \subset \mathbb{R}^n$ be a measurable set. For $x \in \mathbb{R}^{n-1}$ set $E_x := \{y \in \mathbb{R} : (x, y) \in E\}$ and $\ell(x) := \mathcal{H}^1(E_x)$. Then the *Steiner symmetrization of $E$* with respect to the hyperplane $\{x_n = 0\}$ is given by $E^s = \{(x, y) \in \mathbb{R}^{n-1} \times \mathbb{R} : -\ell(x)/2 < y < \ell(x)/2\}$.

The previous definition can be extended in an obvious way to any hyperplane $\pi_\nu$ passing through the origin and orthogonal to a unit vector $\nu$. The resulting Steiner

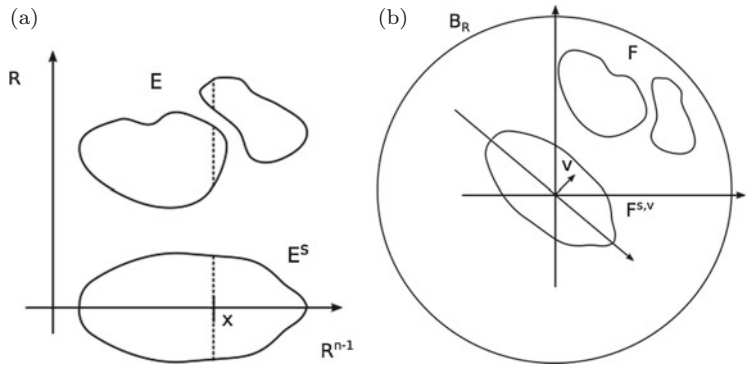


**Fig. 6** Steiner symmetral of a measurable set. (**a**) Symmetrization wrt the plane $\{x_n = 0\}$. (**b**) Symmetrization wrt a plane with normal $\nu \neq e_n$

symmetrization of $E$ with respect to $\pi_\nu$ will be denoted by $E^{s,\nu}$, see Fig. 6. The symmetrization of $E$ with respect to $\{x_n = 0\}$ will be denoted by $E^s$.

From Fubini's theorem we have immediately that $|E| = |E^{s,\nu}|$, while it is not too difficult to show, see for instance [29, Lemma 2, Sect. 2.2], that $\operatorname{diam}(E^{s,\nu}) \leq \operatorname{diam}(E)$. If $E$ is a measurable set, then $\ell$ is a measurable function. Instead, if $E$ is a set of finite perimeter it can be proved that $\ell$ is a function of bounded variation in $\mathbb{R}^{n-1}$ and even a Sobolev function if $\partial^* E$ has no vertical part. However, for the proof of the isoperimetric inequality the relevant fact is that the Steiner symmetrization of a set keeps the volume and decreases the perimeter.

**Theorem 19** *Let $E$ be a set of finite perimeter with $|E| < \infty$. Then the following properties hold:*

(i) $\ell \in BV(\mathbb{R}^{n-1})$;
(ii) $\ell \in W^{1,1}(\mathbb{R}^{n-1})$ *if and only if* $\mathcal{H}^{n-1}(\{z \in \partial^* E : \nu_n^E(z) = 0\}) = 0$;
(iii) $P(E^s) \leq P(E)$;
(iv) *if $P(E^s) = P(E)$ then for $\mathcal{H}^{n-1}$-a.e. $x \in \mathbb{R}^{n-1}$, $E_x$ coincides up to a set of zero $\mathcal{H}^1$ measure with a line segment.*

Inequality (iii) is classical and is proved for smooth sets in the beautiful book of Pólya–Szegö [57]. Property (iv) appears in a weaker form in De Giorgi's original paper on the isoperimetric property of balls [27]. The above statement of (iv) as well as (i) and (ii) are proved in [19, Theorem 1.1, Lemma 3.1, Proposition 1.2].

Note that if $P(E) = P(E^s)$, then $E$ and $E^s$ are not necessarily equal up to a translation, as shown in Fig. 7. In both pictures $P(E) = P(E^s)$, conclusion (iv) of the theorem holds but $E \neq E^s$ up to a translation. However, it is possible to characterize the cases when the equality $P(E) = P(E^s)$ implies that $E$ and $E^s$ coincide up to a translations, see [19] and [17], where a deeper analysis is carried on. We now turn to the proof of the isoperimetric inequality via Steiner symmetrization.

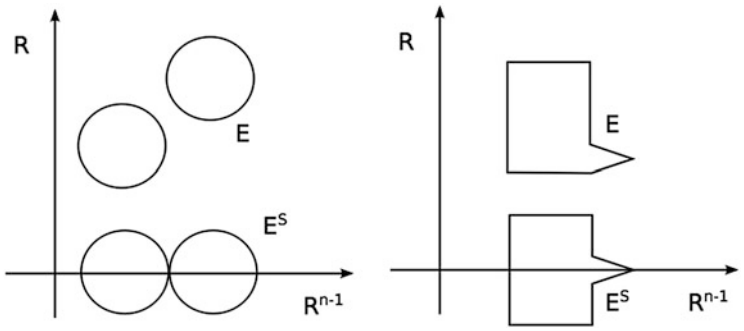


**Fig. 7** In general, symmetrals are not translated of the original sets

*Proof* From the rescaling property (4) it follows that in order to prove (14) it is enough to show that $P(E) \geq P(B)$ for all sets $E$ such that $|E| = |B| = \omega_n$, with the equality holding if and only if $E$ is a ball.

**Step 1.** We first fix $B_R$, with $R > 1$ and consider the minimum problem

$$\inf\{P(E) : E \subset B_R,\ |E| = \omega_n\}.$$

Observe that the above infimum is always attained. In fact, let $E_j \subset B_R$, with $|E_j| = \omega_n$, be a minimizing sequence, i.e., $\lim_j P(E_j) = \inf\{P(E) : E \subset B_R,\ |E| = \omega_n\}$. By the compactness Theorem 15 we may assume, up to a not relabelled subsequence, that $E_j$ converge in measure to some set $F \subset B_R$ with $|F| = \omega_n$. By the lower semicontinuity of the perimeter we have $P(F) \leq \liminf P(E_j)$ and thus $F$ is a minimizer.

We claim that $F$ coincides, up to a set of measure zero, with a convex set. To prove this, fix $\nu \in \mathbb{S}^{n-1}$ and consider the Steiner symmetrization $F^{s,\nu}$ of $F$ with respect to the hyperplane $\pi_\nu$ passing through the origin and orthogonal to $\nu$. Observe that $|F^{s,\nu}| = |F| = \omega_n$ and that $F^{s,\nu} \subset B_R$. Moreover, from part (iii) of Theorem 19 we have that $P(F^{s,\nu}) \leq P(F)$ and thus, by the minimality of $F$, we may conclude that $P(F^{s,\nu}) = P(F)$. Thus, recalling the property (iv) stated in Theorem 19, we have that for $\mathcal{H}^{n-1}$-a.e. $x \in \pi_\nu$, the section $\{t \in \mathbb{R} : x + t\nu \in F\}$ coincides up to a set of $\mathcal{H}^1$ measure zero with an open interval. By the arbitrariness of $\nu$ this property clearly holds for all directions $\nu \in \mathbb{S}^{n-1}$. Notice that if we knew that each section $\{t \in \mathbb{R} : x + t\nu \in F\}$ is an open interval for any $\nu \in \mathbb{S}^{n-1}$ and any $x \in \pi_\nu$, then we could conclude at once that $F$ is a convex set. Although this may be not true, Lemma 20 guarantees that there exists a set equivalent to $F$ up to a set of zero Lebesgue measure which has this property. This set is precisely $F^{(1)}$, the set of all points where $F$ has density 1. Hence, $F^{(1)}$ is an open convex set.

To simplify the notation let us set $F = F^{(1)}$. Our goal now is to show that $F$ is a ball. Denote by $U$ the projection of $F$ on $\mathbb{R}^{n-1}$. Then there exist two functions $y_1, y_2 : U \to \mathbb{R}$, $y_1$ convex and $y_2$ concave, such that $F = \{(x, y) : x \in U,\ y_1(x) <$

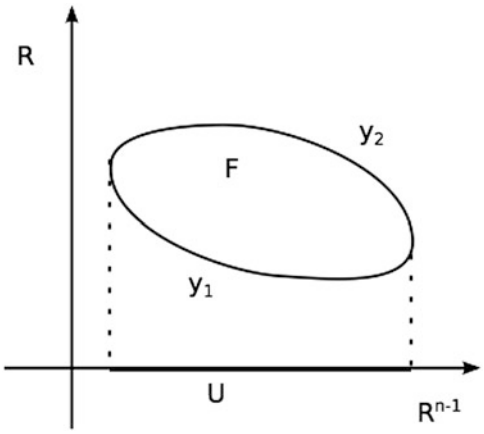


**Fig. 8** Projection of $F$

$y < y_2(x)\}$, see Fig. 8. Moreover, $F^s = \{(x, y) : x \in U,\ (y_1 - y_2)(x)/2 < y < (y_2 - y_1)(x)/2\}$. We have:

$$P(F) = \int_U \sqrt{1 + |\nabla y_1|^2} + \int_U \sqrt{1 + |\nabla y_2|^2}, \qquad P(F^s) = 2 \int_U \sqrt{1 + |\nabla (y_2 - y_1)/2|^2}.$$

Since $F$ is a minimizer, $P(F) = P(F^s)$ and thus by the strict convexity of the function $t \mapsto \sqrt{1 + t^2}$ we get that $\nabla y_2 = -\nabla y_1$, hence $y_2 = -y_1 + c$, thus proving that $F = F^s$ up to a translation. Repeating this argument for all the Steiner symmetrizations $F^{s,\nu}$, with $\nu \in \mathbb{S}^{n-1}$, we finally conclude that $F$ must be a ball. This proves the isoperimetric inequality for a bounded set $E$.

**Step 2.** Let us now consider the case of an unbounded set $E$ with $|E| = \omega_n$. From Theorem 13 we get a sequence of smooth bounded sets $E_j$ such that $E_j$ converge in measure to $E$ in $\mathbb{R}^n$ and $P(E_j) \to P(E)$ as $j \to \infty$. From Step 1 we then have that $P(E_j) \geq P(B_{r_j})$ where $|E_j| = |B_{r_j}|$. From this inequality and using the fact that $|E_j| \to |E|$, letting $j \to \infty$, we have that $P(E) \geq P(B)$. Finally, if $P(E) = P(B)$ we may repeat the same argument used in Step 1 to conclude first that $E^{(1)}$ is an open convex set and then that it is a ball. □

Let us now give the proof of the technical lemma used before. Note that this lemma was not explicitly stated in the original paper [27]. For the proof below I thank Giovanni Alberti with whom I discussed the issue a few years ago. To this aim, given a measurable set $E$, we denote by $\pi(E)^+$ the *essential projection* of $E$ over the first $n - 1$ coordinates plane, that is

$$\pi(E)^+ := \{x \in \mathbb{R}^{n-1} :\ \mathcal{H}^1(E_x) > 0\}.$$

**Lemma 20** *Let $E$ be a measurable set in $\mathbb{R}^n$ such that for $\mathcal{H}^{n-1}$-a.e. $x \in \mathbb{R}^{n-1}$ the section $E_x = \{y \in \mathbb{R} : (x, y) \in E\}$ is equivalent to a segment up to a set of zero $\mathcal{H}^1$*

*measure. Then, denoting by F the set of points of density* 1 *with respect to E, $F_x$ is a segment for every $x \in \mathbb{R}^{n-1}$.*

*Proof* Let $z_1 = (x, y_1)$, $z_2 = (x, y_2)$ be two points in $F_x$ with $y_1 < y_2$. Let us fix $\overline{y} \in (y_1, y_2)$. We claim that $\overline{z} = (x, \overline{y}) \in F_x$. Since $E$ has density 1 at $x_1$ and $x_2$ the same is true also for $F$. Therefore, given $\varepsilon > 0$, there exists $r_\varepsilon$ such that, if $0 < r < r_\varepsilon$, then, see (8),

$$\frac{|F \cap Q_r(z_i)|}{2^n r^n} > 1 - \varepsilon \qquad \text{for } i = 1, 2\,.$$

By Fubini's theorem we have that

$$2^n r^n (1-\varepsilon) < |F \cap Q_r(z_i)| = \int_{\pi(F \cap Q_r(z_i))^+} \mathcal{H}^1(F \cap Q_r(z_i))_x)\, dx \le 2r\mathcal{H}^{n-1}(\pi(F \cap Q_r(z_i))^+)$$

and thus

$$\mathcal{H}^{n-1}(\pi(F \cap Q_r(z_i))^+) > 2^{n-1} r^{n-1}(1-\varepsilon) \qquad \text{for } i = 1, 2\,. \tag{15}$$

Since the essential projections of $F \cap Q_r(z_1)$ and $F \cap Q_r(z_2)$ are both contained in the same $(n-1)$-dimensional cube of edge length $2r$, from (15) we get that

$$\mathcal{H}^{n-1}\big(\pi(F \cap Q_r(z_1))^+ \cap \pi(F \cap Q_r(x_2))^+\big) > 2^{n-1} r^{n-1}(1-2\varepsilon)\,. \tag{16}$$

Now, recall that by assumption for $\mathcal{H}^{n-1}$-a.e. $x \in \pi(F \cap Q_r(z_1))^+ \cap \pi(F \cap Q_r(z_2))^+$ the set $F_x$ is equivalent to a segment such that $\mathcal{H}^1(F_x \cap Q_r(z_i)) > 0$ for $i = 1, 2$. Therefore, if we take $r_\varepsilon < \dfrac{1}{2} \min\{\overline{y} - y_1, y_2 - \overline{y}\}$, we get that

$$\mathcal{H}^1(F_x \cap Q_r(\overline{z})) = 2r\,.$$

This inequality, together with (16) implies that

$$|F \cap Q_r(\overline{x})| > 2^n r^n (1-2\varepsilon), \qquad \text{for all } r < r_\varepsilon\,.$$

Therefore, letting first $r \to 0$ and then $\varepsilon \to 0$, we immediately get that $F$ has density 1 at $\overline{z}$ and thus $\overline{z} \in F$. Hence the result follows. □

An equivalent way of stating the isoperimetric inequality can be obtained noting that if $|E| = |B_r|$ for some $r > 0$, then $|E| = \omega_n r^n$ and $P(B_r) = n\omega_n r^{n-1}$. Therefore (14) becomes

$$P(E) \ge n\omega_n^{1/n} |E|^{1-1/n}. \tag{17}$$

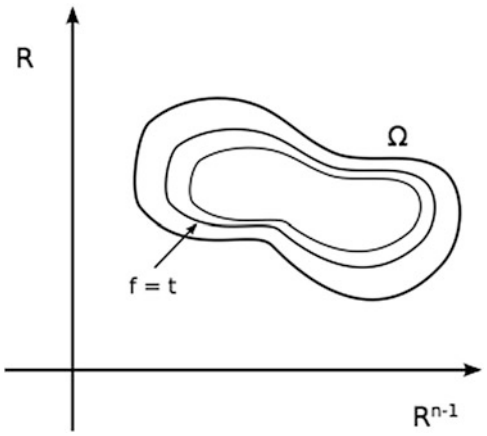


**Fig. 9** The level sets of $f$

Observe that since $P(E) = |D\chi_E|(\mathbb{R}^n)$ and $\|\chi_E\|_{L^{\frac{n}{n-1}}(\mathbb{R}^n)} = |E|^{1-1/n}$ this inequality can be viewed as a particular case of the Sobolev inequality for $W^{1,1}(\mathbb{R}^n)$ or $BV(\mathbb{R}^n)$. To understand better this connection we need to introduce an important formula, first proved by Fleming and Rishel in [37]. As shown in the picture below, it is a sort of curvilinear version of the familiar Fubini theorem, Fig. 9.

**Theorem 21 (Coarea Formula for Lipschitz Function)** *Let $\Omega \subset \mathbb{R}^n$ be an open set and $f : \Omega \to \mathbb{R}$ a Lipschitz function. Then $\{f > t\}$ is a set of finite perimeter for $\mathcal{H}^1$-a.e. $t \in \mathbb{R}$. Moreover, if $g : \Omega \to [0, +\infty]$ is a Borel function,*

$$\int_\Omega g(x)|\nabla f|\,dx = \int_{\mathbb{R}} dt \int_{\partial^*\{f>t\}} g(x)\,d\mathcal{H}^{n-1}(x). \tag{18}$$

The next result shows that the isoperimetric inequality (17) is equivalent to the Sobolev inequality (with the same constant).

**Theorem 22** *The following statements are equivalent:*

(1) *for all measurable set $E$ with finite measure $P(E) \geq C_0|E|^{\frac{n-1}{n}}$;*
(2) *for all $f \in W^{1,1}(\mathbb{R}^n)$ we have that $\|\nabla f\|_{L^1(\mathbb{R}^n)} \geq C_0\|f\|_{L^{\frac{n}{n-1}}(\mathbb{R}^n)}$.*

*Proof* To show that the Sobolev inequality (2) implies the isoperimetric inequality (1), we use mollifiers. For $\varepsilon > 0$ set $f_\varepsilon := \rho_\varepsilon * \chi_E$, where $\rho_\varepsilon(x) = \varepsilon^{-n}\rho(x/\varepsilon)$ is a standard mollifier. Note that $f_\varepsilon \in W^{1,1}(\mathbb{R}^n)$ and that $f_\varepsilon \to \chi_E$ a.e. in $\mathbb{R}^n$. Then, fix $\varphi \in C^1_c(\mathbb{R}^n, \mathbb{R}^n)$ with $||\varphi||_\infty \leq 1$. Using the definition of $f_\varepsilon$, performing a change of variable and recalling Definition 1, we easily get

$$\begin{aligned} -\int_{\mathbb{R}^n} \nabla f_\varepsilon \cdot \varphi\,dx &= \int_{\mathbb{R}^n} f_\varepsilon\,\mathrm{div}\varphi\,dx = \int_{\mathbb{R}^n} dx \int_{\mathbb{R}^n} \rho_\varepsilon(z)\chi_E(x-z)\,\mathrm{div}\varphi(x)dz \\ &= \int_{\mathbb{R}^n} \rho_\varepsilon(z)dz \int_{\mathbb{R}^n} \chi_E(y)\,\mathrm{div}\varphi(y+z)dy \leq P(E)\int_{\mathbb{R}^n} \rho_\varepsilon(z)dz = P(E). \end{aligned}$$

Taking the supremum over all such $\varphi$, from (2) we get

$$P(E) \geq \int_{\mathbb{R}^n} |\nabla f_\varepsilon| \geq C_0 \|f_\varepsilon\|_{\frac{n}{n-1}} .$$

Hence (1) follows, letting $\varepsilon \to 0$ and recalling that $f_\varepsilon(x) \to \chi_E(x)$ for a.e. $x \in \mathbb{R}^n$.

To prove that the isoperimetric inequality implies the Sobolev inequality we are going to use the coarea formula (18). Note that by density it is enough to prove (2) for a function $f \in C^1_c(\mathbb{R}^n)$. Moreover, splitting $f$ in its positive and negative part, we may always assume without loss of generality that $f \geq 0$. Then, for any $t \geq 0$ we truncate $f$ from below by setting $f_t := \min\{f, t\}$. We set also $\phi(t) := \|f_t\|_{\frac{n}{n-1}}$. Note that $\phi$ is an increasing function and that for $h > 0$

$$\phi(t+h) - \phi(t) \leq \|f_{t+h} - f_t\|_{\frac{n}{n-1}} \leq h|\{f > t\}|^{1-1/n}$$

Thus $\phi$ is Lipschitz and $\phi'(t) \leq |\{f > t\}|^{1-1/n}$ for $\mathcal{H}^1$-a.e. $t \in \mathbb{R}$. Furthermore, using the isoperimetric inequality (14), we have

$$\begin{aligned} \|f\|_{\frac{n}{n-1}} &= \lim_{t\to+\infty} \phi(t) = \int_0^{+\infty} \phi'(s)ds \leq \int_0^{+\infty} |\{f > s\}|^{1-1/n} ds \\ &\leq C_0^{-1} \int_0^{+\infty} P(\{f > s\})ds = C_0^{-1} \int_{-\infty}^{+\infty} ds \int_{\partial^*\{f>s\}} d\mathcal{H}^{n-1} = C_0^{-1} \int_{\mathbb{R}^n} |\nabla f|, \end{aligned}$$

where the last equality follows from (18) with $g \equiv 1$. □

## 4 Stability of the Isoperimetric Inequality: Convex and Nearly Spherical Sets

After having proved the isoperimetric inequality we now turn to the next issue, namely the stability of this inequality. In other words, if $E$ is a set such that $|E| = |B_r|$ and $P(E) = P(B_r) + \delta$ for some small $\delta$, can we say that $E$ is somehow close to a ball? And how can we measure the distance from a ball in terms of $\delta$?

The first results in this direction were proven for planar convex sets by Bernstein [8] in 1905 and Bonnesen [11] in 1924. As we shall see in this section, it took some time before the problem was completely solved for convex sets in any dimension.

**Theorem 23 (Bonnesen)** *Given a convex set $E \subset \mathbb{R}^2$, with $|E| = |B|$, there exist two concentric disks $B_{r_1}(x_0) \subset E \subset B_{r_2}(x_0)$ such that*

$$(r_2 - r_1)^2 \leq \frac{P^2(E) - P^2(B)}{4\pi}. \tag{19}$$

**Fig. 10** Bonnesen's theorem

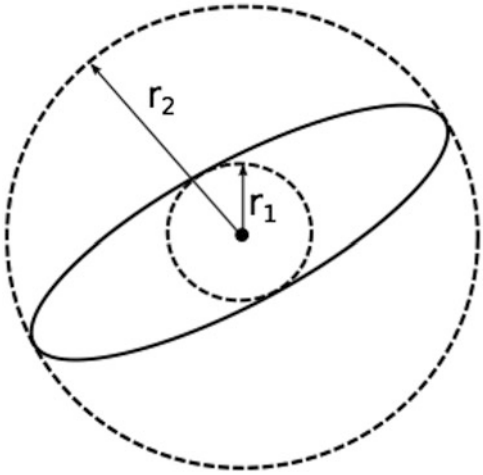


Figure 10 illustrates the statement of the theorem. A remarkable feature of inequality (19) is that the constant appearing on the right hand side is optimal. However, we cannot expect to prove also in higher dimension such a precise inequality. Thus, it may be useful to restate it in a weaker form that we may hope to extend to the general $n$-dimensional case. To this aim, observe that from (19) it follows that if $P(E) - P(B) \leq 1$ there exists $x_0 \in \mathbb{R}^2$ such that

$$d_H^2(E, B(x_0)) \leq C(P(E) - P(B))$$

for some positive constant $C$. Here and in the following we denote by

$$d_H(E, F) := \inf\{\varepsilon > 0 : \ E \subset F + B_\varepsilon, \ F \subset E + B_\varepsilon\}$$

the *Hausdorff distance* between any two sets $E, F \subset \mathbb{R}^n$.

*Remark 24* Let $\Omega \subset \mathbb{R}^n$ be a bounded open set. Set $\mathcal{C}(\overline{\Omega}) := \{K \subset \overline{\Omega} : K \text{ compact}\}$. Then the set $\mathcal{C}(\overline{\Omega})$, endowed with the Hausdorff distance is a compact metric space, see for instance [4, Theorem 6.1]. Moreover the convergence of $K_j$ to $K$ in the metric space $(\mathcal{C}(\overline{\Omega}), d_H)$ is equivalent to the two following conditions

(i) for all $x \in K$ there exist $x_j \in K_j$ such that $x_j \to x$;
(ii) if $x_j \in K_j$, then any limit point of the sequence $\{x_j\}$ belongs to $K$.

The convergence defined by (i) and (ii) is also known as *convergence in the sense of Kuratowski.*

Throughout all this section we shall only deal with sets $E$ of the same volume as $B$. This is not a restriction at all since all the statements that we shall prove under this assumption also apply to sets of any measure, up to a suitable rescaling.

**Definition 25** Let $E \subset \mathbb{R}^n$ be a convex set with $|E| = |B|$. We define the *isoperimetric deficit* and the *asymmetry index* of $E$ by setting

$$\mathcal{D}(E) := P(E) - P(B), \quad \mathcal{A}(E) := \min_{x \in R^n} d_H(E, B(x)),$$

respectively.

The extension of Bonnesen result Theorem 23 to high dimension was obtained by Fuglede in 1989, see [39].

**Theorem 26 (Fuglede)** *Let $n \geq 2$. There exist $\delta, C$, depending only on $n$, such that if $E$ is convex, $|E| = |B|$, and $\mathcal{D}(E) \leq \delta$, then:*

$$\mathcal{A}(E) \leq \begin{cases} C\sqrt{\mathcal{D}(E)}, & n = 2 \\ C\sqrt{\mathcal{D}(E)\log\left(\frac{1}{\mathcal{D}(E)}\right)}, & n = 3 \\ C(\mathcal{D}(E))^{\frac{2}{n+1}}, & n \geq 4. \end{cases} \tag{20}$$

As we already observed, for $n = 2$ the above estimate is just a weaker version of the more precise inequality (19). As shown in [39, Sect. 3] also when $n \geq 3$ the estimates above are optimal. In fact if $n \geq 4$ one cannot replace the power $\frac{2}{n+1}$ by a bigger one and if $n = 3$ one cannot remove the logarithm of $1/\mathcal{D}(E)$ from the right hand side of (20).

Fuglede's theorem is based on the following result for nearly spherical sets, that is sets which are very close to the unit ball, see Fig. 11. It turns out that for such sets one may estimate very precisely the distance from the ball by writing up the Taylor expansion of the perimeter. As we shall see, the next result will be also useful to prove the stability of the isoperimetric inequality for general sets of finite perimeter.

**Theorem 27** *Let $u : \mathbb{S}^{n-1} \to (-1, 1)$ be a Lipschitz function and let*

$$E := \{tz(1 + u(z)) : z \in \mathbb{S}^{n-1}, 0 \leq t < 1\}. \tag{21}$$

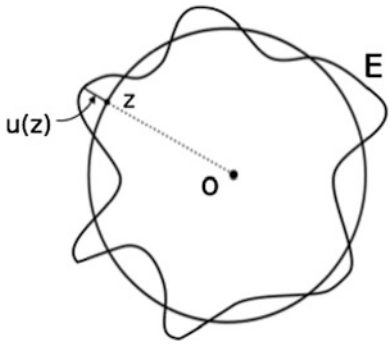


**Fig. 11** A nearly spherical set

*There exists $\varepsilon(n) > 0$ such that if $\|u\|_{W^{1,\infty}(\mathbb{S}^{n-1})} < \varepsilon$, $|E| = |B|$ and the barycenter of $E$ is the origin, then*

$$\mathcal{D}(E) \geq \frac{1}{4}\|\nabla_\tau u\|^2_{L^2(S^{n-1})} \geq \frac{1}{8\omega_n}|E\Delta B|^2. \tag{22}$$

Note that in (22) we have denoted by $\nabla_\tau u$ the tangential gradient of $u$ on $\mathbb{S}^{n-1}$. In the sequel we shall refer to a set $E \subset \mathbb{R}^n$ satisfying (21) as to *a nearly spherical set*. In order to prove Theorem 27 we need the formulas stated in the next lemma.

**Lemma 28** *Let $E$ be as in (21), with $\|u\|_{W^{1,\infty}(\mathbb{S}^{n-1})} < 1$. Then*

$$P(E) = \int_{\mathbb{S}^{n-1}} \sqrt{(1+u)^{2(n-1)} + (1+u)^{2(n-2)}|\nabla_\tau u|^2}\, d\mathcal{H}^{n-1}. \tag{23}$$

*Moreover,*

$$|E| = \frac{1}{n}\int_{\mathbb{S}^{n-1}} (1+u(z))^n\, d\mathcal{H}^{n-1}, \qquad \int_E x\, dx = \frac{1}{n+1}\int_{\mathbb{S}^{n-1}} z(1+u(z))^{n+1}\, d\mathcal{H}^{n-1}. \tag{24}$$

*Proof* We start by proving (24). To this aim we extend $u$ to $\mathbb{R}^n\backslash\{0\}$ by setting $u(x) := u(x/|x|)$ for all $x \neq 0$. In this way we have that $E = \Phi(B)$, where $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is the map $\Phi(x) := x(1+u(x))$, $x \in B$. Note that $D\Phi(x) = (1+u(x))I + x \otimes Du$ and that since $u$ is homogeneous of degree zero, then $x \cdot Du(x) = 0$ for all $x \neq 0$. Thus, recalling (11) we conclude that the Jacobian $J\Phi$ of $\Phi$ is given by $(1+u(x))^n$. Therefore

$$|E| = \int_B J\Phi\, dx = \int_B (1+u(x))^n\, dx = \int_0^1 r^{n-1}\, dr \int_{\mathbb{S}^{n-1}} (1+u(x))^n\, d\mathcal{H}^{n-1}.$$

Hence the first equality in (24) follows. The second one is obtained similarly.

Since $E$ is a bounded open set with Lipschitz boundary, $P(E) = \mathcal{H}^{n-1}(\partial E)$, see Example 10 or [4, Proposition 3.62]. Then, recalling that $\partial E = \Phi(\mathbb{S}^{n-1})$, from the *area formula*, see for instance [4, Theorem 2.92], we have

$$P(E) = \mathcal{H}^{n-1}(\partial E) = \int_{\mathbb{S}^{n-1}} J_{n-1}\Phi\, d\mathcal{H}^{n-1}, \tag{25}$$

where the $(n-1)$-*dimensional Jacobian* $J_{n-1}\Phi$ of the map $\Phi$ is given by

$$J_{n-1}\Phi = \sqrt{\det\big((d\Phi(x))^T \circ d\Phi(x)\big)}.$$

Here the linear map $d\Phi(z) : T_z\mathbb{S}^{n-1} \mapsto \mathbb{R}^n$ is the tangential differential of $\Phi$ defined in (10) and $(d\Phi(z))^T$ is its adjoint. Note that for any $\tau \in T_z\mathbb{S}^{n-1}$ we have $d\Phi(z)(\tau) = \tau(1+u(z)) + zD_\tau u(z)$, where $D_\tau u(z) = \nabla u(z) \cdot \tau$. Therefore the coefficients of

the matrix $d\Phi(z)$ relative to an orthonormal base $\{\tau_1,\dots,\tau_{n-1}\}$ of $T_z\mathbb{S}^{n-1}$ and to the standard base $\{e_1,\dots,e_n\}$ are given by $\tau_i\cdot e_h(1+u(z))+z_hD_{\tau_i}u(z)$, for $i=1,\dots,n-1$, $h=1,\dots,n$. Thus, for all $i,j\in\{1,\dots,n-1\}$, the coefficients $a_{ij}$ of the matrix $(d\Phi(z))^T\circ d\Phi(z)$ are given by

$$a_{ij}=\sum_{h=1}^{n}\big(\tau_i\cdot e_h(1+u)+z_hD_{\tau_i}u\big)\big(\tau_j\cdot e_h(1+u)+z_hD_{\tau_j}u\big)=\delta_{ij}(1+u)^2+D_{\tau_i}uD_{\tau_j}u,$$

where in the last equality we have used the fact that $\tau_i\cdot\tau_j=\delta_{ij}$ and $\tau_i\cdot z=0$ for all $i,j=1,\dots,n-1$. Hence, recalling (11) we have

$$J_{n-1}\Phi=\sqrt{\det(a_{ij})}=\sqrt{(1+u)^{2(n-1)}+(1+u)^{2(n-2)}|\nabla_\tau u|^2}$$

and thus (23) follows immediately from (25). □

We are now in position to give the proof of Theorem 27. The proof below follows closely the one given in [40] which has the advantage of avoiding some heavy computations of the original proof by Fuglede.

*Proof of Theorem 27* **Step 1.** From (23) we have

$$\begin{aligned}P(E)-P(B)&=\int_{\mathbb{S}^{n-1}}\left[(1+u)^{n-1}\sqrt{1+\frac{|\nabla_\tau u|^2}{(1+u)^2}}-1\right]d\mathcal{H}^{n-1}\\&=\int_{\mathbb{S}^{n-1}}\big[(1+u)^{n-1}-1\big]d\mathcal{H}^{n-1}\\&\quad+\int_{\mathbb{S}^{n-1}}(1+u)^{n-1}\left[\sqrt{1+\frac{|\nabla_\tau u|^2}{(1+u)^2}}-1\right]d\mathcal{H}^{n-1}.\end{aligned}$$

From the Taylor expansion of the square root it follows that for $t>0$ sufficiently small $\sqrt{1+t}\ge 1+\frac{t}{2}-\frac{t^2}{7}$. Hence, if $\varepsilon$ is small, from the assumption $\|u\|_{W^{1,\infty}(\mathbb{S}^{n-1})}<\varepsilon$ we get

$$\begin{aligned}P(E)-P(B)&\ge\int_{\mathbb{S}^{n-1}}\big[(1+u)^{n-1}-1\big]d\mathcal{H}^{n-1}\\&\quad+\int_{\mathbb{S}^{n-1}}(1+u)^{n-1}\left[\frac{1}{2}\frac{|\nabla_\tau u|^2}{(1+u)^2}-\frac{1}{7}\frac{|\nabla_\tau u|^4}{(1+u)^4}\right]d\mathcal{H}^{n-1}\\&\ge\int_{\mathbb{S}^{n-1}}\big[(1+u)^{n-1}-1\big]d\mathcal{H}^{n-1}+\left(\frac{1}{2}-C\varepsilon\right)\int_{\mathbb{S}^{n-1}}|\nabla_\tau u|^2d\mathcal{H}^{n-1},\end{aligned}\tag{26}$$

where $C$ is a constant depending only on $n$. From the first equality in (24) it follows that the assumption $|E| = |B|$ is equivalent to

$$\int_{\mathbb{S}^{n-1}} \left[(1+u)^n - 1\right] d\mathcal{H}^{n-1} = 0, \tag{27}$$

that is

$$\int_{\mathbb{S}^{n-1}} \left( nu + \sum_{h=2}^{n} \binom{n}{h} u^h \right) d\mathcal{H}^{n-1} = 0. \tag{28}$$

From this identity, recalling again that $\|u\|_{L^\infty(\mathbb{S}^{n-1})} < \varepsilon$, we have

$$\int_{\mathbb{S}^{n-1}} u d\mathcal{H}^{n-1} \geq -\frac{n-1}{2} \int_{\mathbb{S}^{n-1}} u^2 d\mathcal{H}^{n-1} - C\varepsilon \int_{\mathbb{S}^{n-1}} u^2 d\mathcal{H}^{n-1}.$$

Therefore, using this last inequality and the smallness assumption, we may estimate

$$\begin{aligned}
\int_{\mathbb{S}^{n-1}} [(1+u)^{n-1} - 1] d\mathcal{H}^{n-1} &= (n-1) \int_{\mathbb{S}^{n-1}} u d\mathcal{H}^{n-1} + \sum_{h=2}^{n-1} \binom{n-1}{h} \int_{\mathbb{S}^{n-1}} u^h d\mathcal{H}^{n-1} \\
&\geq (n-1) \int_{\mathbb{S}^{n-1}} u d\mathcal{H}^{n-1} + \frac{(n-1)(n-2)}{2} \int_{\mathbb{S}^{n-1}} u^2 d\mathcal{H}^{n-1} \\
&\quad - C\varepsilon \int_{\mathbb{S}^{n-1}} u^2 d\mathcal{H}^{n-1} \\
&\geq -\frac{n-1}{2} \int_{\mathbb{S}^{n-1}} u^2 d\mathcal{H}^{n-1} - C\varepsilon \int_{\mathbb{S}^{n-1}} u^2 d\mathcal{H}^{n-1}.
\end{aligned}$$

In conclusion, recalling (26), we have proved that if $\|u\|_{W^{1,\infty}(\mathbb{S}^{n-1})} \leq \varepsilon$, then

$$P(E) - P(B) \geq \left( \frac{1}{2} - C\varepsilon \right) \int_{\mathbb{S}^{n-1}} |\nabla_\tau u|^2 d\mathcal{H}^{n-1} - \left( \frac{n-1}{2} + C\varepsilon \right) \int_{\mathbb{S}^{n-1}} u^2 d\mathcal{H}^{n-1}, \tag{29}$$

for some constant $C$ depending only on the dimension $n$.

**Step 2.** Now, for any integer $k \geq 0$, let us denote by $y_{k,i}$, $i = 1, \ldots, G(n,k)$, the spherical harmonics of order $k$, i.e., the restriction to $\mathbb{S}^{n-1}$ of the homogeneous harmonic polynomials of degree $k$, normalized so that $||y_{k,i}||_{L^2(\mathbb{S}^{n-1})} = 1$, for all $k$ and for $i \in \{1, \ldots, G(n,k)\}$. Taking into account the normalization, we have that $y_0 = 1/\sqrt{n\omega_n}$ and $y_{1,i} = z_i/\sqrt{\omega_n}$, for $i = 1, \ldots, n$. Recall that the polynomials $y_{k,i}$ are eigenfunctions of the Laplace-Beltrami operator on $\mathbb{S}^{n-1}$ and that for all $k$ and $i$

$$-\Delta_{\mathbb{S}^{n-1}} y_{k,i} = k(k+n-2) y_{k,i}\,.$$

Therefore if we write

$$u = \sum_{k=0}^{\infty} \sum_{i=1}^{G(n,k)} a_{k,i} y_{k,i}, \quad \text{where} \quad a_{k,i} = \int_{\mathbb{S}^{n-1}} u y_{k,i} d\mathcal{H}^{n-1},$$

we have

$$||u||^2_{L^2(\mathbb{S}^{n-1})} = \sum_{k=0}^{\infty} \sum_{i=1}^{G(n,k)} a^2_{k,i}, \quad ||\nabla_\tau u||^2_{L^2(\mathbb{S}^{n-1})} = \sum_{k=1}^{\infty} k(k+n-2) \sum_{i=1}^{G(n,k)} a^2_{k,i}. \tag{30}$$

Observe that from formula (28) we have

$$a_0 = \frac{1}{\sqrt{n\omega_n}} \int_{\mathbb{S}^{n-1}} u d\mathcal{H}^{n-1} = -\frac{1}{n\sqrt{n\omega_n}} \sum_{h=2}^{n} \binom{n}{h} \int_{\mathbb{S}^{n-1}} u^h d\mathcal{H}^{n-1},$$

hence

$$|a_0| \le C||u||_2^2 \le C\varepsilon||u||_2 .$$

From the assumption that the barycenter of $E$ is at the origin and from the second equality in (24) we have

$$\int_{\mathbb{S}^{n-1}} z(1+u(z))^{n+1} \, d\mathcal{H}^{n-1} = 0 .$$

Then, using the equality $\int_{\mathbb{S}^{n-1}} z = 0$ and arguing as before, we immediately get that for all $i = 1, \dots, n$,

$$|a_{1,i}| = \left| \frac{1}{\sqrt{\omega_n}} \int_{\mathbb{S}^{n-1}} u z_i \, d\mathcal{H}^{n-1} \right| \le C\varepsilon||u||_2.$$

Therefore, from (30) we get

$$||u||_2^2 \le C\varepsilon^2||u||_2^2 + \sum_{k=2}^{\infty} \sum_{i=1}^{G(n,k)} |a_{k,i}|^2 \Longrightarrow ||u||_2^2 \le \frac{1}{1-C\varepsilon} \sum_{k=2}^{\infty} \sum_{i=1}^{G(n,k)} |a_{k,i}|^2.$$

But since for $k \ge 2$, $k(k+n-2) \ge 2n$, from (30) we have

$$||u||_2^2 \le \frac{1}{2n(1-C\varepsilon)} ||\nabla_\tau u||_2^2$$

and thus, recalling (29) and choosing $\varepsilon$ sufficiently small, in dependence on $n$, we get

$$P(E) - P(B) \geq \Big(\frac{1}{2} - C\varepsilon\Big) \int_{\mathbb{S}^{n-1}} |\nabla_\tau u|^2 d\mathcal{H}^{n-1} - \Big(\frac{n-1}{2} + C\varepsilon\Big) \frac{1}{2n(1-C\varepsilon)} ||\nabla_\tau u||_2^2$$
$$\geq \frac{1}{4} \int_{\mathbb{S}^{n-1}} |\nabla_\tau u|^2 d\mathcal{H}^{n-1} \geq \frac{n}{3} ||u||^2_{L^2(\mathbb{S}^{n-1})} \geq \frac{1}{3\omega_n} ||u||^2_{L^1(\mathbb{S}^{n-1})} . \tag{31}$$

This proves the first inequality in (22). To get the second inequality we observe that, choosing again $\varepsilon$ sufficiently small

$$|E \Delta B| = \frac{1}{n} \int_{\mathbb{S}^{n-1}} |(1 + u(x))^n - 1| \, d\mathcal{H}^{n-1} \leq \frac{n+1}{n} \int_{\mathbb{S}^{n-1}} |u| \, d\mathcal{H}^{n-1} .$$

Therefore, from the last inequality of (31) we conclude that

$$P(E) - P(B) \geq \frac{1}{3\omega_n} ||u||^2_{L^1(\mathbb{S}^{n-1})} \geq \frac{n^2}{3(n+1)^2 \omega_n} |E \Delta B|^2 \geq \frac{1}{8\omega_n} |E \Delta B|^2 .$$

□

The theorem we have just proved allows us to estimate the distance in $W^{1,2}$ of a nearly spherical set $E$ from the unit ball with the isoperimetric deficit. Now, an interpolation result will tell us that indeed we may also control the $L^\infty$ distance, hence the Hausdorff distance, between $E$ and $B$. For the proof see [39, Lemma 1.4].

**Lemma 29 (Interpolation Lemma)** *If $v \in W^{1,\infty}(\mathbb{S}^{n-1})$ and $\int_{\mathbb{S}^{n-1}} v = 0$, then*

$$||v||^{n-1}_{L^\infty(\mathbb{S}^{n-1})} \leq \begin{cases} \pi ||\nabla_\tau v||_2, & n = 2 \\ 4||\nabla_\tau v||_2^2 \log \dfrac{8e||\nabla_\tau v||_\infty}{||\nabla_\tau v||_2^2}, & n = 3 \\ C||\nabla_\tau v||_2^2 ||\nabla_\tau v||_\infty^{n-3}, & n \geq 4, \end{cases}$$

*where the constant $C$ depends only on the dimension.*

Combining Lemma 29 with Theorem 27 we immediately get the estimate of the $L^\infty$ distance between a nearly spherical set $E$ and the unit ball.

**Theorem 30** *Under the assumptions of Theorem 27, there exist $\varepsilon, C > 0$ depending only on $n$ such that if $||u||_{W^{1,\infty}(\mathbb{S}^{n-1})} \leq \varepsilon$, then*

$$||u||^{n-1}_{L^\infty(\mathbb{S}^{n-1})} \leq \begin{cases} C\sqrt{\mathcal{D}(E)}, & n = 2 \\ C\mathcal{D}(E) \log\left(\dfrac{1}{\mathcal{D}(E)}\right), & n = 3 \\ C\mathcal{D}(E)||\nabla_\tau u||_\infty^{n-3}, & n \geq 4. \end{cases}$$

*Proof* Set $v := \dfrac{(1+u)^n - 1}{n}$. From the volume constraint $|E| = |B|$ we have, see (27),

$$\int_{\mathbb{S}^{n-1}} v \, d\mathcal{H}^{n-1} = \frac{1}{n} \int_{\mathbb{S}^{n-1}} \big[(1+u)^n - 1\big] d\mathcal{H}^{n-1} = 0.$$

Moreover, since

$$v = u + \frac{1}{n} \sum_{h=2}^{n} \binom{n}{h} u^h,$$

if $\varepsilon > 0$ is small enough we have

$$\frac{1}{2}|u| \le |v| \le 2|u|, \qquad \frac{1}{2}|\nabla_\tau u| \le |\nabla_\tau v| \le 2|\nabla_\tau u|.$$

Then the result follows immediately from Theorem 27 and the interpolation Lemma 29. □

Let us now consider the case of a convex set with small isoperimetric deficit and let us indicate the main steps in the proof of Fuglede's Theorem 26. The first step, see Lemma 32, is to show that a convex set with small isoperimetric deficit is close in the Hausdorff distance to a ball with the same volume. At this stage, however, we are not yet able to quantify how close is the set to the ball in terms of the isoperimetric deficit. Next, we observe that if a convex set is close in the Hausdorff sense to a ball of the same volume, then it is also close to the same ball in $W^{1,\infty}$, see Lemma 33. Then, the final step of the proof consists in combining these observations with the precise estimate provided by Theorem 30.

Let us start with a simple lemma relating the diameter $\mathrm{diam}(E)$ of a convex set $E$ with its volume and perimeter. To this aim, let us recall that

$$P(E) \le P(F) \qquad \text{if } E,\ F \text{ are convex and } E \subset F\,. \tag{32}$$

**Lemma 31** *Let $E \subset \mathbb{R}^n$ be a bounded open convex set. Then*

$$\mathrm{diam}(E) \le c(n) \frac{[P(E)]^{n-1}}{|E|^{n-2}}.$$

*Proof* First observe that if $n = 2$ we trivially have $\mathrm{diam}(E) \le \frac{1}{2}P(E)$.

So let us assume $n \ge 3$. Let $x, y \in \partial E$ be such that $\mathrm{diam}(E) := d = |x-y|$. Then, rotate and translate $E$ so to reduce to the situation shown in Fig. 12.

By Fubini's Theorem, $|E| = \int_0^d \mathcal{H}^{n-1}(E_t)dt$, where $E_t = E \cap \{x_n = t\}$. Observe that there exists $s \in (0, d)$ such that $\mathcal{H}^{n-1}(E_s) \ge |E|/d$. Note that we may always assume that $0 < s \le d/2$ (otherwise we just rotate $E$ upside down). Let $C$ be the

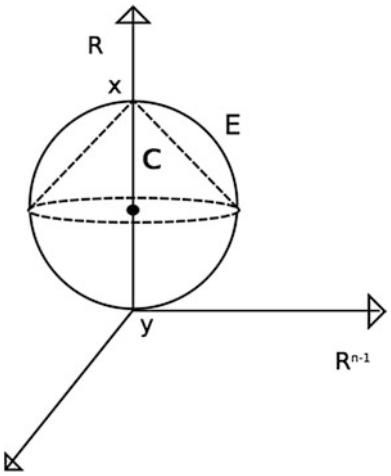


**Fig. 12** The construction in the proof of Lemma 31

cone in Fig. 12 with base $E_s$ and vertex $x$. Using the coarea formula (12) and (32) we may estimate

$$P(E) \geq P(C) \geq \mathcal{H}^{n-1}(\partial C \setminus E_s) = \int_s^d dt \int_{\partial C_t} \frac{1}{|\nu_t^C|}\, d\mathcal{H}^{n-2} \geq \int_s^d \mathcal{H}^{n-2}(\partial C_t)dt$$

$$= \int_s^d \left(\frac{d-t}{d-s}\right)^{n-2} \mathcal{H}^{n-2}(\partial E_s)dt = \frac{(d-s)\mathcal{H}^{n-2}(\partial E_s)}{n-1} \geq \frac{d}{2}\frac{\mathcal{H}^{n-2}(\partial E_s)}{n-1}.$$

From the isoperimetric inequality (17) we get

$$\mathcal{H}^{n-2}(\partial E_s) \geq (n-1)\omega_{n-1}^{1/(n-1)} \left[\mathcal{H}^{n-1}(E_s)\right]^{\frac{n-2}{n-1}} \geq (n-1)\omega_{n-1}^{1/(n-1)} \left(\frac{|E|}{d}\right)^{\frac{n-2}{n-1}}.$$

Thus,

$$P(E) \geq c(n)d \left(\frac{|E|}{d}\right)^{\frac{n-2}{n-1}},$$

whence the result follows. □

Let us now prove that a convex set with small isoperimetric deficit is close in the Hasudorff distance to a ball.

**Lemma 32** *For all $\varepsilon > 0$, there exists $\delta_\varepsilon > 0$ such that if $E$ is convex, $|E| = |B|$, the barycenter of $E$ is the origin and $\mathcal{D}(E) < \delta_\varepsilon$, then there exists a function $u \in W^{1,\infty}(\mathbb{S}^{n-1})$, with $||u||_{L^\infty(\mathbb{S}^{n-1})} \leq \varepsilon$, and such that*

$$E := \{tz(1+u(z)) : z \in \mathbb{S}^{n-1},\ 0 \leq t < 1\}.$$

*Proof* We argue by contradiction. Assume that there exist $\varepsilon_0 > 0$ and a sequence of closed convex sets $E_j$ such that $|E_j| = |B|$, the barycenter of $E_j$ is the origin, $\mathcal{D}(E_j) \to 0$, but $||u_j||_{L^\infty(\mathbb{S}^{n-1})} \geq \varepsilon_0$, where $u_j$ is the Lipschitz function representing $E_j$ as in (21). From Lemma 31 it follows that the sets $E_j$ are equibounded and so, recalling Remark 24, we may assume that they converge in the Hausdorff distance to a closed set $E$. Note that $E$ is convex and that the sequence $E_j$ converge to $E$ also in measure. In particular $|E| = |B|$. Since $\mathcal{D}(E_j) \to 0$, we have that $P(E_j) \to P(B)$. Therefore, from the isoperimetric inequality and the lower semicontinuity of the perimeter we get that

$$P(B) \leq P(E) \leq \lim_{j\to\infty} P(E_j) = P(B)\,.$$

Thus $E$ is a ball, actually the unit ball centered at the origin, since all the $E_j$ have barycenter at the origin. This gives a contradiction, since the $E_j$ are converging in the Hausdorff sense to the unit ball $B$, while $||u_j||_{L^\infty(\mathbb{S}^{n-1})} \geq \varepsilon_0$ for all $j$. □

The following lemma shows that the Hausdorff distance of a convex set from a ball controls indeed also its distance in $W^{1,\infty}$.

**Lemma 33** *Let $E$ is a convex set such that*

$$E := \{tz(1 + u(z)) : z \in \mathbb{S}^{n-1},\, 0 \leq t < 1\}\,.$$

*for some Lipschitz function $u : \mathbb{S}^{n-1} \to (-1/2, 1/2)$. Then*

$$\|\nabla_\tau u\|_{L^\infty} \leq 2\sqrt{\|u\|_{L^\infty}}\frac{1 + \|u\|_{L^\infty}}{1 - \|u\|_{L^\infty}}.$$

*Proof* Let us fix $P_z \in \partial E$ and let $z \in \mathbb{S}^{n-1}$ be such that $P_z = z(1 + u(z))$ and $u$ is differentiable at $z$. Recall that the tangent plane $T_x\partial E$ is spanned by the vectors $(1 + u(z))\tau_i + z\nabla u(z) \cdot \tau_i$, where $\{\tau_1, \ldots, \tau_{n-1}\}$ is an orthonormal base for $T_z\mathbb{S}^{n-1}$. Therefore the exterior normal to $E$ at $P_z$ is given by

$$\nu^E(P_z) = \frac{z(1 + u(z)) - \nabla_\tau u(z)}{\sqrt{(1 + u(z))^2 + |\nabla_\tau u(z)|^2}}\,. \tag{33}$$

Since $z \cdot \nabla_\tau u(z) = 0$, we have

$$z \cdot \nu^E(P_z) = \frac{1 + u(z)}{\sqrt{(1 + u(z))^2 + |\nabla_\tau u(z)|^2}}\,.$$

Then, denoting by $H$ the projection of the origin on the tangent plane to $E$ at $P_z$, we have, see Fig. 13, $\frac{\overline{OH}}{\overline{OP_z}} = z \cdot \nu^E(P_z)$. Observe that

$$\overline{OP_z} \leq 1 + ||u||_\infty, \quad \overline{OH} \geq 1 - ||u||_\infty,$$

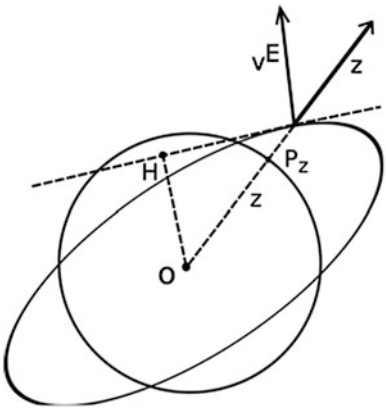


**Fig. 13** The construction in the proof of Lemma 33

where the second inequality follows by the convexity of $E$. Thus,

$$\frac{1-||u||_\infty}{1+||u||_\infty} \le z \cdot \nu^E(P_z) = \frac{1+u(z)}{\sqrt{(1+u(z))^2+|\nabla_\tau u(z)|^2}},$$

from which we get

$$\frac{|\nabla_\tau u(z)|^2}{(1+u(z))^2} \le \left(\frac{1+||u||_\infty}{1-||u||_\infty}\right)^2 - 1 = \frac{4||u||_\infty}{(1-||u||_\infty)^2},$$

thus concluding

$$|\nabla_\tau u(z)|^2 \le 4||u||_\infty \left(\frac{1+||u||_\infty}{1-||u||_\infty}\right)^2,$$

whence the assertion follows. □

Let us conclude this section by giving the

*Proof of Theorem 26* Let as assume $n \ge 4$, since otherwise the proof is similar and even easier.

From Lemmas 32 and 33 it follows that if $E$ is a convex set with $|E| = |B|$ and $\mathcal{D}(E)$ is sufficiently small, then, up to a translation, $E$ is a nearly spherical set as in (21) with barycenter at the origin, satisfying $\|u\|_{W^{1,\infty}} < \varepsilon$, where $\varepsilon > 0$ is the one provided by Theorem 30. Therefore, using this theorem and Lemma 33 again, we get

$$||u||_\infty^{n-1} \le c\mathcal{D}(E)||\nabla_\tau u||_\infty^{n-3} \le c||u||_\infty^{\frac{n-3}{2}} \mathcal{D}(E)\,.$$

hence, $||u||_\infty^{\frac{n+1}{2}} \le c\mathcal{D}(E)$. Thus we may conclude that

$$\mathcal{A}(E) \le d_H(E,B) = ||u||_\infty \le c\,[\mathcal{D}(E)]^{\frac{2}{n+1}}\,. \qquad \square$$

## 5 Stability of the Isoperimetric Inequality: Proof by Symmetrization

We now discuss the quantitative isoperimetric inequality for general sets of finite perimeter. In this case it is clear that we cannot use the Hausdorff distance to measure the distance of a set $E$ from a ball, since a set with the same volume of the ball $B$ and a slightly larger perimeter may have small far away pieces or tiny long tentacles. Taking into account these examples it is then reasonable to introduce the so called *Fraenkel asymmetry* which is defined, for any measurable set $E$ of finite measure, as

$$\alpha(E) := \inf_{x\in\mathbb{R}^n} \left\{ \frac{|E\Delta B_r(x)|}{r^n} \,:\, |E| = |B_r| \right\}.$$

Note that the above infimum is always attained. In the following we shall refer to a minimizer of the right hand side as to an *optimal ball* for $E$. Clearly, optimal balls do not need to be unique. Observe also that, since $|E\Delta B_r(x)|$ is exactly the $L^1$ distance between $\chi_E$ and $\chi_{B_r(x)}$, $\alpha(E)$ can be regarded as the normalized $L^1$ distance of $E$ from its optimal ball. It is convenient to normalize also the isoperimetric deficit by setting

$$D(E) := \frac{P(E) - P(B_r)}{r^{n-1}},$$

where $|B_r| = |E|$.

In 1992 Hall [47], using some previous results proved in collaboration with by Hayman and Weitsman [48], showed that there exists a constant $c(n)$ such that for all measurable sets of finite measure

$$\alpha(E)^4 \le c(n)D(E)\,. \tag{34}$$

Note that the power on the left hand side of (34) is independent of the dimension of the ambient space. Note also that an inequality of this kind becomes critical only when the set $E$ is a small perturbation of the ball. As an example consider for any $n \ge 2$ the ellipsoid

$$E_\varepsilon = \left\{ \frac{x_1^2}{1+\varepsilon} + x_2^2(1+\varepsilon) + x_3^2 + \ldots + x_n^2 \le 1 \right\},$$

with $\varepsilon > 0$. Then $\alpha(E_\varepsilon) = |E_\varepsilon \Delta B|$, see [6, Lemma 5.9] and it is not difficult to show that

$$\frac{D(E_\varepsilon)}{\alpha^2(E_\varepsilon)} \to \gamma > 0, \quad \text{as } \varepsilon \to 0^+.$$

This example led Hall to conjecture in [47] that inequality (34) should hold in any dimension with the (optimal) exponent 2. This was proved by Maggi, Pratelli and the author in [44]. The precise statement goes as follows.

**Theorem 34 (Quantitative Isoperimetric Inequality)** *There exists a constant $\kappa(n)$ such that for any measurable set $E$ of finite measure*

$$\alpha(E)^2 \leq \kappa(n) D(E). \tag{35}$$

In this section we are going to discuss the proof of this result originally given in [44], which relies mostly on symmetrization arguments.

Note that inequality (35) can be rewritten in the following equivalent way: if $|E| = |B_r|$, then

$$P(E) \geq P(B_r)\left(1 + \frac{1}{n\omega_n \kappa(n)} \alpha(E)^2\right).$$

Thus the asymmetry index $\alpha(E)$ estimates from below the second order term in the Taylor expression of $P(E)$ in terms of $P(B_r)$.

Before going into the proof of (35) let us make some preliminary remarks.

First, observe that since both $\alpha(E)$ and $D(E)$ are scale invariant, to prove Theorem 34 we may always assume $|E| = |B|$. Note also that if one proves (35) for a set with small isoperimetric gap, i.e., $D(E) \leq \delta_0$, then the general case follows. As a matter of fact, if $D(E) > \delta_0$ and $|E| = |B_r|$, then

$$\alpha(E) \leq \frac{|E \Delta B_r|}{r^n} \leq 2\omega_n \leq \frac{2\omega_n}{\sqrt{\delta_0}} \sqrt{D(E)}.$$

The strategy of the proof consists in reducing the general case to more and more special classes of sets. Precisely, in the first step one reduces to sets contained in a sufficiently large square, see Lemma 35. Then one wants to reduce to bounded *n-symmetric* sets, i.e., sets which are symmetric with respect to $n$ orthogonal hyperplanes, Theorem 40. These sets have the nice property that the ball centered at their center of symmetry is "almost optimal" in the sense stated in Lemma 38. The last reduction consists in passing from $n$-symmetric to axially symmetric sets whose profile is obtained by rotating a one-dimensional graph. Note that the proof of the quantitative isoperimetric inequality (35) for axially symmetric sets was already contained in Hall's paper [47, Theorem 2]. Different proofs are given in [44, Sect. 4] and in [50, Sect. 7]. The approach to stability issues via symmetrization has been used also used to deal with the Sobolev inequality, the isoperimetric inequality in

Gauss space and with other relevant geometric and functional inequalities, see for instance [6, 7, 13, 21, 22, 32, 33, 35, 43, 45, 46], and also [41, 52, 58].

The first reduction step is provided by the next result, see [44, Lemma 5.1]).

**Lemma 35** *There exist positive constants $L, C, \delta$ depending only on $n$ such that if $|E| = |B|$ and $D(E) \le \delta$ one can find a set $F \subset [-L, L]^n$, with $|F| = |B|$, such that*

$$\alpha(E) \le \alpha(F) + CD(F) \text{ and } D(F) \le CD(E).$$

We will not give the detailed proof of this lemma, which consists in cutting the far away parts of $E$ and rescaling the remaining part of the set. The main ingredients of the proof are the isoperimetric inequality and the strict concavity of the function $t^{\frac{n-1}{n}}$ for $t > 0$, which allows to estimate in a quantitative way the asymmetry created by splitting a set in two parts. To understand how this estimate works observe that for all $\lambda \in (0, 1)$

$$\lambda^{\frac{n-1}{n}} + (1-\lambda)^{\frac{n-1}{n}} - 1 \ge c(n) \min\{\lambda, 1-\lambda\}. \tag{36}$$

Let $E = B_r(x) \cup B_\rho(y)$ the union of two disjoint balls such that $|E| = |B|$ and $r \ge \rho$. Then

$$r^n + \rho^n = 1$$

and from (36) we may estimate the isoperimetric deficit of $E$ by

$$D(E) = P(B_r(x)) + P(B_\rho(y)) - P(B) \ge c(n) \min\{r^n, \rho^n\} \ge c(n)\rho^n .$$

Hence the estimate on the Fraenkel asymmetry of $E$ immediately follows:

$$\frac{1}{2}\alpha(E) \le |B(x) \setminus B_r(x)| = \omega_n(1 - r^n) = \omega_n \rho^n \le \omega_n c(n) D(E) .$$

It is clear how to use Lemma 35. Indeed if the quantitative isoperimetric inequality (35) holds for a bounded set, then, given any set $E$ with $|E| = |B|$ and $D(E) \le 1$, denoting by $F$ the set provided by Lemma 35, we have

$$\alpha(E) \le \alpha(F) + CD(F) \le \sqrt{\kappa(n) D(F)} + CD(F) \le C' \sqrt{D(E)} .$$

for a constant $C'$ depending only on $n$.

Thus, from now one we may assume without loss of generality that the set $E$ has volume $\omega_n$, that $E \subset [-L, L]^n$, for some given $L > 0$, and that $D(E) \le \delta$ for some conveniently small $\delta$.

The advantage of working with bounded sets is that in this case the compactness theorem for sets of finite perimeter Theorem 13 implies that $\alpha(E)$ depends continuously on $D(E)$.

**Lemma 36** *Let $L > 0$. For any $\varepsilon > 0$ there exists $\delta > 0$ such that if $E \subset [-L, L]^n$, $|E| = |B|$, and $D(E) \leq \delta$ then $\alpha(E) \leq \varepsilon$.*

*Proof* The proof is by contradiction. Assume that there exist $\varepsilon > 0$ and a sequence of sets $E_j \subset [-L, L]^n$, with $|E_j| = |B|$, $D(E_j) \to 0$ and $\alpha(E_j) \geq \varepsilon > 0$ for all $j \in \mathbb{N}$. Since the sets $E_j$ are equibounded, by Theorem 15 we may assume that up to a not relabeled subsequence the $E_j$ converge in measure to some set $E_\infty$ of finite perimeter. Thus $|E_\infty| = |B|$, and by the lower semicontinuity of the perimeters $P(E_\infty) \leq P(B)$, so $E_\infty$ is a ball. However the convergence in measure of $E_j$ to $E_\infty$ immediately implies that $|E_j \Delta E_\infty| \to 0$, against the assumption $\alpha(E_j) \geq \varepsilon$. The contradiction concludes the proof. □

Next step in the proof of the quantitative isoperimetric inequality is to reduce to the simpler case of an $n$-symmetric set.

**Definition 37** We say that $E \subset \mathbb{R}^n$ is *$n$-symmetric* if, up to a translation and a rotation, $E$ is symmetric about each coordinate plane.

Note that even if $E$ is $n$-symmetric it is not true in general that the optimal ball is the one centered at the center of symmetry of $E$, as shown in Fig. 14. However, the next lemma shows that for $n$-symmetric sets this ball is optimal "up to a constant".

**Lemma 38** *Let $E$ be $n$-symmetric with centre of symmetry at the origin, $|E| = |B|$. Then*

$$\alpha(E) \leq |E \Delta B| \leq 3\alpha(E)$$

*Proof* Let $B(x_0)$ be an optimal ball for $E$, i.e. $\alpha(E) = |E \Delta B(x_0)|$. Then by the triangular inequality we have

$$|E \Delta B| \leq |E \Delta B(x_0)| + |B(x_0) \Delta B|.$$

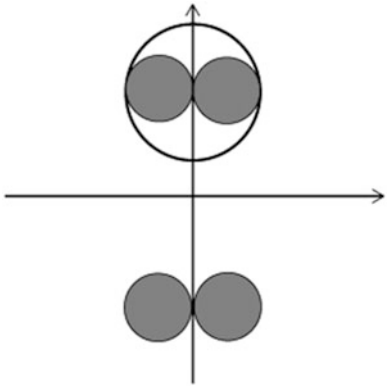

**Fig. 14** An optimal ball not centered at the origin

Note that since $E$ is $n$-symmetric $B(-x_0)$ is optimal as well, i.e., $\alpha(E) = |E\Delta B(-x_0)|$. Therefore from the inequality above we have

$$\begin{aligned}\alpha(E) \le |E\Delta B| &\le \alpha(E) + |B(x_0)\Delta B| \le \alpha(E) + |B(x_0)\Delta B(-x_0)| \\ &\le \alpha(E) + |E\Delta B(x_0)| + |E\Delta B(-x_0)| = 3\alpha(E).\end{aligned}$$

□

The next step is to reduce the proof of the quantitative isoperimetric inequality (35) to $n$-symmetric bounded sets. But before discussing how this can be done let us first introduce a few definitions.

Given a direction $\nu \in \mathbb{S}^{n-1}$ and a measurable set $E$ of finite measure, let us consider the affine hyperplane $\pi_\nu$ orthogonal to $\nu$ splitting $E$ into two parts of equal measure. We denote by $E'$ the part of $E$ contained in the open half space $H_\nu^+$ with inner normal $\nu$ and by $E''$ the part of $E$ contained in the open half space $H_\nu^-$ with inner normal $-\nu$. Then, we set $E_\nu^+ := E' \cup r_\nu(E')$, where $r_\nu$ is the reflection about the hyperplane $\pi_\nu$ and $E_\nu^- := E'' \cup r_\nu(E'')$. See Fig. 15 where, to simplify the notation, we dropped the subscript $\nu$. We claim that

$$P(E_\nu^+) + P(E_\nu^-) \le 2P(E) \tag{37}$$

with the inequality being possibly strict. To see this observe that from the definition of density we easily have that

$$\left[E^{(0)} \cup E^{(1)}\right] \cap \pi_\nu \subset \left[(E_\nu^+)^{(0)} \cup (E_\nu^+)^{(1)}\right] \cap \left[(E_\nu^-)^{(0)} \cup (E_\nu^-)^{(1)}\right] \cap \pi_\nu\,.$$

Therefore, from the definition of measure theoretic boundary given in (9) we deduce in particular that $\partial^M E_\nu^\pm \cap \pi_\nu \subseteq \partial^M E \cap \pi_\nu$ and thus, recalling Theorem 9, $\mathcal{H}^{n-1}(\partial^* E_\nu^\pm \cap \pi_\nu) \le \mathcal{H}^{n-1}(\partial^* E \cap \pi_\nu)$. Hence, (37) follows, since

$$\begin{aligned}P(E_\nu^+) + P(E_\nu^-) &= 2P(E \cap H_\nu^+) + \mathcal{H}^{n-1}(\partial^* E_\nu^+ \cap \pi_\nu) + 2P(E \cap H_\nu^-) \\ &\quad + \mathcal{H}^{n-1}(\partial^* E_\nu^- \cap \pi_\nu) \\ &\le 2P(E \cap H_\nu^+) + 2P(E \cap H_\nu^-) + 2\mathcal{H}^{n-1}(\partial^* E \cap \pi_\nu) = 2P(E)\,.\end{aligned}$$

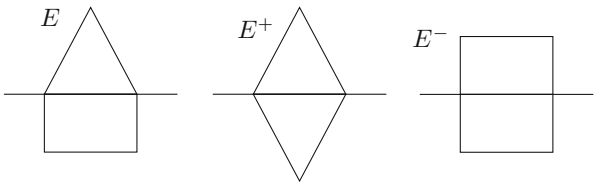


**Fig. 15** The sets $E^+$ and $E^-$ are obtained by reflecting the upper and lower half of $E$ with respect to the horizontal plane

Observe that inequality (37) implies that

$$D(E_\nu^+) + D(E_\nu^-) \le 2D(E)\,. \tag{38}$$

Therefore if we could prove that for some positive constant $C_0$ depending only on $n$

$$\alpha(E) \le C_0[\alpha(E_\nu^+) + \alpha(E_\nu^-)], \tag{39}$$

we would conclude that, setting $F$ either equal to $E_\nu^+$ or $E_\nu^-$, then

$$\alpha(E) \le 2C_0\alpha(F), \qquad D(F) \le 2D(E)\,.$$

Then by applying this argument to all coordinate directions we would find a $n$-symmetric set $G$ with the same volume of $E$ such that

$$\alpha(E) \le 2^n C_0^n \alpha(G), \qquad D(G) \le 2^n D(E)$$

and from these inequalities we would conclude that in order to prove (35) for $E$ it is enough to prove it for the $n$-symmetric set $G$.

Inequality (39) is not true in general as we can see looking at the set $E$ in Fig. 16. In fact, by reflecting the upper and lower halves of $E$ with respect to the horizontal plane we get that $E^\pm$ are both balls, hence $\alpha(E^\pm) = 0$. However, if we symmetrize the same set with respect to the vertical direction the asymmetry index may even increase, as one can see in Fig. 17.

The following lemma shows that the phenomenon illustrated by this example is a general fact. Indeed, if for some $\nu$ the asymmetry of $E_\nu^+$ and $E_\nu^-$ is much lower than the one of $E$, then given any other orthogonal direction $\nu'$, at least one of the two sets $E_{\nu'}^\pm$ has a larger asymmetry than $E$, up to a multiplicative constant depending only on the dimension.

**Lemma 39** *There exist $\delta, C$, depending only on $n$, such that if $E \subset [-L, L]^n$, $|E| = |B|$ and $D(E) \le \delta$, given any two orthogonal direction $\nu_1, \nu_2$ and the four sets*

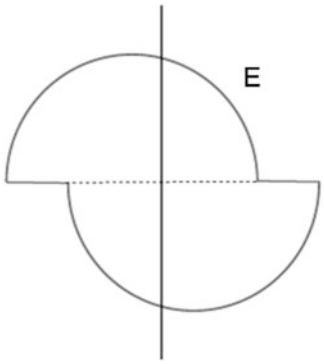


**Fig. 16** A set for which (39) is not true

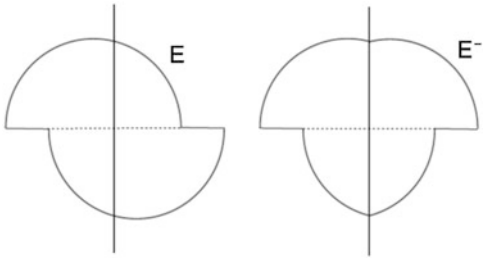


**Fig. 17** A different symmetrization may give a bigger asymmetry

*$E^+_{\nu_1}, E^-_{\nu_1}, E^+_{\nu_2}, E^-_{\nu_2}$, we have that $D(E^\pm_{\nu_i}) \le 2D(E)$, for $i = 1, 2$. Moreover, at least one of them, call it $F$, satisfies the estimate*

$$\alpha(E) \le C\alpha(F).$$

We are not giving the proof of this lemma, for which we refer to (see [44, Lemma 2.5]). Instead we show how to use it in order to reduce the proof of (35) to $n$-symmetric sets.

**Theorem 40** *There exist $\delta_1$ and $C_1$ depending only on $n$ such that if $E \subset [-L, L]^n, |E| = |B|, \delta(E) \le \delta_1$, then there exists an $n$-symmetric set $F$ such that $F \subset [-2L, 2L]^n, |F| = |B|$ and*

$$\alpha(E) \le C_1\alpha(F), \quad D(F) \le 2^n D(E). \tag{40}$$

*Proof* Take $\delta_1 = 2^{-(n-1)}\delta$, where $\delta$ is the constant of Lemma 39. By applying the lemma $n-1$ times to different pairs of orthogonal directions we find a set $\widetilde{E} \subset [-L, L]^n$ with $n-1$ symmetries, $|\widetilde{E}| = |B|$ and such that

$$\alpha(E) \le C^{n-1}\alpha(\widetilde{E}), \quad D(\widetilde{E}) \le 2^{n-1}D(E),$$

where $C$ is the constant given by Lemma 39. Without loss of generality we may assume that $\widetilde{E}$ is symmetric with respect to the first $n-1$ directions $e_1, \dots, e_{n-1}$. Let us consider a hyperplane $\pi_{e_n}$ orthogonal to $e_n$ and dividing $\widetilde{E}$ into two parts of equal measure, $\widetilde{E}', \widetilde{E}''$, and the corresponding sets $\widetilde{E}^\pm_{e_n}$. From (38) we have that

$$D(\widetilde{E}^\pm_{e_n}) \le 2D(\widetilde{E}) \le 2^n D(E).$$

To control the asymmetry of $\widetilde{E}^\pm_{e_n}$ observe that since $\widetilde{E}$ is symmetric with respect to the first $n-1$ directions, the sets $\widetilde{E}^\pm_{e_n}$ are both $n$-symmetric. Moreover, by suitably translating $\widetilde{E}$ if necessary, we may also assume that they are both symmetric around the origin. Thus we may apply Lemma 38 to estimate

$$\alpha(\widetilde{E}) \le |\widetilde{E}\Delta B| = \frac{1}{2}\big[|\widetilde{E}^+_{e_n}\Delta B| + |\widetilde{E}^-_{e_n}\Delta B|\big] \le \frac{3}{2}\big[\alpha(\widetilde{E}^+_{e_n}) + \alpha(\widetilde{E}^-_{e_n})\big].$$

Thus, at least one of the sets $\widetilde{E}^{\pm}_{e_n}$ has asymmetry index greater than $\frac{1}{3}\alpha(\widetilde{E})$. Therefore, denoting by $F$ this set, we have

$$D(F) \le 2D(\widetilde{E}) \le 2^n D(E)$$

and

$$\alpha(E) \le C^{n-1}\alpha(\widetilde{E}) \le 3C^{n-1}\alpha(F). \qquad \square$$

Having proved Theorem 40, from now on we may assume that $E$ is an $n$-symmetric set such that $E \subset [-L, L]^n$ for some $L$ depending only on $n$, $|E| = |B|$. We now want to pass from $n$-symmetric sets to *axially symmetric* sets, i.e., sets $E$ having an axis of symmetry such that every non-empty cross-section of $E$ perpendicular to this axis is a $(n-1)$-dimensional ball.

In order to perform this further simplification, let us recall the definition of *Schwartz symmetrization* of a set $E$ (Fig. 18). To this aim, given a measurable set $E$, for all $t \in \mathbb{R}$ we set

$$E_t = \{x \in \mathbb{R}^{n-1} : (x, t) \in E\}.$$

A result due to Vol'pert states that if $E$ is a set of finite perimeter then $E_t$ is a set of finite perimeter in $\mathbb{R}^{n-1}$ for a.e. $t \in \mathbb{R}$. For a proof of this important property see for instance [6, Theorem 2.4].

**Definition 41** Given a measurable set $E \subset \mathbb{R}^n$, its *Schwartz symmetrization* is defined as

$$E^* = \{(x, t) \in \mathbb{R}^{n-1} \times \mathbb{R} : t \in \mathbb{R},\ |x| < r_E(t)\},$$

where $\omega_{n-1} r_E^{n-1}(t) = \mathcal{H}^{n-1}(E_t)$.

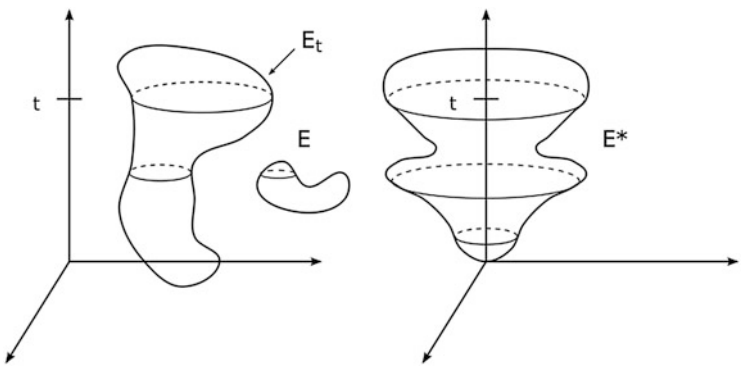


**Fig. 18** The Schwartz symmetrization of the set $E$

Note that $|E^*| = |E|$. Moreover, as for Steiner symmetrization, also Schwartz symmetrization decreases the perimeter. The next result (see [44, Lemma 3.3]) provides a useful formula for the perimeter of an axially symmetric set whose boundary has no horizontal flat parts. To this aim, given a measurable set $E \subset \mathbb{R}^n$, for $\mathcal{H}^1$-a.e. $t \in \mathbb{R}$ we set

$$v_E(t) := \mathcal{H}^{n-1}(E_t), \qquad p_E(t) := P_{n-1}(E_t),$$

where $P_{n-1}(\cdot)$ denotes the perimeter of a subset of $\mathbb{R}^{n-1}$. Observe that this definition makes sense since for $\mathcal{H}^1$-a.e. $t \in \mathbb{R}$ the slice $E_t$ is a set of finite perimeter in $\mathbb{R}^{n-1}$. Note also that from Definition 41 we have $v_E(t) = v_{E^*}(t)$ for all $t$. Moreover, the isoperimetric inequality in $\mathbb{R}^{n-1}$ yields that $p_{E^*}(t) \le p_E(t)$, since $(E^*)_t$ is a ball with the same measure of $E_t$.

**Theorem 42** *Let $E \subset \mathbb{R}^n$ be a set of finite perimeter and let $E^*$ be its Steiner symmetrization. Then*

$$P(E^*) \le P(E)\,. \tag{41}$$

*Moreover, if*

$$\mathcal{H}^{n-1}(\partial^* E \cap \{v_t^E = \pm 1\}) = 0, \tag{42}$$

*then $v_E$ belongs to $W^{1,1}(\mathbb{R})$ and the following formulas hold:*

$$P(E) \ge \int_{\mathbb{R}} \sqrt{v_E'^2 + p_E^2}\, dt, \qquad P(E^*) = \int_{\mathbb{R}} \sqrt{v_E'^2 + p_{E^*}^2}\, dt\,.$$

The next step in the proof of the quantitative isoperimetric inequality is given by the following theorem, which states that we may eventually reduce to the case of axially symmetric sets.

**Theorem 43** *Let $E \subset [-L, L]^n$ be an $n$-symmetric set satisfying (42) such that $|E| = |B|$ and $D(E) \le 1$. If $n = 2$ or if $n \ge 3$ and the quantitative isoperimetric inequality (35) holds true in $\mathbb{R}^{n-1}$, there exists a constant $C$ depending only on $n$ such that*

$$\alpha(E) \le \alpha(E^*) + C\sqrt{D(E)}, \quad \text{and} \quad D(E^*) \le D(E). \tag{43}$$

We shall give the proof of this theorem at the end of this section. First, we show how to conclude the proof of the quantitative isoperimetric inequality (35) by combining this result with a final estimate for axially symmetric sets. This estimate is provided by the next theorem, which is a particular case of a more general one proved by Hall

in [47, Theorem 2] for general axially symmetric sets. As we already mentioned, two different proofs of Theorem 44 below are given in [44, Sect. 4] and in [50, Sect. 7].

**Theorem 44** *Let $E \subset [-L, L]^n$ be an axially and n-symmetric set with center of symmetry at the origin, such that $|E| = |B|$. Then*

$$|E \Delta B(x_0)| \leq C' \sqrt{D(E)}, \tag{44}$$

*for some constant $C'$ depending only on the dimension n.*

The two previous theorems immediately yield the proof of (35).

*Proof of Theorem 34* We argue by induction on the dimension $n$ assuming that either $n = 2$ or $n \geq 3$ and the isoperimetric inequality (35) holds in $\mathbb{R}^{n-1}$.

As we already observed, in order to prove (35) it is enough to consider a set $E \subset [-L, L]^n$, such that $|E| = |B|$ and that $D(E) \leq \delta$ for some conveniently small $\delta \in (0, 1)$. Moreover, since the set of directions $\nu \in \mathbb{S}^{n-1}$ such that $\mathcal{H}^{n-1}(\partial^* E \cap \{\nu_t^E = \pm 1\}) > 0$ is at most countable, by rotating $E$ if necessary we may always assume that (42) holds. Recall that Theorem 40 allows us to replace $E$ by a $n$-symmetric set $F \subset [-2L, 2L]^n$ satisfying (40). And observe that from the proof of Theorem 40 and the statement of Lemma 39 it is clear that also $F$ satisfies (42). Therefore, by replacing $E$ with the $n$-symmetric set $F$ if necessary, we may always reduce the proof of (35) to the case of a set $E$ satisfying all the assumptions of Theorem 41.

Thus, recalling (43) and applying (44) to $E^*$ we conclude, assuming without loss of generality that the center of symmetry of $E^*$ is at the origin,

$$\begin{aligned} \alpha(E) &\leq \alpha(E^*) + C\sqrt{D(E)} \leq |E^* \Delta B| + C\sqrt{D(E)} \\ &\leq C' \sqrt{D(E^*)} + C\sqrt{D(E)} \leq C'' \sqrt{D(E)}\,. \end{aligned}$$

where the constant $C''$ depends only on the dimension $n$. □

We now turn to the proof of Theorem 43.

*Proof of Theorem 43* Denoting by $B(x_0)$ an optimal ball for $E^*$, we have

$$\alpha(E) \leq |E \Delta B(x_0)| \leq |E^* \Delta B(x_0)| + |E \Delta E^*| = \alpha(E^*) + |E \Delta E^*|.$$

Hence, in order to prove the first inequality in (43) it is enough to show that

$$|E \Delta E^*| \leq c(n) \sqrt{D(E)}, \tag{45}$$

for some positive constant $c$ depending only on $n$. The second inequality in (43) follows immediately from (41). To prove (45) we use again Theorem 42 to estimate

$$
\begin{aligned}
D(E) = P(E) - P(B) \geq P(E) - P(E^*) &\geq \int_{\mathbb{R}} \sqrt{v_E'^2 + p_E^2} - \sqrt{v_E'^2 + p_{E^*}^2}\, dt \\
&= \int_{\mathbb{R}} \frac{p_E^2 - p_{E^*}^2}{\sqrt{v_E'^2 + p_E^2} + \sqrt{v_E'^2 + p_{E^*}^2}}\, dt \\
&\geq \left( \int_{\mathbb{R}} \sqrt{p_E^2 - p_{E^*}^2}\, dt \right)^2 \frac{1}{\int_{\mathbb{R}} \sqrt{v_E'^2 + p_E^2} + \sqrt{v_E'^2 + p_{E^*}^2}\, dt} \\
&\geq \left( \int_{\mathbb{R}} \sqrt{p_E^2 - p_{E^*}^2}\, dt \right)^2 \frac{1}{P(E) + P(E^*)},
\end{aligned}
$$

where the inequality before the last one follows from Hölder's inequality. Since $D(E) \leq 1$, we have $P(E^*) \leq P(E) \leq P(B) + 1$. Therefore from the above estimate we get, recalling that $p_E \geq p_{E^*}$,

$$
\begin{aligned}
\sqrt{D(E)} &\geq c_n \int_{\mathbb{R}} \sqrt{p_E^2 - p_{E^*}^2}\, dt \\
&= c_n \int_{\mathbb{R}} \sqrt{p_E + p_{E^*}} \sqrt{p_{E^*}} \sqrt{(p_E - p_{E^*})/p_{E^*}}\, dt \\
&\geq \sqrt{2} c_n \int_{\mathbb{R}} p_{E^*} \sqrt{(p_E - p_{E^*})/p_{E^*}}\, dt\,.
\end{aligned} \tag{46}
$$

Now assume that $n \geq 3$ and observe that since $(E^*)_t$ is an $(n-1)$–dimensional ball of radius $r_E(t)$ with $\omega_{n-1} r_E^{n-1}(t) = \mathcal{H}^{n-1}(E_t)$, the ratio

$$
\frac{p_E(t) - p_{E^*}(t)}{r_E^{n-2}(t)}
$$

is precisely the isoperimetric gap of $E_t$ in $\mathbb{R}^{n-1}$. Since by assumption, the quantitative isoperimetric inequality (35) holds true in $\mathbb{R}^{n-1}$, we have

$$
\kappa(n-1) \sqrt{\frac{p_E(t) - p_{E^*}(t)}{r_E^{n-2}(t)}} \geq \alpha_{n-1}(E_t),
$$

where $\alpha_{n-1}(E_t)$ is the $(n-1)$-dimensional Fraenkel asymmetry of $E_t$. But $E_t$ is an $(n-1)$-symmetric set in $\mathbb{R}^{n-1}$ and $(E^*)_t$ is the ball centered at the center of symmetry of $E_t$. Therefore from Lemma 38 we get

$$
\kappa(n-1) \sqrt{\frac{p_E(t) - p_{E^*}(t)}{r_E^{n-2}(t)}} \geq \alpha_{n-1}(E_t) \geq \frac{1}{3} \frac{\mathcal{H}^{n-1}(E_t \Delta (E^*)_t)}{r_E^{n-1}(t)}.
$$

Inserting this inequality in (46) we then get

$$\sqrt{D(E)} \geq c \int_{\mathbb{R}} r_E^{n-2}(t) \sqrt{\frac{p_E(t) - p_{E^*}(t)}{r_E^{n-2}(t)}}\, dt \geq c \int_{\mathbb{R}} \frac{\mathcal{H}^{n-1}(E_t \Delta E_t^*)}{r_E(t)}\, dt$$
$$\geq \frac{c}{L} \int_{-L}^{L} \mathcal{H}^{n-1}(E_t \Delta E_t^*)\, dt = \frac{c}{L} |E_t \Delta E_t^*|,$$

where the inequality before the last one follows from the inclusion $E \subset [-L, L]^n$ and the last equality is just Fubini's theorem. This proves (45). Hence the assertion follows when $n \geq 3$.

If $n = 2$, since $E$ is 2-symmetric, either $E_t$ is a symmetric interval (and thus $E_t = E_t^*$) or $E_t$ is the union of at least two essentially disjoint intervals and thus $p_E(t) \geq 4$, while $p_{E^*}(t) = 2$. Note also that since $E \subset [-L, L]^2$, then $\mathcal{H}^1(E_t \Delta E_t^*) \leq 2L$ for all $t \in \mathbb{R}$. Therefore, from (46) we easily get

$$\sqrt{D(E)} \geq \sqrt{2} c_2 \int_{\mathbb{R}} p_{E^*} \sqrt{(p_E - p_{E^*})/p_{E^*}}\, dt = 2c_2 \int_{\{t:\, E_t \neq E_t^*\}} \sqrt{p_E - p_{E^*}}\, dt$$
$$\geq 2c_2 \int_{\{t:\, E_t \neq E_t^*\}} \sqrt{2}\, dt \geq \frac{\sqrt{2} c_2}{L} \int_{\{t:\, E_t \neq E_t^*\}} \mathcal{H}^1(E_t \Delta E_t^*) dt = \frac{\sqrt{2} c_2}{L} |E \Delta E^*|,$$

thus concluding the proof also in this case. □

## 6 Alternative Proofs of the Quantitative Isoperimetric Inequality

In this section we discuss two different approaches to the quantitative isoperimetric inequality, the first one via the regularity theory of sets of finite perimeter and the second one via mass transportation. The latter approach will provide us with the extension of (35) to the anisotropic perimeter, a result that cannot be achieved via symmetrization techniques. In the final part of this section we shall give an account of a stronger version of (35) which is very much in the spirit of the estimate (22) proved for nearly spherical sets.

We start by presenting the approach to the quantitative isoperimetric inequalities introduced by Cicalese and Leonardi in [23] with some further simplifications due to Acerbi, Morini and the author, see [1]. In comparing this new proof with the one that we have seen in the previous section one can see two main differences. The proof by symmetrization is more elementary since it relies on some geometric ideas that do not require the use of deep previous results. But that proof is quite long. The approach of Cicalese and Leonardi to the quantitative isoperimetric inequality is based on the deep results of De Giorgi's regularity theory for area minimizing sets of finite perimeter, but it has the advantage of providing a quicker proof. Moreover this

approach has proved to be useful in the study of the stability of other inequalities, see [1, 9, 10, 12, 25].

As we said before we need a regularity result on area minimizing sets of finite perimeter or, more generally, of area almost minimizers.

**Definition 45** Let $\omega, r$ be positive numbers. A set $E$ of finite perimeter is an $(\omega, r)$-*area almost minimizer* if, for all balls $B_\varrho(x_0)$ with $\varrho < r$ and all measurable sets $F$ such that $E\Delta F \subset\subset B_\varrho(x_0)$, we have

$$P(E) \leq P(F) + \omega\varrho^n.$$

So, an almost minimizer minimizes the perimeter with respect to local variations of the set up to a higher order volume term. De Giorgi's regularity theory, originally established only for minimizers, readily extends to almost minimizers, see [60, Sects. 1.9 and 1.10] and [51, Theorems 26.5 and 26.6].

**Theorem 46** *If $E$ is an $(\omega, r)$-area almost minimizer, then $\partial^* E$ is a manifold of class $C^{1,1/2}$, $\partial E \setminus \partial^* E$ is relatively closed in $\partial E$ and $H^s(\partial E \setminus \partial^* E) = 0$ for all $s > n - 8$.*

*Moreover, if $E_j$ is a sequence of equibounded $(\omega, r)$-area almost minimizers converging in measure to an open set $E$ of class $C^2$, then for $j$ large each $E_j$ is of class $C^{1,1/2}$ and the sequence $E_j$ converges to $E$ in $C^{1,\alpha}$ for all $0 < \alpha < 1/2$.*

Next lemma is a simple consequence of the isoperimetric inequality.

**Lemma 47** *If $\Lambda > n$, the unique solution up to translations of the problem*

$$\min\left\{P(F) + \Lambda\big||F| - |B|\big| : F \subset \mathbb{R}^n\right\} \tag{47}$$

*is the unit ball.*

*Proof* By the isoperimetric inequality it follows that in order to minimize the functional in (47), we may restrict to the balls $B_r$. Thus the above problem becomes

$$\min_{r>0}\left\{n\omega_n r^{n-1} + \Lambda\omega_n|r^n - 1|\right\},$$

which has a unique minimum for $r = 1$, if $\Lambda > n$. □

Lets us now describe how the new proof of the isoperimetric inequality works. The main idea in Cicalese and Leonardi approach was to reduce the proof of (35) to nearly spherical sets via a contradiction argument. They start by assuming that there exists a sequence of sets $E_j$ with infinitesimal isoperimetric deficits for which the quantitative inequality does not hold. Then they replace it with a different sequence of sets $F_j$, still not satisfying the quantitative inequality and converging to $B$ in $C^1$, thus contradicting Fuglede's Theorem 27 for nearly spherical sets. The sets $F_j$ are constructed as the solutions of certain minimum problems and their convergence in $C^1$ to the unit ball is a consequence of Theorem 46.

In the following we shall set for any measurable set of finite measure

$$A(E) := \min_{x\in\mathbb{R}^n} \big\{|E\Delta B(x)|\big\}.$$

Clearly $A(E) = \alpha(E)$ if $|E| = |B|$.

*Proof of* (35) *by regularity* **Step 1.** Fix $R > 0$ so that the ball $B_R$ contains the cube $[-L, L]^n$ given by Lemma 35. As we observed in the previous section, it is enough to prove (35) for a set $E \subset B_R$, with $|E| = |B|$ and with $D(E) \le \delta$ for some fixed $\delta > 0$. Thus, let us argue by contradiction assuming that there exists a sequence $E_j \subset B_R$, $|E_j| = |B|$, with $D(E_j) \to 0$ and

$$D(E_j) \le C_0 \alpha(E_j)^2, \tag{48}$$

for some constant $C_0$ to be chosen later. Observe that Lemma 36 implies that $A(E_j) = \alpha(E_j) \to 0$. Let us now introduce a new sequence $F_j$, where for each $j$ the set $F_j$ is a minimizer of the following problem

$$\min \big\{P(F) + |A(F) - A(E_j)| + \Lambda||F| - |B|| : F \subset B_R\big\},$$

where $\Lambda > n$ is a fixed constant. Note that the penalization term $\Lambda||F| - |B||$ forces the minimizers $F_j$ to have almost the same volume of the unit ball, while the presence of $|A(F) - A(E_j)|$ has the effect that the asymmetry of $F_j$ is very close to the one of $E_j$, hence converges to zero.

Since the perimeters of the $F_j$ are equibounded, the compactness Theorem 15 implies that, up to a not relabeled subsequence, they converge in measure to some set $F_\infty$. Moreover, the lower semicontinuity of the perimeter immediately yields that $F_\infty$ is a minimizer of the problem: $\min \big\{P(E) + A(E) + \Lambda||E| - |B|| : E \subset B_R\big\}$. Therefore, for every set $E$ of finite perimeter, from Lemma 47 we have

$$P(F_\infty) + A(F_\infty) + \Lambda||F_\infty| - |B|| \le P(B) \le P(E) + \Lambda||E| - |B||\,.$$

In particular, $F_\infty$ is a minimizer of the problem in (47), hence Lemma 47 implies that $F_\infty$ is a ball and thus that the $F_j$ converge in measure to some ball $B_1(x_0)$.

We now want to show that the there exists $\omega > 0$ such that all sets are $F_j$ are $(\omega, R)$-area almost minimizers. This fact, thanks to Theorem 46, will imply the convergence to $B_1(x_0)$ in $C^1$. To prove the almost minimality of the $F_j$, let us fix a set $F$ such that $F_j\Delta F \subset\subset B_r(x)$ for some ball $B_r(x)$ with radius $r < R$ and let us consider two cases.

First. let us assume that $F \subset B_R$. Then, by the minimality of $F_j$ we get

$$\begin{aligned} P(F_j) &\le P(F) + |A(F) - A(E_j)| - |A(F_j) - A(E_j)| + \Lambda\big[||F| - |B|| - ||F_j| - |B||\big] \\ &\le P(F) + |A(F) - A(F_j)| + \Lambda\,\big||F| - |F_j|\big| \\ &\le P(F) + (\Lambda + 1)|F\Delta F_j| \le P(F) + (\Lambda + 1)\omega_n r^n. \end{aligned}$$

If instead $|F \setminus B_R| > 0$, we split $F$ in two parts, one inside and the other one outside $B_R$. Hence,

$$P(F_j) - P(F) = [P(F_j) - P(F \cap B_R)] + [P(F \cap B_R) - P(F)].$$

Since $F \cap B_R \subset B_R$, as before we have

$$P(F_j) - P(F \cap B_R) \leq (\Lambda + 1)\omega_n r^n,$$

while

$$P(F \cap B_R) - P(F) = P(B_R) - P(F \cup B_R) \leq 0$$

by the isoperimetric inequality. Therefore we may conclude that the sets $F_j$ are all $((\Lambda + 1)\omega_n, R)$–almost minimizers and that the sequence $F_j$ converges to $B_1(x_0)$ in $C^{1,\alpha}$ for all $\alpha < 1/2$.

**Step 2.** By the minimality of the $F_j$, recalling (48) and using Lemma 47, we get

$$\begin{aligned} P(F_j) + \Lambda||F_j| - |B|| + |A(F_j) - A(E_j)| &\leq P(E_j) \qquad (49)\\ &\leq P(B) + C_0 A(E_j)^2 \leq P(F_j) + \Lambda||F_j| - |B|| + C_0 A(E_j)^2. \end{aligned}$$

Therefore, $|A(F_j) - A(E_j)| \leq C_0 A(E_j)^2$. Since $A(E_j) \to 0$, we get that $A(F_j)/A(E_j) \to 1$.

To conclude the proof we need only to rescale the sets $F_j$ to the same volume of the unit ball by setting $\widetilde{F}_j = \lambda_j F_j + x_j$, where $\lambda_j^n |F_j| = |B|$ and $x_j$ is chosen so that $\widetilde{F}_j$ has the baricenter at the origin. Note that $\lambda_j \to 1$ since the $F_j$ converge in $C^1$ to a unit ball. Observe also that, since $P(F_j) \to P(B)$ and $\Lambda > n$, for $j$ large we have $P(F_j) < \Lambda |F_j|$. Thus for $j$ large we have also

$$\big|P(\widetilde{F}_j) - P(F_j)\big| = P(F_j)|\lambda_j^{n-1} - 1| \leq P(F_j)|\lambda_j^n - 1| \leq \Lambda |\lambda_j^n - 1||F_j| = \Lambda \big||\widetilde{F}_j| - |F_j|\big|.$$

From this estimate, recalling (49) we get that

$$P(\widetilde{F}_j) \leq P(F_j) + \Lambda \big||\widetilde{F}_j| - |F_j|\big| = P(F_j) + \Lambda \big||F_j| - |B|\big| \leq P(B) + C_0 A(E_j)^2. \qquad (50)$$

However, since $A(F_j)/A(E_j) \to 1$ as $j \to \infty$, we have $A(E_j)^2 < 2A(\widetilde{F}_j)^2$ for $j$ large. Therefore, from (50) we obtain

$$P(\widetilde{F}_j) - P(B) < 2C_0 A(\widetilde{F}_j)^2,$$

which leads to contradiction to (22) if $C_0 < 1/(16\omega_n)$, since the $\widetilde{F}_j$ are converging in $C^1$ to $B$, This contradiction concludes the proof. □

In the remaining part of this section we will present two extensions of the isoperimetric inequality (35). The first one deals with the *anisotropic perimeter*. We recall that if $\gamma : \mathbb{R}^n \to [0, \infty)$ is a positively 1-homogeneous function such that $\gamma(x) > 0$ for all $x \neq 0$ the *anisotropic perimeter* associated with $\gamma$ is defined for any set $E$ of locally finite perimeter by setting

$$P_\gamma(E) := \int_{\partial^* E} \gamma(\nu^E(x)) d\mathcal{H}^{n-1}(x).$$

It is well known that the isoperimetric sets with respect to this perimeter are the homothetic and translated of the so called *Wulff shape set* associated to $\gamma$, see [38] and also [26] for two-dimensional case, which is given by

$$W_\gamma := \{x \in \mathbb{R}^n : \langle x, \nu \rangle - \gamma(\nu) < 0 \text{ for all } \nu \in \mathbb{S}^{n-1}\}.$$

Then, the anisotropic isoperimetric inequality states that

$$P_\gamma(E) \geq P_\gamma(W_\gamma)$$

for all sets of finite perimeter such that $|E| = |W_\gamma|$, with equality holding if and only if $E$ is a translated of the Wulff shape set $W_\gamma$.

The quantitative version of the anisotropic isoperimetric inequality was proved by Figalli, Maggi and Pratelli in [34]. It states that there exists a constant $C$, depending only on $n$, such that for any set of finite perimeter $E$ such that $|E| = r^n |W_\gamma|$

$$\alpha_\gamma(E)^2 \leq C D_\gamma(E), \tag{51}$$

where

$$\alpha_\gamma(E) := \min_{x \in \mathbb{R}^n} \left\{ \frac{|E \Delta (x + rW_\gamma)|}{r^n} \right\}, \qquad D_\gamma(E) := \frac{P_\gamma(E) - P_\gamma(rW_\gamma)}{r^{n-1}}$$

denote the *anisotropic asymmetry index* and the *anisotropic isoperimetric deficit*, respectively.

Since the Wulff shape $W_\gamma$ can be any bounded open convex set, it is clear that no symmetrization argument can be used to prove the anisotropic isoperimetric inequality or its quantitative counterpart (51). An extra difficulty is also due to the extreme rigidity of the anisotropic perimeter which is not invariant by rotation. Moreover, even the equality $P_\gamma(E) = P_\gamma(\mathbb{R}^n \setminus E)$ holds true only if $\gamma$ is symmetric with respect to the origin. Observe also that since the Wulff shape set $W_\gamma$ is in general a non smooth convex set, no strategy based on regularity may ever work. And in fact a completely different strategy was devised in [34] to prove inequality (51), based on optimal mass transportation and on the proof of the isoperimetric inequality given by Gromov in [55].

The idea of this proof, that we present in the simpler case of the standard perimeter, is to use a *transport map* from the set $E$ to the an isoperimetric set of the same volume. Though the original proof of Gromov used the *Knothe map*, which has the advantage of being defined by an explicit construction, it is more convenient to use the so called *Brenier map* whose properties are stated in the following theorem, see [14], and also [53] and [54].

**Theorem 48** *Let $E$ be a set of finite perimeter with $|E| = |B|$. There exists a convex function $\varphi : \mathbb{R}^n \to \mathbb{R}$ such that if we set $T = \nabla\varphi$, then $T(x) \in B$ for a.e. $x \in \mathbb{R}^n$ and* $\det \nabla T(x) = 1$ *for a.e. $x \in E$.*

Let us now give the

*Gromov's proof of the isoperimetric inequality* Being the gradient of a convex function, $T$ is a $BV$ map, see [29, Sect. 6.3, Theorem 2]. However, in order to avoid unnecessary technical difficulties, let us assume that $T$ is Lipschitz. For every $x \in E$ let us denote by $\lambda_i(x)$, $i = 1, \dots, n$, the eingenvalues of the symmetric matrix $\nabla T(x)$. Using the arithmetic–geometric mean inequality, we have

$$P(B) = n\omega_n = n\int_{B_1} dy = n\int_E (\det \nabla T)^{1/n} dx = n\int_E (\lambda_1 \dots \lambda_n)^{1/n} dx$$
$$\leq \int_E (\lambda_1 + \dots + \lambda_n) dx = \int_E \operatorname{div} T\, dx = \int_{\partial E} T \cdot \nu^E d\mathcal{H}^{n-1} \leq P(E).$$

Moreover, since $\det \nabla T(x) = 1$, if $P(E) = P(B)$ we have that $\lambda_1(x) = \lambda_2(x) = \dots = \lambda_n(x) = 1$ for a.e. $x \in E$. Therefore, $T$ is a translation and $E$ is a ball. □

Beside being extremely simple, this argument gives some non trivial quantitative information. In fact, by subtracting the last and the first terms in the above chain of inequalities we get that

$$\int_E \left[(\lambda_1 + \dots \lambda_n)/n - (\lambda_1 \dots \lambda_n)^{1/n}\right] \leq \frac{1}{n} D(E), \tag{52}$$

$$\int_{\partial E} (1 - T \cdot \nu^E) d\mathcal{H}^{n-1} \leq D(E). \tag{53}$$

The first inequality (52) is telling us that if the isoperimetric deficit $D(E)$ is small the eigenvalues of $T(x)$ are almost equal, at least in an integral sense. From this inequality one can deduce, see [34, Corollary 2.4]), that there exists a constant $c$ depending only on $n$ such that

$$\int_E |\nabla T - I| \leq c\sqrt{D(E)}\,. \tag{54}$$

Let us assume, without loss of generality, that $B$ is the optimal ball for $E$ and let us observe, as proved in [34, Lemma 3.5], that

$$|E\Delta B| \leq c(n) \int_{\partial^* E} \big|1 - |x|\big|\, d\mathcal{H}^{n-1}.$$

Then, in order to prove (51) one should control the right hand side of the previous inequality with $\sqrt{D(E)}$. To this aim, using (53) we have

$$\begin{aligned}\int_{\partial^* E} \big|1 - |x|\big|\, d\mathcal{H}^{n-1} &\leq \int_{\partial^* E} \big[\big|1 - |T(x)|\big| + \big||T(x)| - |x|\big|\big]\, d\mathcal{H}^{n-1} \\ &\leq \int_{\partial^* E} \big[\big(1 - T(x)\cdot \nu^E(x)\big) + |T(x) - x|\big]\, d\mathcal{H}^{n-1} \leq D(E) + \int_{\partial^* E} |T(x) - x|\, d\mathcal{H}^{n-1}.\end{aligned}$$

The difficult part of the proof of Figalli, Maggi and Pratelli consists in showing that if the isoperimetric deficit $D(E)$ is small one may always reduce to the case when a Poincaré type inequality for the boundary traces holds with a constant $c(n)$ depending only on $n$. If this is true, recalling (54), one gets

$$\int_{\partial^* E} |T(x) - x|\, d\mathcal{H}^{n-1} \leq c(n) \int_E |\nabla T - I| \leq c\sqrt{D(E)}\,.$$

Beside providing an alternative proof of the quantitative isoperimetric inequality in the wider framework of anisotropic perimeter, the paper by Figalli, Maggi and Pratelli contains several interesting results. In particular, Theorem 3.4 in [34] states that given any set of finite perimeter $E$ with small deficit one may always extract from $E$ a maximal set for which a trace inequality holds with a universal constant. This is a new and deep result that may have several applications. Moreover, the mass transportation approach used in [34] has been also successfully used to obtain the quantitative versions of other important inequalities (see [21, 33, 35]).

At the beginning of this section we observed how the proof of the isoperimetric inequality of Cicalese and Leonardi shows that one may always reduce to the case of a nearly spherical set and thus to Fuglede's Theorem 27. However, the two sets $E$ and $F$ in Fig. 19 have the same measure, the same asymmetry index, but $D(E) << 1$, while $D(F) >> 1$. Therefore the quantitative isoperimetric inequality (35) gives a sharp information on the set $E$ while gives no information at all on the set $F$. The reason is that while the asymmetry index looks only at the distance in measure of a set from a ball, the isoperimetric gap encodes also an information on the oscillation of the boundary of the set.

This suggests that we should introduce a more precise index which takes into account also the oscillation of the normals to the boundary of the set $E$. To this aim, given a set of finite perimeter $E$ and a ball $B_r(y)$ with the same volume of $E$, we are

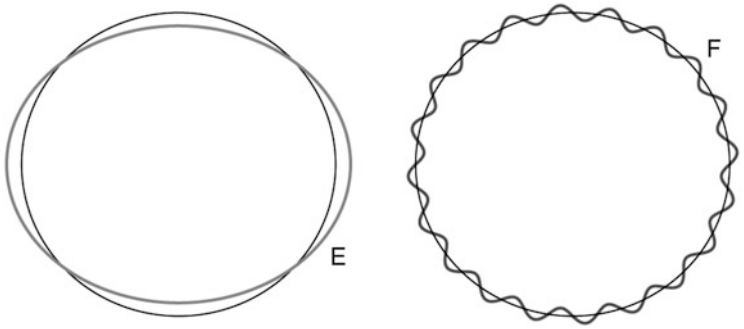


**Fig. 19** $E$ and $F$ have the same measure and the same Fraenkel asymmetry

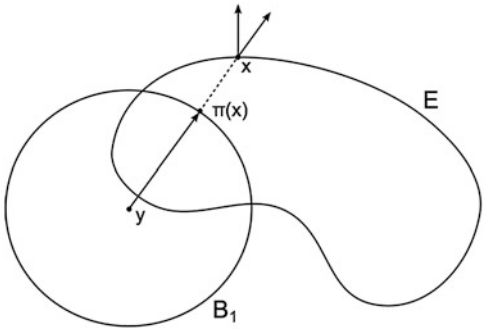


**Fig. 20** The construction of the asymmetry index $\beta(E)$

going to measure the distance from $E$ to the ball in the following way (see Fig. 20). For every point $x \in \partial^* E$ we take the projection $\pi_{y,r}(x)$ of $x$ on the boundary of $\partial B_r(y)$ and consider the distance $|\nu^E(x) - \nu^{r,y}(\pi_{y,r}(x))|$ between the exterior normal to $E$ at the point $x$ and the exterior normal to $B_r(y)$ at the projection point $\pi_{y,r}(x)$. Then, we take the $L^2$ norm of this distance and minimize the resulting norm among all possible balls, thus getting

$$\beta(E) := \min_{y\in\mathbb{R}^n} \left\{ \left( \frac{1}{2r^{n-1}} \int_{\partial^* E} |\nu^E(x) - \nu^{r,y}(\pi_{y,r}(x))|^2 \, d\mathcal{H}^{n-1}(x) \right)^{1/2} \right\}.$$

We shall refer to $\beta(E)$ as to *the oscillation index* of $E$. Observe that Fuglede's Theorem 27 provides an estimate for both the asymmetry and the oscillation index. In fact, if $E$ is a nearly spherical set satisfying (21) with a sufficiently small $\varepsilon$, recall, see (33), that for every point $x \in \partial^* E$ the exterior normal to $E$ is given by

$$\nu^E(x) = \frac{z(1 + u(z)) - \nabla_\tau u(z)}{\sqrt{(1 + u(z))^2 + |\nabla_\tau u(z)|^2}},$$

where $z = x/|x|$ and thus $x = z(1 + u(z))$. Thus, from (22) we have

$$\begin{aligned}
\alpha(E)^2+\beta(E)^2 &\le |E\Delta B|^2 + \frac{1}{2}\int_{\partial^* E}\left|\nu^E(x)-\frac{x}{|x|}\right|^2 d\mathcal{H}^{n-1}\\
&= |E\Delta B|^2 + \int_{\partial^* E}\left(1-\nu^E(x)\cdot\frac{x}{|x|}\right)d\mathcal{H}^{n-1}\\
&\le c\int_{\mathbb{S}^{n-1}}|u|^2 d\mathcal{H}^{n-1} + c\int_{\mathbb{S}^{n-1}}\left(1-\frac{1+u(z)}{\sqrt{(1+u)^2+|\nabla u|^2}}\right)d\mathcal{H}^{n-1}\\
&= c\int_{\mathbb{S}^{n-1}}|u|^2 d\mathcal{H}^{n-1} + c\int_{\mathbb{S}^{n-1}}\frac{\sqrt{(1+u)^2+|\nabla u|^2}-(1+u)}{\sqrt{(1+u)^2+|\nabla u|^2}}\,d\mathcal{H}^{n-1}\\
&\le c\int_{\mathbb{S}^{n-1}}|u|^2 d\mathcal{H}^{n-1} + c\int_{\mathbb{S}^{n-1}}|\nabla u|^2\,d\mathcal{H}^{n-1} \le cD(E).
\end{aligned}$$

Next result, proved by Julin and the author in [42], is an improved version of the quantitative isoperimetric inequality.

**Theorem 49** *There exists a constant $\gamma(n)$ such that for any set of finite perimeter $E$*

$$\beta(E)^2 \le \gamma D(E).$$

Note that the inequality above is stronger than the quantitative isoperimetric inequality (35) since it can be shown (see [42, Proposition 1.2]) that there exists a constant $C(n)$ such that

$$\alpha(E) + \sqrt{D(E)} \le C\beta(E).$$

## References


1. E. Acerbi, N. Fusco, M. Morini, Minimality via second variation for a nonlocal isoperimetric problem. Commun. Math. Phys. **322**, 515–557 (2013)
2. A.D. Aleksandrov, Uniqueness theorems for surfaces in the large. V, (Russian) Vestn. Leningr. Univ. **13**, 5–8 (1958)
3. A. Alvino, V. Ferone, C. Nitsch, A sharp isoperimetric inequality in the plane. J. Eur. Math. Soc. **13**, 185–206 (2011)
4. L. Ambrosio, N. Fusco, D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems* (Oxford University Press, Oxford, 2000)
5. M. Barchiesi, A. Brancolini, V. Julin, Sharp dimension free quantitative estimates for the Gaussian isoperimetric inequality. Ann. Probab. (2015, to appear). arXiv:1409.2106v1
6. M. Barchiesi, F. Cagnetti, N. Fusco, Stability of the Steiner symmetrization of convex sets. J. Eur. Math. Soc. **15**, 1245–1278 (2013)
7. M. Barchiesi, G.M. Capriani, N. Fusco, G. Pisante, Stability of Pólya-Szegö inequality for log-concave functions. J. Funct. Anal. **267**, 2264–2297 (2014)

8. F. Bernstein, Über die isoperimetriche Eigenschaft des Kreises auf der Kugeloberflache und in der Ebene. Math. Ann. **60**, 117–136 (1905)
9. V. Bögelein, F. Duzaar, N. Fusco, A sharp quantitative isoperimetric inequality in higher codimension. Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl. **26**, 309–362 (2015)
10. V. Bögelein, F. Duzaar, N. Fusco, A quantitative isoperimetric inequality on the sphere. Adv. Calc. Var., to appear. Also available at http://cvgmt.sns.it/paper/2146/
11. T. Bonnesen, Über die isoperimetrische Defizit ebener Figuren. Math. Ann. **91**, 252–268 (1924)
12. L. Brasco, G. De Philippis, B. Velichkov, Faber-Krahn inequalities in sharp quantitative form. Duke Math. J. **164**, 1777–1831 (2015)
13. L. Brasco, A. Pratelli, Sharp stability of some spectral inequalities. Geom. Funct. Anal. **22**, 107–135 (2012)
14. Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions. Commun. Pure Appl. Math. **44**, 375–417 (1991)
15. Y.D. Burago, V.A. Zalgaller, *Geometric Inequalities* (Springer, Berlin, 1988)
16. X. Cabré, Partial differential equations, geometry and stochastic control, (Catalan) Butll. Soc. Catalana Mat. **15**, 7–27 (2000)
17. F. Cagnetti, M. Colombo, G. De Philippis, F. Maggi, Rigidity of equality cases in Steiner's perimeter inequality. Anal. PDE **7**, 1535–1593 (2014)
18. E.A. Carlen, A. Figalli, Stability for a GNS inequality and the log-HLS inequality, with application to the critical mass Keller-Segel equation. Duke Math. J. **162**, 579–625 (2013)
19. M. Chlebík, A. Cianchi, N. Fusco, The perimeter inequality under Steiner symmetrization: cases of equality. Ann. Math. **162**, 525–555 (2005)
20. A. Cianchi, L. Esposito, N. Fusco, C. Trombetti, A quantitative Pólya-Szegö principle. J. Reine Angew. Math. **614**, 153–189 (2008)
21. A. Cianchi, N. Fusco, F. Maggi, A. Pratelli, The sharp Sobolev inequality in quantitative form. J. Eur. Math. Soc. **11**, 1105–1139 (2009)
22. A. Cianchi, N. Fusco, F. Maggi, A. Pratelli, On the isoperimetric deficit in the Gauss space. Am. J. Math. **133**, 131–186 (2011)
23. M. Cicalese, G.P. Leonardi, A selection principle for the sharp quantitative isoperimetric inequality. Arch. Rat. Mech. Anal. **206**, 617–643 (2012)
24. M. Cicalese, G.P. Leonardi, Best constants for the isoperimetric inequality in quantitative form. J. Eur. Math. Soc. **15**, 1101–1129 (2013)
25. M. Cicalese, G.P. Leonardi, F. Maggi, Sharp stability inequalities for planar double-bubbles (2012). arXiv:1211.3698v2
26. B. Dacorogna, C.E. Pfister, Wulff theorem and best constant in Sobolev inequality. J. Math. Pures Appl. **71**, 97–118 (1992)
27. E. De Giorgi, Sulla proprietà isoperimetrica dell'ipersfera, nella classe degli insiemi aventi frontiera orientata di misura finita. Atti Accad. Naz. Lincei Mem. Cl. Sci. Fis. Mat. Nat. Sez. I **8**, 33–44 (1958)
28. E. De Giorgi, in *Selected Papers*, ed. by L. Ambrosio, G. Dal Maso, M. Forti, S. Spagnolo (Springer, New York, 2005)
29. L.C. Evans, R.F. Gariepy, *Lecture Notes on Measure Theory and Fine Properties of Functions* (CRC Press, Boca Raton, 1992)
30. H. Federer, *Geometric Measure Theory*. Die Grundlehren der mathematischen Wissenschaften, vol. 153 (Springer, New York, 1969)
31. A. Figalli, N. Fusco, F. Maggi, V. Millot, M. Morini, Isoperimetry and stability properties of balls with respect to nonlocal energies. Commun. Math. Phys. **336**, 441–507 (2015)
32. A. Figalli, F. Maggi, A. Pratelli, A note on Cheeger sets. Proc. Am. Math. Soc. **137**, 2057–2062 (2009)
33. A. Figalli, F. Maggi, A. Pratelli, A refined Brunn-Minkowski inequality for convex sets. Ann. Inst. H. Poincaré Anal. Non Linéaire **26**, 2511–2519 (2009)
34. A. Figalli, F. Maggi, A. Pratelli, A mass transportation approach to quantitative isoperimetric inequalities. Invent. Math. **182**, 167–211 (2010)

35. A. Figalli, F. Maggi, A. Pratelli, Sharp stability theorems for the anisotropic Sobolev and log-Sobolev inequalities on functions of bounded variation. Adv. Math. **242**, 80–101 (2013)
36. A. Figalli, R. Neumayer, Gradient stability for Sobolev inequality: the case $p \geq 2$ (2015). arXiv:1510.02119
37. W.H. Fleming, R. Rishel, An integral formula for total gradient variation. Arch. Math. **11**, 218–222 (1960)
38. I. Fonseca, S. Müller, A uniqueness proof for the Wulff theorem. Proc. Roy. Soc. Edinb. **119A**, 125–136 (1991)
39. B. Fuglede, Stability in the isoperimetric problem for convex or nearly spherical domains in $\mathbb{R}^n$, Trans. Am. Math. Soc. **314**, 619–638 (1989)
40. N. Fusco, The quantitative isoperimetric inequality and related topics. Bull. Math. Sci. **5**, 517–607 (2015)
41. N. Fusco, M.S. Gelli, G. Pisante, On a Bonnesen type inequality involving the spherical deviation. J. Math. Pures Appl. **98**, 616–632 (2012)
42. N. Fusco, V. Julin, A strong form of the quantitative isoperimetric inequality. Calc. Var. Partial Differential Equations **50**, 925–937 (2014)
43. N. Fusco, F. Maggi, A. Pratelli, The sharp quantitative Sobolev inequality for functions of bounded variation. J. Funct. Anal. **244**, 315–341 (2007)
44. N. Fusco, F. Maggi, A. Pratelli, The sharp quantitative isoperimetric inequality. Ann. Math. **168**, 941–980 (2008)
45. N. Fusco, F. Maggi, A. Pratelli, Stability estimates for certain Faber–Krahn, isocapacitary and Cheeger inequalities. Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **8**, 51–71 (2009)
46. N. Fusco, V. Millot, M. Morini, A quantitative isoperimetric inequality for fractional perimeters. J. Funct. Anal. **261**, 697–715 (2011)
47. R.R. Hall, A quantitative isoperimetric inequality in $n$-dimensional space. J. Reine Angew. Math. **428**, 161–176 (1992)
48. R.R. Hall, W.K. Hayman, A.W. Weitsman, On asymmetry and capacity. J. d'Analyse Math. **56**, 87–123 (1991)
49. A. Hurwitz, Sur le problème des isopérimètres. C. R. Acad. Sci. Paris **132**, 401–403 (1901)
50. F. Maggi, Some methods for studying stability in isoperimetric type problems. Bull. Am. Math. Soc. **45**, 367–408 (2008)
51. F. Maggi, *Sets of Finite Perimeter and Geometric Variational Problems. An Introduction to Geometric Measure Theory*. Cambridge Studies in Advanced Mathematics, vol. 135 (Cambridge University Press, Cambridge, 2012)
52. F. Maggi, M. Ponsiglione, A. Pratelli, Quantitative stability in the isodiametric inequality via the isoperimetric inequality. Trans. Am. Math. Soc. **366**, 1141–1160 (2014)
53. R.J. McCann, Existence and uniqueness of monotone measure-preserving maps. Duke Math. J. **80**, 309–323 (1995)
54. R.J. McCann, A convexity principle for interacting gases. Adv. Math. **128**, 153–179 (1997)
55. V.D. Milman, G. Schechtman, *Asymptotic Theory of Finite-dimensional Normed Spaces. With an appendix by M. Gromov*. Lecture Notes in Mathematics, vol. 1200 (Springer, Berlin, 1986)
56. R. Neumayer, A strong form of the quantitative Wulff inequality, arXiv:1503.06705 (2015)
57. G. Pólya, G. Szegö, *Isoperimetric Inequalities in Mathematical Physics*. Annals of Mathematics Studies, vol. 27 (Princeton University Press, Princeton, 1952)
58. K. Rajala, X. Zhong, Bonnesen's inequality for John domains in $\mathbb{R}^n$. J. Funct. Anal. **263**, 3617–3640 (2012)
59. J. Steiner, Einfacher Beweis der isoperimetrischen Hauptsätze. J. Reine Angew. Math. **18**, 281–296 (1838); Gesammelte Werke, vol. 2, 77–91 (Reimer, Berlin, 1882)
60. I. Tamanini, Regularity Results for Almost Minimal Oriented Hypersurfaces Quaderni del Dipartimento di Matematica dell'Università di Lecce, Lecce, 1984. Also available at http://cvgmt.sns.it/paper/1807/

# Mathematical Problems in Thin Elastic Sheets: Scaling Limits, Packing, Crumpling and Singularities

**Stefan Müller**

## 1 Introduction

### *1.1 Thin Objects are Different*

Thin elastic objects have fascinated mathematicians and engineers for centuries and more recently have also become an object of intense study in theoretical physics, biology and material design. While there have been a number of mathematical theories for thin elastic objects for a long time, a new rigorous variational approach has only emerged more recently. In these lectures I will review some of the variational tools which have emerged and discuss a number of open and challenging problems.

From a mathematical point of view the deformation of thin objects is interesting because it leads to more complex, more nonlinear and at the same time more universal behaviour. Geometry, rather than specific material dependent constitutive relations, drives complex behaviour. Here are some typical features of thin objects:

- Small forces can lead to large deformations
- Large rotations arise easily
- Compression leads to instability (buckling)
- Even small forces are sufficient to push the system into a deeply nonlinear regime, beyond classical bifurcation analysis

S. Müller (✉)
Hausdorff Center for Mathematics & Department of Applied Mathematics, University of Bonn, Bonn, Germany
e-mail: sm@hcm.uni-bonn.de



J. Ball, P. Marcellini (eds.), *Vector-Valued Partial Differential Equations and Applications*, Lecture Notes in Mathematics 2179,
DOI 10.1007/978-3-319-54514-1_3

- Energy may concentrate on irregular, lower dimensional sets (e.g., in crumpling)
- Often fine scale structure in the form of wrinkles arises, frequently on multiple scales

Thin objects also often have optimal features, such as maximum strength at given weight. Therefore their performance and their, sometimes dramatic, failure have been studied for a long time in structural engineering. More recently thin elastic sheets have also created a lot of interest in the physics community as model systems for the concentration of energy (see, e.g., Witten's review [116]) and in biology as a model for the formation of spatial patterns, see, e.g., Sharon et al. [109]. For a general introduction with many interesting examples see the book by Audoly and Pomeau [8]. New experimental techniques to prescribe the intrinsic geometry of a thin elastic sheets have led to new questions about predicting the three dimensional shapes such sheets will occupy and a broad spectrum of potential applications ranging from micromachines through programmable (meta)materials to new architectural designs [6, 38, 56, 102, 110].

In these notes we will consider the limiting behaviour of thin elastic objects as the thickness $h$ goes to zero. To organise the wealth of interesting phenomena and mathematical questions one can broadly distinguish two types of problems:

(a) problems where the solution has a well-defined limit as $h \to 0$ and the natural goal is characterise the limit; or
(b) problems where the solution develops increasing complexity (e.g., finer and finer wrinkles) and the goal is to understand the scale and geometry of this complexity.

Problems of type (b) are familiar in many other contexts where one has the competition between a nonconvex energy which favours the formation of fine scale microstructure and a higher order singular perturbation which limits the fineness of the microstructure. In these notes I will mostly focus on (a). Problems of type (b) are briefly discussed in Sect. 6.1 in connection with crumpled sheets and in the final Sect. 7 which essentially consists of a list of some interesting open problems and pointers to the literature.

For problems where the solution has a well-defined limit one can further distinguish between

(a1) problems where the limiting solutions have localized defects which play an important role in the limiting theory and
(a2) problems where the limiting solutions do not have defects, or where the presence of defects is not particularly important for the limiting theory.

There is a large literature on problems of type (a1) in the context of Ginzburg-Landau theories where the defects take the form of vortices (or point singularities) in two dimensions and of vortex lines in three dimensions. In these notes I will mostly focus on (a2), with the exception of Sect. 5. In that section I will discuss the effect of point defects in thin elastic sheets which either are induced by external forces (d-cones) or by an intrinsic defect in the underlying metric (r-cones). In

this case one expects a logarithmic term in the scaling of the energy just as in the Ginzburg-Landau theories. The situation is, however, more subtle since in contrast to GL theories the basic field is not a general scalar (or complex-valued) field but a gradient field, so that one cannot apply cut-and-paste arguments.

## *1.2 Formulation of Mathematical Theories for Lower Dimensional Objects*

There are two general approaches to formulate mathematical theories for lower dimensional objects. The first approach is to formulate new theories for one and two dimensional objects from scratch based on suitable kinematic restrictions and lower dimensional versions of the fundamental balance laws. The second approach is to derive such theories from the three dimensional theory of nonlinear elasticity. Examples of the first approach go back to Euler ('Euler elastica' as unstretchable objects which only have bending stiffness) and the Cosserat brothers. An excellent account of the modern version of this approach can be found in Antman's book [5].

The second approach (derivation from three dimensional elasticity) has also been followed for a very long time. The classic implementation is to make a certain *ansatz* for the expected form of the three dimensional deformation (based on physics or engineering intuition) and to carry out a formal expansion into the small thickness parameter. Since one can make many different reasonable ansatzes this has led to a large variety of plate and shell theories which have been very useful for certain applications but which may lead contradicting predictions and whose range of validity remains somewhat unclear. In particular the von Kármán plate theory has been both in wide use and been faced with serious criticism.

Truesdell writes about von Kármán's theory: 'Analysts seems to love it, and it makes no sense to critical students of mechanics'. He then discusses five specific objections to the theory (which he attributes to S.S. Antman) and concludes 'These objections do not prove that anything is wrong with von Kármán's strange theory. They merely suggest that it would be difficult to prove that anything is right with it' [113, pp. 601–602].

Ciarlet writes in his three-volume treatise on elasticity: 'The two-dimensional von Kármán equations play an almost mythical role in applied mathematics' [26, p. 367].

As we will see the mysteries in the derivation of lower dimensional theories such as the von Kármán theory disappear if one follows a variational approach based on minimization of the energy as we will in these lectures. This approach is ansatz-free. A special asymptotic form of the underlying deformation in the limit of vanishing thickness emerges as rigorous mathematical conclusion and is not an assumption of the theory. One key ingredient is a precise quantitative version of the idea that three dimensional elastic deformations with low energy are very close to rigid motions. This idea goes back to work of F. John in the 1960s but a version necessary for

the derivation of lower dimensional theories appeared only about 15 year ago (see Theorem 2 below).

## 1.3 Mathematical Questions

We consider elastic objects whose reference configuration is a thin three dimensional domain of the form

$$\Omega_h = S \times (-\frac{h}{2}, \frac{h}{2}) \subset \mathbb{R}^3$$

where $S \subset \mathbb{R}^2$ is a bounded domain with Lipschitz boundary. To a deformation

$$u : \Omega_h \to \mathbb{R}^3$$

we associate the elastic energy per unit height

$$E^h(u) = \frac{1}{h} \int_{\Omega_h} W(\nabla u)\, dx.$$

The stored energy density $W$ describes the specific properties of a given elastic material. We will only need the following very general properties of $W : \mathbb{R}^{3\times 3} \to (-\infty, \infty]$:

$$W(RF) = W(F) \quad \forall R \in \mathrm{SO}(3), \quad \text{(frame indifference)} \tag{1}$$

$$W(\mathrm{Id}) = \min W = 0, \quad \text{(normalization)} \tag{2}$$

$$W(F) \geq c\, \mathrm{dist}^2(F, \mathrm{SO}(3)), \quad \text{for some } c > 0 \text{ and all F}, \quad \text{(coercivity)} \tag{3}$$

$$W \text{ is } C^2 \text{ near } \mathrm{SO}(3). \tag{4}$$

Condition (1) expresses the fact that the elastic energy is independent under postmultiplication by a rigid motion or, equivalently, independent under the change to another observer who uses another (oriented) orthonormal frame (frame indifference). Condition (2) is just a normalization, while condition (3) expresses non-degeneracy of the energy near its minimum point Id and for very large deformations. Finally (4) will allow us to use Taylor expansion for deformations with small strain (it could be replaced by the slightly weaker hypothesis that $W$ has a second order Taylor expansion at Id, but we will not pursue this here). We allow that $W(F)$ takes the value $\infty$. In this way one can impose constraints like the condition $\det \nabla u > 0$ which guarantees that $\nabla u$ is (infinitesimally) orientation preserving.

Frame indifference (1) implies that there exists a unique function $\tilde{W} : \mathbb{R}^{3\times 3}_{sym,+} \to (-\infty, \infty)$ defined on positive definite symmetric matrices such that

$$W(F) = \tilde{W}(F^T F) \quad \text{if } \det F > 0. \tag{5}$$

This identity holds in particular in a neighbourhood of SO(3) and by (3) the function $\tilde{W}$ has a unique minimum at Id. Condition (4) implies that $\tilde{W}$ is $C^2$ in a neighbourhood of Id and

$$D^2 W(\mathrm{Id})(G, G) = 4D^2 \tilde{W}(\mathrm{Id})(G, G) \tag{6}$$

for all symmetric matrices $G$.

Natural mathematical questions are:

- What is the scaling of the minimal energy (subject to certain boundary conditions on $u$ or additional energy contributions $-\int_\Omega f \cdot u \, dx$ from applied forces $f$)?
- After suitable rescaling, is there a limiting two dimensional theory?
- What can we say about the solution to specific problems?

The rigorous derivation of lower dimensional theories in the limit $h \to 0$ will be discussed in Sects. 3 and 4 below, the behaviour of specific solutions will be discussed in Sect. 5 on conical singularities and Sect. 6 on packing, crumpling and origami.

## *1.4 Heuristics for Scaling Laws*

To develop some intuition what limiting theories to expect we motivate and explore certain ansatzes to guess the scaling of the energy as $h \to 0$. As a warm-up we first consider the reduction from two to one dimensions.

### 1.4.1 From 2d to 1d

To simplify, we first consider the analogous question of passing from a two dimensional to a one dimensional theory. Thus we consider the previous energy functional on a strip $\Omega_h = (0, L) \times (-\frac{h}{2}, \frac{h}{2})$. A key step is to understand the structure of deformations with low energy. For a map $u : (0, L) \times (-\frac{h}{2}, \frac{h}{2}) \to \mathbb{R}^2$ the energy density $W(\nabla u(x))$ is zero at a point $x$ if and only if

$$|\partial_1 u(z)| = 1, \quad |\partial_2 u(z)| = 1, \quad (\partial_1 u(z), \partial_2 u(z)) = 0$$

and if the pair $(\partial_1 u, \partial_2 u)$ is positively oriented with respect to the standard orientation of $\mathbb{R}^2$. Here $(\cdot, \cdot)$ denotes the standard scalar product in $\mathbb{R}^2$.

It is easy to see that the above conditions can be satisfied for all points $x$ on the midline $(0, L) \times \{0\}$ of the strip if $u$ has the form

$$u(z_1, z_2) = \gamma(z_1) + \nu(z_1) z_2 \tag{7}$$

with

$$|\gamma'| = 1, \tag{8}$$

$$\nu = D^{90}\gamma' \tag{9}$$

where $D^{90}$ denotes a rotation by 90°. These conditions have an easy interpretation. The first condition says that the midline is mapped isometrically to $\mathbb{R}^2$. The second condition asserts that fibres perpendicular to the midline are mapped to lines of equal length, perpendicular to the image of the midline (and with the correct orientation). If we assume that $\gamma$ and $\nu$ are $C^1$ we can compute $\nabla u$ at any point.

$$\nabla u = \underbrace{(\gamma', \nu)}_{\in \mathrm{SO}(2)} + (\nu', 0) \underbrace{z_2}_{\mathcal{O}(h)}.$$

Here we write $(a, b)$ for the matrix $a \otimes e_1 + b \otimes e_2$. Since $\nu$ is a unit vector the vector $\nu'$ is perpendicular to $\nu$ and hence parallel to $\gamma'$. In fact

$$\nu' = -\kappa\gamma' \qquad \text{where } \kappa \text{ is the curvature of } \gamma.$$

This yields

$$\nabla u = ([1 - \kappa z_2]\gamma', \nu) = (\gamma', \nu) \begin{pmatrix} 1 - \kappa z_2 & 0 \\ 0 & 1 \end{pmatrix}$$

and using frame indifference (2) we get

$$W(\nabla u) \sim h^2 \kappa^2 = h^2 |\gamma''|^2.$$

Thus we arrive heuristically at the variational problem that Euler proposed for one dimensional elastica

$$\text{Minimize} \int_0^L |\gamma''|^2 \quad \text{subject to} \quad |\gamma'| = 1.$$

More precisely, if we make the ansatz (7) with (8) and (9) then we get

$$\lim_{h \to 0} \frac{1}{h^2} \frac{1}{h} \int_{(0,L) \times (-\frac{h}{2}, \frac{h}{2})} W(\nabla u)\, dz = \frac{1}{24} E \int_0^L (\gamma''(z_1))^2\, dz_1 \tag{10}$$

with

$$E = D^2W(\mathrm{Id})(e_1 \otimes e_1, e_1 \otimes e_1) \tag{11}$$

Interestingly, a rigorous (ansatz-free) argument based on $\Gamma$-convergence (see below for the definition and properties) shows that the limit functional has indeed the form (10), but that the coefficient $E$ in (11) does not give the correct $\Gamma$-limit. The ansatz (7) is too rigid and misses an important pathway by which the system can reduce its energy. A posteriori one can see that an ansatz of the form

$$u(z_1, z_2) = \gamma(z_1) + \nu(z_1)z_2 + a(z_1)z_2^2$$

would have been sufficient. Initially, however, it is by no means obvious that one needs to include a term of order $\mathcal{O}(h^2)$ in the ansatz.

### 1.4.2 From 3d to 2d

Now we return to the original problem to study objects whose reference configuration is a thin three dimensional domain

$$\Omega_h = S \times (-\frac{h}{2}, \frac{h}{2}).$$

Again we can obtain low energy deformations if we make the following assumptions:

- The midplane $S \times \{0\}$ is mapped isometrically;
- fibres perpendicular to the midplane are mapped to straight segments perpendicular to the image of the midplane and their orientation is preserved.

These assumptions are sometimes referred to as 'Kirchhoff's hypothesis' since Kirchhoff made them in his fundamental paper [55] on geometrically nonlinear plate theory. We shall see later that it is not necessary to make these assumptions; energy bounds guarantee that the deformations must have such a behaviour asymptotically as $h \to 0$, see (23) in Theorem 4 (note that $y$ is defined on the rescaled domain $\Omega \times (-\frac{1}{2}, \frac{1}{2})$ and $\nabla_h y(x) = \nabla u(z)$). Thus we look for an ansatz of the form

$$u(z_1, z_2, z_3) = w(z') + \nu(z')z_3 \qquad \text{where } z' = (z_1, z_2) \tag{12}$$

subject to the constraints

$$(\nabla w)^T \nabla w = \mathrm{Id}_{2\times 2}$$
$$\nu = \partial_1 w \wedge \partial_2 w$$

where $a \wedge b$ denotes the vector product of two vectors in $\mathbb{R}^3$. Both conditions can be summarized by saying that $\partial_1 w, \partial_2 w, \nu$ is an orthonormal basis with the standard orientation or, equivalently, that $(\partial_1 w, \partial_2 w, \nu) \in \mathrm{SO}(3)$.

Since $\nu$ is a unit vector, any derivative $\partial_i \nu$ is perpendicular to $\nu$ and hence can be written as a linear combination of $\partial_1 w$ and $\partial_2 w$. Since $\partial_1 w$ and $\partial_2 w$ are orthonormal the coefficients are just given by the second fundamental form $A$:

$$\partial_i \nu = A_{i1} \partial_1 w + A_{i2} \partial_2 w \quad \text{for } i \in \{1, 2\}.$$

Thus

$$\begin{aligned} \nabla u &= (\partial_1 w, \partial_2 w, \nu) + (\partial_1 \nu, \partial_2 \nu, 0) x_3 \\ &= (\partial_1 w, \partial_2 w, \nu) \begin{pmatrix} \mathrm{Id}_{2\times 2} + x_3 A & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

Thus using frame indifference we get

$$\frac{1}{h} \int_{\Omega_h} W(\nabla u)\, dz \sim h^2 \int_S Q(A)\, dz'$$

where $Q$ is a suitable quadratic form on symmetric $2 \times 2$ matrices. In this way we are led to a variational problem of the form

$$\text{Minimize } \int_S Q(A)\, dz' \quad \text{subject to } (\nabla^T w) \nabla w = \mathrm{Id}_{2\times 2} \tag{13}$$

where $A$ is the second fundamental form of the map $w : S \subset \mathbb{R}^2 \to \mathbb{R}^3$. We will see below that a problem of this form arises indeed as the rigorous $\Gamma$-limit of the three dimensional problem. As in the reduction from 2d to 1d the naive ansatz (12) predicts, however, the wrong form of $Q$, see Theorem 4 below for the correct form of $Q$.

### 1.4.3 From 3d to 1d

Curves in $\mathbb{R}^3$ have a richer geometric structure than planar curves and are characterised by curvature and torsion rather than merely curvature. One can consider a rigorous reduction from three dimensional elasticity to one dimensional rods in $\mathbb{R}^3$ by fixing a cross section $S \subset \mathbb{R}^2$ and considering the tube like domain

$$\Omega_h = (0, L) \times hS \subset \mathbb{R}^3$$

maps $u : \Omega_h \to \mathbb{R}^3$ and the scaled energy

$$E^h(u) = \frac{1}{h^2} \int_\Omega W(\nabla u)\, dx.$$

Here the rescaling by $h^{-2}$ reflects the fact that the volume of $\Omega_h$ scales like $h^2$. In this case the ansatz

$$u(z_1, z_2, z_3) = \gamma(z_1) + \nu(z_1) z_2 + b(z_1) z_3 \tag{14}$$

with

$$R(z_1) := (\gamma'(z_1), \nu(z_1), b(z_1)) \in \mathrm{SO}(3)$$

a Frenet frame of the curve $\gamma$ yields

$$W(\nabla u) \sim h^2 Q(R^T R')$$

where $Q$ is a quadratic form which depends on $D^2 W(\mathrm{Id})$. In terms of the curvature $\kappa$ and the torsion $\tau$ of the curve the matrix $R^T R'$ is given by

$$R^T R' = \begin{pmatrix} 0 & -\kappa & 0 \\ \kappa & 0 & -\tau \\ 0 & \tau & 0 \end{pmatrix}$$

and thus the limiting energy can be expressed as a quadratic form in $\kappa$ and $\tau$. Again the ansatz (14) predicts the right form of the limiting energy but not the correct quadratic form. Indeed that ansatz misses important phenomena such as the energy contribution due to warping of the deformed cross-section if the cross-section is not a disc. The rigorous limiting theory for the scaling $h^{-2}E^h$ was derived in [78]. Similar results were obtained independently by Pantz [98].

At lower energies further rescaling leads to a rod theory which is the one dimensional counterpart of the von Kárman theory for plates [79]. A limiting theory for strings without bending stiffness was established earlier by Acerbi et al. [3]. This theory was the precursor of the 2d membrane limit discussed in Sect. 4.1. Mielke [76] has used a centre manifold approach to compare solutions in a thin tube to a one dimensional problem. His approach already works for finite thickness $h$, but requires that the nonlinear strain $(\nabla u)^T \nabla u$ is uniformly close to the identity (and the approach cannot easily be extended to include applied forces).

### 1.4.4 Two Dimensional Theories for $h > 0$

We will see below in Theorems 4 and 5 that the limit problem (13) for the reduction from 3d to 2d can be rigorously justified in the regime where the energy per unit height scales like $h^2$. One disadvantage of the limit problem is that it is very rigid. In particular only boundary conditions which are compatible with exact isometric immersions of the midplane $S$ are admissible. Also it may well happen that globally the energy is not bounded by $h^2$, but that this is only due to localized singularities (see Sect. 5 for further discussion). In this case no argument by $\Gamma$-convergence is available at the moment, but we would still expect that solutions to the three dimensional behave like the ansatz (12) but with a map $w$ which may slightly deviate from an isometric immersion.

So let $w : S \to \mathbb{R}^3$ be a (smooth) map such that $(\nabla w)^T \nabla w - \mathrm{Id}_2 = \mathcal{O}(h)$ and let

$$\nu := \frac{\partial_1 w \wedge \partial_2 w}{|\partial_1 w \wedge \partial_2 w|}, \quad A_{ij} := (\partial_i \nu, \partial_j w).$$

$$u(z) = w(z_1, z_2) + z_3 \nu(z_1, z_2). \tag{15}$$

Then

$$\partial_i u = \partial_i w + z_3 \partial_i \nu, \quad \partial_3 u = \nu \qquad \forall i \in \{1, 2\}.$$

Now $[(\nabla u)^T \nabla u]_{\alpha\beta} = (\partial_\alpha u, \partial_\beta u)$ for $\alpha, \beta \in \{1, 2, 3\}$. Since $\nu$ is perpendicular to $\partial_i w$ we get that $A_{ij} = (-\partial_i \partial_j w, \nu)$ and in particular $A_{ij} = A_{ji}$. Moreover $\nu$ is perpendicular to $\partial_i \nu$ since $|\nu|^2 = 1$. Thus

$$[(\nabla u)^T \nabla u]_{3i} = 0, \quad [(\nabla u)^T \nabla u]_{33} = 1 \qquad \forall i \in \{1, 2\}$$

and

$$[(\nabla u)^T \nabla u]_{ij} = (\partial_i w, \partial_j w) + 2 z_3 A_{ij} + \mathcal{O}(h^2)$$

Together with (5) and (6) and using that $\int_{-\frac{h}{2}}^{\frac{h}{2}} \frac{h}{2} z_3^2 = \frac{1}{12} h^3$ we find

$$\frac{1}{h} \int_{-\frac{h}{2}}^{\frac{h}{2}} W(\nabla u)\, dz_3 \approx \frac{1}{8} Q_3((\nabla w)^T \nabla w - \mathrm{Id}_2) + \frac{h^2}{24} Q_3(A)$$

since the integral over the term linear in $z_3$ vanishes.

Thus we expect

$$\frac{1}{h} \int_{\Omega_h} W(\nabla u)\, dz \approx \int_S \underbrace{Q\left[(\nabla w)^T \nabla w - \mathrm{Id}_2\right]}_{\text{stretching energy}} + \underbrace{h^2 Q'(A)}_{\text{bending energy}} \, dz' \tag{16}$$

where $Q$ and $Q'$ are suitable positive definite quadratic forms on symmetric $2 \times 2$ matrices. The preceeding calculation suggests that $Q = \frac{1}{8}Q_3$ and $Q' = \frac{1}{24}Q_3$ but once more the ansatz (15) gives the right structure of the two dimensional functional but is too rigid to predict the precise quadratic form correctly.

The first term on the right hand side of (16) corresponds to a stretch of the midplane (which extends more or less uniformly across the thickness) and the second term reflects the effect of curvature: even when the midplane is unstretched positive curvature leads to a stretch on parallel planes above the midplane and a compression on parallel planes below the midplane.

At the moment there is no result in the spirit of $\Gamma$-convergence which guarantees that minimization of the integral on the left hand side and the integral on the right hand side yield asymptotically the same result. One can nonetheless use minimization of the right hand side as a starting point. Once one has some information about the minimizers of this simpler problem it is usually possible to construct test functions with similar energy for the full three dimensional problem. To show that minimizers of the left hand side cannot have substantially lower energy than those on the right hand side is usually harder, but in some cases can be achieved as well, once we have a good understanding of the simpler problem on the right hand side of (16).

## 1.5 Convergence of Minimizers and $\Gamma$-convergence

The results on limiting variational problems are most naturally expressed in terms of $\Gamma$-convergence, even though $\Gamma$-convergence is not really needed to understand and prove them. Consider the following general setting: let $X$ be a metric space and for $k \in \mathbb{N}$ consider functionals $I^k : X \to [-\infty, \infty]$. We would like to define a notion of convergence of the $I^k$ which guarantees that the infimum of $I^k$ converges to the infimum of the limit functional $I$. Pointwise convergence is not enough, not even when $X = [-1, 1]$. Indeed consider the functions

$$I^k(x) = \min(k \left| x - \frac{2}{k} \right|, 1) \qquad \forall x \in [-1, 1].$$

Then the $I^k$ converge pointwise to $I \equiv 1$ and $\min I^k = 0$ for all $k \geq 2$, but $\min I = 1$. De Giorgi found the right notion of convergence which combines a lower bound for *all* sequences which approximate a point with an upper bound for a *particular* sequence.

**Definition 1 (De Giorgi and Franzoni [34])** We say that a sequence $I^k : X \to [-\infty, \infty]$ $\Gamma$-converges to $I : X \to [-\infty, \infty]$ if the following two conditions hold.

(i) For all $x \in X$ and all sequences with $x_k \to x$ we have $\liminf_{k\to\infty} I^k(x_k) \geq I(x)$;

(ii) For all $x \in X$ there exist $y_k \to x$ such that $\limsup_{k\to\infty} I^k(y_k) \le I(x)$.

We write $I^k \overset{\Gamma}{\to} I$ to denote $\Gamma$-convergence.

Using a diagonalization argument one can easily check that a $\Gamma$-limit is always lower semicontinuous. Together with a mild compactness condition on the sublevel sets of $I^k$ $\Gamma$-convergence implies the convergence of minimizers.

**Theorem 1** *Assume that the sequence $I^k$ $\Gamma$-converges to $I$ and satisfies the following compactness condition*

$$\forall t \in \mathbb{R} \quad I^k(y^k) \le t \implies \{y^k : k \in \mathbb{N}\} \text{ is precompact in } X. \tag{17}$$

*Then the minimum of $I$ is attained and*

$$\inf I^k \to \min I.$$

*If, in addition, $I \not\equiv \infty$ then every sequence of approximate minimizers of $I^k$ contains a subsequence which converges to a minimizers of $I$.*

This result is an immediate consequence of the definitions. To illustrate this we recall the short proof.

*Proof* First assume that $\liminf_{k\to\infty} \inf I^k > -\infty$ and $\limsup_{k\to\infty} \inf I^k < \infty$. Then there exist $x_k$ such that $I^k(x_k) \le \inf I^k + \frac{1}{k}$. In particular, the sequence $I^k(x_k)$ is uniformly bounded from above. Now first select a subsequence such that $\lim_{j\to\infty} \inf I^{k_j} = \liminf_{k\to\infty} \inf I^k$. By (17) there exists a further subsequence (not relabelled) which has a limit $x_*$ and the lower bound in $\Gamma$-convergence gives

$$I(x_*) \le \liminf_{j\to\infty} I^{k_j}(x_{k_j}) = \lim_{j\to\infty} \inf I^{k_j} = \liminf_{k\to\infty} \inf I^k.$$

By the upper bound there exist $y_k$ such that $\limsup_{k\to\infty} I^k(y_k) \le I(x_*)$. Thus in particular

$$\limsup_{k\to\infty} \inf I^k \le I(x*).$$

Hence $\limsup_{k\to\infty} \inf I^k = \liminf_{k\to\infty} \inf I^k = I(x_*)$.

If $\liminf_{k\to\infty} \inf I^k = -\infty$ then one can show similarly that there exists $x_*$ with $I(x_*) = -\infty$. The upper bound in $\Gamma$-convergence implies that there exist $y_k$ such that $\limsup_{k\to\infty} I^k(y_k) = -\infty$. Thus $\limsup_{k\to\infty} \inf I^k = -\infty$.

Finally if $\limsup_{k\to\infty} \inf I^k = \infty$ then the upper bound in the definition of $\Gamma$-convergence implies that $I \equiv \infty$. We claim that the lower bound in the definition of $\Gamma$-convergence implies that $\liminf_{k\to\infty} \inf I^k = \infty$. Indeed, if $\liminf_{k\to\infty} \inf I^k < M < \infty$ then there exists a subsequence $k_j$ and points $x_{k_j}$ such that $I^{k_j}(x_{k_j}) < M$. By (17) a further subsequence has a limit point $x_*$. Thus $I(x_*) \le M$, a contradiction.

This finishes the proof of the first statement. A short inspection of the proof shows that the same arguments also prove the second statement.

One obvious, but useful, property of $\Gamma$-convergence is that it is stable under continuous perturbations.

**Proposition 1** *Suppose that* $I^k : X \to [-\infty, \infty]$ *and* $I^k \overset{\Gamma}{\to} I$. *Let* $F : X \to [-\infty, \infty]$ *be continuous. Then*

$$I^k + F \overset{\Gamma}{\to} I + F.$$

*Proof* This follows directly from the definition of $\Gamma$-convergence because $F(x_k) \to F(x)$ and $F(y_k) \to F(x)$ for the sequences $x_k$ and $y_k$ which appear in the definition.

To verify $\Gamma$-convergence it is often useful that it suffices to verify the upper bound for $x$ in a suitable dense set, and up to a small error.

**Lemma 1** *Let* $D \subset X$ *be a dense set with the following additional property*

$$\forall x \in X \ \exists x_j \in D \quad x_j \to x, \quad \limsup_{j\to\infty} I(x_j) \le I(x).$$

*Assume that*

$$\forall \delta > 0 \ \forall x \in D \ \exists x_k \in X \quad x_k \to x, \quad \limsup_{k\to\infty} I^k(x_k) \le I(x) + \delta.$$

*Then property (ii) in Definition 1 holds.*

*Proof* This can be proved by a diagonalization argument.

$\Gamma$-convergence has many additional natural and useful properties. For surveys on $\Gamma$-convergence see Alberti [4], Braides [20] and Dal Maso [32].

We often deal with functionals $I^h$ depending on a continuous parameter $h > 0$. We say that $I^h$ $\Gamma$-converges to $I$ if and only if for every sequence $h_k \to 0$ with $h_k > 0$ the functional $J^k := I^{h_k}$ converges to $I$.

## *Notation*

We use the usual notation $L^p$ and $W^{k,p}$ for the Lebesgue and Sobolev spaces, respectively. By $\rightharpoonup$ we denote weak convergence. We will always take limits $h \to 0$, with $h > 0$. To avoid clumsy notation for subsequences we say that '$a^h \to a$ for a subsequence' if there exists a sequence $h_k$ with $h_k > 0$ and $\lim_{k\to\infty} h_k = 0$ such that $\lim_{k\to\infty} a^{h_k} = a$.

## 2 A Key Ingredient: The Quantitative Rigidity Estimate or Nonlinear Korn Inequality

In the previous section we have seen the construction of certain low energy test functions on thin domains. They have the feature that their gradient is a rotation which varies on a scale of order 1 plus a small perturbation. Our goal is to show the converse: if the energy per unit height of a deformation $u$ scales like $h^2$ then $\nabla u$ must be close to a rotation which only depends on the in-plane variables and the difference quotient of $\nabla u$ is controlled, see Theorem 3 below.

The key ingredient is the following rigidity result for $n$ dimensional maps. It states that if a gradient field is $L^2$ close to the *set* of all rotations then it is close to a *single* rotation, with a linear bound between the two errors. This is a quantitative version of a classical result in geometry and mechanics, often refereed to as Liouville's theorem: a map whose gradient is a rotation at (almost) every point is a rigid motion and in particular affine (for a proof of Liouville's theorem for Sobolev maps see Reshetnyak [105]).

**Theorem 2 ([40, Thm. 3.1])** *Let $n \geq 2$ and let $U \subset \mathbb{R}^n$ be a bounded domain with Lipschitz boundary. There exists a constant $C(U)$ with the following property: for each $v \in W^{1,2}(U;\mathbb{R}^n)$ there is an associated rotation $R$ such that*

$$\|\nabla v - R\|_{L^2(U)} \leq C(U)\| \operatorname{dist}(\nabla u, \mathrm{SO}(n))\|_{L^2(U)}. \tag{18}$$

*Moreover $C(U)$ is invariant under dilation and translation: $C(a + \lambda U) = C(U)$ for all $a \in \mathbb{R}^n$ and $\lambda > 0$.*

Similarly one can obtain estimates in $L^p$ (for $1 < p < \infty$) and even in Lorentz space $L^{p,q}$ with $1 < p < \infty$ and $1 \leq q \leq \infty$, see [30], but we will not need such estimates in these notes.

The invariance under translation and dilation is obvious. If we set $v_{a,\lambda}(x) = \lambda^{-1} v(a + \lambda x)$ then both sides of (18) scale in the same way. We will later use this invariance to apply the estimate to cubes of size $h$.

The estimate (18) can be see as a nonlinear version of Korn's inequality. This inequality states that there exists a constant $C'(U)$ such for each $w \in W^{1,2}(U;\mathbb{R}^n)$ there exists a skew-symmetric matrix $W$ such that

$$\|\nabla w - W\|_{L^2(U)} \leq C'(U)\, \|\operatorname{sym} \nabla w\|_{L^2(U)} \tag{19}$$

where $\operatorname{sym} F = (F^T + F)/2$ is the symmetric part of $F$. Korn's inequality follows from Theorem 2 by linearization. It suffices to apply the theorem to $\mathrm{id} + \delta w$, to note that the skew symmetric matrices are the tangent space of $\mathrm{SO}(n)$ and to pass to the limit $\delta \to 0$.

We now briefly discuss related earlier results. John [52, 53] proved the counterpart of estimate (18) under the additional condition that $u \in C^1$ and that $\operatorname{dist}(\nabla u, \mathrm{SO}(n))$ is uniformly small. In particular he shows that in this case for any

cube $Q \subset U$ there exists a rotation $R_Q$ such that

$$\int_Q |\nabla u - R_Q|^2 \leq \mathcal{L}^n(Q) \sup_Q |\operatorname{dist}(\nabla u, \mathrm{SO}(n))|^2 .$$

Division by $\mathcal{L}^n(Q)$ shows that $\nabla u$ is in the space BMO; in fact John's paper [52] was the birth of BMO. Reshetnyak [107] obtained related results for almost quasiconformal, rather then almost conformal, maps.

It follows from the work of Reshetnyak [105, 106] that if $\operatorname{dist}(\nabla u^{(k)}, \mathrm{SO}(n))$ converges to zero in $L^n(U)$ then a subsequence of $\nabla u^{(k)}$ converges strongly in $L^n(U)$ to a constant $R \in \mathrm{SO}(n)$. Kohn [58] showed an $L^p$ (or $C^0$) estimate for $u$ (rather than $\nabla u$) which follows formally by combining the $L^p$ version of (18) with the Poincaré inequality (see his Theorem (1.4) and estimate (1.5)). Actually he uses a quantity $E_u(x)$ on the right hand side to measure the deviation from a rotation which is a bit stronger than $\operatorname{dist}(\nabla u(x), \mathrm{SO}(n))$.

## 3 Kirchhoff's Geometrically Nonlinear Plate Theory

We return to the study of the energy functional

$$E^h(u) = \frac{1}{h} \int_{\Omega_h} W(\nabla u(z))\, dz$$

for maps

$$u : \Omega_h = S \times (-\frac{h}{2}, \frac{h}{2}) \to \mathbb{R}^3, \qquad S \subset \mathbb{R}^2.$$

In Sect. 1.4 we saw that the ansatz (12) yields maps $u^{(h)}$ which satisfy

$$E^h(u^{(h)}) \leq Ch^2.$$

We now want to show that *every* sequence of maps $u^h$ with $E^h(u^{(h)}) \leq Ch^2$ looks asymptotically like the ansatz and to compute the $\Gamma$-limit of $\frac{1}{h^2}E^h$. The maps $u^h$ are defined on $h$-dependent domains $\Omega_h$. To study the limit it is more convenient to rescale all maps to a fixed domain. We set

$$\Omega = \Omega_1 = S \times (-\frac{1}{2}, \frac{1}{2}),$$
$$z = (z_1, z_2, z_3) = (x_1, x_2, hx_3), \qquad y(x) := u(z).$$

Then $y : \Omega \to \mathbb{R}^3$ and

$$\nabla u(z) = \nabla_h y(x) := (\partial_1 y, \partial_2 y, \frac{1}{h}\partial_3 y).$$

$$E^h(u) = \int_\Omega W(\nabla_h y(x))\, dx =: I^h(y).$$

The following compactness result is crucial.

**Theorem 3 (Compactness [40, Thm. 4.1])** *Assume that $W$ satisfies (3). There exists a constant $C$ which depends only on $S$ and the constant $c$ in (3) with the following property. If $h \in (0, 1]$ then there exists a map $R^{(h)} : S \to \mathrm{SO}(3)$ such that*

*(i)* $\|\nabla_h y^{(h)} - R^{(h)}\|_{L^2(\Omega)} \le C I^h(y^{(h)})$,

*(ii)* $\int_{S'} |R^{(h)}(x' + \xi) - R^{(h)}(x')|^2\, dx' \le C \frac{I^h(y^{(h)})}{h^2}(|\xi|^2 + h^2)$ *for all $S' \subset S$ with $\mathrm{dist}(S', \partial S) > 2(|\xi| + h)$ and all $\xi$.*

*Moreover, if $h_k \to 0$ and*

$$\limsup_{k\to\infty} \frac{1}{h_k^2} I^{h_k}(y^{(h_k)}) < \infty$$

*then*

*(iii)* $\nabla_{h_k} y^{(h_k)}$ *is precompact in $L^2(\Omega; \mathbb{R}^{3\times 3})$ and every cluster point $R$ belongs to $W^{1,2}(\Omega; \mathbb{R}^{3\times 3})$ with $\partial_3 R = 0$ and satisfies $R \in \mathrm{SO}(3)$ a.e.*

*Remark 1 (Loss of Compactness by Wrinkling)* The assumption $\limsup_{h\to 0} h^{-2} I^h(y^{(h)}) < \infty$ is the weakest assumption on $I^h(y^{(h)})$ which implies compactness (if $W$ satisfies (3) and (4)). Indeed, given any function $\omega : (0, \varepsilon) \to (0, \infty)$ with $\limsup_{h\to 0} h^{-2}\omega(h) = \infty$ there exist $y^{(h)}$ such that $I^h(y^{(h)}) \le C\omega(h)$ for all $h \in (0, h_0)$ and a sequence $h_k \to 0$ such that $y^{(h_k)}$ is not compact. One may take the following deformations which correspond to fine scale wrinkling in the $x_2$ direction.

$$y^{(h)}(x_1, x_2, x_3) = \begin{pmatrix} x_1 \\ \gamma^h(x_2) \end{pmatrix} + h x_3 \begin{pmatrix} 0 \\ \nu^h(x_2) \end{pmatrix}$$

with

$$(\gamma^h)'(x_2) = \begin{pmatrix} \cos p[h^{-1}\omega^{1/2}(h)x_2] \\ \sin p[h^{-1}\omega^{1/2}(h)x_2] \end{pmatrix}, \quad \nu^h(x_2) = \begin{pmatrix} -\sin p[h^{-1}\omega^{1/2}(h)x_2] \\ \cos p[h^{-1}\omega^{1/2}(h)x_2] \end{pmatrix}.$$

where $p : \mathbb{R} \to \mathbb{R}$ is a smooth periodic function. Then there exist rotations $R^{(h)}(x_2)$ such that $|\nabla_h y^{(h)} - R^{(h)}| \le C\omega^{1/2}(h)$ and this gives the desired bound for $I^h(y^{(h)})$. One the other hand there exist $h_k \to 0$ with $\omega(h_k)/h_k^2 \to \infty$ and $\nabla_{h_k} y^{(h_k)}$ is not

precompact in $L^2$. The weak $L^2$-limit of $\nabla_{h_k} y^{(h_k)}$ is a constant matrix $(e_1, a, 0)$ and $|a| < 1$. Thus the weak limit does not correspond to an isometric immersion of the mid-plane $S$. See [40, Section 5] for further discussion.

*Proof (Idea of Proof)* We use the following strategy to prove (i) and (ii)

- Rescale back to $\Omega_h$
- Divide $\Omega_h$ into cubes of size $h$ (up to a thin boundary layer)
- Apply the rigidity estimate (18) in each cube
- This yields a piecewise constant map $R^{(h)} : S \to \mathrm{SO}(3)$ which satisfies (i)
- To get the difference quotient estimate (ii) for $\xi = \lambda e_i$ with $|\lambda| \leq h$ apply the rigidity estimate to two neighbouring cubes. By the rigidity estimate there is a single rotation for the union; hence the rotations for the two neighbouring cubes must be close.
- For general $\xi$ iterate the previous estimate and use the triangle inequality.

Now we prove (iii). Let $R^{(h_k)}$ be as in (i) and (ii). By (i) the sequence $\nabla_{h_k} y^{(h_k)}$ is precompact in $L^2(\Omega; \mathbb{R}^{3\times 3})$ if and only if $R^{(h_k)}$ is precompact in $L^2(S; \mathbb{R}^{3\times 3})$ and the cluster points of the two sequences agree (up to the identification of maps $Q : \Omega \to \mathbb{R}^{3\times 3}$ which satisfy $\partial_3 Q = 0$ with maps $\tilde{Q} : S \to \mathbb{R}^{3\times 3}$.

The precompactness of $R^{(h_k)}$ follows from (ii) and the Frechet-Kolmogorov compactness criterion. Here we use that $R^{(h_k)}$ is uniformly bounded to ensure the $R^{(h_k)}$ cannot concentrate near $\partial S$, see [40] for the details. If $R$ is a cluster point of $R^{(h_k)}$ then $R \in \mathrm{SO}(3)$ a.e. by $L^2$ convergence. Moreover $R$ satisfies the difference quotient estimate

$$\int_{S'} |R(x' + \xi) - R(x')|^2 \, dx' \leq CL|\xi|^2 \quad \text{if } \operatorname{dist}(S', \partial S) > 2(|\xi|)$$

for all $\xi \neq 0$ where $L = \limsup_{k\to\infty} h_k^{-2} I^h(y^{(h_k)})$. It follows that $R \in W^{1,2}(S, \mathbb{R}^{3\times 3})$.

Now we can easily establish $\Gamma$-convergence. Define the admissible set of isometric maps

$$\mathcal{A} := \{y \in W^{2,2}(\Omega; \mathbb{R}^3) : \partial_3 y = 0, \ (\partial_i y, \partial_j y) = \delta_{ij} \text{ for } i, j \in \{1, 2\}\}$$

and the limit functional

$$I_{Ki}(y) = \begin{cases} \frac{1}{24} \int_S Q_2(A) \, dx' & \text{if } y \in \mathcal{A}, \\ \infty & \text{else.} \end{cases} \tag{20}$$

Here

$$A_{ij} := (\partial_i \nu, \partial_j y), \quad \text{where } \nu = \partial_1 y \wedge \partial_2 y$$

are the second fundamental form and the normal, respectively, associated to the map $y_{|S}$. The quadratic form $Q_2$ is defined implicitly by minimization. Consider first the quadratic form

$$Q_3(H) := D^2W(\mathrm{Id})(H,H) \tag{21}$$

on $3\times 3$ matrices. Then define for $2\times 2$ matrices $A=(a_{ij})_{i,j\in\{1,2\}}$

$$Q_2(A) := \min_{b\in\mathbb{R}^3} Q_3\begin{pmatrix} a_{11} & a_{12} & b_1\\ a_{21} & a_{22} & b_2\\ b_1 & b_2 & b_3\end{pmatrix}. \tag{22}$$

In the context of $\Gamma$-convergence it is convenient to view $\mathcal{A}$ as a set of maps from $\Omega$ to $\mathbb{R}^3$. Because of the constraint $\partial_3 y = 0$ we can of course view $\mathcal{A}$ equivalently as a set of maps from the two dimensional midplane $S$ to $\mathbb{R}^3$.

**Theorem 4 ($\Gamma$-convergence [40, Thm. 6.1])** *Assume that $W$ satisfies (1)–(4). Then the functionals $h^{-2}I^h$ are $\Gamma$-convergent to $I_{Ki}$. In particular the following assertions hold.*

*(i) (ansatz free lower bound) If $y^{(h)}\to\bar y$ in $W^{1,2}(\Omega;\mathbb{R}^3)$ as $h\to 0$ and*

$$\liminf_{h\to 0}\frac{1}{h^2}I^h(y^{(h)})<\infty$$

*then $\bar y\in\mathcal{A}$ and, for a subsequence,*

$$\nabla_h y^{(h)}\to(\partial_1\bar y,\partial_2\bar y,\nu)\quad \text{with } \nu=\partial_1\bar y\wedge\partial_2\bar y. \tag{23}$$

*Moreover*

$$\liminf_{k\to\infty}\frac{1}{h^2}I^h(y^{(h)})\geq I_{Ki}(\bar y). \tag{24}$$

*(ii) (upper bound) Given $\bar y\in\mathcal{A}$ there exist $\hat y^{(h)}$ such that $\hat y^{(h)}\to\bar y$ in $W^{1,2}(\Omega;\mathbb{R}^3)$ and*

$$\limsup_{h\to 0}\frac{1}{h^2}I^h(\hat y^{(h)})\leq I_{Ki}(\bar y).$$

*Remark 2* If $\liminf_{h\to 0}h^{-2}I^h(y^{(h)})=\infty$ then (24) in (i) holds trivially. In view of (i) in Theorem 3 it suffices to assume that $y^{(h)}$ converges to $\bar y$ strongly in $L^2$ (or weakly in $W^{1,2}$).

Pantz [97, 99] proved similar upper and lower bounds under the additional restriction that the maps $y^{(h)}$ are $C^1$ diffeomorphisms and that $\mathrm{dist}(\nabla y^{(h)},\mathrm{SO}(3))$ is *uniformly* small. Under these assumptions he can use the earlier results of John

[52, 53] instead of the rigidity result in Theorem 2. Of course uniform estimates on $\operatorname{dist}(\nabla y^{(h)}, \mathrm{SO}(3))$ do not follow from mere control of the scaled energy $h^{-2}I^h(y^{(h)})$.

*Proof* *Step 1. Upper bound*

This is easy, up to a little twist. For $\bar{y} \in \mathcal{A} \cap C^2(\bar{\Omega}; \mathbb{R}^3)$ one can use the functions

$$\hat{y}^{(h)}(x', x_3) := \bar{y}(x') + hx_3 \nu(x') + \frac{h^2}{2} x_3^2 d(x')$$

with $\nu = \partial_1 \bar{y} \wedge \partial_2 \bar{y}$ and a suitable choice of $d$. The term involving $d$ (which is absent in the naive ansatz (12)) allows one to really achieve the upper bound with the quadratic form $Q_2$ as defined above. If one drops this term one only gets an upper bound with the quadratic form

$$\tilde{Q}(A) = Q_3 \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

For $\bar{y} \in \mathcal{A} \cap W^{2,\infty}(\Omega; \mathbb{R}^3)$ one can argue in the same way. Finally if we only have $\bar{y} \in \mathcal{A}$ one can carefully approximate $\bar{y}$ by $W^{2,\infty}$ maps (not necessarily isometric immersions) in such a way that the approximation agrees on a very large set with $\bar{y}$ ('truncation of gradients': see [74, 118] for the general approximation scheme and [40] for the application to our concrete situation). Alternatively one can also use the non trivial fact that smooth isometric immersion are dense in $W^{2,2}$ isometric immersions. This was shown by Pakzad [96] for convex sets and by Hornung [49] for bounded sets with Lipschitz boundary. One cannot simply use convolution since this would destroy the isometry constraint. Instead one uses that isometric immersions in $W^{2,2}$ are developable maps and then smoothes the Frénet frame.

*Step 2. Lower bound, overview*

The lower bound is the main point. Let $L := \liminf_{h \to 0} h^{-2} I^h(y^{(h)})$. By assumption $L < \infty$. Thus, for a subsequence,

$$h^{-2} I^h(y^{(h)}) \to L.$$

For ease of notation we assume that this convergence holds along the whole sequence (for the general case one should replace $h$ by $h_k$ with $h_k \to 0$ in the following argument).

Then there are three main ingredients in the argument, the most important one being compactness.

(i) Approximation by maps on $S$ and compactness: there exist $R^{(h)} : S \to \mathrm{SO}(3)$ such that $h^{-1} \|\nabla_h y^{(h)} - R^{(h)}\|_{L^2} \leq C$ and

$$\nabla_h y^{(h)} \to R \quad \text{in } L^2(\Omega; \mathbb{R}^{3\times 3})$$

as $h \to 0$ where

$$R = (\partial_1 \bar{y}, \partial_2 \bar{y}, \nu), \quad \text{with } \nu = \partial_1 \bar{y} \wedge \partial_2 \bar{y}, \tag{25}$$

$$\partial_3 \bar{y} = 0, \qquad \bar{y} \in W^{2,2}. \tag{26}$$

(ii) Identification of the limiting strain: let $R^{(h)}$ be as in (i) and suppose that for a subsequence

$$G^{(h)} := \frac{(R^{(h)})^T \nabla_h y^{(h)} - \mathrm{Id}}{h} \rightharpoonup G \quad \text{in } L^2(\Omega; \mathbb{R}^{3\times 3}). \tag{27}$$

Let $G''$ be the $2 \times 2$ submatrix which consists of the first two rows and columns of $G$. Then

$$G''(x', x_3) = G_0(x') + x_3 A(x') \tag{28}$$

where $A$ is the second fundamental form of $\bar{y}_{|S}$.

(iii) Careful Taylor expansion of $W(\mathrm{Id} + hG^{(h)})$: let $G^{(h)}$ be as in (ii). Then

$$\liminf_{h\to\infty} \frac{1}{h^2} \int_\Omega W(\mathrm{Id} + hG^{(h)})\, dx \geq \frac{1}{24} \int_S Q_2(A)\, dx'.$$

The desired lower bound follows from (i), (ii) and (iii) since

$$W(\nabla_h y^{(h)}) = W(R^{(h)}(\mathrm{Id} + G^{(h)})) = W(\mathrm{Id} + G^{(h)}).$$

Most assertions in (i) follow directly from Theorem 3. It only remains to show that for each sequence $h_k \to 0$ and each $L^2$ cluster point $R$ of $\nabla_{h_k} y^{(h_k)}$ we have

$$R = (\partial_1 \bar{y}, \partial_2 \bar{y}, \partial_1 \bar{y} \wedge \partial_2 \bar{y}). \tag{29}$$

and that (26) holds.

By Theorem 3(iii) every cluster point $R$ of $\nabla_{h_k} y^{(h_k)}$ satisfies $R \in \mathrm{SO}(3)$. By assumption we have for $i \in \{1, 2\}$ that $\nabla_h y^{(h)} e_i = \partial_i y^{(h)} \to \partial_i \bar{y}$ in $L^2$. Thus $Re_i = \partial_i \bar{y}$. Since $R \in \mathrm{SO}(3)$ this implies that $Re_3 = \partial_1 \bar{y} \wedge \partial_2 \bar{y}$ and this proves (29).

To prove the first identity in (26) note that $\partial_3 y^{(h)} = h \nabla_h y^{(h)} e_3 \to 0$ in $L^2$ since $\nabla_h y^{(h)}$ is bounded in $L^2$. Finally by Theorem 3(iii) we have $R \in W^{1,2}$ and hence $\bar{y} \in W^{2,2}$.

The identification of the limiting strain will be discussed below. Finally (iii) follows from Lemma 2 below.

*Step 3. Identification of the limiting strain*

We have

$$\nabla_h y^h = R^{(h)}(\mathrm{Id} + hG^{(h)}). \tag{30}$$

and

$$\lim_{h\to 0} R^{(h)} = \lim_{h\to 0} \nabla_h y^{(h)} = (\partial_1 \bar{y}, \partial_2 \bar{y}, \nu)$$

in $L^2(\Omega; \mathbb{R}^{3\times 3})$. The key ingredient in the proof of (28) is the compatibility condition for gradients, i.e., the fact that second derivatives commute (in the sense of distributions)

$$\frac{1}{h}\partial_3(\nabla y^{(h)})e_i = \frac{1}{h}\partial_3\partial_i y^{(h)} = \partial_i \frac{1}{h}\partial_3 y^{(h)} = \partial_i(\nabla_h y^{(h)})e_3 \quad \forall i \in \{1,2\}.$$

Applying this to (30) and using that $\partial_3 R^{(h)} = 0$ we get

$$R^{(h)}\partial_3 G^{(h)} e_i = \partial_i(R^{(h)}e_3 + hG^{(h)}e_3). \tag{31}$$

If we can pass to the limit $h \to 0$ (in the sense of distributions) we get

$$R\partial_3 G e_i = \partial_i(Re_3) = \partial_i \nu.$$

Let $j \in \{1,2\}$ and take the scalar product with $Re_j = \partial_j \bar{y}$. Then

$$(e_j, \partial_3 G e_i) = (Re_j, R\partial_3 G e_i) = (\partial_j \bar{y}, \partial_i \nu) = A_{ij}.$$

This is the desired conclusion. Now we cannot quite pass to the limit $h \to 0$ on the left hand side of (31) because $\partial_3 G^{(h)}$ converges only weakly in $W^{-1,2}$ and we only know that $R^{(h)}$ converges strongly in $L^2$. This difficulty can be easily overcome if we work with difference quotients in $x_3$ rather than derivatives.

Let $s > 0$ and for $x_3 \in (-1, 1-s)$ define

$$H^{(h)}(x', x_3) := \frac{1}{s}[G^{(h)}(x', x_3 + s) - G^{(h)}(x', x_3)].$$

Now we apply (30) to $e_i$ with $i \in \{1,2\}$, divide by $h$, take the difference quotient, express the difference quotient in $y^{(h)}$ in terms of an integral over $\partial_3 y^{(h)}$ and use that $R^{(h)}$ is independent of $x_3$. This yields

$$\partial_i \frac{1}{s}\int_0^s \frac{1}{h}\partial_3 y^{(h)}(x', x_3 + \sigma)\, d\sigma = R^{(h)}(x')H^{(h)}(x', x_3)e_i$$

where the outer derivative $\partial_i$ is understood in the sense of distributions. The integrand is just $\nabla_h y^{(h)} e_3$ and converges in $L^2$ to $\nu$ as $h \to 0$. Now we can pass to the limit on the right hand side because $H^{(h)}$ converges weakly in $L^2$ and $R^{(h)}$ converges strongly in $L^2$. Since $\nu$ is independent of $x_3$ we thus obtain

$$\partial_i \nu(x') = R(x')H(x', x_3)e_i.$$

It follows that $He_i$ is independent of $x_3$. Thus $Ge_i$ is affine in $x_3$ and $R\partial_3 Ge_i(x', x_3) = \partial_i \nu(x')$ for $x_3 \in (-1, 1-s)$). Since $s > 0$ was arbitrary we get the same identity for $x_3$ and the proof is finished as above by taking the scalar product with $Re_j = \partial_j \bar{y}$.

**Lemma 2 (Taylor Expansions for Lower Bounds)** *Assume that* $W : \mathbb{R}^{3\times 3} \to [0, \infty]$ *is* $C^2$ *in a neighbourhood of* Id *and satisfies* $W(\mathrm{Id}) = 0$. *Assume further that* $h_k > 0$ *with* $h_k \to 0$ *and*

$$G^k \rightharpoonup G \quad \text{in } L^2(\Omega; \mathbb{R}^{3\times 3}) \tag{32}$$

*as* $k \to \infty$. *Then*

$$\liminf_{k\to\infty} \frac{1}{h_k^2} \int_\Omega W(\mathrm{Id} + h_k G^k)\, dx \geq \frac{1}{2} \int_\Omega Q_3(G)\, dx. \tag{33}$$

*where* $Q_3(F) := D^2 W(\mathrm{Id})(F, F)$. *If, in addition,* $W$ *satisfies the frame indifference condition* (2) *and if the* $2 \times 2$ *submatrix* $G''$ *of* $G$ *satisfies*

$$G''(x', x_3) = G_0(x') + x_3 G_1(x')$$

*then*

$$\liminf_{k\to\infty} \frac{1}{h_k^2} \int_\Omega W(\mathrm{Id} + h_k G^k)\, dx \geq \frac{1}{2} \int_S Q_2(\mathrm{sym}\, G_0)\, dx' + \frac{1}{24} \int_S Q_2(\mathrm{sym}\, G_1)\, dx' \tag{34}$$

*where* $Q_2$ *is defined by* (22).

*Proof* The main difficulty is that $h_k G^k$ may not converge uniformly to 0. We will circumvent this problem by restricting to a large set where $h_k G_k$ is uniformly small.

Since $W$ is $C^2$ in a neighbourhood of Id and has a minimum at Id we have $DW(\mathrm{Id}) = 0$ and there exists an increasing function $\omega : [0, \infty) \to [0, \infty)$ with $\lim_{t\to 0} \omega(t) = 0$ such that

$$W(\mathrm{Id} + F) \geq \frac{1}{2} Q_3(F) - \omega(|F|)\, |F|^2.$$

Let

$$\Omega_k := \{x \in \Omega : |G^k| \leq h_k^{-1/2}\}$$

and let $\chi_k$ be the characteristic function of $\Omega_k$. Then

$$\mathcal{L}^3(\Omega \setminus \Omega_h) \leq h_k \int_\Omega |G^k|^2\, dx \to 0.$$

Thus for every $g \in L^2(\Omega)$ we have $\chi_k g \to g$ in $L^2(\Omega)$ as $k \to \infty$. This implies that

$$\chi_k G^k \rightharpoonup G \quad \text{in } L^2(\Omega;\mathbb{R}^{3\times 3}). \tag{35}$$

Since $W \geq 0$ we get

$$\begin{aligned}\int_\Omega W(\mathrm{Id} + h_k G^k)\,dx &\geq \int_\Omega \chi_k W(\mathrm{Id} + h_k G^k)\,dx \\ &\geq \int_\Omega \frac{1}{2}\chi_k Q_3(G_k) - \int_\Omega \chi_k\, \omega(h_k|G^k|)\, h_k^2 |G_k|^2\,dx.\end{aligned}$$

Now $\chi_k \omega(h_k|G^k|) \leq \omega(\sqrt{h_k}) \to 0$. Since $\chi_k$ only takes the values 0 and 1 we also have $\chi_k Q_3(G^k) = Q_3(\chi_k G^k)$. Thus

$$\liminf_{k\to\infty} \frac{1}{h_k^2}\int_\Omega W(\mathrm{Id} + h_k G^k)\,dx = \liminf_{k\to\infty} \frac{1}{2}\int_\Omega Q_3(\chi_k G^k)\,dx.$$

Since $W \geq 0$ the quadratic form $Q_3$ is positive semidefinite and hence convex. Since $\chi_k G^k \rightharpoonup G$ standard lower semicontinuity results imply that

$$\liminf_{k\to\infty} \frac{1}{2}\int_\Omega Q_3(\chi_k G^k)\,dx \geq \frac{1}{2}\int_\Omega Q_3(G)\,dx.$$

This finishes the proof of (32).

To prove (33) note that frame indifference (2) implies that $Q_3(G) = Q_3(\mathrm{sym}\, G)$ and the definition of $Q_2$ implies that $Q_3(\mathrm{sym}\, G) \geq Q_2(\mathrm{sym}\, G'')$. Now (33) follows from (32) by expanding $Q_2(\mathrm{sym}\, G_0 + x_3\, \mathrm{sym}\, G_1)$ and using that for $I = (-\frac{1}{2}, \frac{1}{2})$ we have $\int_I x_3\,dx_3 = 0$ and $\int_I x_3^2\,dx_3 = \frac{1}{12}$.

From the upper and lower bounds in Theorem 4 (i.e., $\Gamma$-convergence of the energy) one easily obtains convergence of minimizers for problems with additional forcing terms.

**Theorem 5 (Convergence of Minimizers in the Presence of Applied Forces)** *Assume that $W$ satisfies (1)–(4). Let $f \in L^2(\Omega;\mathbb{R}^3)$ and assume that $f$ is independent of $x_3$ and the total applied force vanishes, i.e.,*

$$\int_\Omega f\,dx = 0. \tag{36}$$

*Consider the functionals*

$$J^h(y) := \int_\Omega W(\nabla_h y) - \int_\Omega h^2 f\cdot y\,dx,$$

$$J_{Ki}(y) := I_{Ki}(y) - \int_\Omega f\cdot y\,dx,$$

*where $I_{Ki}$ is given by (20). Then the minimum of $J_{Ki}$ is attained and*

$$\lim_{h\to 0} \frac{1}{h^2} \inf J^h = \min J_{Ki}.$$

*Moreover if $y^{(h)}$ is a sequence of almost minimizers of $J^h$, i.e., if $h^{-2}[J^h(y^{(h)}) - \inf J^h] \to 0$ then there exist constants $c^{(h)} \in \mathbb{R}^3$ such that a subsequence of $y^{(h)} - c^{(h)}$ converges strongly in $W^{1,2}(\Omega;\mathbb{R}^3)$ to a minimizer of $J$. We may take $c^{(h)}$ as the average of $y^{(h)}$,*

$$c^{(h)} = \frac{1}{\mathcal{L}^3(\Omega)} \int_\Omega y\, dx.$$

Note that without the assumption (36) we have $\inf J^h = -\infty$. Indeed if $y$ is any deformation with $J^h(y) < \infty$ and $a \in \mathbb{R}^3$ then

$$J^h(a + y) = J^h(y) - a \cdot \int_\Omega f\, dx.$$

Thus if $\int_\Omega f\, dx \neq 0$ optimization in $a$ shows that $\inf J^h = -\infty$.

For convergence of (almost) minimizers subject to certain boundary conditions on $\partial S \times (-\frac{1}{2}, \frac{1}{2})$ (or a large enough subset thereof) see [40, Thm. 6.2].

*Proof* Condition (36) implies that for all $a \in \mathbb{R}^3$ we have $J^h(a + y) = J^h(y)$ and $J_{Ki}(a + y) = J_{Ki}(y)$. Hence we may restrict the problem to the space

$$X = \{y \in W^{1,2}(\Omega;\mathbb{R}^3) : \int_\Omega y\, dx = 0\}.$$

The Poincaré inequality shows that

$$\|y\|_{L^2} \le C\|\nabla y\|_{L^2} \le C\|\nabla_h y\|_{L^2} \quad \forall y \in X \ \forall h \in (0, 1]. \tag{37}$$

It follows from Theorem 4 and Proposition 1 that $J^h \xrightarrow{\Gamma} J_{Ki}$. In view of Theorem 1 it only remains to show that if $h_k \to 0$ and $J^{h_k}(y^{(h_k)}) \le t$ then $y^{(h_k)}$ is precompact in $X$. This follows from the following key estimate. There exists a constant $C$ such that

$$J^h(y) \ge \frac{1}{2}\frac{1}{h^2} I^h(y) - C \tag{38}$$

for all $y$ and all $h \in (0, 1]$. Indeed this estimate and Theorem 3 imply that $\nabla_{h_k} y^{(h_k)}$ is precompact in $L^2$ and precompactness of $y^{(h_k)}$ in $X$ then follows from the Poincaré inequality (37).

To prove (38) note that (3) implies that $W(F) \geq c|F^2| - C$ for some $c > 0$. Thus

$$c\|\nabla_h y\|_{L^2}^2 \leq \int_\Omega W(\nabla_h y)\, dx + C \leq I^h(y) + C$$

and in combination with the Poincaré inequality (37) we get

$$\int_\Omega f \cdot y\, dx \leq \frac{1}{4\delta}\|f\|_{L^2}^2 + \delta\|\nabla_h y\|_{L^2}^2 \leq C_\delta + C\delta I^h(y).$$

Choosing $\delta = \frac{1}{2C}$ we get (38).

## 4 A Hierarchy of Theories Ordered by the Scaling of the Elastic Energy

The previous section gives a very satisfactory theory if the elastic energy per unit height $I^h(y)$ scales like $h^2$, where $h$ is the thickness. Depending on the applied loads and the boundary conditions other scalings of the energy may arise, too.

### *4.1 The Regime $I^h(y) \gg h^2$ and Relaxed Membrane Energies*

Historically the case $I^h(y) \gg h^2$ was the first case in which a rigorous $\Gamma$-limit was established. In this case we cannot expect compactness in $W^{1,2}$ (see Remark 1). We only have weak convergence of (a subsequence of) $\nabla_h y^{(h)}$ in $L^2$ and the limiting theory involves a minimization over all possible oscillations (see the definition of $Qf$ below). More precisely LeDret and Raoult [64–66] considered the scaling $I^h(y) \sim 1$ and (under some technical growth and coercivity conditions on $W$) showed that the $\Gamma$-limit of $I^h$ (with respect to $L^2$, not $W^{1,2}$) is given by the membrane energy

$$I_{me}(\bar{y}) := \int_S (QW_2)(\partial_1 \bar{y}, \partial_2 \bar{y})\, dx',$$

where, as before, $\bar{y}$ satisfies the constraint $\partial_3 \bar{y} = 0$. Here $W_2$ is defined on $3 \times 2$ matrices by

$$W_2((a,b)) = \min_{c \in \mathbb{R}^3} W((a,b,c)) \qquad \forall a \in \mathbb{R}^3, b \in \mathbb{R}^3$$

and where for a function $f : \mathbb{R}^{3\times 2} \to \mathbb{R}$ the symbol $Qf$ denotes its quasiconvexification

$$Qf(G) := \inf \left\{ \int_{(0,1)^2} f(G + \nabla\varphi)\, dx' : \varphi \in C_c^\infty((0,1)^2; \mathbb{R}^3) \right\}. \tag{39}$$

We note in passing that the energies $QW_2$ are necessarily very degenerate. If $W = 0$ on SO(3) and $W \geq 0$ then $QW_2(G) = 0$ for all 'short' linear maps, i.e., for all $G$ such that $|Gx| \leq |x|$ for all $x$. Thus the limiting theory has no resistance to compression. For the connection of the limiting theory with the classical tension field theory in mechanics [103, 104, 115] see Pipkin [100, 101].

One can also consider the minimization of functionals with applied loads

$$J^h(y) = I^h(y) - \int_\Omega h^\alpha f \cdot y\, dx.$$

As in the previous section we assume that the total force vanishes, i.e., $\int_\Omega f\, dx = 0$. For $\alpha = 0$ a subsequence of (almost) minimizers of $J^h$ converges weakly (after possible substraction of a constant) to minimizers of $J_{me}$ given by

$$J_{me}(y) = I_{me}(y) - \int_\Omega f \cdot y\, dx.$$

For $0 < \alpha < 2$ Conti [27] showed that a subsequence of (almost) minimizers of $h^{-\alpha} J^h$ converges weakly (after possible substraction of a constant) to a minimizer of the constrained functional $J_{mc}$

$$J_{mc}(\bar{y}) = I_{mc}(\bar{y}) - \int_\Omega f \cdot \bar{y}\, dx$$

where the constrained membrane energy $I_{mc}$ is defined by

$$I_{mc}(\bar{y}) = \begin{cases} 0 & \text{if } QW_2(\nabla'\bar{y}) = 0 \text{ a.e.} \\ +\infty & \text{else.} \end{cases}$$

Here $\nabla'\bar{y} = (\partial_1\bar{y}, \partial_2\bar{y})$ and we impose the constraint $\partial_3\bar{y} = 0$ as before. Thus to minimize $J_{mc}$ one maximizes the work done by the force $f$ among all deformations $y$ which have zero relaxed membrane energy. Under typical assumptions on $W$ this will be all short maps, i.e., all maps with $(\nabla' y)^T \nabla' y < \mathrm{Id}$.

In the following we mostly focus on the scaling regimes $I^h(y) \ll h^2$. A summary of the results for all scaling regimes $I^h \sim h^\beta$ is given at the end of this section, see Tables 1 and 2.

## 4.2 The Regime $I^h(y) \ll h^2$ and von Kármán Like Theories

This subsection follows very closely [42].

### 4.2.1 Convergence to Affine Isometries

We consider again the energy with an applied force with strength $h^\alpha$.

$$J^h(y) = I^h(y) - \int_\Omega h^\alpha f \cdot y \, dx = \int_\Omega W(\nabla_h y) \, dx - \int_\Omega h^\alpha f \cdot y \, dx.$$

We assume again that the force is independent of $x_3$,

$$\partial_3 f = 0, \tag{40}$$

and that the total force is zero

$$\int_\Omega f \, dx = 0. \tag{41}$$

In Sect. 3 we have seen that for $\alpha = 2$ the (almost) minimizers $y^{(h)}$ of $J^h$ satisfy $J^h(y^{(h)}) \sim h^2$ and $I^h(y^{(h)}) \sim h^2$ and that a subsequence converges (up to substraction of constants) to a minimizer of Kirchhoff's geometrically nonlinear plate theory.

Now we are interested in the behaviour of (almost) minimizers $y^{(h)}$ for weaker forces, i.e., the case $\alpha > 2$. Using the Poincaré inequality (37) it is easy to see that $h^{-2} I^h(y^{(h)}) \to 0$. Indeed, let $y^{(h)}$ be an almost minimizers of $J^h$. Using the trivial comparison function $y^{(h)}(x) = (x', hx_3)^T$ we see that $J^h(y^{(h)}) \leq Ch^\alpha$. Hence by (38) we have $I^h(y^{(h)}) \leq Ch^2$ and thus $\|\nabla_h y^{(h)}\|_{L^2} \leq C$. The Poincaré inequality (37) implies that $\int_\Omega y \cdot h^\alpha f \, dx \leq Ch^\alpha$. Thus $I^h(y^{(h)}) \leq Ch^\alpha \ll h^2$.

Theorems 3 and 4 still apply and we get that a subsequence of the $y^{(h)}$ (after subtraction of suitable constants) converges in $W^{1,2}$ to $\bar{y} \in \mathcal{A}$ with $I_{Ki}(\bar{y}) = 0$. Now (3) implies that the quadratic form $Q_3 = D^2 W(\mathrm{Id})$ is positive definite on symmetric matrices. Thus $Q_2$ is positive definite on symmetric matrices and therefore the second fundamental form $A$ of $\bar{y}_{|S}$ vanishes. We recall the classical argument that this implies that all second derivatives of $\bar{y}$ vanish.

**Proposition 2** *Let $y \in W^{2,2}(S; \mathbb{R}^3)$ and assume that*

$$(\partial_i y, \partial_j y) = \delta_{ij} \quad \forall i, j \in \{1, 2\}.$$

*Set $\nu = \partial_1 y \wedge \partial_2 y$ and $A_{ij} = (\partial_i \nu, \partial_j y)$. Then*

$$\partial_i \partial_j y = -A_{ij} \nu \quad \forall i, j \in \{1, 2\}.$$

*In particular*

$$|\nabla^2 y|^2 = |A|^2$$

*where* $|\cdot|$ *denotes the Euclidean norm.*

*Proof* We first show that

$$(\partial_i \partial_j y, \partial_k y) = 0 \quad \forall i, j, k \in \{1, 2\}. \tag{42}$$

Indeed

$$0 = \partial_j (\partial_i y, \partial_i y) = (\partial_i \partial_j y, \partial_i y).$$

Taking $(i, j) = (1, 2)$ and $(i, j) = (2, 1)$ we get $(\partial_1 \partial_2 y, \partial_1 y) = (\partial_1 \partial_2 y, \partial_2 y) = 0$ while the choice $i = j$ gives $(\partial_i \partial_i y, \partial_i y) = 0$. Finally for $(i, j) = (1, 2)$ or $(2, 1)$ we also have

$$(\partial_i \partial_i y, \partial_j y) = \partial_i (\partial_i y, \partial_j y) - (\partial_i y, \partial_i \partial_j y) = 0.$$

Thus (42) holds and we get $\partial_i \partial_j y = (\partial_i \partial_j y, \nu)\nu$. Now

$$(\partial_i \partial_j y, \nu) = \partial_i \underbrace{(\partial_j y, \nu)}_{=0} - (\partial_j y, \partial_i \nu) = -A_{ij}.$$

Thus the limit $\bar{y}_{|S}$ is an affine isometry. Therefore there exists a rotation $R \in \mathrm{SO}(3)$ such that

$$R^T \bar{y}(x) = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} = \begin{pmatrix} \mathrm{id}_2(x') \\ 0 \end{pmatrix}.$$

It follows that there exist rotations $\bar{R}^{(h)} \in \mathrm{SO}(3)$ and constants $c^{(h)} \in \mathbb{R}^3$ such that

$$\tilde{y}^{(h)} := (\bar{R}^{(h)})^T y^{(h)} - c^{(h)} \to \begin{pmatrix} \mathrm{id}_2 \\ 0 \end{pmatrix}.$$

We would like to show that a suitable rescaling of $\tilde{y}^{(h)} - \binom{\mathrm{id}_2}{0}$ converges to a nontrivial limit and that the limit is the minimizer of a suitable functional. To get an idea which rescalings and which limit functionals to look for we briefly return to the heuristic ansatz based approach in Sect. 1.4. We will then show that the scalings suggested by the ansatz lead indeed to a rigorous convergence result, see Theorem 6.

### 4.2.2 Heuristic Arguments for the Form of the Limit Functional

In Sect. 1.4 we have seen that an ansatz of the form

$$y^{(h)}(x) = w(x_1, x_2) + hx_3 \nu(x_1, x_2)$$

where $w : S \to \mathbb{R}^3$ a smooth map with $|(\nabla w)^T \nabla w - \mathrm{Id}_2| \leq Ch$ and where $\nu = \frac{\partial_1 w \wedge \partial_2 w}{|\partial_1 w \wedge \partial_2 w|}$ yields

$$\int_\Omega W(\nabla_h y)\, dx \approx \frac{1}{8} \int_S Q_3 \left[ (\nabla w)^T \nabla w - \mathrm{Id}_2 \right] dx' + \frac{1}{24} \int_S Q_3(A)\, dx' \tag{43}$$

where $A_{ij} = (\partial_i \nu, \partial_j w)$. Here we translated the results in Sect. 1.4 into the rescaled setting using the rescalings $(z_1, z_2, z_3) = (x_1, x_2, hx_3)$ and $y(x) = u(z)$ which imply $\nabla_h y(x) = \nabla u(z)$. Now we look for near isometries $w$ which are in addition close to the trivial map $x' \mapsto \binom{x'}{0}$. We make the ansatz

$$w(x_1, x_2) = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} + \begin{pmatrix} h^\gamma u_1(x_1, x_2) \\ h^\gamma u_2(x_1, x_2) \\ h^\delta v(x_1, x_2) \end{pmatrix}.$$

The reason to choose different scaling exponents $\gamma > 0$ and $\delta > 0$ for the in-plane components $w_1, w_2$ and the out-of-plane component $w_3$, respectively, will become clear when we compute the terms in (43). We get

$$\partial_1 w \wedge \partial_2 w = \begin{pmatrix} -h^\delta \partial_1 v \\ -h^\delta \partial_2 v \\ 1 \end{pmatrix} + \mathcal{O}(h^{\gamma+\delta} + h^{2\gamma})$$

and $|\partial_1 w \wedge \partial_2 w| = 1 + \mathcal{O}(h^\delta)$. Neglecting the higher order terms we conclude that

$$\nu \approx \begin{pmatrix} -h^\delta \partial_1 v \\ -h^\delta \partial_2 v \\ 1 \end{pmatrix}, \qquad A \approx -h^\delta \nabla^2 v.$$

Thus we get the following approximations for $y^{(h)}$ and $(\nabla w)^T \nabla w$:

$$y^{(h)}(x', x_3) \approx \begin{pmatrix} x' \\ hx_3 \end{pmatrix} + \begin{pmatrix} U(x') \\ V(x') \end{pmatrix} - hx_3 \begin{pmatrix} \nabla V(x') \\ 0 \end{pmatrix}, \quad \text{with } U = h^\gamma u \text{ and } V = h^\delta V. \tag{44}$$

$$\begin{aligned} [(\nabla w)^T \nabla w]_{ij} &= \left( e_i + \begin{pmatrix} h^\gamma \partial_i u \\ h^\delta \partial_i v \end{pmatrix}, e_j + \begin{pmatrix} h^\gamma \partial_j u \\ h^\delta \partial_j v \end{pmatrix} \right) \\ &= \delta_{ij} + h^\gamma (\partial_i u_j + \partial_j u_i) + h^{2\delta} \partial_i v \partial_j v + \mathcal{O}(h^{2\gamma}) \end{aligned} \tag{45}$$

The right hand side of (44) is commonly called a Kirchhoff-Love ansatz and often used as a starting point for a formal expansion.

We proceed from (45). Neglecting again the higher order terms we get

$$(\nabla w)^T \nabla w - \mathrm{Id}_2 \approx 2h^\gamma \operatorname{sym} \nabla u + h^{2\delta} \nabla v \otimes \nabla v$$

where $\operatorname{sym} G = \frac{1}{2}(G^T + G)$ denotes the symmetric part. Thus for our ansatz

$$\int_\Omega W(\nabla_h y^{(h)})\, dx \approx \frac{1}{8} \int_S Q_3(2h^\gamma \operatorname{sym} \nabla u + h^{2\delta} \nabla v \otimes \nabla v)\, dx' + \frac{h^2}{24} \int_S Q_3(h^\delta \nabla^2 v)\, dx' \tag{46}$$

We focus on applied forces perpendicular to the unperturbed plate, i.e., forces with $f_1 = f_2 = 0$. Then we get for the loading term

$$-\int_\Omega h^\alpha f \cdot y\, dx = -h^{\alpha+\delta} \int_S f_3 v\, dx'.$$

The sum of the right hand side of (46) and the loading term becomes

$$\begin{aligned} &\frac{1}{8} \int_S Q_3(2h^\gamma \operatorname{sym} \nabla u + h^{2\delta} \nabla v \otimes \nabla v) dx' \\ &\quad + h^{2+2\delta} \frac{1}{24} \int_S Q_3(\nabla^2 v)\, dx' - h^{\alpha+\delta} \int_S f_3 v\, dx' \end{aligned} \tag{47}$$

We now explore the different scaling regimes. The last two terms can balance only if $2 + 2\delta = \alpha + \delta$, i.e., if

$$\delta = \alpha - 2. \tag{48}$$

The first term scales like $h^{\min(2\gamma, 4\delta)}$. Hence there is a special case in which all terms scale in the same way namely $2\gamma = 4\delta = 2 + \delta = \alpha + \delta$, i.e.,

$$\delta = 1, \quad \gamma = 2, \quad \alpha = 3.$$

This is the scaling of the von Kármán theory and in this case the approximate functional becomes

$$h^4 \left( \frac{1}{8} \int_S Q_3(2 \operatorname{sym} \nabla u + \nabla v \otimes \nabla v) dx' + \frac{1}{24} \int_S Q_3(\nabla^2 v)\, dx' - \int_S f_3 v\, dx' \right).$$

If $\alpha > 3$ then (48) implies that $\delta > 1$ and thus the term $h^{4\delta}$ becomes negligible compared to $h^{2+2\delta}$. Thus for $\gamma = (2 + 2\delta)/2 = \alpha - 1$ the functional behaves

asymptotically like the functional

$$h^{2\alpha-2}\left(\frac{1}{8}\int_S Q_3(2\,\mathrm{sym}\,\nabla u)dx' + \frac{1}{24}\int_S Q_3(\nabla^2 v)\,dx' - \int_S f_3 v\,dx'\right).$$

This functional is quadratic and the contributions of $u$ and $v$ are decoupled. Since $Q_3$ is positive definite we get $\mathrm{sym}\,\nabla u = 0$ and thus by Korn's inequality $u$ is an infinitesimal isometry, i.e., $u$ is affine and $\nabla u$ is a constant skew-symmetric matrix. The Euler-Lagrange equation for $v$ is a linear fourth order partial differential equation, in the simplest case $\frac{1}{12}\Delta^2 v = f_3$.

Finally, if $\alpha \in (2,3)$ then there are two possibilities. If there exist pairs $(u,v)$ such that $\int_S fv \neq 0$ (in particular this requires that $v$ is non-constant) then

$$2\,\mathrm{sym}\,\nabla u + \nabla v \otimes \nabla v = 0\,, \tag{49}$$

and the energy is determined by the balance of the last to terms. This leads again to $\delta = \alpha - 2$ and the minimal energy among maps which satisfy the constraint (49) is negative and of order $h^{2\alpha-2}$.

On the other hand if $\int_S fv = 0$ for all $(u,v)$ which satisfy (49) then the energy is determined by the competition between the first and the last term in (47). This yields $2\gamma = 4\delta = \alpha + \delta$ and thus $\delta = \frac{1}{3}\alpha$. The minimal energy is again negative and of order $h^{\frac{4}{3}\alpha}$. Since $\alpha < 3$ we have $(2\alpha - 2) < \frac{4}{3}\alpha$ and thus $-h^{2\alpha-2} < -h^{\frac{4}{3}\alpha}$ for $0 < h \ll 1$.

Thus the first regime $\delta = \alpha - 2$ in which nontrivial solutions of (49) exist gives the lower energy. We focus on this regime first and will briefly return to the other regime in Sect. 4.4.

We take $\delta = \alpha - 2$ and rescale the energy by $h^{\alpha+\delta} = h^{2\alpha-2}$. Then the first term explodes as $h \to 0$ unless it is exactly zero, i.e., unless $\gamma = 2\delta$ and (49) holds. Thus we expect that the limit problem consists in minimizing

$$\frac{1}{24}\int_S Q_3(\nabla^2 v)\,dx' - \int_S f_3 v\,dx'$$

subject to the constraint

$$2\,\mathrm{sym}\,\nabla u + \nabla v \otimes \nabla v = 0.$$

We will now state a theorem which shows that the heuristic reasoning indeed captures the correct behaviour of (almost) minimizers as $h \to 0$. The only difference is that the quadratic form $Q_3$ has to be replaced by $Q_2$ defined by (22). The reason for the change from $Q_3$ to $Q_2$ is as before that the ansatz we made does not allow for sufficient relaxation of strains in the direction perpendicular to the midplane.

### 4.2.3 Rigorous Limit Functionals

Let $y^{(h)}$ be a sequence of almost minimizers of $J^h$. We want to analyse the deviation of

$$\tilde{y}^{(h)}(x) := (\bar{R}^{(h)})^T y^{(h)}(x) - c^{(h)} \tag{50}$$

from its limit $\binom{x'}{0}$, for suitably chosen rotations $\bar{R}^{(h)} \in \mathrm{SO}(3)$ and translations $c^{(h)} \in \mathbb{R}^3$. We set $I = (-\frac{1}{2}, \frac{1}{2})$ and consider the averaged in-plane and out-of-plane displacements

$$U^{(h)}(x') := \int_I \begin{pmatrix} \tilde{y}_1^{(h)} \\ \tilde{y}_2^{(h)} \end{pmatrix}(x', x_3) - \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} dx_3, \quad V^{(h)}(x') := \int_I \tilde{y}_3^{(h)}\, dx_3 \tag{51}$$

and their rescalings

$$u^{(h)} = \frac{1}{h^\gamma} U^{(h)}, \quad v^{(h)} = \frac{1}{h^\delta} V^{(h)} \tag{52}$$

defined by parameters $\gamma, \delta \in \mathbb{R}$.

For $u \in W^{1,2}(S, \mathbb{R}^2)$ and $v \in W^{2,2}(S)$ we introduce the generalized von Kármán functional

$$\begin{aligned} I_{vK,\alpha}(u, v) := {} & \frac{\Lambda_\alpha}{2} \int_S Q_2\Big(\frac{1}{2}[\nabla u + (\nabla u)^T + \nabla v \otimes \nabla v]\Big)\, dx' \\ & + \frac{1}{24} \int_S Q_2(\nabla^2 v)\, dx' \end{aligned} \tag{53}$$

where

$$\Lambda_\alpha := \begin{cases} \infty \text{ if } 2 < \alpha < 3, \\ 1 \text{ if } \alpha = 3, \\ 0 \text{ if } \alpha > 3 \end{cases}$$

with the convention that $0 \cdot \infty = 0$. In other words for $\alpha = 3$ we have the usual von Kármán functional

$$I_{vK}(u, v) := \frac{1}{2} \int_S Q_2\Big(\frac{1}{2}[\nabla u + (\nabla u)^T + \nabla v \otimes \nabla v]\Big) + \frac{1}{24} Q_2(\nabla^2 v)\, dx', \tag{54}$$

for $\alpha > 3$ we have the "linearized" von Kármán functional

$$I_{vK,\,\mathrm{lin}}(v) = \frac{1}{24} \int Q_2(\nabla^2 v)\, dx',$$

and for $2 < \alpha < 3$ we also have $I^{vK}_{\text{lin}}$ but subject to the nonlinear constraint

$$\nabla u + (\nabla u)^T + \nabla v \otimes \nabla v = 0. \tag{55}$$

A symmetrized gradient $e = \operatorname{sym} \nabla u$ satisfies $\partial_2\partial_2 e_{11} + \partial_1\partial_1 e_{22} - 2\partial_1\partial_2 e_{12} = 0$ (in the sense of distributions). Thus if (55) holds with $v \in W^{2,2}(S)$ we must have

$$\det(\nabla^2 v) = 0. \tag{56}$$

Conversely (56) is sufficient for the existence of a map $u$ such that (55) holds, see [42, Proposition 9].

Geometrically, (56) is exactly the condition that the Gauss curvature of the graph of $v$ vanishes. Thus, at least for sufficiently smooth functions, (56) is equivalent to existence of an isometric map from the graph of $v$ to $\mathbb{R}^2$, see [42, Theorem 7] for a precise statement.

**Theorem 6 (von Kármán Like Theories [42, Thm. 2])** *Suppose that $W$ satisfies (1)–(4) and the applied forces independent of $x_3$, normal, i.e.,*

$$f_1 = f_2 = 0$$

*and have vanishing total force and total moment*

$$\int_\Omega f_3 \, dx = 0, \quad \int_\Omega x' f_3 \, dx = 0.$$

*Then the following assertions hold.*

*(i) (linearized isometry constraint) Suppose $2 < \alpha < 3$ and set $\beta = 2\alpha - 2$, $\gamma = 2(\alpha - 2)$, $\delta = \alpha - 2$. If $\alpha \in (2, \frac{5}{2})$ suppose in addition that $S$ is simply connected. Then $0 \geq \inf J^h \geq -Ch^\beta$. If $y^{(h)}$ is a $\beta$-minimizing sequence (i.e., if $h^{-\beta}(J^h(y^{(h)}) - \inf J^h) \to 0$) then there exists constants $\bar{R}^{(h)} \in SO(3)$ and $c^{(h)} \in \mathbb{R}^3$ such that $\bar{R}^{(h)} \to \bar{R}$ and $\tilde{y}^{(h)}$ and the scaled in-plane and out-of-plane deformations given by (50)–(52) satisfy (for a subsequence)*

$$\nabla_h \tilde{y}^{(h)} \to Id \quad \text{in } L^2(\Omega; \mathbb{R}^{3\times 3}), \tag{57}$$

$$u^{(h)} \to \bar{u} \quad \text{in } W^{1,2}(S; \mathbb{R}^2), \qquad v^{(h)} \to \bar{v} \quad \text{in } W^{1,2}, \tag{58}$$

*Eq. (55) holds and $\bar{v} \in W^{2,2}$. Moreover the pair $(\bar{v}, \bar{R})$ minimizes the functional*

$$J_{vK,\,lin}(v, R) = \frac{1}{24} \int Q_2(\nabla^2 v) \, dx' - R_{33} \int_S f_3 \cdot v \, dx', \tag{59}$$

*subject to*

$$\det(\nabla^2 v) = 0. \tag{60}$$

*(ii) (vK theory)* *Suppose that $\alpha = 3$ and set $\beta = 4$, $\gamma = 2$, $\delta = 1$. Then $0 \geq J^h \geq -Ch^\beta$ and for a (subsequence of a) $\beta$-minimizing sequence (57) and (58) hold and the limit $(\bar{u}, \bar{v}, \bar{R})$ minimizes the usual von Kármán functional*

$$J_{vk}(u, v, R) = I_{vk}(u, v) - R_{33} \int_S f_3 \cdot v \, dx'.$$

*(iii) (linearized vK theory)* *Suppose $\alpha > 3$ and set $\beta = 2\alpha - 2$, $\gamma = \alpha - 1$ and $\delta = \alpha - 2$. Then $0 \geq \inf J^h \geq -Ch^\beta$ and for a (subsequence of a) $\beta$-minimizing sequence (57) and (58) hold with $\bar{u} = 0$ and the pair $(\bar{v}, \bar{R})$ minimizes the linearized von Kármán functional*

$$J_{vK,\,lin}(v, R) = \frac{1}{24} \int_S Q_2(\nabla^2 v) \, dx' - R_{33} \int_S f_3 \cdot v \, dx'.$$

*In all cases we have convergence of the scaled energy $h^{-\beta} J^h(y^{(h)})$ to the minimum of the limit functional. Moreover for $f_3 \not\equiv 0$ we have $\bar{R}_{33} = 1$ or $\bar{R}_{33} = -1$.*

*Remark 3* If $\bar{R}_{33} = 1$ then $\bar{R}$ is an in-plane rotation and $y^{(h)}$ is close to $\bar{R}\binom{x'}{0}$ (up to translation). If $\bar{R}_{33} = -1$ then $\bar{R}$ is an in-plane rotation followed by a 180° out-of-plane rotation $R_0 = \mathrm{diag}(-1, 1, -1)$. Since $J^0$ is invariant under the transformation $(u, v, R) \mapsto (u, -v, R_0 R)$ it suffices to consider the (conventional) situation $R_{33} = 1$.

*Remark 4* Under the *assumption* that the minimizers $y^{(h)}$ of the three dimensional problem admit an asymptotic expansion $y^{(h)} = y^{(0)} + hy^{(1)} + h^2 y^{(2)} + \ldots$ where the coefficients $y^{(k)}$ are bounded in suitable spaces Ciarlet [25] showed that the equations for the leading order non trivial terms are given by the von Kármán equations if the forces are scaled in the natural way.

## 4.3 Strategy of Proof in the von Kármán Scaling

For simplicity we focus on the von Kármán regime: $\alpha = 3$, $\beta = 4$, $\delta = 1$, $\gamma = 2$. The main point is to show the following combined compactness and $\Gamma$-convergence type result for $I^h$. The assertion for the minimizer of $J^h$ then follows from a Poincaré type inequality, similar to, but slightly more subtle than, the one used in the proof of Theorem 5, see [42, Section 7.1, p. 219] for the details. For sequences with bounded scaled energy we obtain in general only weak convergence of $u^{(h)}$ in $W^{1,2}$. For the argument that for *minimizing* sequences this can be improved to strong convergence, see [42, Section 7.2, p. 219]. This argument uses the fact that for minimizing

sequences the integrands are equiintegrable and uses an equiintegrable version of the rigidity estimate in Theorem 2. For the proof of that result see Conti [30].

**Theorem 7 ([42, Thm. 3])** *Let the exponents for the rescaling of the averaged in-plane and out-of-plane deviations by given by $\gamma = 2$ and $\delta = 1$, respectively. Then the following assertions hold.*

*(i) (Compactness and ansatz-free lower bound) If*

$$\liminf_{h\to 0} \frac{1}{h^4} I^h(y^{(h)}) < \infty$$

*then there exist constants $\bar{R}^{(h)} \in \mathrm{SO}(3)$ and $c^{(h)} \in \mathbb{R}^3$ such that, for a subsequence, $\bar{R}^{(h)} \to \bar{R}$ and the maps $\tilde{y}^{(h)} := [\bar{R}^{(h)}]^T y^{(h)} - c^{(h)}$ and the scaled in-plane and out-of plane deformations defined by (51) and (52) satisfy (for a subsequence)*

$$\left.\begin{aligned} \nabla_h y^{(h)} &\to \mathrm{Id} \quad in \quad L^2(\Omega;\mathbb{R}^{3\times 3}),\\ u^{(h)} &\rightharpoonup u \quad in \quad W^{1,2}(S;\mathbb{R}^2),\\ v^{(h)} &\to v \quad in \quad W^{1,2}(S), \quad v \in W^{2,2}(S). \end{aligned}\right\} \tag{61}$$

*Moreover*

$$\liminf_{h\to 0} \frac{1}{h^4} I^h(y^{(h)}) \geq I_{vk}(u, v).$$

*(ii) (Optimality of the lower bound) If $v \in W^{2,2}(S)$ and $u \in W^{1,2}(S;\mathbb{R}^2)$ there exist $\hat{y}^{(h)}$ such that (61) holds and*

$$\lim_{h\to 0} \frac{1}{h^4} I^h(\hat{y}^{(h)}) = I_{vk}(u, v).$$

*Proof* *Proof of optimality*. Assume first that $u$ and $v$ are smooth and use the ansatz

$$\hat{y}^h(x', x_3) = \begin{pmatrix} x' \\ hx_3 \end{pmatrix} + \begin{pmatrix} h^2 u \\ hv \end{pmatrix} - h^2 x_3 \begin{pmatrix} \partial_1 v \\ \partial_2 v \\ 0 \end{pmatrix} + h^3 x_3 d^{(0)} + \frac{h^3}{2} x_3^2 d^{(1)}$$

where $d^{(i)} : S \to \mathbb{R}^3$. Note that for $d^{(0)} = d^{(1)} = 0$ this ansatz agrees with the one used in the heuristic calculation above. Now

$$\nabla_h \hat{y}^{(h)} = \mathrm{Id} + \left(\begin{array}{c|c} h^2 \nabla u & -h(\nabla v)^T \\ \hline h\nabla v & 0 \end{array}\right) - h^2 x_3 \left(\begin{array}{c|c} \nabla^2 v & 0 \\ \hline 0 & 0 \end{array}\right)$$
$$+ h^2 d^{(0)} \otimes e_3 + h^2 x_3 d^{(1)} \otimes e_3 + \mathcal{O}(h^3).$$

A short calculation shows that

$$\lim_{h\to 0} h^{-4} W(\nabla_h \hat{y}^{(h)}) = \lim_{h\to 0} h^{-4} \tilde{W}[(\nabla_h \hat{y}^{(h)})^T \nabla_h \hat{y}^{(h)})] = Q_3(A + x_3 B)$$

where

$$A = \operatorname{sym} \nabla u + \frac{1}{2} \nabla v \otimes \nabla v + \frac{1}{2} |\nabla v|^2 e_3 \otimes e_3 + \operatorname{sym}[d^{(0)} \otimes e_3],$$
$$B = -\nabla^2 v + \operatorname{sym}[d^{(1)} \otimes e_3]$$

and optimizing over $d^{(0)}$ and $d^{(1)}$ we obtain the assertion for smooth $u$ and $v$. For $u \in W^{1,2}$ and $v \in W^{2,2}$ optimality of the bound follows by approximation.

The proof of part (i) of Theorem 7 involves again three steps, as the proof of Theorem 4

- Compactness: proof of (61)
- Identification of the limiting strain
- Use of Lemma 2 to prove the lower bound

### 4.3.1 Outline of the Argument

For the first two points the main thing is to show that using only the condition $I^h(y^{(h)}) \leq Ch^4$ and a good choice of the constant rotations $\bar{R}^{(h)}$ we can show that $u^{(h)}$ and $v^{(h)}$ converge to $u$ and $v$, respectively and that $\tilde{y}^{(h)}$ looks essentially like the ansatz used in the heuristic argument, i.e.,

$$\nabla_h \tilde{y}^{(h)} \approx \mathrm{Id} + \left( \begin{array}{c|c} h^2 \nabla u - x_3 h^2 \nabla^2 v & -h \nabla v + \mathcal{O}(h^2) \\ \hline h(\nabla v)^T & \mathcal{O}(h^2) \end{array} \right) \tag{62}$$

This implies

$$(\nabla_h \tilde{y}^{(h)})^T \nabla_h \tilde{y}^{(h)} - \mathrm{Id} \approx h^2 \left( \begin{array}{c|c} 2 \operatorname{sym} \nabla u + \nabla v \otimes \nabla v - 2x_3 \nabla^2 v & \mathcal{O}(1) \\ \hline \mathcal{O}(1) & \mathcal{O}(1) \end{array} \right). \tag{63}$$

Since the quadratic form $Q_2$ depends only on the entries $[(\nabla_h \tilde{y}^{(h)})^T \nabla_h \tilde{y}^{(h)}]_{ij}$ for $i, j \in \{1, 2\}$ the limits $u$ and $v$ are thus enough to compute the limiting energy and the unknown terms of order 1 do not matter for the lower bound.

We first sketch the general strategy of the argument and then give precise statements and proofs, see Theorem 8, Lemmas 3 and 4. The starting point is the approximation of $\tilde{y}^{(h)}$ by rotations. In Theorem 3 we saw that there exists maps

$R^{(h)} : S \to \mathrm{SO}(3)$ such that

$$\|\nabla_h \tilde{y}^{(h)} - R^{(h)}\|_{L^2(\Omega)} \leq C\sqrt{I^h(\tilde{y}^{(h)})} = C\sqrt{I^h(y^{(h)})} \leq Ch^2 \tag{64}$$

and that the maps $R^{(h)}$ satisfy a difference quotient estimate. We will show shortly that we can actually construct maps $R^{(h)}$ which are in addition differentiable and satisfy the gradient estimate

$$\|\nabla R^{(h)}\|_{L^2(S)} \leq C\frac{1}{h}\sqrt{I^h(\tilde{y}^{(h)})} \leq Ch. \tag{65}$$

This implies in particular that $R^{(h)}$ is close to a constant (in an $L^q$ sense). By a good choice of the constant rotation $\bar{R}^{(h)}$ in the definition of $\tilde{y}^{(h)}$ we may assume that this constant is the identity matrix. Using the fact that the exponential map maps skew symmetric matrices to SO(3) we get that

$$R^{(h)} = \mathrm{Id} + hW^{(h)} + \frac{h^2}{2}{W^{(h)}}^2 + \mathcal{O}(h^3), \qquad W^{(h)} = \begin{pmatrix} 0 & -W_{21}^{(h)} & -W_{31}^{(h)} \\ W_{21}^{(h)} & 0 & -W_{32}^{(h)} \\ W_{31}^{(h)} & W_{32}^{(h)} & 0 \end{pmatrix}$$

where $W^{(h)}$ is bounded in $L^q$. Thus

$$\nabla_h \tilde{y}^{(h)} \approx \mathrm{Id} + hW^{(h)}(x') + \frac{h^2}{2}(W^{(h)})^2(x') + h^2 G^{(h)}(x', x_3)$$

where $G^{(h)}$ is bounded in $L^2(\Omega)$. Now we can bring $v^{(h)}$ and $u^{(h)}$ into the picture. First it follows from the definition of $v^{(h)}$ that

$$\nabla v^{(h)} = (W_{31}^{(h)}, W_{32}^{(h)}) + \mathcal{O}(h) = \frac{1}{h}(R_{31}^{(h)}, R_{32}^{(h)}) + \mathcal{O}(h).$$

Since $\frac{1}{h}\nabla R^{(h)}$ is bounded in $L^2$ this shows that $\nabla v^{(h)} \to \nabla v$ in $L^2$ and that $\nabla^2 v \in L^2$. For the averaged in-plane components we obtain

$$\nabla u = \frac{1}{h}\begin{pmatrix} 0 & -W_{21}^{(h)} \\ W_{21}^{(h)} & 0 \end{pmatrix} + \mathcal{O}(1).$$

This first looks bad because the first term is of order $\mathcal{O}(h^{-1})$. This first term is, however, skew-symmetric and we thus conclude that $\operatorname{sym} \nabla u^{(h)}$ is bounded in $L^2$. It follows from Korn's inequality that there exist *constant* skew-symmetric $2 \times 2$ matrices $\bar{W}^{(h)}$ such that $\nabla u^{(h)} - \bar{W}^{(h)}$ is bounded in $L^2$. Now by a good choice of the rotations $\bar{R}^{(h)}$ we can actually assume that the constant matrices $\bar{W}^{(h)}$ are zero. Thus we get (for a subsequence) weak convergence $\nabla u^{(h)} \rightharpoonup \nabla u$ in $L^2$. So far we have

shown that

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \nabla_h \tilde{y}^{(h)}\, dx_3 = \mathrm{Id} + \left(\begin{array}{c|c} h^2\nabla u^{(h)} & -h\nabla v^{(h)} + \mathcal{O}(h^2) \\ \hline h(\nabla v^{(h)})^T & \mathcal{O}(h^2) \end{array}\right)$$
$$\approx \mathrm{Id} + \left(\begin{array}{c|c} h^2\nabla u & -h\nabla v + \mathcal{O}(h^2) \\ \hline h(\nabla v)^T & \mathcal{O}(h^2) \end{array}\right) \tag{66}$$

To get some information on the $x_3$ dependence of $\tilde{y}^{(h)}$ we proceed as for the Kirchhoff functional and use the (distributional) compatibility relation

$$\frac{1}{h}\partial_3(\nabla_h \tilde{y}^{(h)})e_i = \frac{1}{h}\partial_3\, \partial_i \tilde{y}^{(h)} = \partial_i \frac{1}{h}\partial_3 \tilde{y}^{(h)} = \partial_i(\nabla_h \tilde{y}^{(h)})e_3 \quad \text{for } i \in \{1,2\}. \tag{67}$$

Since

$$\nabla_h \tilde{y}^{(h)} = R^{(h)} + \mathcal{O}(h^2) = \mathrm{Id} + \left(\begin{array}{c|c} 0 & -h\nabla v^{(h)} \\ \hline h(\nabla v^{(h)})^T & 0 \end{array}\right) + \mathcal{O}(h^2)$$

we get from the convergence $\nabla v^{(h)} \to \nabla v$ and from (67) for $i = 1, 2$

$$\partial_3(\nabla_h \tilde{y}^{(h)})e_i \approx \begin{pmatrix} -h^2\nabla^2 v \\ 0 \end{pmatrix} + \mathcal{O}(h^3). \tag{68}$$

Here the approximate identity $\approx$ should strictly speaking be understood in the space $W^{-1,2}$. Moreover (64) implies that

$$(\nabla_h \tilde{y}^{(h)} - R^{(h)})e_3 - \int_{-\frac{1}{2}}^{\frac{1}{2}} (\nabla_h \tilde{y}^{(h)} - R^{(h)})e_3\, dx_3 = \mathcal{O}(h^2).$$

Since $R^{(h)}$ is independent of $x_3$ the terms in $R^{(h)}$ drop out on the left hand side and we get

$$\nabla_h \tilde{y}^{(h)} e_3 - \int_{-\frac{1}{2}}^{\frac{1}{2}} \nabla_h \tilde{y}^{(h)} e_3\, dx_3 = \mathcal{O}(h^2) \tag{69}$$

Now (69), (68) and (66) imply (62). This finishes the outline of the argument.

### 4.3.2 Detailed Proof

The first ingredient in the detailed proof of the lower bound in Theorem 7 is the following refinement of Theorem 3 which provides an approximation by more regular rotations.

**Theorem 8 ([42, Thm. 6])** *(Approximation by regular rotations) Let $S \subset \mathbb{R}^2$ be a bounded domain with Lipschitz boundary and let $\Omega = S \times (-\frac{1}{2}, \frac{1}{2})$. There exists a constant $\delta_0 > 0$ and a constant $C > 0$ with the following property. If $y \in W^{1,2}(\Omega)$,*

$$E := \int_\Omega \operatorname{dist}^2(\nabla_h y, \mathrm{SO}(3))\, dx$$

*and*

$$E \leq \delta_0 h^2 \tag{70}$$

*then there exists a map $R \in W^{1,2}(S; \mathrm{SO}(3))$ and a constant $\bar{Q} \in \mathrm{SO}(3)$ with*

$$\|\nabla_h y - R\|^2_{L^2(\Omega)} \leq CE, \tag{71}$$

$$\|\nabla R\|^2_{L^2(S)} \leq \frac{C}{h^2} E, \tag{72}$$

$$\|R - \bar{Q}\|^2_{L^p} \leq \frac{C_p}{h^2} E, \quad \forall p < \infty, \tag{73}$$

$$\|\nabla_h y - \bar{Q}\|^2_{L^2} \leq \frac{C}{h^2} E. \tag{74}$$

*Proof* We first show (71) and (72). Then (73) will follow from the Poincaré inequality and (74) follows from (73) and (71).

In the proof of Theorem 3 we used the rigidity estimate to construct an approximation by rotations $R_h : S \to \mathrm{SO}(3)$ which are constant on squares of size $h$ and which satisfy a difference quotient estimate. Let $S'$ be compactly contained in $S$. By mollification at the scale $h$ we obtain a smooth map $\tilde{R}_h : S \to \mathbb{R}^{3\times 3}$ which satisfies the estimates $\|\nabla \tilde{R}_h\|^2_{L^2(S')} \leq Ch^{-2}E$ and the estimate $\|\tilde{R}_h - R_h\|_{L^2(S')} \leq CE$. Condition (70) and the difference quotient estimate guarantee that $R_h$ only changes by $C\sqrt{\delta_0}$ between two neighbouring squares of size $h$ on which $R_h$ is constant. It follows that $\sup_{S'} |\tilde{R}_h - R_h| \leq C\sqrt{\delta_0}$. Thus $\tilde{R}_h$ is uniformly close to SO(3). Now there exists a smooth projection $\pi$ from a tubular neighbourhood of SO(3) to SO(3). Then $R := \pi \circ \tilde{R}_h$ has the desired properties on each subset $S'$ compactly contained in $S$. By using a similar mollification argument near the boundary and a partition of unity we can construct a map $R : S \to \mathrm{SO}(3)$ which satisfies (71) and (72), see [42], pp. 200–203 for the details.

Actually, for the proof of the lower bound in Theorem 7 an estimate on each compactly contained subset $S'$ would be sufficient. For the $u^{(h)}$ and $v^{(h)}$ we only get

convergence in compactly contained subsets, but this suffices to prove a lower bound for the von Kármán functional restricted to $S'$. Since $S'$ was arbitrary we recover the desired lower bound by the monotone convergence theorem.

To prove (73) let $\tilde{Q}$ denote the average of $R$ over $S$. Then by the Poincaré inequality

$$\|R - \tilde{Q}\|_{L^p(S)} \leq C_p \|\nabla R\|_{L^p(S)} \tag{75}$$

We will show that $\tilde{Q}$ is close to SO(3). Since the function $F \mapsto \operatorname{dist}(F, \mathrm{SO}(3))$ is 1-Lipschitz we have for all $x' \in S$

$$|\operatorname{dist}(\tilde{Q}, \mathrm{SO}(3))| = |\operatorname{dist}(\tilde{Q}, \mathrm{SO}(3)) - \operatorname{dist}(R(x'), \mathrm{SO}(3)| \leq |\tilde{Q} - R(x')|.$$

Thus

$$\mathcal{L}^2(S)^{\frac{1}{2}} |\operatorname{dist}(\tilde{Q}, \mathrm{SO}(3))| = \|\operatorname{dist}(\tilde{Q}, \mathrm{SO}(3))\|_{L^2(S)} \leq \|\tilde{Q} - R\|_{L^2(S)} \leq C_2 \|\nabla R\|_{L^2(S)}. \tag{76}$$

By definition of the distance function there exists $\bar{Q} \in \mathrm{SO}(3)$ such that $|\tilde{Q} - \bar{Q}| = \operatorname{dist}(\tilde{Q}, \mathrm{SO}(3))$. Thus (73) follows from (75) and (76).

Now we use the approximation by more regular rotations to establish the compactness properties (61).

**Lemma 3 ([42, Lemma 1])** *Suppose that*

$$\limsup_{h \to 0} \frac{1}{h^4} I^h(y^{(h)}) < \infty.$$

*Then there exist maps $R^{(h)} : S \to \mathrm{SO}(3)$ and constants $\bar{R}^{(h)} \in \mathrm{SO}(3)$ and $c^{(h)} \in \mathbb{R}^3$ such that the maps*

$$\tilde{y}^{(h)} := (\bar{R}^{(h)})^T y^{(h)} - c^{(h)}$$

*and the in-plane and out-of-plane displacements*

$$U^{(h)}(x') := \int_I \begin{pmatrix} \tilde{y}_1^{(h)} \\ \tilde{y}_2^{(h)} \end{pmatrix} (x', x_3) - \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} dx_3, \quad V^{(h)}(x') := \int_I \tilde{y}_3^{(h)} \, dx_3$$

*and their rescalings*

$$u^{(h)} = \frac{1}{h^2} U^{(h)}, \quad v^{(h)} = \frac{1}{h} V^{(h)}$$

*have the following properties.*

$$\|\nabla_h \tilde{y}^{(h)} - R^{(h)}\|_{L^2(\Omega)} \leq Ch^2, \tag{77}$$

$$\|R^{(h)} - \mathrm{Id}\|_{L^q(S)} \leq C_q h \quad \forall q < \infty, \tag{78}$$

$$\|\nabla R^{(h)}\|_{L^2(S)} \leq Ch. \tag{79}$$

*Moreover, for a subsequence,*

$$\frac{1}{h}(R^{(h)} - \mathrm{Id}) \to A \quad \text{in } L^q(S;\mathbb{R}^{3\times 3}) \quad \forall q < \infty, \tag{80}$$

$$\frac{1}{h}(\nabla_h \tilde{y}^{(h)} - \mathrm{Id}) \to A \quad \text{in } L^2(\Omega;\mathbb{R}^{3\times 3}) \tag{81}$$

$$\frac{1}{h^2}\,\mathrm{sym}[R^{(h)} - \mathrm{Id}] \to \frac{A^2}{2} \quad \text{in } L^q(\Omega;\mathbb{R}^{3\times 3}) \quad \forall q < \infty. \tag{82}$$

$$v^{(h)} \to v \quad \text{in } W^{1,2}(S), \quad v \in W^{2,2}(S), \tag{83}$$

$$u^{(h)} \rightharpoonup u \quad \text{in } W^{1,2}(S;\mathbb{R}^2), \tag{84}$$

*with*

$$\partial_3 A = 0, \quad A \in W^{1,2}(S,\mathbb{R}^{3\times 3}), \tag{85}$$

$$A = e_3 \otimes \nabla v - \nabla v \otimes e_3. \tag{86}$$

*Proof* This is taken almost verbally from [42].

*Step 1. Normalization*
Apply first Theorem 8 to $y^{(h)}$ and let $\bar{Q}_h$ be the constant for which (73) and (74) hold. If we take $\bar{R}^{(h)} = \bar{Q}_h$ we get (77)–(79). By Jensen's inequality the average deformation gradient $\bar{F}^{(h)} := \mathcal{L}^3(\Omega)^{-1}\int_\Omega \nabla_h \tilde{y}^{(h)}$ then satisfies $|\bar{F}^{(h)} - \mathrm{Id}| \leq Ch$. Thus if we change $\bar{R}^{(h)}$ and $R^{(h)}$ simultaneously by an additional in-plane rotation of order $h$ then (77)–(79) still hold and we can assume in addition that

$$\int_\Omega \partial_2 \tilde{y}_1^{(h)} - \partial_1 \tilde{y}_2^{(h)}\, dx = 0. \tag{87}$$

By choosing $c^{(h)}$ appropriately we may also assume

$$\int_\Omega \tilde{y}^{(h)} - \begin{pmatrix} x' \\ hx_3 \end{pmatrix} dx = 0. \tag{88}$$

*Step 2. Convergence of* $A^{(h)} := \frac{1}{h}(R^{(h)} - \mathrm{Id})$

By (78) there exists a subsequence (not relabelled) such that $A^{(h)}$ converges weakly

$$A^{(h)} \rightharpoonup A \quad \text{in } W^{1,2}(S;\mathbb{R}^{3\times 3}).$$

Using the compact Sobolev embedding we get (80). Together with (77) we also get (81).
*Step 3. Convergence of* $\frac{1}{h^2}\operatorname{sym}(R^{(h)} - \mathrm{Id})$
Since $(R^{(h)})^T R^{(h)} = \mathrm{Id}$ we have $A^{(h)} + (A^{(h)})^T = -h(A^{(h)})^T A^{(h)}$. Hence $A + A^T = 0$ and after division by $h$ we get (82) from the strong convergence of $A^{(h)}$.
*Step 4. Convergence of the scaled normal and in-plane deviations*
Considering the (31) and (32) component of (80) we get $\nabla v^{(h)} \to \nabla v = (A_{31}, A_{32})$. The normalization (88) implies that $v^{(h)}$ has average zero. Thus the convergence in (83) holds. Moreover $v \in W^{2,2}$ as $A \in W^{1,2}$. From (77) and (82) we see that $\operatorname{sym}\nabla u$ is bounded in $L^2$. Using Korn's inequality and the normalizations (87) and (88) we get (84)
*Step 5. Identification of A.*
By Steps 3 and 4 the matrix $A$ is skew-symmetric, $A_{31} = \partial_1 v$ and $A_{32} = \partial_2 v$. It only remains to show that $A_{12} = 0$. Integrating (81) over $x_3$ and using (84) we get

$$A_{12} = \lim_{h\to 0} h\partial_2 u_1^{(h)} = 0$$

in $L^2(S)$.

We now can identify the limiting strain. We use the same notation as in Lemma 3.

**Lemma 4 (Identification of the Limiting Strain, [42, Lemma 2])** *Let the assumptions in Lemma 3 hold. Let*

$$G^{(h)} := \frac{(R^{(h)})^T \nabla_h \tilde{y}^{(h)} - \mathrm{Id}}{h^2}$$

*Then, for a subsequence,*

$$G^{(h)} \rightharpoonup G \quad \text{in } L^2(\Omega;\mathbb{R}^{3\times 3})$$

*and the* $2\times 2$ *submatrix* $G''$ *given by* $G''_{ij} = G_{ij}$ *for* $i,j \in \{1,2\}$ *satisfies*

$$G''(x', x_3) = G_0(x') + x_3 G_1(x')$$

*where*

$$\operatorname{sym} G_0 = \operatorname{sym}\nabla u + \frac{1}{2}\nabla v \otimes \nabla v, \qquad G_1 = -\nabla^2 v.$$

*Remark 5* Note that this rigorous conclusion is slightly weaker than the relation (62) we obtained by formal reasoning. We can uniquely identify $G_1$ (i.e. $\partial_3 G''$) but for $G_0 = \int_I G''$ we can only identify the symmetric part. This is, however, enough to prove the lower bound because $Q_2(G'')$ depends only on the symmetric part of $G''$.

*Proof* Weak convergence of a subsequence of $G^{(h)}$ follows from (77).

*Step 1. Proof that* $\partial_3 G'' = -\nabla^2 v$.

As in the Kirchhoff scaling regime this is based on the compatibility relations for gradients. For $s > 0$ and $x_3 \in (-1, 1-s)$ consider the difference quotients

$$H^{(h)}(x', x_3) := \frac{1}{s}[G^{(h)}(x', x_3 + s) - G^{(h)}(x', x_3)].$$

Multiply the definition of $G^{(h)}$ by $R^{(h)}$ take the difference quotient and express the difference quotient acting on $y$ by an integral over $\partial_3 y$ (note that $R^{(h)}$ is independent of $x_3$ and drops out when taking difference quotients). This yields for $i, j \in \{1, 2\}$

$$(R^{(h)} H^{(h)})_{ij}(x', x_3) = \partial_j \left( \frac{1}{h} \frac{1}{s} \int_0^s \frac{1}{h} \partial_3 \tilde{y}_i^{(h)}(x', x_3 + \sigma)\, d\sigma \right).$$

The integrand is $(\nabla_h \tilde{y}^{(h)})_{i3}(x', x_3 + \sigma)$. Hence by (81) the right hand side converges in $W^{-1,2}(S \times (-1, 1-s))$ to $\partial_j(A_{i3}) = -\partial_j \partial_i v$. Since $R^{(h)} \to \mathrm{Id}$ in $L^q$ and hence boundedly a.e. the left hand side converges weakly to

$$H_{ij}(x, x_3) = \frac{1}{s}[G_{ij}(x', x_3 + s) - G_{ij}(x', x_3)].$$

Thus

$$H_{ij}(x', x_3) = -\partial_j \partial_i v(x').$$

In particular $H_{ij}$ is independent of $x_3$. Hence $G_{ij}$ is affine in $x_3$ and $\partial_3 G'' = -\nabla^2 v$.

*Step 2. Identification of* $G_0$.

It suffices to consider the averages

$$G_0^{(h)}(x') := \int_{-\frac{1}{2}}^{\frac{1}{2}} G^{(h)}(x', x_3)\, dx_3.$$

Using the relation $R^T F - \mathrm{Id} = R^T(F - R) = (R^T - \mathrm{Id})(F - R) + F - R$ for $R \in SO(3)$ in the definition of $G^{(h)}$ we get

$$G^{(h)} = (R^{(h)} - \mathrm{Id})^T \frac{\nabla_h \tilde{y}^{(h)} - R^{(h)}}{h^2} + \frac{\nabla_h \tilde{y}^{(h)} - \mathrm{Id}}{h^2} + \frac{\mathrm{Id} - (R^{(h)})^T}{h^2}.$$

Now the first term converges to zero in $L^1(\Omega)$ by (77) and (78). Thus after integration over the $x_3$ variable and passing to the weak limit in $L^2$ we get from (84) and (82)

$$\operatorname{sym} G_0'' = 0 + \operatorname{sym} \nabla u - \frac{(A^2)''}{2}.$$

Now $A = e_3 \otimes \nabla v - \nabla v \otimes e_3$ and thus $-(A^2)'' = \nabla v \otimes \nabla v$. This finishes the proof.

## 4.4 Overview of the Hierarchy of Models

The different limiting theories and the regimes of applied forces in which they arise are summarized in Table 1, taken from [42]. Our discussion so far has been focussed on natural boundary conditions. The results can often be extended to prescribed Dirichlet type boundary conditions on all or part of $\partial S \times (-\frac{1}{2}, \frac{1}{2})$, see [40, 42] for details.

For certain rather rigid boundary conditions new scaling regimes can arise. One example are the fully clamped boundary conditions

$$y^{(h)}(x', x_3) = \begin{pmatrix} x' \\ hx_3 \end{pmatrix} \quad \text{on } \partial S \times (-\frac{1}{2}, \frac{1}{2}). \tag{89}$$

This corresponds to the condition that the deformation is the identity map $\partial S \times (-\frac{h}{2}, \frac{h}{2})$ in the original variables. The relevance of these boundary conditions is that they imply that the scaled in-plane deviations $u$ satisfy (for $\bar{R}^{(h)} = \mathrm{Id}$)

$$\int_S \operatorname{tr} \operatorname{sym} \nabla u \, dx' = \int_S \operatorname{div} u \, dx' = 0.$$

Thus if

$$\operatorname{sym} \nabla u + \nabla v \otimes \nabla v = 0 \tag{90}$$

then

$$\int_S |\nabla v|^2 \, dx' = \int -\operatorname{tr} \operatorname{sym} \nabla u \, dx' = 0.$$

Hence the infinitesimal isometry constraint (90) can only be satisfied if $\nabla v = 0$. Thus the heuristic discussion following (47) suggest that for $0 < \alpha < 3$ the optimal scaling is given by $\beta = \frac{4}{3}\alpha$, $\gamma = \frac{2}{3}\alpha$ and $\delta = \frac{1}{3}\alpha$. Moreover the limit problem

**Table 1** Relation between the scaling exponents $\alpha$ of the applied forces, $\beta$ of the energy $\gamma$ of the in-plane deformation and $\delta$ of the out-of-plane deformation

| $\alpha$ | $\beta$ | $\gamma$ | $\delta$ | | |
|---|---|---|---|---|---|
| Applied force | Energy | In-plane | Out-of-plane | Limit model | Reference |
| $\alpha = 0$ | 0 | 0 | 0 | Membrane | Le Dret and Raoult [65] |
| $0 < \alpha < 1$ | $\alpha$ | 0 | 0 | Constrained membrane | Conti [27] |
| $\alpha = 2$ | $\alpha$ | 0 | 0 | Bending, isometric midplane | Friesecke et al. [40] |
| $2 < \alpha < 3$ | $2\alpha - 2$ | $2(\alpha - 2)$ | $\alpha - 2$ | Linearized isometry constraint | Friesecke et al. [42] |
| $\alpha = 3$ | $2\alpha - 2$ | $2(\alpha - 2)$ | $\alpha - 2$ | von Kármán | Friesecke et al. [42], Monneau [77] |
| $\alpha > 3$ | $2\alpha - 2$ | $\alpha - 1$ | $\alpha - 2$ | Linearized vK | Friesecke et al. [42] |

For $\alpha > 2$ we assume that the limit force is normal

should consist in minimizing

$$\frac{1}{2}\int_S Q_3(\operatorname{sym}\nabla u + \frac{1}{2}\nabla v \otimes \nabla v)\,dx' - \int_S f_3 v\,dx' \tag{91}$$

subject to

$$u = v = 0 \quad \text{on } \partial S.$$

This energy was proposed by Föppl. It is similar to the von Kármán energy but without the term in $\nabla^2 v$ which represents bending stiffness. Experimentally, the predictions of Föppl's theory (for clamped boundary conditions) have been confirmed over a range of forces that covers four orders of magnitude by Head and Sechler [48]. Nonetheless the precise status of Föppl's theory as a limiting theory had been unclear until recently. In [29] it was shown that if one incorporates the boundary conditions (89) into the functional by setting

$$\tilde{I}^h(y) := \begin{cases} \int_\Omega W(\nabla_h y) & \text{if (89) holds,} \\ \infty & \text{else} \end{cases}$$

and if one formally extends $\tilde{I}_h$ to $L^2(\Omega;\mathbb{R}^3)$ by setting it to $\infty$ on $L^2(\Omega;\mathbb{R}^3) \setminus W^{1,2}(\Omega;\mathbb{R}^3)$ then the functionals

$$\tilde{J}^h(y) := h^{-\frac{4}{3}\alpha}\left[\tilde{I}^h(y) - h^\alpha \int_\Omega f_3 y_3\,dx\right]$$

$\Gamma$-converge (with respect to the $L^2$ metric) to

$$I_{Fp,\mathrm{rel}}(u,v) - \int_S f_3 v\,dx'$$

where $I_{Fp,\mathrm{rel}}(u,v)$ is the lower semicontinuous envelope (or relaxation) of the Föppl functional

$$I_{Fp}(u,v) := \frac{1}{2}\int_S Q_3(\operatorname{sym}\nabla u + \frac{1}{2}\nabla v \otimes \nabla v)\,dx'.$$

Note that in view of the coercivity condition $W(F) \geq c|F|^2 - C$ strong convergence in $L^2$ is essentially the same as weak convergence in $W^{1,2}$. The latter convergence, however, is not metrizable and hence not so suitable for the general set-up of $\Gamma$-convergence.

The full range limiting theories for the clamped boundary conditions (89) is summarized in Table 2, also taken from [42].

**Table 2** Relation between the scaling exponents for a *clamped* plate, assuming normal forces

| $\alpha$ | $\beta$ | $\gamma$ | $\delta$ | | |
|---|---|---|---|---|---|
| Applied force | Energy | In-plane | Out-of-plane | Limit model | |
| $\alpha = 0$ | 0 | 0 | 0 | Membrane | |
| $0 < \alpha < 3$ | $(4/3)\alpha$ | $(2/3)\alpha$ | $(1/3)\alpha$ | Relaxed Föppl<br>= linearized membrane | Conti et al. [29], Föppl [39] |
| $\alpha = 3$ | $2\alpha - 2$ | $2(\alpha - 2)$ | $\alpha - 2$ | von Kármán | Friesecke et al. [42] |
| $\alpha > 3$ | $2\alpha - 2$ | $\alpha - 1$ | $\alpha - 2$ | Linearized vK | Friesecke et al. [42] |

Föppl's theory (or more precisely its relaxed version) can be seen as a geometrically linear version of membrane theory. Von Kármán's theory which has both membrane and bending contributions lies in between Föppl's theory (capturing only membrane energy) and the linear theory (capturing only bending energy)

### 4.5 Extension to Shells and Non Euclidean Geometries

It is natural at ask whether the $\Gamma$-convergence results can be extended to the case when the underlying two dimensional domain $S$ is not flat, but curved (a shell) or more generally an abstract surface with a Riemannian metric. An early easy example is given in [41], but in general the situation is much more subtle, see, e.g., Efrati et al. [37], Lewicka et al. [71], Lewicka et al. [70], Lewicka and Pakzad [69], Kupferman and Solomon [63], Lewicka et al. [72] and Bhattacharya et al. [17]. For an even more general setting that includes dislocations and disclinations see, e.g., Yavari and Goriely [117], Kupferman and Maor [62] as well as Moshe et al. [83].

## 5 Conical Singularities in Elastic Sheets

In most $\Gamma$-limit functionals only the stretching energy or only the bending energy appears explicitly in the energy. The only exception is the von Kármán theory where both terms enter the energy. The von Kármán theory applies rigorously, however, only when the applied forces and the induced strains are rather small (even though it can often be a good guide to explore more nonlinear regimes). In a number of interesting concrete problems the interaction of stretching energy and bending energy plays an important role at small, but finite, thickness $h$ and these problems can thus not be analyzed by using only the limit models discussed above. We will use an energy functional of the form (16) to explore some of these problems. In fact we will focus on the simplest energy which captures the competition of stretching and bending, i.e., the functional

$$I^h(y) = \int_S \left|(\nabla y)^T \nabla y) - \mathrm{Id}\right|^2 + h^2 |\nabla^2 y|^2 \, dx \tag{92}$$

where $y : S \subset \mathbb{R}^2 \to \mathbb{R}^3$ is now a map defined only on the midplane $S$. Here we have replaced the general quadratic forms by the square of the Euclidean norm and the norm of the second fundamental form $|A|$ by $|\nabla^2 v|$. The latter replacement is motivated by the fact that $|A|^2 = |\nabla^2 v|^2$ if $(\nabla v)^T \nabla v = \mathrm{Id}$, see Proposition 2.

### 5.1 Conical Singularities in a Sheet Pushed into a Hollow Cylinder, d-Cones

This subsection follows closely [84] and [86].

Imagine, or even better perform, the following experiment. Put a hollow cylinder on a table and put a thin sheet on top of the cylinder. Push the sheet into the cylinder along the middle axis of the cylinder. The sheet will form a fold and partially lift

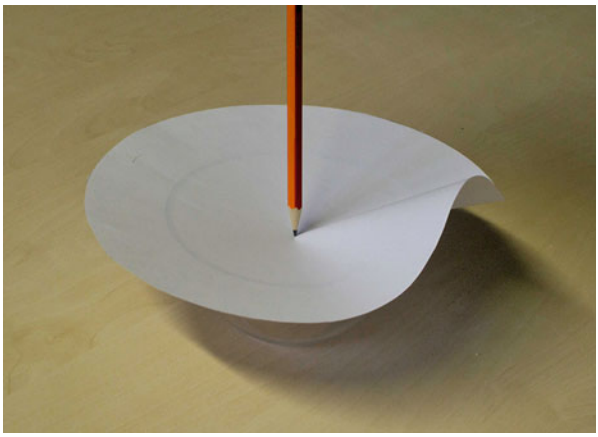

**Fig. 1** When a sheet is pushed into a hollow cylinder a developable cone, or d-cone, forms. Reprinted with permission from [86]

off the cylinder, see Fig. 1. We try to understand what shape the sheet forms and how the shape depends on its thickness $h$, in the limit $h \to 0$. This problem has been discussed extensively in the physics literature, see, e.g., Cerda and Mahadevan [22, 23], Liang and Witten [73] as well as Witten [116], and several remarkable features have been predicted. The angle corresponding to the region where the sheet lifts off the rim of the hollow cylinder is a universal constant ($\approx 139°$), independent of the indentation, the thickness and the material (for small indentations, see Cerda and Mahadevan [23]). The tip of the ridge consists of a crescent-shaped ridge where stress and energy focus. There are predictions based on numerical simulations and certain formal asymptotic arguments that the radius of the crescent scales like $R_{cres} \sim h^{1/3} R_{cyl}$ where $h$ is the thickness of the sheet and $R_{cyl}$ the radius of the cylinder, but in Witten [116] it is argued that the scaling behaviour is not really understood and cannot be derived from the dominant energy scaling.

As a first step towards a better rigorous understanding we try to understand the scaling of the total energy of the sheet as a function of $h$. For the problem as stated no rigorous lower bound for the energy is known. Following Brandman et al. [21] we consider a slightly modified problem for which one can obtain rigorous upper and lower bounds. For $h \to 0$ we expect that the sheet forms a cone which is locally isometric to the plane and hence developable (such cones are often referred to as d-cones in the physics literature). We now first consider the problem of finding upper and lower bounds for the energy (92) when the two dimensional set $S$ is the unit disc $B_1$ and $y$ agrees on $\partial B_1$ with a developable cone $\tilde{y}$. Here we ignore the fact that the sheet is constrained by the container. In a second step one can then minimize over all isometric cones $\tilde{y}$, subject to the constraints that the indentation $d > 0$ is given and that the cone cannot penetrate the boundary of the container. If we move the

vertex of the cone to the origin then any cone $\tilde{y}$ is determined by a curve

$$\gamma : S^1 = \partial B_1 \to \mathbb{R}^3$$

via

$$\tilde{y}(x) = |x|\gamma\left(\frac{x}{|x|}\right).$$

One can easily check that $\tilde{y}$ is an isometric immersion if and only if

$$|\gamma(x)| = 1, \quad |\gamma'(x)| = 1 \qquad \forall x \in S^1, \tag{93}$$

i.e., if $\gamma$ is a unit speed curve on the sphere $S^2$. Indeed both conditions are clearly necessary: the first condition guarantees that radial rays from the origin have the right length while the second condition guarantees that the boundary of the unit disc is mapped to a curve of the correct length $2\pi$. Conversely, it is easy to check that the conditions imply that $(\nabla\tilde{y})^T\nabla\tilde{y} = \mathrm{Id}$, i.e., that $\tilde{y}$ is an isometric immersion.

**Theorem 9 ([84, Thm. 1] or [86, Thm. 1])** *Assume that $\gamma : S^1 \to \mathbb{R}^3$ satisfies (93) and consider the space of admissible maps*

$$V_\gamma := \{y \in W^{2,2}(B_1;\mathbb{R}^3) : y(0) = 0, y_{|\partial B_1} = \gamma\}$$

*and the functional*

$$I^h(y) = \int_S \left|(\nabla y)^T\nabla y) - \mathrm{Id}\right|^2 + h^2|\nabla^2 y|^2\, dx$$

*If the curve $\gamma$ does not lie in a plane then there exist constants $C_2$ and $C_3$, which only depend on $\gamma$, such that for sufficiently small $h > 0$*

$$C_1 \ln\frac{1}{h} - 4C_1 \ln\left(\ln\frac{1}{h}\right) - C_2 \le \frac{1}{h^2}\min_{y\in V_\gamma} I^h(y) \le C_1 \ln\frac{1}{h} + C_3, \tag{94}$$

*where*

$$C_1 = C_1(\gamma) = \frac{1}{\ln 2}\int_{B_1\setminus B_{1/2}} |\nabla^2\tilde{y}|^2 \mathrm{d}x\,.$$

This result improves on [21]. In particular it is show that the deviation from the leading order logarithmic term is a most of order $\ln\ln h^{-1}$. It is not known whether the double logarithm really arises or whether the correction is simply of order 1 (as is assumed in the physics literature). There are some relations to the large literature on Ginzburg-Landau functionals (where it is known that the correction is of order 1).

The main additional difficulty is that in our setting the basic variable is already a gradient field so that one cannot use cut-and-paste arguments to modify the field.

In a second step one can bring back the constraint that the sheet is pushed into a hollow cylinder by requiring in addition that $\gamma_3 \geq d$, where $d > 0$ is the amount of indentation and minimize $C_1(\gamma)$ over all curves which satisfy this additional constraint. This is a one dimensional problem, so one can even hope for explicit solutions. In the physics literature this minimization has been carried out by Cerda and Mahadevan [22, 23]. More precisely in [22] they consider the minimization of $C_1(\gamma)$ in the small deflection approximation (which corresponds to the von Kármán regime studied above). In [23] the equilibrium equations are derived for developable cones in geometrically nonlinear theory (large deflections). This is equivalent to considering the Euler-Lagrange equation for the minimization of $C_1(\gamma)$. In the small deflection approximation the solutions are analyzed in detail. Under the assumption that the set of angles for which the cone lifts off the cylinder is a finite number of intervals the solutions are determined, solutions for different number of intervals (of equal length) are compared and solutions with one interval (of $\approx 139°$) are found to be global minimizers. Olbermann [91] has recently proved that the small deflection regime arises rigorously as a $\Gamma$-limit and has shown that for local minimizers of the small deflection problem the detachment set is indeed always a finite union of intervals (without any apriori assumption on the detachment set).

*Proof (Upper Bound)* This is easy. It suffices to take $y^{(h)} = h\phi\left(\frac{|x|}{h}\right)\gamma\left(\frac{x}{|x|}\right)$ where $\phi \in C^2(\mathbb{R})$ and $\phi(t) = t$ for $t \geq 1$ while $\phi(t) = 0$ for $t \leq \frac{1}{2}$.

The main point is to prove the lower bound. The difficulty is that the stretching energy

$$I_{st}(y) = \int_{B_1} W(\nabla y)\, dx \quad \text{with } W(F) = |F^T F - \mathrm{Id}|^2 \tag{95}$$

is not lower semicontinuous (with respect to weak convergence in the natural energy space $W^{1,4}$). The lower semicontinuous envelope (or relaxation) is given by

$$I_{rel}(y) = \int_{B_1} QW(\nabla y)\, dx \tag{96}$$

where $QW$ is the quasiconvexification of $W$ (see (39) for the definition). Now $W(F) = 0$ for $F^T F = \mathrm{Id}$ implies that $QW(F) = 0$ whenever $F^T F \leq Id$, i.e. for all maps which do not involve stretch. Thus every map $y$ with $\nabla^T y \nabla y \leq \mathrm{Id}$ can be approximated (weakly in $W^{1,4}$) by maps $y^{(k)}$ such that $I_{st}(y^{(k)}) \to 0$.

Thus, roughly speaking, we can expect a good lower bound for $I_{st}$ only if there are regions where stretch occurs. Here the Dirichlet boundary conditions come to our aid. We have

$$|y(z) - y(0)| = |\tilde{y}(z) - 0| = 1 \quad \text{for all } z \in \partial B_1. \tag{97}$$

Thus the endpoints of the curve $\alpha : [0,1] \to \mathbb{R}^3$ with $\alpha(t) = y(tz)$ have distance 1. Therefore the length of $\alpha$ is bigger or equal than 1 and it equals 1 only if $\alpha$ is a straight line, i.e., if $y = \tilde{y}$ on the radial segment $\{tz : t \in [0,1]\}$. If $y$ deviates from $\tilde{y}$ anywhere on this segment then $\alpha$ has length larger than 1 and hence stretching must occur somewhere. In other words, moving a chord which is pulled tight involves some stretching and hence has an energy cost. This idea is quantified in Lemma 6 below.

We now verbally follow [84], pp. 2235–2236 to sketch the main ideas of the proof of the lower bound. For details of the argument see [86]. The proof of the lower bound will be an easy consequence of the following three lemmata. We fix $\gamma$ and will assume that

$$I^h(y) \le C_1 \frac{1}{h^2} \ln \frac{1}{h}\,. \tag{98}$$

We first show that $y$ remains very small close to zero. This will allow us to use an approximate version of (97) where the origin 0 is replaced by a point close to the origin.

**Lemma 5** $\sup_{B_h} |y| \le Ch \ln \frac{1}{h}$.

**Lemma 6** *Assume that* $h \ln \frac{1}{h} \le r_0 \le 1$ *and set* $A_{r_0} = B_{r_0} \setminus \overline{B_{r_0/2}}$. *Then*

$$\int_{A_{r_0}} |y - \tilde{y}|^2 \,\mathrm{d}x \le C r_0^3 h \ln \frac{1}{h}\,.$$

**Lemma 7** *Assume that* $h \ln \frac{1}{h} \le r_0 \le 1$. *Then*

$$\left| \int_{A_{r_0}} \nabla^2(y - \tilde{y}) : \nabla^2 \tilde{y} \,\mathrm{d}x \right| \le C \left(\frac{r_0}{h}\right)^{1/8} \left(\ln \frac{1}{h}\right)^{1/2}.$$

Lemma 5 follows from the scale invariant estimate

$$\sup_{x \in B_h} \left| y(x) - y(0) - x \cdot \frac{1}{|B_h|} \int_{B_h} \nabla y \right| \le C \|\nabla^2 y\|_{L^2(B_h)}\,,$$

the estimate $\left| \int_{B_1} \nabla y \right| = \left| \int_{\partial B_1} y \otimes \nu \right| \le 2\pi$ and the BMO-type estimate

$$\left| \frac{1}{|B_h|} \int_{B_h} \nabla y - \frac{1}{|B_1|} \int_{B_1} \nabla y \right| \le C \left(\ln \frac{1}{h}\right)^{1/2} \|\nabla^2 y\|_{L^2(B)}$$

(to get the optimal exponent $1/2$ in the logarithm one can use e.g. the Trudinger-Moser inequality).

To prove Lemma 6 set $e = y - \tilde{y}$ and denote by $e' = \partial_r e$ the derivative in the radial direction. On a.e. segment $r \mapsto (r\cos\theta, r\sin\theta)$ we have

$$|e(r,\theta) - e(h,\theta)|^2 \le r \int_h^r |e'|^2(\rho,\theta)\,\mathrm{d}\rho \tag{99}$$

and using that $y(1,\theta)\cdot\gamma(\theta) = |\gamma(\theta)|^2 = 1$ we get

$$\int_h^1 |e'(\rho,\theta)|^2\,\mathrm{d}\rho = \int_h^1 \left(|\partial_r y|^2 - 1\right)\,\mathrm{d}\rho - 2h + 2y(h,\theta)\cdot\gamma(\theta)\,. \tag{100}$$

To finish the proof we integrate (99) with respect to $r\,\mathrm{d}r\,\mathrm{d}\theta$, use the pointwise estimate $|(\nabla y)^T\nabla y - Id|^2 \ge (|\partial_r y|^2 - 1)^2$, the Cauchy-Schwarz inequality with respect to $\mathrm{d}\rho\,\mathrm{d}\theta$ and Lemma 5.

To prove Lemma 7 we use integration by parts, Lemma 6, the simple estimate $\|\nabla^2 e\|^2_{L^2(A_r)} \le \ln\frac{1}{h}$ and standard interpolation estimates for $\|\nabla e\|_{L^2(A_r)}$ and $\|\nabla e\|_{L^2(\partial A_r)}$ as well as the homogeneity properties of $\nabla^2\tilde{y}$ and $\nabla^3\tilde{y}$.

*Proof of the Lower Bound in Theorem 9* Let $M \in \mathbb{N}$ with $M \approx \log_2\frac{1}{h} - 4\log_2\ln\frac{1}{h}$. Then by Lemma 7

$$\begin{aligned}
\int_{B_1\setminus B_{2-M}} |\nabla y|^2\,\mathrm{d}x &\ge \int_{B_1\setminus B_{2-M}} |\nabla\tilde{y}|^2\,\mathrm{d}x - 2\int_{B_1\setminus B_{2-M}} \nabla^2(y-\tilde{y}) : \nabla^2\tilde{y}\,\mathrm{d}x \\
&\ge C_1 M\ln 2 - 2C\sum_{k=0}^{M-1}(2^k h)^{1/8}\left(\ln\frac{1}{h}\right)^{1/2} \\
&\ge C_1 M\ln 2 - 2C(2^M h)^{1/8}\left(\ln\frac{1}{h}\right)^{1/2}
\end{aligned}$$

and the assertion follows from the choice of $M$.

## 5.2 *Another Conical Singularity, Regular Cones*

Another method to produce a conical singularity is as follows. Take a two dimensional disc $B_R$ of radius $R$, cut out a sector of angle $2\pi\delta^2$ and reglue the edges. If we ignore bending energy, the sheet will form a radially symmetric cone given by the image of the disc under the map $y_\delta(x) = (\sqrt{1-\delta^2}x, \delta|x|)^T$. In the physics literature this cone is sometimes referred to as a regular cone, or r-cone for short.

Geometrically the cutting-and-glueing procedure is equivalent to a change of metric. Instead of the standard Euclidean metric on the unit disc we use the metric

$$g_\delta(x) = e_r\otimes e_r + (1-\delta^2)e_\theta\otimes e_\theta = \mathrm{Id} - \delta^2 e_\theta\otimes e_\theta \quad \text{where } e_r = \frac{x}{|x|}, e_\theta = \frac{x^\perp}{|x|}.$$

A natural energy for a map $y : D \to \mathbb{R}^3$ is now

$$I_\delta^h(u) = \int_{B_R} |(\nabla y)^T \nabla y - g_\delta|^2 + h^2 |\nabla^2 y|^2 \, dx.$$

If we take $y = y_\delta$ then the first term in the integrand vanishes identically, but the second term behaves like $h^2|x|^{-2}$ and hence the integral diverges logarithmically near 0. As for the developable cone discussed in the previous subsection we expect that the minimizer of $I_\delta^h$ is close to $y_\delta$ but regularized at scale $h$ and that the minimal energy behaves like $h^2(C_1 \ln h^{-1} + \mathcal{O}(1))$ where $C_1$ is a specific constant, computed using the cone $y_\delta$.

### 5.2.1 The Radially Symmetric Case

One advantage of the functional $I_\delta^h$ is that it allows for interesting radially symmetric competitors while the only radially symmetric developable cone is the flat disc. Thus it is natural to test our expectations on the minimum of $I_\delta^h$ and the shape of the minimizers by first minimizing only over radially symmetric maps $y$, i.e., maps which satisfy $y(x) = (a(|x|)x_1, a(|x|)x_2, b(|x|))^T$. In [85] it is shown that for the von Kármán approximation of the energy functional the energy of minimizers among radially symmetric maps behaves indeed like $h^2(C_1 \ln h^{-1} + \mathcal{O}(1))$, with $C_1 = 2\pi\delta^2$, and that minimizers converge almost exponentially to a radially symmetric cone as $|x| \to \infty$.

To describe the results more precisely we first consider the von Kárman approximation in this setting. We make the von Kármán ansatz

$$y(x) = \begin{pmatrix} x \\ 0 \end{pmatrix} + \begin{pmatrix} \delta^2 \hat{U}(x) \\ \delta \hat{V}(x) \end{pmatrix}$$

and get $(\nabla y)^T \nabla y \approx \mathrm{Id} + \delta^2(2 \operatorname{sym} \nabla \hat{U} + \nabla \hat{V} \otimes \nabla \hat{V})$. Neglecting terms which are formally of higher order and using the identity $e_\theta \otimes e_\theta = \mathrm{Id} - e_r \otimes e_r$ we get

$$I_\delta^h(y) \approx \delta^4 \int_{B_R} \left| 2 \operatorname{sym} \nabla \hat{U} + \mathrm{Id} + \nabla \hat{V} \otimes \nabla \hat{V} - e_r \otimes e_r \right|^2 + \frac{h^2}{\delta^2} |\nabla^2 \hat{V}|^2 \, dx. \tag{101}$$

We will from now on work with the functional on the right hand side of (101). It is easy to see that for the standard cone given by $\hat{V}(x) = |x|$ (and the choice $\hat{U}(x) = -\frac{1}{2}x$) the first term vanishes and the second term diverges like $\frac{h^2}{\delta^2}|x|^{-2}$ leading to a logarithmic divergence of the energy.

We now restrict attention for radially symmetric functions, i.e., we assume that

$$\hat{V}(x) = W(r), \quad \hat{U}(x) = \frac{1}{2}(\hat{u}(r) - r)\frac{x}{r}, \quad \text{where } r = |x|,$$

and we set

$$\hat{w}(r) := W'(r).$$

Here the factor $\frac{1}{2}$ in the formula for $\hat{U}$ is included to ensure consistency with the notation in [85]. Then the integral on the right hand side of (101) becomes up to factor of $2\pi$

$$I_R^\lambda(\hat{u}, \hat{v}) = \int_0^R \rho_\lambda(r)\, r\, dr, \quad \text{where } \lambda = \frac{h}{\delta} \text{ and} \tag{102}$$

$$\rho_\lambda(r) := (\hat{u}' + \hat{w}^2 - 1)^2 + \left(\frac{\hat{u}}{r}\right)^2 + \lambda^2 \left(\hat{w}'^2 + \frac{\hat{w}^2}{r^2}\right). \tag{103}$$

We are interested in the limit $h \to 0$ or, equivalently, $\lambda \to 0$. Using the rescaling $\hat{u}_\lambda(r) = \frac{1}{\lambda} u(\lambda r)$ and $\hat{w}_\lambda(r) = \hat{w}(\lambda r)$, where $\hat{u}_\lambda$ and $\hat{w}_\lambda$ are defined on $[0, R/\lambda]$, we see that this is the same as taking $\lambda = 1$ and considering the limit $R \to \infty$. In particular we have

$$\min \frac{1}{\lambda^2} I_R^\lambda = \min I_{R/\lambda}^1. \tag{104}$$

We thus assume from now on

$$\lambda = 1 \quad \text{and we set} \quad \rho(r) := \rho_1(r).$$

We expect that $w(r) \approx 1$ for $r \gg 1$. Thus the integral in (102) diverges logarithmically in $R$. The first key observation in [85] is that we can obtain a well defined limit functional for $R \to \infty$ if we renormalize the energy density $\rho$ by substracting $r^{-2}$ for $r \geq 1$. More precisely consider a cut-off function $\psi \in C^\infty([0, \infty))$ with

$$\psi = 0 \quad \text{on } [0, \frac{1}{2}], \quad \psi = 1 \quad \text{on } [1, \infty)$$

and define formally

$$\begin{aligned} &\hat{E} : \mathcal{W} \to \mathbb{R} \cup \{\infty\} \\ &(\hat{u}, \hat{w}) \mapsto \lim_{R\to\infty} \int_0^R \left(\rho_1(r) - \frac{\psi(r)}{r^2}\right) r\, dr, \end{aligned} \tag{105}$$

where

$$\mathcal{W} := \left\{ (\hat{u}, \hat{w}) \in W^{1,2}_{\text{loc}}((0, \infty); \mathbb{R}^2) : \int_0^1 \rho_1(r)\, r\, dr < \infty \right\}.$$

**Theorem 10 ([85, Thm. 1.1])** *The functional $\hat{E}$ is well-defined, i.e., the limit in (105) exists as an element of $\mathbb{R} \cup \{\infty\}$, and is bounded from below. Moreover $\hat{E}$ possesses minimizers $(\hat{u}, \hat{w}) \in \mathcal{W}$ with $\hat{w} \geq 0$ and $\hat{E}(\hat{u}, \hat{w}) < \infty$. In addition, each minimizer with $\hat{w} \geq 0$ satisfies for any $\sigma < 2$*

$$\hat{u}(r) = \frac{1}{2r} + o\left(e^{-\sigma\sqrt{r}}\right)$$
$$\hat{w}(r) = 1 + o\left(e^{-\sigma\sqrt{r}}\right)$$

*as $r \to \infty$.*

*Remark 6* The restriction $\hat{w} \geq 0$ was only imposed to break the obvious non-uniqueness arising from the symmetry $\hat{w} \mapsto -\hat{w}$.

*Proof (Sketch of Proof)* The main point is to prove a lower bound for $\int_1^R (\rho(r) - r^{-2})\, r\, dr$, uniformly in $R$. The key observation is that the integrand can be split into a positive term, a term with rapid decay and a null Lagrangian, i.e., a term which only depends on the values of $(\hat{u}, \hat{w})$ at 1 and $R$. Indeed with the substitutions $\hat{u} = \frac{1}{2r} + u$ and $\hat{w} = 1 + w$ we get

$$\rho - \frac{1}{r^2} = (2w + w^2 + u')^2 + \left(\frac{u}{r}\right)^2 + w'^2 + \frac{1}{2r^4} - \frac{1}{r}\left(\frac{u}{r}\right)'. \tag{106}$$

The first three terms are positive, the fourth term is integrable against $r\, dr$ and the integral of the last term against $r\, dr$ is $u(R)/R - u(1)$. Careful interpolation arguments then show that $u(R)/R$ can be controlled by the integral of the positive terms on the right hand side of (106) and $u(1)$ can be controlled by $\int_0^1 \rho(r)\, r\, dr$. Thus we obtain the desired uniform lower bound. In view of the rescaling (104) it follows (see [85, Cor 3.5]) that there exists a constant $C$ such that for all $\lambda \leq 1$ the unrenormalized energy on the unit disc $I_1^\lambda$ defined in (102) satisfies

$$\ln \lambda^{-1} - C \leq \inf \frac{1}{\lambda^2} I_1^\lambda \leq \ln \lambda^{-1} + C \quad \text{where } \lambda = \frac{h}{\delta}.$$

This is the counterpart of the bound (94) for the d-cone, but without the ln ln term.

To establish the asymptotics for $\hat{u}$ and $\hat{w}$ we first show by an energy argument that $\lim_{r\to\infty} w(r) = 0$ and then study the Euler-Lagrange equations in the $u, w$ variables. It turns out that after the change of variables $r = s^2$ the linear part of the Euler-Lagrange equation is autonomous up to corrections of order $\frac{1}{s}$ or smaller in the coefficients of the lower order derivatives. A careful analysis of the ODE yields the desired decay for $u$ and $w$.

### 5.2.2 The General Case

We return to the study of minimizers of

$$I_\delta^h(u) = \int_{B_R} |(\nabla y)^T \nabla y - g_\delta|^2 + h^2 |\nabla^2 y|^2 \, dx, \tag{107}$$

where

$$g_\delta(x) = e_r \otimes e_r + (1-\delta^2) e_\theta \otimes e_\theta = \mathrm{Id} - \delta^2 e_\theta \otimes e_\theta$$

with $e_r = \frac{x}{|x|}$, $e_\theta = \frac{x^\perp}{|x|}$, and we drop the assumption of radial symmetry.

For the d-cone discussed in Sect. 5.1 the main argument for a lower bound on the energy was that 'moving a cord which is pulled tight costs energy'. This argument relies heavily on the fact that we described Dirichlet boundary condition. Without a Dirichlet boundary condition the only lower bound which is known in the setting of Sect. 5.1 is that

$$\lim_{h \to 0} \frac{1}{h^2} \min I^h = \infty.$$

Indeed, if this failed, we could find $y^{(h)}$ such that a subsequence of $y^{(h)}$ (after substraction of constants) converges strongly in $W^{1,2}$ to a limit $y \in W^{2,2}$ which satisfies $(\nabla y)^T \nabla y = \mathrm{Id}$. Thus $y$ is developable (see the proof of Proposition 3 below). This is incompatible with the assumptions that the sheet is indented and cannot penetrate the boundary of the cylinder.

For the r-cone problem, i.e., the minimization of (107) we want to derive a lower bound without imposing additional Dirichlet conditions. To achieve this Olbermann [92, 93] pursued an approach which is based on intrinsic geometric quantities, in particular the pull-back metric $g = (\nabla y)^T \nabla y$ and its Gaussian curvature.

His starting point is that the prescribed metric $g_\delta$ is non-Euclidean and its (generalized) Gaussian curvature is a multiple of the Dirac mass at the origin, more precisely $K_\delta = \pi \delta^2 \delta_0$. Now $g$ has to be $L^2$ close to $g_\delta$ to keep the energy small and thus we expect that the Gaussian curvature $K$ of $g$ is close to $K_\delta$ in a suitable negative norm, or equivalently $K$ and $K_\delta$ are close after suitable smoothing. Now one can use the fact that $K$ is the Jacobian of the Gauss map $\nu_y : D \to S^2$ which maps a point $x \in D$ to the normal $\nu_y(x) = \partial_1 y \wedge \partial_2 y / |\partial_1 y \wedge \partial_2 y|$. Assuming for a moment that $y$ is sufficiently regular and the Gauss map is injective (and its image is contained in a half-sphere) we can now bring the isoperimetric inequality into the picture. First we have

$$\mathcal{H}^2(\nu_y(B_r)) = \int_{B_r} K \, dx \approx \int_{B_r} K_\delta \, dx = \pi \delta^2.$$

Thus the isoperimetric inequality gives

$$\mathcal{H}^1(\partial v_y(B_r)) \gtrsim 2\pi\delta.$$

Hence by the Cauchy-Schwarz inequality

$$2\pi r \int_{\partial B_r} |\nabla v_y(x)|^2 \, d\mathcal{H}^1 \geq \left( \int_{\partial B_r} |\nabla v_y(x)|^2 \, d\mathcal{H}^1 \right)^2 \geq (\mathcal{H}^1(\partial v_y(B_r)))^2.$$

Now we expect $|\nabla v_y|^2 \lesssim |\nabla^2 y|^2$ (compare Proposition 2). Putting these estimates together we arrive at

$$\int_{\partial B_r} |\nabla^2 y|^2 \, d\mathcal{H}^1 \gtrsim \frac{4\pi^2\delta^2}{2\pi r} = \frac{2\pi\delta^2}{r}. \tag{108}$$

We cannot expect this to hold for arbitrarily small $r$ since the approximate equality of $K$ and $K_\delta$ involves some smoothing. The natural scale for the smoothing is $h$. If we assume that the previous reasoning can be turned into rigorous estimates for $r \geq Ch$ then integration of (108) from $r = Ch$ to 1 gives the desired lower bound. The above reasoning is only a cartoon of the real argument, but Olbermann showed in [93] that under a mild global assumption one can use a reasoning based on proximity of $K$ and $K_\delta$, the Gauss map, the isoperimetric inequality on $S^2$ and suitable interpolation inequalities to get bounds for the modified functional

$$J_h^\infty(y) := \|g - g_\delta\|^2_{L^\infty(B_1 \setminus B_h)} + h^2 \|Dv_y\|^2_{L^2(B_1)}$$

where, as above, $v_y$ denotes the normal to the surface $y$. He shows that there exist constants $C_\delta$ such that for all $h \in (0, e^{-1})$

$$2\pi\delta^2 \ln h^{-1} - \frac{3}{2} \ln \ln h^{-1} - C_\delta \leq \frac{1}{h^2} \inf J_{h,\delta}^\infty \leq 2\pi\delta^2 \ln h^{-1} + C_\delta. \tag{109}$$

Note that the quantity $m_0^2$ in [93] is related to $\delta$ by $m_0^2 = 1 - \delta^2$.

The estimate (109) represents very important progress because it is based on the intrinsic properties of the prescribed geometry rather than externally imposed boundary conditions. Nonetheless, it still requires a modification of the original energy functional (107) and some global conditions. Recently Olbermann achieved a breakthrough which allows him to treat the original functional without any additional conditions [92] . The key idea is, roughly speaking, to replace the Gaussian curvature in the previous argument by its linearization, i.e., the expression $-\frac{1}{2}(\partial_2\partial_2 g_{11} + \partial_1\partial_1 g_{22} - 2\partial_1\partial_2 g_{12})$. Since

$$g_{ij} = (\partial_i y, \partial_j y) = \sum_{k=1}^{3} \partial_i y^k \partial_j y^k$$

a short calculation shows that

$$-\frac{1}{2}(\partial_2\partial_2 g_{11} + \partial_1\partial_1 g_{22} - 2\partial_1\partial_2 g_{12}) = \sum_{k=1}^{3} \det \nabla^2 y^k.$$

Then an argument which essentially involves the isoperimetric inequality for the sets $w^k(B_r)$, where $w^k = \nabla y^k$, shows that [92, Thm. 1]

$$2\pi\delta^2 \ln h^{-1} - \frac{3}{2}\ln\ln h^{-1} - C_\delta \le \frac{1}{h^2}\inf I_\delta^h \le 2\pi\delta^2 \ln h^{-1} + C_\delta. \tag{110}$$

Actually, with a slight variation of the argument one can even remove the doubly logarithmic term and one can show that minimizers $y^{(h)}$ converge to the cone $y_\delta$ as $h \to 0$ [94]. Technically instead of the isoperimetric inequality for $w^k(B_r)$ one uses a Sobolev estimate for the degree of $w_k$ which is essentially equivalent to the isoperimetric inequality but properly accounts for possible multiple coverage of the image.

# 6 Crumpling, Packing and Origami

## 6.1 Crumpling and the $h^{5/3}$ Conjecture

Let $D \subset \mathbb{R}^2$ be the unit disc and let

$$\Omega_h = D \times (-\frac{h}{2}, \frac{h}{2})$$

be a cylinder of height $h$ with $0 < h \ll 1$. We are interested in the minimum energy (per unit height) needed to pack the thin cylinder into a three dimensional ball $B_{1/4}$ of radius $\frac{1}{4}$ and the corresponding minimizers or minimizing sequences. With our notation

$$E^h(u) = \frac{1}{h}\int_{\Omega_h} W(\nabla u)\, dx$$

were are thus interested in the quantity

$$e(h) := \inf\{E_h(u) : u : \Omega_h \to B_{1/4} \subset \mathbb{R}^3\} \tag{111}$$

This problem has been discussed widely in the physics literature, see, e.g., Lobkovsky et al. [75], Kramer and Witten [60] and Witten's survey [116]. It is believed that the minimizers for small $h$ correspond to crumpled structures with a fine network of rounded ridges which become sharper as $h \to 0$. In fact this problem

is seen as a model problem to study the concentration of the energy on complex lower dimensional sets. Based on scaling arguments, an assumed equipartition of stretching and bending energy and numerical simulations Lobkovsky et al. [75] conjectured that the minimal energy satisfies the scaling law

$$e(h) \sim h^{5/3}. \tag{112}$$

The best mathematical results known are the following. As before, we assume that $W$ satisfies the conditions (1)–(4).

**Proposition 3**

$$\liminf_{h\to 0} h^{-2} e(h) = \infty.$$

**Theorem 11 (Conti and Maggi [28, Thm. 1.2])**

$$\limsup_{h\to 0} h^{-5/3} e(h) < \infty.$$

Conti and Maggi also prove a lower bound of order $h^{5/3}$ for a deformation which is close to a single ridge. Under slightly stronger assumption the $h^{5/3}$ scaling law for deformations close a single ridge had earlier been established by Venkataramani [114]. The main mathematical problem in proving a lower bound of order $h^{5/3}$ for the energy is to show that there is not a deformation which looks completely different from an almost origami (or ridge-like) pattern which gives a much lower energy.

*Proof (Proof of Proposition 3)* This is an easy consequence of Theorem 4. Indeed if there existed a sequence $h_k \to 0$ and maps $u_k$ with $u_k(\Omega_h) \subset B_{1/4}$ such that $h_k^{-2} E_{h_k}(u_k)$ remains bounded then Theorem 4 would imply that (a subsequence of) the rescaled maps $y_k(x) = u_k(x', h_k x_3)$ converged strongly in $W^{1,2}(\Omega; \mathbb{R}^m)$ to a map $\bar{y} \in W^{2,2}$ with $\partial_3 \bar{y}$ and with $\bar{y}_{|D}$ an isometric immersion.

This implies that $\bar{y}_{|D}$ is a developable map, i.e., for each point $x \in D$ the map $\bar{y}$ is constant in a neighbourhood of $x$ or $\bar{y}$ is affine on a line segment through $x$ with endpoints in on $\partial D$. This is a classical results for smooth isometric immersions, for isometric immersions $W^{2,2}$ on convex domains see Pakzad [96] who extended earlier work of Kirchheim for $W^{2,\infty}$ maps [54]; for nonconvex domains see also Hornung [49]. It follows that there is a straight line segment through 0 with endpoints on $\partial D$ on which $y$ is affine or that $y$ is affine on a polygon with corners on $\partial D$. In particular $\bar{y}(D)$ contains a line segment of length one. On other hand the constraint $u_k(\Omega_h) \subset B_{1/4}$ implies that $\bar{y}(D) \subset \overline{B_{1/4}}$. Hence $\bar{y}(D)$ cannot contain such a line segment. This contradiction finishes the proof of Proposition 3.

The upper bound by Conti and Maggi requires a much more delicate analysis and makes use of the striking results by by Nash [90] and Kuiper [61] that $C^1$ isometric

immersions are much more flexible than $C^2$ isometric immersions. The proof of the upper bound proceeds in three steps.

- Use the Nash-Kuiper embedding theorem to approximate the contraction $U(x) = \frac{1}{8}x$ in $C^0$ by a $C^1$ isometric embedding.
- Use a careful construction to approximate an $C^1$ isometric embedding by a piecewise affine isometric embedding, a so called origami map.
- Smooth the edges of the piecewise affine isometric embedding in an optimal way

Origami maps and origami structures have recently attracted a lot of attention, e.g., as building structures or metamaterials [38, 102, 110].

## 6.2 Packing of Biomembranes

There are also very interesting packing problems for biomembranes. Here a membrane is modelled as compact surface $\Sigma \subset \mathbb{R}^3$ and one considers the so-called Helfrich-Canham bending energy which in the simplest case is given by

$$E(\Sigma) = \frac{1}{4}\int_\Sigma H^2 \, d\mathcal{H}^2.$$

In this case the energy agrees with the Willmore energy in differential geometry. Here the bending energy depends only on the mean curvature $H$ (the sum of the two principal curvatures) rather the full second fundamental form $A$. This reflects the fact that we are interested in membranes which have no intrinsic shear resistance. One usually assumes also that the membrane is incompressible. Thus the total area $a$ of $\Sigma$ is fixed. Given an open set $\Omega \subset \mathbb{R}^3$ (a 'container') and a number $a > 0$ a natural minimization problem is

$$\text{Minimize } E(\Sigma) \text{ subject to } \Sigma \subset \Omega \text{ and } \operatorname{area}(\Sigma) = a.$$

In [88] the simplest case when $\Omega$ equals the unit ball $B(1)$ is considered. If $a = 4\pi k$ with an integer $k$ then a natural candidate for a minimizer is given by $k$ copies of the unit sphere (more precisely one constructs a minimizing sequence by taking $k$ concentric spheres of radius close to 1 and connecting them by very thin tubes which are almost catenoids and hence have mean curvature almost zero). A calibration argument shows that this construction indeed provides the infimum of the energy. Using the rigidity estimates of [35, 36] one can also analyse the behaviour of the minimal energy for $a = 4\pi + \delta$ for a small $\delta > 0$. The case of general containers $\Omega$ is wide open, even for convex $\Omega$. An interesting question is whether one understand the asymptotic behaviour for large prescribed area $a$.

## 7 Outlook

To close, let me briefly mention a number of related problems, most of which are wide open, with some pointers to the rapidly growing literature.

(i) Blistering in thin films
Consider a thin film deposited on a substrate at high temperature. If the thermal expansion coefficient of the substrate is larger than that of the film, the film is under compressive stress after cooling. The compressive stress may be partially released by a debonding of the film and the formation of wrinkling patterns.

This problem was first studied from a point of global energy minimization by Gioia and Ortiz [44, 95]. A key feature of the problem is that for films with small aspect ratio the system is in the regime well beyond the first unstable eigenmode (this regime is sometimes called the far from threshold regime or FFT regime).

For a prescribed debonded region the optimal scaling laws was identified for von Kármán approximation in [15] and for the full three dimensional problem in [16], see also Jin-Sternberg [51]. An interesting feature is that a finite fraction of the energy is concentrated in a thinner and thinner boundary layer.

If the debonded region is not prescribed, but is included in the minimization problem through a bonding energy per bonded area, then only partial results are known [18, 19]. For compliant substrates see Kohn and Nguyen [59] as well as Bedrossian and Kohn [11].

(ii) Wrinkling under loading by boundary forces
In the blister problem the emergence of wrinkles is driven by an incompatibility at the boundary of the debonded region: the Dirichlet boundary condition prescribes a circumferential length which is shorter than the one preferred by the film. One can also look for wrinkling for softer boundary conditions given by prescribed loads. Davidovitch et al. [33] discuss wrinkling in a thin annulus under axial loading as prototypical example. The corresponding scaling law of the energy has been rigorously established by Bella and Kohn [13].

(iii) Embedding of non Euclidean sheets
Here the formation of singularities or microstructure through wrinkling is not driven by (soft or hard) boundary conditions but by an intrinsic incompatibility of the metric. We already saw a very special case of a prescribed metric with a single conical singularity in Sect. 5.2. One motivation to study more generally the minimization of stretching and bending energy for sheets with a non Euclidean background metric came from experiments by Sharon et al. [108] which showed a buckling cascade with up to six generations of refinements in thin raptured sheets.

The author's explanation is that the rapture introduces plastic deformation near the line of rupture leading a non Euclidean background metric. The self-similar refinement has attracted a lot of attention in the physics literature, see,

e.g., the Fourier approach of Audoly and Boudaoud [7] as well as recent work Gemmer et al. [43] who introduce a new family of singular test function which allow for refinement. Very few mathematically rigorous results are known. Recently interesting progress was achieved by Bella and Kohn [12].

The original work of Sharon et al. [108] already discusses a number of other possible applications including three dimensional pattern formation in biological systems such as leaves. Swelling hydrogels whose metric can be prescribed by photochemistry provide an excellent experimental system to test the effect of specific geometries, see Klein et al. [56, 57].

(iv) Wrinkling in drapes
In the physics literature this has been studied by Cerda et al. [24]. A recent mathematical analysis was carried out by Bella and Kohn [14].

(v) Buckling of thin walled cylinders
Understanding the critical buckling load of thin walled cylinders and the complex folding patterns which appear at the onset of instability is a classical problem on which a huge literature exists. For recent progress and a review of earlier results on rigorous scaling laws of the elastic energy in terms of the compression $\lambda$ and the thickness $h$ see Tobasco [112]. A related problem was studied by Conti et al. [31]. More generally, the critical buckling load of a curved surface (shell) is closely related to the optimal scaling of the constant in Korn's inequality (our reasoning in the proof of Theorem 3 shows that for flat plates the Korn constant scales like $h^2$), see the work by Grabovsky and Haratunyan on axially compressed circular shells [45, 46] and on zero Gauss curvature shells [47] for recent progress.

(vi) Relevance of the $\Gamma$-limit theories for stability
In the context of forces which are consistent with the von Kármán scaling this is discussed in [67].

(vii) Convergence for low-energy equilibria, rather than minimizers
Here the main idea is to replace $\Gamma$-convergence, which is taylored to global minimizers, by the theory of compensated compactness developed by Murat and Tartar [89, 111] . For the reduction from 2d to 1d or from 3d to 1d this can be done in great generality [80, 82]. For the reduction from 3d to 2d convergence of equilibria has so far only been established if the energy per unit volume decays like $h^4$ or stronger, i.e., in the von Kármán or the biharmonic (linear) regime [87].

A main difficulty in going beyond that regime is that there we have no canonical way to approximate 'almost isometric immersions' maps by exact isometric immersions. One cannot use the implicit function theorem because $W^{2,2}$ isometric immersions are characterised by the condition $\det A = 0$ for the second fundamental form and the linearisation of this condition is degenerate.

Indeed the space of isometric immersions (or equivalently the space of developable maps) is very large. This is in sharp contrast with the situation for curves in two or three dimensions which can be described by the angle of the tangent vector with the $x_1$ axis or a Frenet frame, respectively. Even

establishing the Euler-Lagrange equation corresponding to the minimization of $\int_S |A|^2$ or, equivalently, the Willmore energy $\int_S |H|^2$, subject to the isometry constraint $\det A = 0$ is highly nontrivial. Hornung [50] carried out the careful study of the regularity properties of $W^{2,2}$ isometric immersions, derived a suitable forms of the Euler-Lagrange equation and studied the regularity of their solutions.

The convergence result in [87] requires the commonly used assumption that $|DW(F)| \leq C(|F| + 1)$. Unfortunately this assumption is incompatible with the condition that the energy should blow up at infinite compression: $W(F) \to \infty$ as $\det F \downarrow 0$. Mora and Scardia [81] have overcome this difficulty and have shown that convergence of equilibria still holds under the much weaker growth condition $|DW(F)F^T| \leq k(W(F) + 1)$ which is compatible with blow-up at infinite compression. In this setting it is not known whether minimizers satisfy the usual form of the Euler-Lagrange equation. Nonetheless Ball [9, 10] has shown that using outer variations of the form $(\mathrm{Id} + \varepsilon\phi) \circ u$ one can show that minimizers satisfy a physically very natural form of the equilibrium condition (which under a natural invertibility assumption is equivalent to the vanishing of the weak divergence of the Cauchy stress tensor). This equilibrium condition is the starting point of the analysis by Mora and Scardia. Lewicka and Hui [68] have extended the convergence result i to incompressible elastic materials (i.e., $W(F) = \infty$ if $\det F \neq 1$).

Monneau [77] has shown that given a sufficiently smooth (and sufficiently small) solution of the von Kármán equation there exists a nearby solution of the of three dimensional elasticity problem for sufficiently thin domains.

Mielke [76] has used a centre manifold approach to compare solutions in a thin strip to a one dimensional problem. His approach already works for finite thickness $h$, but requires that the nonlinear strain $(\nabla_h y)^T \nabla_h y$ is in $L^\infty$ close to the identity. Applied forces are difficult to handle in this approach.

(viii) Time dependent problems

Here results for the reduction from 3d to 2d are so far only known in von Kármán regime and the biharmonic (linear) regime. In [2] it is shown that (rescaled) solutions of the 3d problem which satisfy the natural energy bounds converge to solutions of the 2d problem (here again linear growth of $DW$ is required). In [1] it is shown that given a sufficiently regular solution of the 2d problem (with periodic boundary conditions) there exists a solution of the 3d problem (on a rescaled time interval) nearby. The main difficulty is to show that a regular 3d solution exists on a sufficiently long time interval and does not develop shocks.

**Acknowledgements** The work reported here would not have been possible without the support and inspiration of many colleagues and friends over a long period of time, in particular S. Conti, G. Friesecke, R.D. James, R.V. Kohn and H. Olbermann. I thank J. M. Ball, P. Marcellini and E. Mascolo for their invitation to prepare these lecture notes and for providing such a delightful and stimulating atmosphere at the C.I.M.E. summer school at Cetraro. I also thank H. Olbermann and

R.V. Kohn for a careful reading of the notes and their very helpful suggestions for improvements. In particular R.V. Kohn suggested the structuring into different classes of problems described at the end of Sect. 1.1. Of course responsibility for all remaining errors, omissions and inadequacies rests solely with me.

My work has been supported by the DFG through the CRC 1060 'The mathematics of emergent effects' and the excellence cluster EXC 59 'Mathematics: foundations, models, applications'.

## References

1. H. Abels, M.G. Mora, S. Müller, Large time existence for thin vibrating plates. Commun. Partial Differ. Equ. **36**(12), 2062–2102 (2011)
2. H. Abels, M.G. Mora, S. Müller, The time-dependent von Kármán plate equation as a limit of 3d nonlinear elasticity. Calc. Var. **41**(1–2), 241–259 (2011)
3. E. Acerbi, G. Buttazzo, D. Percivale, A variational definition of the strain energy for an elastic string. J. Elast. **25**(2), 137–148 (1991)
4. G Alberti, Variational models for phase transitions, an approach via $\Gamma$-convergence, in *Calculus of Variations and Partial Differential Equations (Pisa, 1996)* (Springer, Berlin, 2000), pp. 95–114
5. S.S. Antman, Nonlinear problems of elasticity. *Applied Mathematical Sciences*, 2nd edn., vol. 107. (Springer, New York, 2005)
6. M. Arroyo, L. Heltai, D. Millán, A. DeSimone, Reverse engineering the euglenoid movement. Proc. Natl. Acad. Sci. U. S. A. **109**(44), 17874–17879 (2012)
7. B. Audoly, A. Boudaoud, Self-similar structures near boundaries in strained systems. Phys. Rev. Lett. **91**(8), 086105 (2003)
8. B. Audoly, Y. Pomeau, *Elasticity and Geometry - from Hair Curls to the Non-linear Response of Shells* (Oxford University Press, Oxford, 2010)
9. J.M. Ball, Minimizers and the Euler-Lagrange equations, in *Trends and Applications of Pure Mathematics to Mechanics (Palaiseau, 1983)* (Springer, Berlin, 1984), pp. 1–4
10. J.M. Ball, Some open problems in elasticity, in *Geometry, Mechanics and Dynamics (Marsden Festschrift)* (Springer, New York, 2002), pp. 3–59
11. J. Bedrossian, R.V. Kohn, Blister patterns and energy minimization in compressed thin films on compliant substrates. Commun. Pure Appl. Math. **68**(3), 472–510 (2015)
12. P. Bella, R.V. Kohn, Metric-induced wrinkling of a thin elastic sheet. J. Nonlinear Sci. **24**(6), 1147–1176 (2014)
13. P. Bella, R.V. Kohn, Wrinkles as the result of compressive stresses in an annular thin film. Commun. Pure Appl. Math. **67**(5), 693–747 (2014)
14. P. Bella, R.V. Kohn, The coarsening of folds in hanging drapes (2015). arXiv.org, 1507.08034v1
15. H. Ben Belgacem, S. Conti, A. DeSimone, S. Müller, Rigorous bounds for the Föppl-von Kármán theory of isotropically compressed plates. J. Nonlinear Sci. **10**(6), 661–683 (2000)
16. H. Ben Belgacem, S. Conti, A. DeSimone, S. Müller, Energy scaling of compressed elastic films - three-dimensional elasticity and reduced theories. Arch. Ration. Mech. Anal. **164**(1), 1–37 (2002)
17. K. Bhattacharya, M. Lewicka, M. Schäffner, Plates with incompatible prestrain. Arch. Ration. Mech. Anal. **221**(1), 143–181 (2016)
18. D. Bourne, S. Conti, S. Müller, Folding patterns in partially delaminated thin films (2015). arXiv.org, 1512.06320v1
19. D.P. Bourne, S. Conti, S. Müller, Energy bounds for a compressed elastic film on a substrate. J. Nonlinear Sci. **27**, 453–494 (2017)
20. A. Braides, $\Gamma$-convergence for beginners. *Oxford Lecture Series in Mathematics and Its Applications*, vol. 22. (Oxford University Press, Oxford, 2002)

21. J. Brandman, R.V. Kohn, H.-M. Nguyen, Energy scaling laws for conically constrained thin elastic sheets. J. Elast. **113**(2), 251–264 (2013)
22. E. Cerda, L. Mahadevan, Conical surfaces and crescent singularities in crumpled sheets. Phys. Rev. Lett. **80**(11), 2358–2361 (1998)
23. E. Cerda, L. Mahadevan, Confined developable elastic surfaces: cylinders, cones and the Elastica. Proc. R Soc. A-Math. Phys. Eng. Sci. **461**(2055), 671–700 (2005)
24. E. Cerda, L. Mahadevan, J.M. Pasini, The elements of draping. Proc. Natl. Acad. Sci. U. S. A. **101**(7), 1806–1810 (2004)
25. P.G. Ciarlet, A justification of the von Kármán equations. Arch. Ration. Mech. Anal. **73**(4), 349–389 (1980)
26. P.G. Ciarlet, Mathematical elasticity. Vol. II. *Studies in Mathematics and Its Applications*, vol. 27 (North-Holland, Amsterdam, 1997)
27. S. Conti, Low energy deformations of thin elastic plates: isometric embeddings and branching patterns. Habilitation thesis, University Leipzig, 2003
28. S. Conti, F. Maggi, Confining thin elastic sheets and folding paper. Arch. Ration. Mech. Anal. **187**(1), 1–48 (2008)
29. S. Conti, F. Maggi, S. Müller, Rigorous derivation of Föppl's theory for clamped elastic membranes leads to relaxation. SIAM J. Math. Anal. **38**(2), 657–680 (2006)
30. S. Conti, G. Dolzmann, S. Müller, Korn's second inequality and geometric rigidity with mixed growth conditions. Calc. Var. **50**(1–2), 437–454 (2014)
31. S. Conti, H. Olbermann, I. Tobasco, Symmetry breaking in indented elastic cones. Math. Models Methods Appl. Sci. **27**(2), 291–321 (2017)
32. G. Dal Maso, *An Introduction to $\Gamma$-convergence*. Progress in Nonlinear Differential Equations and their Applications, vol. 8 (Birkhäuser, Boston, 1993)
33. B. Davidovitch, R.D Schroll, D. Vella, M. Adda-Bedia, E.A. Cerda, Prototypical model for tensional wrinkling in thin sheets. Proc. Natl. Acad. Sci. U. S. A. **108**(45), 18227–18232 (2011)
34. E. De Giorgi, T. Franzoni, Su un tipo di convergenza variazionale. Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Nat. Rend. **58**(6), 842–850 (1975)
35. C. De Lellis, S. Müller, Optimal rigidity estimates for nearly umbilical surfaces. J. Differ. Geom. **69**, 75–110 (2005)
36. C. De Lellis, S. Müller, A $C^0$ estimate for nearly umbilical surfaces. Calc. Var. **26**(3), 283–296 (2006)
37. E. Efrati, E. Sharon, R. Kupferman, Elastic theory of unconstrained non-Euclidean plates. J. Mech. Phys. Solids **57**(4), 762–775 (2009)
38. E.T. Filipov, T. Tachi, G.H. Paulino, Origami tubes assembled into stiff, yet reconfigurable structures and metamaterials. Proc. Natl. Acad. Sci. U. S. A. **112**(40), 12321–12326 (2015)
39. A. Föppl, *Vorlesungen über technische Mechanik*, vol. 5. (B.G. Teubner, Leipzig, 1907)
40. G. Friesecke, R.D. James, S. Müller, A theorem on geometric rigidity and the derivation of nonlinear plate theory from three-dimensional elasticity. Commun. Pure Appl. Math. **55**(11), 1461–1506 (2002)
41. G. Friesecke, R.D. James, M.G. Mora, S. Müller, Derivation of nonlinear bending theory for shells from three-dimensional nonlinear elasticity by Gamma-convergence. C.R. Math. Acad. Sci. Paris **336**(8), 697–702 (2003)
42. G. Friesecke, R.D. James, S. Müller, A hierarchy of plate models derived from nonlinear elasticity by gamma-convergence. Arch. Ration. Mech. Anal. **180**(2), 183–236 (2006)
43. J. Gemmer, E. Sharon, T. Shearman, S.C. Venkataramani, Isometric immersions, energy minimization and self-similar buckling in non-Euclidean elastic sheets (2016). arXiv.org, 1601.06863v2
44. G. Gioia, M. Ortiz, Delamination of compressed thin films. Adv. Appl. Mech. **33**, 119–192 (1997)
45. Y. Grabovsky, D. Harutyunyan, Exact scaling exponents in Korn and Korn-type inequalities for cylindrical shells. SIAM J. Math. Anal. **46**(5), 3277–3295 (2014)

46. Y. Grabovsky, D. Harutyunyan, Rigorous derivation of the formula for the buckling load in axially compressed circular cylindrical shells. J. Elast. **120**(2), 249–276 (2015)
47. Y. Grabovsky, D. Harutyunyan, Korn inequalities for shells with zero Gaussian curvature (2016). arXiv.org, 1602.03601v1
48. R.M. Head, E.J. Sechler, Normal pressure tests on unstiffened flat plates. Technical Report, National Advisory Committee for Aeronautics, 1944 Available from the NASA technical reports server, http://ntrs.nasa.gov/search.jsp?R=19930086088.
49. P. Hornung, Approximation of flat $W^{2,2}$ isometric immersions by smooth ones. Arch. Ration. Mech. Anal. **199**(3), 1015–1067 (2011)
50. P. Hornung, Euler-Lagrange equation and regularity for flat minimizers of the Willmore functional. Commun. Pure Appl. Math. **64**(3), 367–441 (2011)
51. W. Jin, P. Sternberg, Energy estimates for the von Kármán model of thin-film blistering. J. Math. Phys. **42**(1), 192–199 (2001)
52. F. John, Rotation and strain. Commun. Pure Appl. Math. **14**, 391–413 (1961)
53. F. John, Bounds for deformations in terms of average strains, in *Inequalities, III (Proc. Third Sympos., Univ. California, Los Angeles, Calif., 1969; dedicated to the memory of Theodore S. Motzkin)* (Academic Press, New York, 1972), pp. 129–144
54. B. Kirchheim, Rigidity and geometry of microstructures. Habilitation thesis, University Leipzig, 2001; See also Lecture Notes MPI Mathematics in the Sciences, vol. 16, Leipzig, 2003 http://www.mis.mpg.de/publications/other-series/ln/lecturenote-1603.html.
55. G. Kirchhoff, Über das Gleichgewicht und die Bewegung einer elastischen Scheibe. J. Reine Angew. Math. [Crelle's J.] **40**, 55–88 (1850)
56. Y. Klein, E. Efrati, E. Sharon, Shaping of elastic sheets by prescription of non-Euclidean metrics. Science **315**(5815), 1116–1120 (2007)
57. Y. Klein, S. Venkataramani, E. Sharon, Experimental study of shape transitions and energy scaling in thin non-Euclidean plates. Phys. Rev. Lett. **106**(11) (2011)
58. R.V. Kohn, New integral estimates for deformations in terms of their nonlinear strains. Arch. Ration. Mech. Anal. **78**(2), 131–172 (1982)
59. R.V. Kohn, H.-M. Nguyen, Analysis of a compressed thin film bonded to a compliant substrate: the energy scaling law. J. Nonlinear Sci. **23**(3), 343–362 (2013)
60. E.M. Kramer, T.A. Witten, Stress condensation in crushed elastic manifolds. Phys. Rev. Lett. **78**(7), 1303–1306 (1997)
61. N.H. Kuiper, On $C^1$-isometric imbeddings. I, II. Nederl. Akad. Wetensch. Proc. Ser. A. **17**, 545–556, 683–689 (1955); (Indag. Math. vol. 58).
62. R. Kupferman, C. Maor, Limits of elastic models of converging Riemannian manifolds. Calc. Var. Partial Differ. Eqn. **55**(2), Article ID 40, 22 p. (2016). doi:10.1007/s00526-016-0979-6
63. R. Kupferman, J.P. Solomon, A Riemannian approach to reduced plate, shell, and rod theories. J. Funct. Anal. **266**(5), 2989–3039 (2014)
64. H. Le Dret, A. Raoult, Le modèle de membrane non linéaire comme limite variationnelle de l'élasticité non linéaire tridimensionnelle. C. R. Seances Acad. Sci. D. Sér. I. Math. **317**(2), 221–226 (1993)
65. H. Le Dret, A. Raoult, The nonlinear membrane model as variational limit of nonlinear three-dimensional elasticity. J. Math. Pures Appl. Neuvième Série **74**(6), 549–578 (1995)
66. H. Le Dret, A. Raoult, The membrane shell model in nonlinear elasticity: a variational asymptotic derivation. J. Nonlinear Sci. **6**(1), 59–84 (1996)
67. M. Lecumberry, S. Müller, Stability of slender bodies under compression and validity of the von Kármán theory. Arch. Ration. Mech. Anal. **193**(2), 255–310 (2009)
68. M. Lewicka, H. Li, Convergence of equilibria for incompressible elastic plates in the von Kármán regime. Commun. Pure Appl. Anal. **14**(1), 143–166 (2015)
69. M. Lewicka, M.R. Pakzad, The infinite hierarchy of elastic shell models: some recent results and a conjecture, in *Infinite Dimensional Dynamical Systems*. Fields Institute Communications, vol. 64 (Springer, New York, 2013), pp. 407–420

70. M. Lewicka, L. Mahadevan, M.R. Pakzad, The Föppl-von Kármán equations for plates with incompatible strains. R. Soc. Lond. Proc. Ser A. Math. Phys. Eng. Sci. **467**(2126), 402–426 (2011)
71. M. Lewicka, M.G. Mora, M.R. Pakzad, The matching property of infinitesimal isometries on elliptic surfaces and elasticity of thin shells. Arch. Ration. Mech. Anal. **200**(3), 1023–1050 (2011)
72. M. Lewicka, P. Ochoa, M.R. Pakzad, Variational models for prestrained plates with Monge-Ampère constraint. Differ. Integral Equ. **28**(9–10), 861–898 (2015)
73. T. Liang, T.A. Witten, Crescent singularities in crumpled sheets. Phys. Rev. E. Statistical, Nonlinear Soft Matter Phys. **71**(1), 016612 (2005)
74. F.C. Liu, A Luzin type property of Sobolev functions. Indiana Univ. Math. J. **26**(4), 645–651 (1977)
75. A. Lobkovsky, S. Gentges, H. Li, D. Morse, T.A. Witten, Scaling properties of stretching ridges in a crumpled elastic sheet. Science **270**(5241), 1482–1485 (1995)
76. A. Mielke, Saint-Venant's problem and semi-inverse solutions in nonlinear elasticity. Arch. Ration. Mech. Anal. **102**(3), 205–229 (1988)
77. R. Monneau, Justification of the nonlinear Kirchhoff-Love theory of plates as the application of a new singular inverse method. Arch. Ration. Mech. Anal. **169**(1), 1–34 (2003)
78. M.G. Mora, S. Müller, Derivation of the nonlinear bending-torsion theory for inextensible rods by *Gamma*-convergence. Calc. Var. **18**(3), 287–305 (2003)
79. M.G. Mora, S. Müller, A nonlinear model for inextensible rods as a low energy *Gamma*-limit of three-dimensional nonlinear elasticity. Ann. Inst. H. Poincaré Anal. Non Linéaire **21**(3), 271–293 (2004)
80. M.G. Mora, S. Müller, Convergence of equilibria of three-dimensional thin elastic beams. Proc. R. Soc. Edinb. Sec. A. Math. **138**(4), 873–896 (2008)
81. M.G. Mora, L. Scardia, Convergence of equilibria of thin elastic plates under physical growth conditions for the energy density. J. Differ. Equ. **252**(1), 35–55 (2012)
82. M.G. Mora, S. Müller, M.G. Schultz, Convergence of equilibria of planar thin elastic beams. Indiana Univ. Math. J. **56**(5), 2413–2438 (2007)
83. M. Moshe, I. Levin, H. Aharoni, R. Kupferman, E. Sharon, Geometry and mechanics of two-dimensional defects in amorphous materials. Proc. Natl. Acad. Sci. U. S. A. **112**(35), 10873–10878 (2015)
84. S. Müller, H. Olbermann, Energy scaling for conical singularities in thin elastic sheets. Oberwolfach Rep. **9**(3), 2233–2236 (2012); Abstracts from the workshop held July 22–28, 2012, Organized by Camillo De Lellis, Gerhard Huisken and Robert Jerrard
85. S. Müller, H. Olbermann, Almost conical deformations of thin sheets with rotational symmetry. SIAM J. Math. Anal. **46**(1), 25–44 (2014)
86. S. Müller, H. Olbermann, Conical singularities in thin elastic sheets. Cal. Var. **49**(3–4), 1177–1186 (2014)
87. S. Müller, M.R. Pakzad, Convergence of equilibria of thin elastic plates–the von Kármán case. Commun. Partial Differ. Equ. **33**(4–6), 1018–1032 (2008)
88. S. Müller, M. Röger, Confined structures of least bending energy. J. Differ. Geom. **97**(1), 109–139 (2014)
89. F. Murat, Compacité par compensation. Ann. Sc. Norm. Super. Pisa Cl. Sci. Ser. IV **5**(3), 489–507 (1978)
90. J. Nash, $C^1$ isometric imbeddings. Ann. of Math. (2) **60**, 383–396 (1954)
91. H. Olbermann, The one-dimensional model for d-cones revisited. Adv. Calc. Var. **9**(3), 201–215 (2016)
92. H. Olbermann, Energy scaling law for a single disclination in a thin elastic sheet. Arch. Ration. Mech. Anal. **224**(3), 985–1019 (2017)
93. H. Olbermann, Energy scaling law for the regular cone. J. Nonlinear Sci. **26**(2), 287–314 (2016)
94. H. Olbermann, The shape of low energy configurations of a thin sheet with a single disclination (2017). arXiv.org, 1702.06468v1

95. M. Ortiz, G. Gioia, The morphology and folding patterns of buckling-driven thin-film blisters. J. Mech. Phys. Solids **42**(3), 531–559 (1994)
96. M.R. Pakzad, On the Sobolev space of isometric immersions. J. Differ. Geom. **66**(1), 47–69 (2004)
97. O. Pantz, Une justification partielle du modèle de plaque en flexion par $\Gamma$-convergence. C. R. Seances Acad. Sci. Ser. I. Math **332**(6), 587–592 (2001)
98. O. Pantz, Le modèle de poutre inextensionnelle comme limite de l'elasticité non-linéaire tridimensionnelle, pp. 1–16. Preprint 2002
99. O. Pantz, On the justification of the nonlinear inextensional plate model. Arch. Ration. Mech. Anal. **167**(3), 179–209 (2003)
100. A.C. Pipkin, Continuously distributed wrinkles in fabrics. Arch. Ration. Mech. Anal. **95**(2), 93–115 (1986)
101. A.C. Pipkin, The relaxed energy density for isotropic elastic membranes. IMA J. Appl. Math. **36**(1), 85–99 (1986)
102. P.M. Reis, F.L. Jimenez, J. Marthelot, Transforming architectures inspired by origami. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 112(40):12234–12235, 2015.
103. E. Reissner, On tension field theory, in *5th International Congress for Applied Mechanics* (1938), pp. 88–92
104. E. Reissner, *Selected Works in Applied Mechanics and Mathematics* (Jones and Bartlett, London, 1996)
105. Y.G. Reshetnyak, Liouville's theorem on conformal mappings for minimal regularity assumptions. Sib. Math. J. **8**, 631–634 (1967)
106. Y.G. Reshetnyak, On the stability of conformal mappings in multidimensional spaces. Sib. Math. J. **8**, 69–85 (1967)
107. Y.G. Reshetnyak, Stability estimates in Liouville's theorem and the $L^p$ integrability of the derivatives of quasi-conformal mappings. Sib. Math. J. **17**, 653–674 (1976)
108. E. Sharon, B. Roman, M. Marder, G.S. Shin, H.L. Swinney, Mechanics: buckling cascades in free sheets - wavy leaves may not depend only on their genes to make their edges crinkle. Nature **419**(6907), 579–579 (2002)
109. E. Sharon, M. Marder, H.L. Swinney, Leaves, flowers and garbage bags: making waves. Am. Sci. **92**(3), 254–261 (2004)
110. J.L. Silverberg, A.A. Evans, L. McLeod, R.C. Hayward, T. Hull, C.D. Santangelo, I. Cohen, Using origami design principles to fold reprogrammable mechanical metamaterials. Science **345**(6197), 647–650 (2014)
111. L. Tartar, Compensated compactness and applications to partial differential equations, in *Nonlinear Analysis and Mechanics: Heriot-Watt Symposium, Vol. IV*. Research Notes in Mathematics, vol. 39 (Pitman, Boston, 1979), pp. 136–212
112. I. Tobasco, Axial compression of a thin elastic cylinder: bounds on the minimum energy scaling law (2016). arXiv.org, 1604.08574v2
113. C. Truesdell, Some challenges offered to analysis by rational thermomechanics. in *Contemporary Developments in Continuum Mechanics and Partial Differential Equations (Proc. Internat. Sympos., Inst. Mat., Univ. Fed. Rio de Janeiro, Rio de Janeiro)* (North-Holland, Amsterdam, 1978), pp. 495–603
114. S.C. Venkataramani, Lower bounds for the energy in a crumpled elastic sheet—a minimal ridge. Nonlinearity **17**(1), 301–312 (2003)
115. H. Wagner, Ebene Blechwandträger mit sehr dünnem Steigblech. Z. Flugtechnik u. Motorluftschiffahrt **20**, 200–207, 227–233, 256–262, 279–284, 306–314 (1929)
116. T.A. Witten, Stress focusing in elastic sheets. Rev. Mod. Phys. **79**(2), 643–675 (2007)
117. A. Yavari, A. Goriely, Riemann-Cartan geometry of nonlinear dislocation mechanics. Arch. Ration. Mech. Anal. **205**(1), 59–118 (2012)
118. W.P. Ziemer, Weakly differentiable functions. *Graduate Texts in Mathematics*, vol. 120 (Springer, New York, 1989)

# Aspects of PDEs Related to Fluid Flows

**Vladimír Šverák**

## 1 Introduction

These notes loosely follow the lectures given by the author at the International Mathematical Summer Center in Cetraro in the summer of 2013. There are two main themes. The first concerns the long-time behavior of solutions of the 2d incompressible Euler equations, and other Hamiltonian equations. For 2d Euler one observes (numerically and experimentally) a tendency for a certain "order" to appear from seemingly chaotic data. Well-known works which gave insights into this phenomenon include papers by Onsager [46] and Kraichnan [34]. Subsequent contributions by many other researchers are mentioned in the corresponding sections of the notes.

Ultimately this theme can be related to a general phenomenon discovered early on by the founders of statistical mechanics: in phase spaces of systems with infinitely many degrees of freedom (such as electro-magnetic fields), there is always "a lot of room" at the degrees of freedom corresponding to small scales (or high spatial Fourier modes). For the purposes of fundamental physics, there is in fact too much room, and this leads to the classical "thermodynamical death" paradoxes which were only resolved by quantum mechanics.

Here we will not be concerned with foundational issues, but will use this phenomenon for making conjectures about the behavior of various infinite-dimensional Hamiltonian systems. It should be emphasized that rigorous results proving that the phenomenon indeed happens, based on actual dynamics given by various PDEs, are rare. Most of the time the rigorous results are fairly far away from what is

V. Šverák (✉)
School of Mathematics, University of Minnesota, 206 Church Street S.E., Minneapolis, MN 5545, USA
e-mail: sverak@math.umn.edu



J. Ball, P. Marcellini (eds.), *Vector-Valued Partial Differential Equations and Applications*, Lecture Notes in Mathematics 2179,
DOI 10.1007/978-3-319-54514-1_4

conjectured, and the problems are difficult, even for the simplest models. As an example, let us take the de-focusing cubic non-linear wave equation on the circle

$$u_{tt} - u_{xx} + u^3 = 0\,, \qquad x \in \mathbf{S}^1\,. \tag{1}$$

One can construct simple solutions by considering functions independent of $x$. These will be of the form $u(x,t) = U(t)$, where $U'' + U^3 = 0$. These are periodic functions (related to elliptic functions). What happens when we slightly perturb these solutions, let the perturbed function evolve exactly according to the equation, and wait for a long time? In Sect. 2.1 we show that based on the thermodynamical principles one should conjecture that, generically, the perturbed solution will converge weakly in $H^1$ to zero. (Of course, the time scales involved might be enormous.) There may be obstacles from the KAM theory, pioneered for PDEs by Kuksin [35]. Many non-generic perturbations might stay on KAM tori, and the generic solutions would have to find their way between a possibly rich family of these tori ("Arnold diffusion").

For other equations, which include the 2d incompressible Euler equation, various non-linear Schrödinger equations, and generalized KdV equations, for example, the behavior is modified by conserved quantities which are continuous with respect to relevant weak topologies, and prevent the escape of the full solution to the high spatial Fourier modes. The combination of the (sufficiently continuous) conserved quantities and the large amount of room at high frequencies gives rise to the "order from chaos" phenomenon mentioned above, at least at the level of statistical considerations (and generic solutions). The actual dynamics can again exhibit KAM tori and various non-generic solutions which behave in a special way, and it is difficult to say what really happens for general solutions.

It is important to realize that the problems we consider may have several natural time-scales. For example, in the famous Fermi-Ulam-Pasta numerical experiments [22] convergence to statistical equilibria ("thermalization") of the type considered here was not observed, and the search for explanations famously lead to the discovery of the complete integrability of the KdV equation. However, it turns out there is a second, much longer, time scale on which thermalization is indeed observed, see for example [4]. In the context of the issues discussed here, one always has in mind the longest possible time scales.

The main benefit of the statistical mechanics methods applied to the problems considered here is that we can relatively easily make conjectures which are hard to prove or disprove.

The statistical mechanics picture suggests a sharp distinction between the behavior of finite-dimensional Hamiltonian systems and sufficiently generic[1] infinite-dimensional Hamiltonian systems at finite energy when the longest time-scales are concerned. Whereas in the former one will have (under some reasonable boundedness assumptions) the Poincaré recurrence, and the complexity of a given

[1] We should exclude the completely integrable systems, for example.

solution has nowhere to escape, in infinite dimensions there is always enough room for the complexity to escape to high modes, and the "visible part" of the solution can potentially simplify.[2]

None of this is new, of course, these conclusions have been known for a long time. Our goal here is to present a relatively simple approach which hopefully will make the issues more accessible to mathematicians who are not familiar with the methods of statistical mechanics. There are many other excellent sources, some of which be quoted in the text.

It should be mentioned that recently there has been progress in understanding the appearance of the small scales from the actual dynamics for *special* solutions, see for example [3, 14, 23, 31, 44]. While these works do not directly address the problems above, where information about generic solutions is needed, they do improve our understanding of the issues involved.

The second theme of the lectures is related to the problem of uniqueness of the Leray-Hopf weak solutions with $L^2$ initial data. Recent developments concerning scale-invariant solutions have led to plausible scenario of non-uniqueness for compactly supported initial date with finite energy. This is discussed in Sect. 4.

## 2 Motivation and Examples

For finite-dimensional Hamiltonian systems we have the Liouville theorem about volume-preservation in the phase space by the evolution, and Poincaré recurrence (under appropriate assumptions). Therefore the system will typically "oscillate", in the sense that, generically, its behavior on a time interval of any fixed length will be almost exactly repeated if we wait long enough. On the other hand, in typical infinite-dimensional systems such behavior is likely not generic, as there is always more "room" in the phase space and there may be no non-degenerate Liouville measure.

Let us start with some examples.

### *2.1 1d Non-linear Wave Equation*

On the one-dimensional circle $\mathbf{S}^1 = \mathbf{R}/2\pi\mathbf{Z}$ let us consider the non-linear wave equation

$$u_{tt} - u_{xx} + \varkappa u^3 = 0\,, \tag{2}$$

[2]This may not be the case for infinite-energy solutions which may have enough energy to fill all the available phase-space, even though it is infinite-dimensional.

where $\varkappa$ is a parameter. For our example we will consider only the de-focusing case $\varkappa > 0$. This is of course a Hamiltonian system, with the Hamiltonian

$$H = \int_{\mathbf{S}^1} \left( \frac{1}{2} v^2 + \frac{1}{2} u_x^2 + \frac{\varkappa}{4} u^4 \right) dx \,, \tag{3}$$

(where $v = u_t$ is considered as an independent variable) and "canonical form"

$$\begin{aligned} \dot{u} &= \tfrac{\delta H}{\delta v} \,, \\ \dot{v} &= -\tfrac{\delta H}{\delta u} \,. \end{aligned} \tag{4}$$

The global well-posedness of the Cauchy problem for (2) with $u(0) = u_0 \in H^1(\mathbf{S}^1)$ and $u_t(0) = v_0 \in L^2(\mathbf{S}^1)$ is not hard to prove and we will take it for granted (as well as preservation of $H^s$ regularity of the initial data by the evolution). What is the long-time behavior of the solution? We have the energy conservation

$$\frac{d}{dt} H(v, u) = 0 \tag{5}$$

and, in addition, we have the conservation of the momentum: letting

$$P(v, u) = \int_{\mathbf{S}^1} v u_x \, dx \,, \tag{6}$$

we have

$$\frac{d}{dt} P(v, u) = \int_{\mathbf{S}^1} \left( \frac{\delta H}{\delta u} \, u_x - v \left( \frac{\delta H}{\delta v} \right)_x \right) = 0 \,. \tag{7}$$

This is of course a consequence of the translational symmetry of our problem and Noether's theorem. The evolution generated by $P$ via

$$\begin{aligned} \dot{v} &= \tfrac{\delta P}{\delta u} \\ \dot{u} &= -\tfrac{\delta P}{\delta v} \end{aligned} \tag{8}$$

is the translation of $(v, u)$.

The solution $(v, u)$ starting from $(v_0, u_0)$ with

$$H(v_0, u_0) = E \,, \qquad P(v_0, u_0) = p \tag{9}$$

will therefore satisfy $H(v, u) = E$ and $P(v, u) = p$ for all time. We do not know about any other constraints a "generic" solution would satisfy. The first guess at the

long-time behavior therefore could be:

Ergodicity guess, nlw
*After a long time the solution* $(v(t), u(t))$ *looks like a "random element" of the manifold given by* $H = E$ , $P = p$. (10)

One problem with this statement is that we have no natural probability measure on the infinite-dimensional manifold

$$\Sigma_{E,p} = \left\{(v,u) \in L^2 \times H^1 \ , \ H(v,u) = E \, , \, P(v,u) = p \right\} \, . \tag{11}$$

Modulo this difficulty, which will be addressed later by taking suitable limits of finite-dimensional subspaces, the guess above amounts to replacing the equation of motion by a postulate of statistical mechanics. How good is the guess? In general, if we make such guesses and do not forget to take into account all known conserved quantities, we get statements which at the level of generic solutions (and sufficiently general equations) are hard to prove or disprove. (The word "generic" in the last statements is important, as the statement would typically not be true for all solutions, due to obstacles from KAM theory, for example.) Numerical verification can be tricky, as it is difficult to simulate Hamiltonian dynamics in high dimensions with high precision over long time-scales. Therefore from a purely mathematical point of view, the above considerations cannot replace the study of the actual dynamics. From a more practical point of view of predicting the future of physical systems the benefits are not as limited, thought, similar to the situation with statistical mechanics. This is because the equations are never completely precise and in the end the conservation laws may be more fundamental than the equations themselves.

There are various non-generic solutions which do not obey the above principle (P). These include periodic solutions, [39], and also small quasi-periodic solutions with many frequencies constructed by KAM techniques, [11, 55]. A simple class of non-trivial periodic solutions are travelling waves[3]

$$u(x,t) = h(x - ct) \, , \tag{12}$$

where $c$ is a real number with $|c| > 1$ and $h$ is a smooth function on $\mathbf{S}^1$ satisfying

$$(c^2 - 1)h'' + \varkappa h^3 = 0 \, . \tag{13}$$

The last equation has many solutions (and they can be expressed in terms of elliptic functions).

If we believe statement (10), all the travelling waves above should be unstable, and under generic small perturbations they should "disintegrate" (perhaps after a long time) to solutions described in (P).

[3] In the context of this example these were drawn to the author's attention by Jalal Shatah.

One can view (4) as a dynamical system in the space $L^2(\mathbf{S}^1) \times H^1(\mathbf{S}^1)$. If $H(v_0, u_0) = E$, then the solution will stay in the set

$$X_E = \{(v, u) \in L^2(\mathbf{S}^1) \times H^1(\mathbf{S}^1)\,,\ H(v, u) \le E\,.\} \tag{14}$$

When we equip $X_E$ with the weak topology $L^2(\mathbf{S}^1) \times H^1(\mathbf{S}^1)$, we can think of it as a compact metric space, as the function $H$ is weakly lower-semicontinuous. We will denote this metric space by $(X_E, w)$, to emphasize the weak topology.

**Lemma 1** *Equation (2) (or, equivalently, Eq. (4)) defines a dynamical system on* $(X_E, w)$. *In particular, the solution map*

$$((v_0, u_0), t) \to (v(t), u(t)) \tag{15}$$

*is continuous as a map from* $(X_E, w) \times \mathbf{R}$ *to* $(X_E, w)$.

*Proof* This follows easily from the known well posedness results for (2), see for example [52]. For the specific equation considered here there are many ways to do the proof, including quite elementary ones, which the reader may do as an exercise. ■

We can now apply the standard dynamical system considerations. In particular, we can define the $\omega$-limit sets for each solution $(v(t), u(t))$ as

$$\Omega = \Omega(v_0, u_0) = \cap_{t>0} \overline{\{(v(s), u(s)) \in X_E, s \ge t\}}\,, \tag{16}$$

where the overbar denotes the closure in $(X_E, w)$.

As we already mentioned, the energy $H$ is weakly lower-semi-continuous on $(X_E, w)$ and hence we have to have $H \le H(v_0, u_0)$ on $\Omega$. The other conserved quantity, $P$, does not have good continuity properties on $(X_E, w)$, and therefore there are no obvious constraints from it on the set $\Omega$.

Let us represent the solution by a Fourier series:

$$u(x, t) = \frac{1}{2\pi} \sum_{k \in \mathbf{Z}} \hat{u}(k, t) e^{ikx}\,. \tag{17}$$

For the real-valued solutions which we consider here we have to have

$$\hat{u}(-k, t) = \overline{\hat{u}(k, t)}\,. \tag{18}$$

where the overbar denotes complex conjugation. The independent degrees of freedom are therefore determined by the coefficients

$$\hat{u}(0, t), \hat{u}(1, t), \hat{u}(2, t), \dots\,, \tag{19}$$

with $\hat{u}(0, t)$ real and $\hat{u}(1, t), \hat{u}(2, t), \dots$ complex. The classical way of thinking about the system is as follows: for each $k = 1, 2 \dots$ the linear part of the wave

equation defines a 2d harmonic oscillator with frequency $k$ (which can also be thought of a two independent 1d oscillators). The frequency $k = 0$ can be thought of as a degenerate oscillator. For the linear equation there is no interaction of these oscillators, they evolve independently of each other. The non-linear term introduces a complex interaction, and once it is turned on, the oscillations can spread from one frequency to another. The interaction is conjectured to be sufficiently complex so that in the generic case, after a sufficiently long time, the energy will be distributed between many frequencies and the energy in any single frequency will be small, and approaching zero as $t \to \infty$. In the space $(X_E, w)$ this will mean that the solution $(v(t), u(t))$ will weakly converge to zero, and the omega limit set $\Omega$ will be

$$\Omega = \{(0,0)\}\,. \tag{20}$$

We emphasize again that this cannot be expected to be true for every solution, but only for "generic solutions". In addition to the travelling wave (12), a class of simple non-trivial solution can be obtained by considering solutions independent of $x$, i.e. $u(x,t) = U(t)$. These solutions are also periodic in time, and it would be interesting if any small generic perturbation would, after a long time, cause a complete "disintegration" of these solutions.

It seems to be beyond the possibilities of existing methods to prove or disprove such statements rigorously. However, it is possible to study rigorously the connection between such conjectures and the statement (10). This will be our main goal.

The same issues can be considered for many other equations. We will consider two other examples.

## 2.2 Non-linear Schrödinger Equation

For functions on $\mathbf{S}^1$ let us consider the de-focusing non-linear Schrödinger equation

$$iu_t + u_{xx} - \varkappa |u|^{2\sigma} u = 0\,, \tag{21}$$

where $\sigma > 0, \varkappa > 0$ are parameters. For simplicity we can think of $\sigma \sim 1$, but not exactly $\sigma = 1$, as for $\sigma = 1$ the equation is completely integrable and has infinitely many conserved quantities, see for example [20]. We can also think of the more general form

$$iu_t + u_{xx} - \varkappa f(u\bar{u})u = 0, \tag{22}$$

where $f(u) = F'(u)$ with $F$ convex and satisfying some growth conditions at for $u \to \infty$. Letting $u = u_1 + iu_2$, we can write (22) in the real form as

$$\begin{aligned} \dot{u}_1 &= \tfrac{\delta H}{2\delta u_2} \\ \dot{u}_2 &= -\tfrac{\delta H}{2\delta u_1}\,, \end{aligned} \tag{23}$$

or in complex form as

$$\dot{u} = -i\frac{\delta H}{\delta \bar{u}}\,, \tag{24}$$

where the Hamiltonian $H$ is given by

$$H = H(u) = \int_{\mathbf{S}^1} \left(u_x \bar{u}_x + F(u\bar{u})\right)\, dx\,. \tag{25}$$

In addition to conservation of $H$, Noether's theorem provides two other conserved quantities: the momentum

$$P = P(u) = \int_{\mathbf{S}^1} -\frac{i}{2}\left(\bar{u}u_x - u\bar{u}_x\right)\,, \tag{26}$$

and the mass

$$M = M(u) = \int_{\mathbf{S}^1} u\bar{u}\, dx\,. \tag{27}$$

It is also easy to check directly that (22) implies

$$\frac{d}{dt}H = 0\,, \qquad \frac{d}{dt}P = 0\,, \qquad \frac{d}{dt}M = 0\,. \tag{28}$$

As an exercise, the reader can check that—as expected from Noether's theorem—the equations

$$\dot{u} = -i\frac{\delta P}{\delta u} \qquad \text{and} \qquad \dot{u} = -i\frac{\delta M}{\delta u} \tag{29}$$

generate symmetries of the equation.

For (21) and $\sigma = 1$ there are many more conserved quantities, see for example [20], but here we focus on the cases the set of known conserved quantities is exhausted by (28).

For our purposes here the natural "phase space" for the evolution is $H^1$. Similarly to the previous example, we can consider the set

$$X_E = \{u \in H^1\,, H(u) \le E\} \tag{30}$$

equipped with the weak topology. This is a compact metric space, which we will denote by $(X_E, w)$. The well-posedness results for (22) imply (under some mild continuity and growth assumptions on $f$) that the equation defines a good dynamical system on $(X_E, w)$, similarly to Lemma (1). We can therefore again consider the $\omega$-limit sets $\Omega$.

The Hamiltonian $H$ will again be lower-semicontinuous on $(X_E, w)$. The main difference with the previous example now is the following:

**Lemma 2** *The functions $P$ and $M$ defined by (26) and (27) respectively are continuous on $(X_E, w)$.*

This is of a standard statement, and we leave the proof as an exercise for the reader.

An obvious corollary of the lemma is the following. Assume that $\Omega = \Omega(u_0)$ is the $\omega$-limit set for the initial condition $u_0$ and let $m = M(u_0), p = P(u_0)$. Then $M = m$ and $P = p$ on $\Omega$.

If we think of the long-time behavior of the solution in terms of the Fourier modes, as in the last section, the non-linear interaction argument would still suggest that each Fourier mode should approach zero for generic solutions, but from the conservation and weak continuity of $M$ and $P$ we see that this is not possible. To accommodate both the "spreading of the solution across Fourier modes" and the conservation of $M$ and $P$, we can guess the following standard conjecture:

**Conjecture 1 (Variational Characterization of Long-Time Behavior)** *Over long-time, the solutions will approach (in the sense of $(X_E, w)$)) the set of minimizers of $H$ under the constraints $M = m$ and $P = p$.*

This heuristics has been well-known in many contexts, see for example [18, 41] and references therein.

Another point of view would be an analogy of (10). Denoting by $u_0$ the initial conditions and letting $E = H(u_0), p = P(u_0), m = M(u_0)$ it seems reasonable to guess (unless there are additional conserved quantities which we do not know about) the following:

Ergodicity guess, nls
*After a long time, the solution $u(t)$ of (22) looks (generically) like a "random function" from the "manifold" $H = E, P = p, M = m$.* (31)

Proving or disproving this statement (after making it more precise by considering suitable limits of finite-dimensional subspaces) seems beyond reach of current methods. However, we will see that it is possible to link it to conjecture (1) above.

The non-linear Schrödinger equation can also be considered in higher dimension. The initial value problem has been studied in depth by many authors, see for example [6, 13, 30, 32]. In the context of our focus here we should mention the topic of wave turbulence, see for example [57], and the cubic non-linear Schrödinger equation on the 2d torus. The considerations above concerning the long-time behavior apply with the obvious adjustments to higher-dimensional situation, too.

### 2.3 The Generalized Korteweg-de Vries Equation (gKdV) on $\mathbf{S}^1$

The equation is

$$u_t + f(u)_x - u_{xxx} = 0\,, \tag{32}$$

where $f = F'$ with $F$ smooth and some mild assumptions on the growth at $u \to \infty$. The classical cases of KdV and modified KdV (mKdV), when the equation is completely integrable and has infinitely many conserved quantities, correspond respectively to

$$F(u) = \frac{u^3}{6} \qquad \text{or} \qquad F(u) = \frac{u^4}{12}\,. \tag{33}$$

The equation again is well-known to have a Hamiltonian structure. Letting

$$H(u) = \int_{\mathbf{S}^1} \left(\frac{1}{2}u_x^2 + F(u)\right) dx\,, \tag{34}$$

we can write

$$\dot{u} = -\frac{\partial}{\partial x}\frac{\delta H}{\delta u}\,. \tag{35}$$

In this case the symplectic form behind the equation is

$$\Omega(u, v) = \int_{\mathbf{S}^1} -(\partial_x^{-1}u)v \; dx\,, \tag{36}$$

which is well defined on $H_0^1(\mathbf{S}^1) = \{u \in H^1(\mathbf{S}^1)\,,\ \int_{\mathbf{S}^1} u\,dx = 0\}$ (and, in fact also on the larger space $H^{\frac{1}{2}}$). Noether's theorem gives the conserved quantity

$$I(u) = \int_{\mathbf{S}^1} \frac{1}{2}u^2\,. \tag{37}$$

We emphasize again that in the famous completely integrable cases given by (33) we have many more conserved quantities. Here our focus is on the situation when $H$ and $I$ are the only known conserved quantities. Assuming we have not missed any conserved quantities, we can again take a guess at the long time behavior. First, based on the idea of energy spreading "as much as possible over the Fourier modes" in a way consistent with the conservation of $I$, it is natural to guess that after long time the generic solutions should be related to minimizers of $H$ under the constraint of a given value of $I$. Another point of view would be that after a long time a generic solution for $H(u_0) = E$ and $I(u_0) = p$ will look as a "random function" from the manifold $\{H = E\,,\ I = p\}$, which again can be made more precise by considering suitable limits of finite dimensional subspaces.

## 2.4 Critical Points of Hamiltonians on Invariant Submanifolds

We see from the examples above that the problem of minimizing the Hamiltonian under constraints given by conservation laws appears naturally in the context of the long-time behavior. This leads to "solitons", as we now recall. For simplicity let us consider a finite dimensional Hamiltonian system which we will write as

$$\dot{x} = JH'(x)\,, \tag{38}$$

where $J$ is the matrix describing the underlying symplectic structure. We can think of $J$ as constant in $x$, although the considerations apply equally well to the case when $J$ depends on $x$. Let $f_1, \ldots, f_m$ be conserved quantities for (38). Let us consider a critical point $\bar{x}$ of $H$ on the submanifold

$$\Sigma_{c_1,\ldots,c_m} = \{f_1 = c_1\,, f_2 = c_2,\, , \ldots, f_m(x) = c_m\}\,. \tag{39}$$

Let us assume that $\Sigma_{c_1,\ldots,c_m}$ is smooth near $\bar{x}$. We have

$$H'(\bar{x}) = \lambda_1 f_1'(\bar{x}) + \cdots + \lambda_m f_m'(\bar{x}) \tag{40}$$

for some real numbers $\lambda_1, \ldots, \lambda_m$. If $H'(\bar{x}) = 0$, then $\bar{x}$ is a rest point of the system. We will consider the more interesting case when $H'(\bar{x})$ does not vanish. Let

$$f = \lambda_1 f_1 + \lambda_2 f_2 + \ldots \lambda_m f_m\,, \quad c = \lambda_1 c_1 + \cdots + \lambda_m c_m \quad E = H(\bar{x})\,. \tag{41}$$

Let $\phi^t$ be the flow induced by (38). Both $H$ and $f$ are preserved by the flow, in the sense that $H(\phi^t(x)) = H(x)$ and $f(\phi^t(x) = f(x)$. Hence the condition $H'(\bar{x}) = f'(\bar{x})$ will be preserved along the trajectory passing through $\bar{x}$. This means that the trajectory is given also by

$$\dot{x} = Jf'(x)\,, \qquad x(0) = \bar{x}\,. \tag{42}$$

By Noether's theorem, the flow $\psi^t : x_0 \to x(t)$ given by solving

$$\dot{x} = Jf' \tag{43}$$

with the initial condition $x(0) = x_0$ is a symmetry of the system. Hence the motion of $\bar{x}$ is given by the 1-parameter symmetry group $\psi^t$.

As an example we can consider the classical Kepler problem of a motion of a planet. The conserved quantities will be the energy and the angular momentum. If we minimize the energy subject to the constraint of a given momentum, we get circular orbits. The solutions of (22) and (32) obtained from the constrained minimizations are of similar nature: for (22) we get a modulated travelling wave,

whereas for (32) we get a travelling wave. The travelling wave (12) for the nonlinear wave equation also belongs to this category, although it cannot be obtained by direct constrained minimization, as the function $P$ is not weakly continuous.

## 2.5 2d Incompressible Euler

On the 2d torus $\mathbf{T}^2 = \mathbf{T}^2_{a,b} = \mathbf{R}^2/a\mathbf{Z} \oplus b\mathbf{Z}$, where $a, b > 0$ are parameters, we consider the 2d incompressible Euler equation in the vorticity form:

$$\omega_t + u\nabla\omega = 0\,, \tag{44}$$

where $u$ is determined from $\omega$ via the equations

$$\Delta\psi = \omega, \qquad u = \nabla^\perp\psi. \tag{45}$$

Here $\nabla^\perp\psi$ denotes the field $(-\psi_{x_2}, \psi_{x_1})$ and we assume

$$\int_{\mathbf{T}^2} \omega(x,t)\,dx = 0\,, \tag{46}$$

which is clearly preserved by the evolution.

Strictly speaking, Eq. (44) is a Poisson system (rather than Hamiltonian). It arises from a larger Hamiltonian system by a symmetry reduction. A classical finite-dimensional example of this is the following: in $\mathbf{R}^2$ consider a version of the classical Kepler problem with Hamiltonian

$$H(y,x) = \frac{1}{2}|y|^2 + V(r)\,, \qquad r = |x|\,. \tag{47}$$

This problem has an obvious $SO(2)$ symmetry: if $(y(t)\,,\, x(t))$ is a solution and $R$ is a rotation, then $(Ry(t), Rx(t))$ is again a solution. We can write the equations of motion in terms of the Poisson bracket

$$\{f,g\} = \sum f_{y_k} g_{x_k} - f_{x_k} g_{y_k}\,, \tag{48}$$

as

$$\dot{f} = \{H,f\}\,. \tag{49}$$

The Poisson bracket preserves the class of functions $f$ on the phase space $\mathbf{R}^2 \times \mathbf{R}^2$ which are invariant under the action $(y,x) \to (Rx, Ry)$ of $SO(2)$. The set of orbits of this action can be though of as a three-dimensional manifold $X$ (with singular points corresponding to the projections of the point of the form $(0,x)$ and $(y,0)$)

and the invariant functions on $\mathbf{R}^2 \times \mathbf{R}^2$ can be thought of as functions on $X$. The equation of motion (49) descends to $X$, and hence we now have an equation on $X$. Since $\dim X = 3$, the system on $X$ given by (49) is not symplectic. The function

$$M(x, y) = x_1 y_2 - x_2 y_1 \tag{50}$$

(angular momentum) on $\mathbf{R}^2 \times \mathbf{R}^2$ which generates the $SO(2)$ symmetry via

$$\dot{f} = \{M, f\} \tag{51}$$

is itself invariant (because $\{M, M\} = 0$, due to the anti-symmetry of the bracket), it also descends on $X$. As the invariant functions can be defined exactly by $\{M, f\} = 0$, we see that

$$\{M, f\} = 0 \qquad \text{for any (sufficiently regular) function of } X. \tag{52}$$

The manifold $M$ is foliated into two-dimensional manifolds $\{M = c\}$ which will be invariant under the evolution on $X$ given by

$$\dot{f} = \{\tilde{H}, f\} \tag{53}$$

for any Hamiltonian $\tilde{H}$. The systems (53) will be hamiltonian on each leaf $\{M = c\}$. The functions $C$ on $X$ with the property that $\{C, f\} = 0$ for each $f$ are called Casimir functions. They coincide with functions which are locally constant on each symplectic leaf (and in our case they can be factored through $M$ near the points where the differential of $M$ does not vanish).

This structure has many consequences. For example, the equilibria of the system on $M$ correspond to the critical points of $H$ restricted to the leaves, and hence can be expected to come in one-dimensional families, is indeed the case generically.

Euler equation also can also be thought about in this way. We consider incompressible fluids in $\mathbf{T}^2$. From the point of view of classical mechanics, our goal is to determine the motion of the "fluid particles". At the level of the continuum description, our natural configuration space is therefore the group of volume preserving diffeomorphisms of $\mathbf{T}^2$, or, more precisely, its connected component containing the identity, which we will denote by $G$. The Hamiltonian is given by the kinetic energy. If $\phi = \phi^t$ is a trajectory in $G$, the Hamiltonian is

$$H(\phi, \dot{\phi}) = \int_{\mathbf{T}^2} \frac{1}{2} |\dot{\phi}(x)|^2 \, dx \,. \tag{54}$$

The tangent space to $G$ at the identity element is the space of divergence-free fields $u$ on $\mathbf{T}^2$, and on those fields the Hamiltonian is of course

$$\int_{\mathbf{T}^2} \frac{1}{2} |u(x)|^2 \, dx \,. \tag{55}$$

The phase space can be identified with the pairs $(\phi, \dot{\phi})$ where $\phi \in G$ and $\dot{\phi} \in T_\phi G$ (the tangent space to $G$ at $\phi$, which is identified with the co-tangent space via the $L^2$ scalar product) and $G$ acts on the phase space by the symmetries

$$(\phi, \dot{\phi}) \to (\phi \circ \psi, \dot{\phi} \circ \psi)\,, \tag{56}$$

which can be interpreted as re-labeling of the particles. The "reduced phase space", analogous to the space $X$ in the above example, can be identified with the space of (smooth) div-free fields (by taking $\psi = \phi^{-1}$ in (56), which essentially represents passing from the Lagrangian description to the Eulerian description). The reader can consult for example [42] for details.

There are of course many ways to introduce coordinates for all the relevant objects. The description in terms of the vorticity $\omega$ and the stream function $\psi$ in (45) has the advantage that there are no constraints on $\omega$ other then the trivial one given by (46), and therefore it identifies the genuine degrees of freedom (which is not the case with the velocity field $u$, for example, or the vorticity field in the case of three dimensions). We can view the space of the scalar vorticities $\omega$ satisfying the condition (46) as another incarnation of our reduced space, analogous to $X$ above. Strictly speaking the correspondence between $\omega$ and $u$ works only if we assume that $\int_{\mathbf{T}^2} u(x)\,dx = 0$, which is a condition preserved by the Euler equation written in terms of $u$, and is also a condition required for the introduction of the stream function $\psi$, but it is not automatically satisfied in general. The difficulty is caused by the fields constant in $x$. It is not hard to extend the analysis so that this possibility could be included. Here we will restrict ourself to the case $\int_{\mathbf{T}^2} u(x)\,dx = 0$. It captures the most interesting features of the general case.

From the point of view of applying Statistical Mechanics considerations it is important to analyze the conserved quantities. The Hamiltonian is of course conserved, and in terms of $\omega$ and $\psi$ we have

$$H = \int_{\mathbf{T}^2} -\frac{1}{2}\,\omega\psi\,dx\,. \tag{57}$$

All the other known conserved quantities are Casimir functions whose conservation is derived from the fact that the Euler evolution leaves invariant the sets

$$\mathcal{O}_{\omega_0} = \{\omega = \omega_0 \circ \phi\,,\ \phi \in G\}\,. \tag{58}$$

The invariance of $\mathcal{O}_{\omega_0}$ is obvious, we can see it directly without the considerations above. This also gives the conservation of the quantities

$$I_f(\omega) = \int_{\mathbf{T}^2} f(\omega(x))\,dx\,, \tag{59}$$

where $f$ is any continuous function. These are Casimir functions, generated from the full phase space $\{\phi, \dot{\phi}\}$ by the symmetries (56) by which we factored to get the

reduced space. Therefore $I_f$ do not generate non-trivial symmetries of the reduced space (similarly to the function $M$ not generating any non-trivial symmetries of $X$ in the simple Kepler-type example above). The orbits $\mathcal{O}_{\omega_0}$ can also be thought of (formally) as the symplectic leaves of the Poisson structure induced on the space of the vorticities from the symplectic reduction. This should only be taken as a useful heuristics. Finding a rigorous framework which would preserve the heuristics is non-trivial. For some results in this direction see for example [2, 12, 26]. For the expressions for the Poisson brackets see [45].

A natural space for the vorticities in which the equation is globally well-posed is $L^\infty$. A well-known result of Yudovich, see [56], says that the initial value problem for (44) is uniquely solvable for all time (in a natural class of functions) when the initial condition $\omega_0$ is in $L^\infty$. With non-smooth vorticities the corresponding flows will not generate smooth diffeomorphism, but volume-preserving homeomorphism. We will still denote the set of such mappings by $G$, slightly abusing the notation.

For any $C \geq 0$ the set of all vorticities $\omega$ satisfying (46) and $||\omega||_{L^\infty} \leq C$ is compact in the weak* topology of $L^\infty$ (induced by considering $L^\infty$ as the dual space of $L^1$). In the rest of this section we will use the notation

$$X_C = \{\omega \in L^\infty(\mathbf{T}^2)\,,\ ||\omega||_{L^\infty} \leq C\ ,\ \int_{\mathbf{T}^2} \omega(x)\,dx = 0\}\,. \tag{60}$$

The space $(X_C, w^*)$ can be considered as a compact metric space. The dynamics is still well-behaved with respect to the weak* topology:

**Lemma 3** *The Euler equation (44) defined a dynamical system on* $(X_c, w^*)$. *In particular, the solution map*

$$\omega_0 \to \omega(t) \tag{61}$$

*is continuous as a map from* $(X_C, w^*) \times \mathbf{R} \to (X_C, w^*)$.

The proof is not difficult, essentially one can just check that the arguments usually used in the proof of the Yudovic theorem give after minor adjustments the proof of the continuity. For a proof of the Yudovic theorem which works very well in this respect see [40], for example.

The energy $H$ given by (57) is easily seen to be a continuous function on $(X_C, w^*)$. On the other hand, the functionals $I_f$ given by (59) are not weakly* continuous on $X_C$, except for the trivial case when $f$ is affine. However, when $f$ is convex, $I_f$ is lower semi-continuous on $(X_C, w^*)$ by standard arguments.

We will denote by $\overline{\mathcal{O}}_{\omega_0}$ the closure of $\mathcal{O}_{\omega_0}$ is $(X_C, w^*)$.

**Lemma 4** *For any* $\omega_0 \in X_C$ *the set* $\overline{\mathcal{O}}_{\omega_0}$ *is convex.*

Heuristically this is not difficult to understand, as we can use "micro-structures" to approximate convex combinations, see for example [51], where the reader can find a full proof.

*Example* Assume that $\omega_0 = \mathbf{1}_A - \mathbf{1}_{\mathbf{T}^2 \setminus A}$ where $A$ is a measurable set with $|A| = \frac{1}{2}|\mathbf{T}^2|$. Then

$$\overline{\mathcal{O}}_{\omega_0} = \{\omega \in L^\infty(\mathbf{T}^2) \ ||\omega||_{L^\infty} \leq 1 \, , \int_{\mathbf{T}^2} \omega \, dx = 0\} \, . \tag{62}$$

What is a good ergodic guess for the long-time behavior? This question goes back to the classical paper by Onsager [46]. In that paper Onsager considered the system of finitely many point vortices, i.e. the situation when $\omega_0$ is a linear combination of finitely many Dirac masses. This leads to a finite-dimensional system with Liouville measure, but its phase-space does not approximate the phase-space $(X_C, w*)$ very well. Nevertheless, it already captures some key phenomena. Statistical mechanics of point vortices has been further pursued for example in [19].

A classical approach, due to Kraichnan, see [34], is to take the Fourier truncation of (44) together with two natural conserved quantities. The Fourier truncation preserves the natural Lebesgue measure on the Fourier coefficients, the energy $H$ and also the enstrophy

$$I_2(\omega) = \int_{\mathbf{T}^2} \omega^2 \, dx \, . \tag{63}$$

Assume

$$I_2(\omega_0) = c_2 \, . \tag{64}$$

Kraichnan's original calculation is with the Gibbs measures, but one can also take the following version:

> Approximate ergodicity guess, 2d Euler (after Kraichnan)
> *After a long time the solution $\omega(t)$ looks like a "random element" of the manifold given by $H = E$, $I_2 = c_2$.* (65)

The exact meaning and the mathematical consequences of this hypothesis can be worked our completely, see Example 1 after Theorem 1 in Sect. 3.6. Although it is easy to come up with examples where the actual dynamics cannot follow this model (e. g. by considering $\omega_0$ for which the prediction $\omega$ based on (65) does not satisfy $||\omega||_{L^\infty} \leq ||\omega_0||_{L^\infty}$), the model already captures the famous downward cascade.

There is a corresponding variational principle, which follows from (65) (see the above mentioned Example in Sect. 3.6), and which can also be conjectured directly based on other heuristic considerations, such as the "spreading of enstrophy in Fourier modes", or more sophisticated arguments, see or example [41].

A guess which takes into account all known conserved quantities is the following. We refer the reader to [5, 41, 43, 49, 53] for more discussion.

Ergodicity guess, 2d Euler (after Miller and Robert-Sommeria)
*After a long time the solution $\omega(t)$ looks like a "random element" of the manifold given by $\{H = E\} \cap \mathcal{O}_{\omega_0}$.* (66)

It less obvious how to make this mathematically precise. One approach, essentially taken in [43] is as follows. Divide $\mathbf{T}^2$ into $N \times N$ rectangles. On each rectangle replace $\omega_0$ by its average over the rectangle (corresponding to taking the $L^2$-projection to functions which are constant on the rectangles). Denote this new function by $\omega_0^{(N)}$. Act on the function $\omega_0^{(N)}$ by permuting the rectangles. The group involved in this action is the symmetric group $S_{N^2}$ of all possible permutations of the $N^2$ rectangles. Let $\mathcal{O}_{\omega_0}^{(N)}$ be the orbit of this group.[4] It is not clear to what degree the actual dynamics of Euler equation can achieve this type of mixing, especially when $N$ is very large (see for example the discussion in [53]), but the results one obtains from this model are not unreasonable.

The natural measure to take on $\mathcal{O}_{\omega_0}^{(N)}$ is the counting measure, which we will denote by $\gamma_N$ (instead of the more precise $\gamma_{\omega_0,N}$). For any set $A$ the measure $\gamma_N(A)$ is just the number of elements in $\mathcal{O}_{\omega_0}^{(N)} \cap A$. If we decide to work with the canonical ensemble, one should work with the measures

$$\nu_{E,\omega_0;N} = Z^{-1} e^{-\beta H(\omega)} \gamma_N \,, \qquad Z = \int e^{-\beta H(\omega)} d\gamma_N(\omega) \,. \tag{67}$$

where $\beta$ is chosen so that

$$\int H(\omega)\, d\nu_{E,\omega_0;N}(\omega) = E \,. \tag{68}$$

Another approach, relying on the micro-canonical ensemble, and perhaps reflecting better the Euler dynamics (although ultimately both approaches may lead to similar conclusions), is the following. Let $\mathcal{O}_{\omega_0}^{(N)}$ and $\gamma$ be as above, take $\varepsilon > 0$, and consider the measure

$$\mu^{\varepsilon}_{E,\omega_0,N} = Z^{-1} \mathbf{1}_{\{E-\varepsilon<H(\omega)<E+\varepsilon\}}\ \gamma_N \,, \quad Z = \int \mathbf{1}_{\{E-\varepsilon<H(\omega)<E+\varepsilon\}}\, d\gamma_N(\omega) \,. \tag{69}$$

We are interested in the limits

$$\lim_{\varepsilon\to 0} \lim_{N\to\infty} \mu_{E,\omega_0,N} \tag{70}$$

or perhaps

$$\lim_{N\to\infty} \mu^{\varepsilon_N}_{E,\omega_0,N} \,, \tag{71}$$

[4] We avoid the more logical but unwieldy notation $\mathcal{O}^{(N)}_{\omega_0^{(N)}}$.

where $\varepsilon_N \to 0$ as $N \to \infty$ is a suitable chosen sequence. It is not clear to the author if one can find in the literature a completely rigorous version of such calculations, although one can find references (such as those above) where the calculations are done, perhaps from a slightly different angle, with the physicist's level of rigor, or with the use of some heuristics.

Another approach (which does not quite address the problem of the exact calculations above, but it is related to it) is to use some elementary combinatorics and/or heuristics for deriving a notion of entropy adopted to a given orbit $\mathcal{O}_{\omega_0}$. Such a function $S = S_{\omega_0}$ should have the following properties. We use the notation

$$L_0^\infty = L^\infty(\mathbf{T}^2) \cap \{\omega\,,\ \int_{\mathbf{T}^2} \omega\, dx = 0\}\,. \tag{72}$$

(i) $S: L_0^\infty \to \mathbf{R} \cup -\infty$ is concave and upper semi-continuous on $(X_C, w^*)$ for each $C > 0$.
(ii) $\{\omega \in L_0^\infty\,,\ S(\omega) > -\infty\}$ is dense subset of $\overline{\mathcal{O}}_{\omega_0}$.

For instance, in the example following Lemma 4 a natural entropy function (which can be derived from a certain quite natural "state counting", see for instance [51]) is

$$S(\omega) = \int_{\mathbf{T}^2} \left( -\frac{1+\omega}{2} \log \frac{1+\omega}{2} - \frac{-1+\omega}{2} \log \frac{-1+\omega}{2} \right) dx\,. \tag{73}$$

This function comes up often in statistical mechanics, in connections fermions or spin systems, see for example [17]. Each rectangle in our $N \times N$ mesh used above can be occupied exactly once, either by 1 or $-1$. There is an extra constraint that the total number of 1 is the same as the total number of $-1$. In some sense, the incompressibility condition introduces a variant of the exclusion principle. This is exactly the point which is not quite captured by simple point-vortex models.

The selection principle for the equilibria on which (variants of) our measures should concentrate now is the maximization of the entropy subject to the constraint of a given energy. This leads to equations of the type

$$\Delta\psi = f(\psi)\,. \tag{74}$$

These themes are discussed (in a somewhat different technical setup) for example in [41, 43, 49, 51, 53].

In [50] Shnirelman introduced another interesting idea, which we explain here in a slightly different language. Let us denote by $\overline{\mathcal{O}}_{\omega_0,E}$ the set $\overline{\mathcal{O}}_{\omega_0} \cap \{H = E\}$. One can introduce a partial order on the set $\overline{\mathcal{O}}_{\omega_0,E}$ be defining

$$\omega_1 \prec \omega_2 \quad \Longleftrightarrow \quad \overline{\mathcal{O}}_{\omega_2,E} \subset \overline{\mathcal{O}}_{\omega_1,E}\,. \tag{75}$$

Heuristically one can argue that $\omega_2$ is "more mixed" than $\omega_1$. In [50] one can find very good explanations for this, from a slightly different angle. Shnirelman's form of the "ergodic guess" then is that over long time, a generic solution converges weakly* towards a maximal element. (Maximal elements are easily seen to exists by standard set-theoretical considerations and compactness.) One can show, see [50], that any maximal element satisfies an equation of the form (74), although this time, unlike in the previous case, the new form of the "ergodic guess" does not say much about the function $f$ beyond the fact that it is monotone. The advantage of this approach is that in some sense it is not trying to "guess too much". Instead, we just insist that the solutions will "mix" as much as they can, without trying to specify the entropy functions which quantifies the amount of mixing. As the exact form of this function depends on the details of the model and sometimes may be in doubt (see [53], for example) this approach seems to have its advantages.

## 3 Choosing Random Functions with Constraints

### 3.1 Measures $\delta(f(x) - b)\,dx$

We first recall some standard notation. For more general definition involving composition of distributions with smooth maps see for example [25].

Let $n > m$ consider a sufficiently regular $f\colon \mathbf{R}^n \to \mathbf{R}^m$. Let $\delta = \delta_{\mathbf{R}^m}$ be the Dirac mass in $\mathbf{R}^m$ and let $\delta^{(\varepsilon)}$ be its approximation by smooth functions, e. g.

$$\delta^{(\varepsilon)} = \frac{1}{\varepsilon^m}\phi\left(\frac{x}{\varepsilon}\right)dx\,, \tag{76}$$

where $\phi$ is a suitable mollifying function. Let $b \in \mathbf{R}^m$ and let us consider the measure

$$\delta^{(\varepsilon)}(f(x) - b)\,dx\,. \tag{77}$$

This is clearly a well-defined measure in $\mathbf{R}^n$. If assume that the limit

$$\lim_{\varepsilon \searrow 0} \int_{\mathbf{R}^n} \varphi(x)\delta^{(\varepsilon)}(f(x) - b)\,dx \tag{78}$$

exists and is finite for each smooth, compactly supported continuous function $\varphi$, we can define the measure $\delta(f(x) - b)\,dx$ as

$$\int \varphi(x)\delta(f(x) - b)\,dx = \lim_{\varepsilon \searrow 0} \int_{\mathbf{R}^n} \varphi(x)\delta^{(\varepsilon)}(f(x) - b)\,dx\,. \tag{79}$$

Some assumptions on $f$ are needed for this limit to exist. The nature of these assumptions is probably best seen from the co-area formula, which we now recall.

Let us treat $b$ as a variable. We note that, trivially,

$$\int_{\mathbf{R}^m} \varphi(x)\delta(f(x)-b)\,db = \varphi(x). \tag{80}$$

An additional integration over $x$ gives

$$\int_{\mathbf{R}^n}\int_{\mathbf{R}^m} \varphi(x)\,\delta(f(x)-b)\,db\,dx = \int_{\mathbf{R}^n} \varphi(x)\,dx \tag{81}$$

Changing the order of integration in the double integral on the left, we obtain

$$\int_{\mathbf{R}^m}\left[\int_{\mathbf{R}^n} \varphi(x)\delta(f(x)-b)\,dx\right] db = \int_{\mathbf{R}^n} \varphi(x)\,dx\,. \tag{82}$$

Let us compare this formula with the classical co-area formula (see Federer [21], Theorem 3.2.12, p. 249). Let $J_f(x)$ be defined as the square root of the sum of squares of all $m\times m$ sub-determinants of $\nabla f(x)$. In particular, if $m=1$, then $J_f(x)=|\nabla f(x)|$. The co-area formula says that for any Lipschitz $f$ we have

$$\int_{\mathbf{R}^n} \varphi(x)J_f(x)\,dx = \int_{\mathbf{R}^m}\int_{f^{-1}(b)} \varphi(x)d\mathcal{H}^{n-m}(x)\,db\,, \tag{83}$$

where $\mathcal{H}^k$ denotes the $k$-dimensional Hausdorff measure. Comparing (82) and (83), we see that, at least formally,

$$\delta(f(x)-b)\,dx = \frac{1}{J_f(x)}\;\mathcal{H}^{n-m}|_{f^{-1}(b)}\,. \tag{84}$$

In particular, for $m=1$ we have

$$\delta(f(x)-b) = \frac{1}{|\nabla f(x)|}\;\mathcal{H}^{n-1}|_{\{f(x)=b\}}\,. \tag{85}$$

Although this formula is useful, it has the disadvantage of bringing into the consideration structures which are not part of the original setup. Note that the definition of $\delta(f(x)-b)$ involves only a set with a measure and a function on it, whereas the expression on the right-had side of (85) involves also expressions which need a metric.

We see that $\delta(f(x)-b)$ is definitely well-defined when $f$ is a $C^1$-function and $J_f(x)$ does not vanish on $\{f(x)=b\}$. In the case when $J_f(x)$ vanishes at some points of the set $\{f(x)=b\}$ the question of existence of $\delta(f(x)-b)$ near those

points has to be investigated in more detail.[5] In general, questions concerning the existence of $\delta(f(x) - b)$ in the presence of degenerate points (where $J_f$ vanishes) can be subtle, but the measure is well-defined in many cases when the degeneracies of $J_f(x)$ are relatively mild. One can also consider definitions when the measures $\delta^{(\varepsilon)}(f(x) - b)$ are suitably normalized before taking the limit $\varepsilon \searrow 0$. However, with such definitions the limit measure may concentrate at the set of the degenerate points, if the degeneracies are significant. Moreover, the limit may not be unique.

In what follows we will sometimes use the notation $\delta(f(x) - b)\, dx$ even when the existence of this object is not completely clarified. This will be done at an intermediate stage, where we are not yet formulating the exact statements and our discussion is at a heuristic level. Of course, in the formulation of our rigorous results we should be more careful.

Recalling that

$$\delta_{\mathbf{R}^2}(x_1, x_2) = \delta_{\mathbf{R}}(x_1)\delta_{\mathbf{R}}(x_2)\,, \tag{86}$$

we see that for $f = (f_1, \ldots, f_m)$ and $b = (b_1, \ldots, b_m)$ we can write

$$\delta(f(x) - b)\, dx = \delta(f_1(x) - b_1)\delta(f_2(x) - b_2) \ldots \delta(f_m(x) - b_m)\, dx\,. \tag{87}$$

## 3.2 Elementary Example

Let us start with a simple example. On the unit circle $\mathbf{S}^1$ consider set of real-valued functions in $H^1(\mathbf{S}^1)$ with

$$\int_{\mathbf{S}^1} u\, dx = 0, \qquad J(u) = \int_{\mathbf{S}^1} u_x^2\, dx = E\,, \tag{88}$$

where $E > 0$ is given. How does a "random function" satisfying these conditions look? Unless we specify some probability measure on our set, the question is of course not well-defined. Let us set $H_0^1 = H_0^1(\mathbf{S}^1) = \{u \in H^1(\mathbf{S}^1)\,, \int u\, dx = 0\}$. It is natural to work with finite-dimensional subspaces $V \subset H_0^1$. On such subspaces we have a natural measure $\mathcal{L}_V$ invariant under translations (which is unique up to a multiplicative constant). In many cases is coincides with the Liouville measure of a truncation of Hamiltonian systems of interest. A natural probability measure corresponding to our question (when $u$ is in $V$) is the measure

$$Z^{-1}\delta(J - E)\, \mathcal{L}_V\,. \tag{89}$$

[5] Consider for example $n = 2, m = 1$ and $f(x) = x_1x_2$. The reader can check that the measure $\delta(f(x))$ is well-defined in $\mathbf{R}^2 \setminus \{0\}$, but not in $\mathbf{R}^2$.

where

$$Z = \int_V \delta(J(u) - V)\, d\mathcal{L}_V(u) \tag{90}$$

is a normalizing factor. We can now look at "typical properties" of functions with probability distribution given by (89) in suitable limits $V \nearrow H^1_0$. For example, we have

**Proposition 1** *For any $s < 1$*

$$\int_V ||u||^2_{H^s}\, Z^{-1}\delta(J(u) - E)\, d\mathcal{L}_V(u) \to 0 \quad \text{as } \dim V \to \infty\,. \tag{91}$$

We see that the measures (89) will concentrate at 0 when in spaces $H^s$ when $s < 0$. In particular, when the dimension of $V$ is sufficiently high, the "random function" from $V$ with $J(u) = E$ will be close to 0 in the sup-norm. On the compact metric space $(X_E, w)$ where $X_E = \{u \in H^1_0\,,\, J(u) \le E\}$ and $w$ is the weak topology of $H^1$ the measures (90) will also concentrate at 0 as $\dim V \nearrow \infty$.

*Proof of the proposition:* We can choose coordinates $u_1, \ldots, u_n$ on $V$ in which

$$J(u) = u_1^2 + \cdots + u_n^2\,. \tag{92}$$

and

$$J_s(u) = ||u||^2_{\dot{H}^s} = a_1 u_1^2 + \cdots + a_n u_n^2\,. \tag{93}$$

Then

$$\int_V u_k^2\, Z^{-1}\delta(J(u) - E)\, d\mathcal{L}_V(u) = \frac{1}{n}\,, \quad k = 1, 2, \ldots, n \tag{94}$$

and hence

$$\int_V J_s(u)\, Z^{-1}\delta(J(u) - E)\, d\mathcal{L}_V(u) = \frac{1}{n}(a_1 + \cdots + a_n) \tag{95}$$

and the claim follows once we establish that the coefficients that for large $n$ most of the coefficients $a_j$ have to be small. For some specific choices of increasing family of subspaces $V$, such as Fourier truncations, or some other common approximations this is trivial. The general case is left to the reader as an exercise. ■

*Remark* The simple lemma above already can be used to explain some classical variational principles from the point of view of Statistical Mechanics, without invoking "friction" or other type of artificially introduced dissipation. Instead, at the level of Statistical Mechanics, one can see from the above the effect of

the "inviscid dissipation", or "thermodynamical death" (discovered by physicist studying foundations of thermodynamics in nineteenth century).

Let us consider for example the equation

$$u_{tt} - u_{xx} = f(x)\,, \tag{96}$$

on the unit circle $\mathbf{S}^1$, where $u$ and $f$ are real-valued, with

$$\int f(x)\,dx = 0 \qquad \text{and} \qquad \int u(x,t)\,dx = 0\,. \tag{97}$$

Note that the general case when (97) is not satisfied can be reduced to the case when (97) by subtracting solutions of $u_{tt} = f$ for suitable $u$ and $f$ which are constant in $x$. The Eq. (97) is Hamiltonian, with the Hamiltonian

$$H(u_t, u) = \int_{\mathbf{S}^1} \left( \frac{1}{2} u_t^2 + \frac{1}{2} u_x^2 - fu \right) dx\,. \tag{98}$$

The solutions of this linear problem can of course be written quite explicitly, for example in terms of the Fourier coefficients, and they are quasi-periodic. There is no dissipation effect in that case. The standard argument in statistical mechanics is the following. Even if we assume that our system is closed, the Hamiltonian is in reality more complicated than (98). There will be a small extra term $\epsilon H_1$ which will change the long-time dynamics so that the long-time evolution will be towards equi-distribution of energy on the surface $\{H + \epsilon H_1 = E\}$. In the limit of very small (but non-zero) $\epsilon$ the system will evolve with good approximation towards the equi-distribution of energy on the surface $\{H = E\}$. From the lemma above we see (after a simple change of coordinates) that these assumptions imply that over sufficiently long time, the solution should approach weakly in $H^1$ the solution of the steady state equation $-u_{xx} = f$. All this is just another expression of the classical rule of the statistical mechanics that the system is trying to distribute energy between the degrees of freedom to the maximal degree allowed by conservation laws. ■

### 3.3 *Heuristics from Probability for Two Quadratic Constraints*

Let now modify the previous example and consider functions in $H^1(\mathbf{S}^1)$ with

$$\int_{\mathbf{S}^1} u\,dx = 0\,, \quad H(u) = \int_{\mathbf{S}^1} u_x^2\,dx = E\,, \quad I(u) = \int_{\mathbf{S}^1} u^2\,dx = b. \tag{99}$$

We again consider finite-dimensional subspaces

$$V \subset H_0^1 = \{u \in H^1(\mathbf{S}^1)\,,\ \int_{\mathbf{S}^1} u\,dx = 0\}\,, \tag{100}$$

and this time the measure we are interested in will be

$$\mu_{E,b;V} = Z^{-1}\delta(H(u) - E)\delta(I(u) - b)\,\mathcal{L}_V\,, \tag{101}$$

where the normalizing factor $Z$ is given by

$$Z = \int_V \delta(H(u) - E)\,\delta(I(u) - b)\,d\mathcal{L}_V(u)\,. \tag{102}$$

To keep things as simple as possible, we will work out the case when the spaces $V$ are given by Fourier truncation. The space $V$ corresponding to Fourier truncation to frequencies $\leq N$ can be identified with $\mathbf{C}^N$. We will use coordinates $z_k = k\hat{u}(k)$, so that the two functions $H, I$ become

$$H(z) = |z_1|^2 + |z_2|^2 + \cdots + |z_N|^2\,, \tag{103}$$

$$I(z) = a_1|z_1|^2 + a_2|z_2|^2 + \cdots + a_N|z_N|^2\,, \tag{104}$$

where $a_j = 1/j^2$. We will write $\mu_{E,b;N}$ for $\mu_{E,b;V}$ (defined by (101)) in this situation.

There are several heuristic arguments which can help us to see what we can expect as $N \to \infty$. Here we will present one using probability theory. We first simplify the problem slightly and consider the coefficients $a_1, a_2, \ldots$ for which $a_k = 0$ when $k > l$, where $l$ is a fixed (possibly large, but independent of $N$). A classical result in probability is that as $N \to \infty$, the ($\mathbf{C}$-valued) functions $z_1, z_2, \ldots, z_l$ considered on the sphere $|z_1|^2 + \cdots + |z_N|^2 = E$ with surface measure normalized to one increasingly behave as independent normally distributed random variables with mean zero and variance $E/N$, see for example [16]. In other words, we can write, with increasingly good approximation as $N \to \infty$

$$z_k = \sqrt{\frac{E}{N}}\,Z_k\,, \qquad k = 1, 2, \ldots, l\,, \tag{105}$$

where $Z_k$ are independent (complex-valued) variables with normal distribution. This takes (approximately) into the account the constraint $H = E$. To take into account the constraint $I = b$, we restrict out attention to "events"

$$a_1 Z_1^2 + a_2 Z_2^2 + \cdots + a_l Z_l^2 \sim b\,\frac{N}{E}\,. \tag{106}$$

(Strictly speaking, we should take suitable limit measure on events when the last sum belonging to $(bN/E - \varepsilon, bN/E + \varepsilon)$ as $\varepsilon \searrow 0$.). Assume $a_1 > a_2 > \ldots a_l > 0$. For large $N$ the event (106) happens only with extremely low probability, and from the properties of the normal distribution it is clear that the best chance for (106) to happen is that

$$a_1 Z_1^2 \sim b\,\frac{N}{E}\,. \tag{107}$$

All the other possibilities are exponentially less likely. We see from this argument that we can expect that as $N \to \infty$, the behavior of the measures $\mu_{E,b;N}$ will be such that their push-forward on any finite number of coordinates $z_1, \ldots, z_n$ will concentrate in the first one, where it will approach the circle $|z_1|^2 = \frac{bN}{a_1 E}$.

In other words, if we denote by $(X_E, w)$ the space $\{u \in H_0^1(\mathbf{S}^1)\,,\, H(u) \le E\}$ with the weak topology of $H^1$, the measures $\mu_{E,b;N}$ (assuming they are well-defined) will concentrate in this space on the first Fourier mode, i.e. the set of minimizers of $H$ under the constraint $I = b$.

This phenomenon is at heart of all the conjectures about the long-time behavior discussed in these notes. It is very closely related to Bose-Einstein condensation (we will comment more on this later), and to the phenomenon of "thermodynamical death" as discovered in the nineteenth century (as we mentioned earlier).

The behavior of the measures $\mu_{E,b;N}$ should be contrasted with the behavior of the Gibbs measures

$$Z^{-1} e^{-\beta H(u)}\, d\mathcal{L}_V(u)\,, \qquad Z = \int_V e^{-\beta H(u)}\, d\mathcal{L}_V(u) \tag{108}$$

for a fixed $\beta > 0$, which corresponds to fixed positive temperature. In the limit of $V \nearrow H_0^1$ these measures will approach (in a suitable sense) the Wiener measure, which "lives" on functions with regularity below $H^{\frac{1}{2}}$. This corresponds to the situation when, on average, all modes will be excited with a non-zero amount of energy. In particular, the total energy will be infinite. Such measures are relevant for the study of low-regularity (infinite-energy) solutions of non-linear Schrödinger equation and other Hamiltonian equations, as pioneered in [7, 37].

The situation with finite energy can be also related to Gibbs measures with $\beta$ changing as $\beta \sim \beta_0 \dim V$ as the dimension of $V$ increases.

Studying low probability events such as (106) in our situation above is the topic of the Large Deviations Theory, see for example [15]. The techniques developed in that area are applicable to problems we study, see for example [9, 18, 54]. The most natural setup for that theory is in terms of Gibbs measures corresponding to temperatures $1/(\beta_0 \dim V)$. Some work is still needed to handle the constraints, though.

## 3.4 Laplace Principle

Consider a compact metric space $X$ with a probability measure $\mu$. Let $w\colon X \to R$ be a non-negative continuous function which is not identically zero on the support of $\mu$. Consider the probability measures $\nu_n$ given by

$$\int \varphi(x) d\,\nu_n(x) = \frac{\int_X \varphi(x) w^n(x)\, d\mu(x)}{\int_X w^n(x)\, d\mu(x)}\,, \tag{109}$$

for each continuous function $\varphi$.

**Lemma 1** *As $n \to \infty$ the measures $\nu_n$ concentrate in the set $K = K_{\mu,w} = \{x \in \operatorname{supp}\mu\,,\ w(x) = \max_{\operatorname{supp}\mu} w\}$.*

*Proof* Assume $\varphi$ is supported away from $K$ and that $|\varphi| \le C$. Let $M = \max_{\operatorname{supp}\mu} w$ and $M_1 = \max_{\operatorname{supp}\varphi} w\}$. Let $U = \{x\,, w(x) > M_2 = (M + M_1)/2.\}$ and $A = \mu(U)$. Then the expression (109) is clearly estimated by

$$\frac{CM_1^n}{AM_2^n} \tag{110}$$

which approaches zero as $n \to \infty$. ■

More precise forms of the Laplace principle can be considered. For example one can study the asymptotics of the integrals

$$\int_0^1 \varphi(x) x^m e^{\kappa f(x)}\, dx \tag{111}$$

as $\kappa \to \infty$.

In the context of large deviations methods, there is a generalization of the Laplace principle which is called Varadhan's lemma, see [15], which is exactly what is needed if we wish to approach the problems studied here using those methods.

## *3.5 Perturbations Depending on Finitely Many Variables*

There is a very simple heuristics (coming from statistical mechanics) behind all the results discussed which relies only on the Laplace principle, without any references to probability techniques. It is best illustrated in the following example. We will change our notation slightly. We will work with $\mathbf{R}^N$, where we think of $N$ as large (and taking the limit $N \to \infty$. The coordinates in $\mathbf{R}^N$ will be denoted by $X_1, X_2, \dots X_N$. (We emphasize that these are not random variables, but classical plain coordinates.) Assume that we have functions $H, f_1, f_2, \dots, f_r$ of the variables $X$, where we think of $H$ as a Hamiltonian and $f_1, \dots, f_r$ as conserved quantities. We will be interested in the measures

$$\mu_N = Z^{-1}\delta(H - E)\delta(f_1 - c_1)\delta(f_2 - c_2)\dots\delta(f_r - c_r)\,\mathcal{L}_N\,, \tag{112}$$

where, as usual,

$$Z = \int_{\mathbf{R}^N} \delta(H(X) - E)\delta(f_1(X) - c_1)\delta(f_2(X) - c_2)\dots\delta(f_r(X) - c_r)\, d\mathcal{L}_N(X)\,, \tag{113}$$

is a normalizing factor. We assume that all these objects are well defined in the sense of Sect. 3.1. In particular, we assume of course that the conditions $H = E, f_1 = c_1, \ldots, f_r = c_r$ are compatible and can be satisfied for some non-trivial set of states. We will write $f = (f_1, \ldots, f_r)$, $c = (c_1, \ldots, c_r)$ and use the shorter notation

$$\delta_{\mathbf{R}^r}(f - c) = \delta(f_1(X) - c_1)\delta(f_2(X) - c_2) \ldots \delta(f_r(X) - c_r)\,. \tag{114}$$

Finally, we will assume that the measure $\delta_{\mathbf{R}^r}(f(x) - c)\,dx$ is well-defined and its support coincides with the set $\{f(x) = c\}$. This will always be the case when $c$ is a regular value of $f$ (in terms of the Morse-Sard Theorem), and, in fact, under quite weaker assumptions once $m$ is large.

Let us write the Hamiltonian in the form

$$H(X) = \frac{1}{2}(X_1^2 + X_2^2 + \cdots + X_N^2) + f_0(X)\,. \tag{115}$$

Our main assumption in this section will be:

$$\begin{aligned}&\textit{There exists } m \in \mathbf{N} \textit{ independent of } N \textit{ such that } f_0, f_1, \ldots, f_r\\ &\textit{depend only on } X_1, \ldots, X_m.\end{aligned} \tag{116}$$

Assuming that $m, N$ are even and that the symplectic form relevant for the dynamics has constant coefficients, (116) implies for the dynamics that there is no interaction between the variables $X_1, \ldots, X_m$ and $X_{m+1}, \ldots, X_N$. In terms of Statistical Mechanics the system $X_{m+1}, \ldots, X_N$ represents "ideal gas", with very simple dynamics. The assumption that the measures (112) give the correct long-time behavior of the actual dynamics is of course an idealization of what (conjecturally) happens if small terms which can be neglected in this calculation (but still make ergodicity plausible) are present. The usual heuristics is that the system $X_1, \ldots, X_m$ is in contact with the ideal gas represented by $X_{m+1}, \ldots, X_N$

The whole situation can be embedded into the Hilbert space $l^2(\mathbf{N})$ in the obvious way:

$$(X_1, X_2, \ldots, X_N) \to (X_1, X_2, \ldots, X_N, 0, 0, \ldots)\,. \tag{117}$$

In what follows we will assume that $f_0$ is bounded from below. Let $(X_E, w)$ be the compact metric space given by the subset $\{H \leq E\}$ of $l^2(\mathbf{N})$ taken with the weak topology of $l^2(\mathbf{N})$.

**Proposition 2** *In the situation above, as $N \to \infty$, the measures $\mu_N$ considered in $(X_E, w)$ will concentrate on the set of minimizers of $H$ under the constraint $f = c$.*

*Proof* Let $\varphi$ be a continuous function depending on finitely many variables $X_1, \ldots, X_l$. We can assume without loss of generality that $l \geq m + 1$. We will write

$$(X_1, \ldots, X_N) = (x_1, \ldots, x_l, y_1, \ldots, y_n)\,, \qquad l + n = N\,. \tag{118}$$

Then by our assumptions

$$H(X) = H_1(x) + \frac{1}{2}|y|^2\,, \quad f(X) = f(x)\,. \tag{119}$$

Writing the integral $\int dX$ as $\int\int dx\,dy$ and using the easy formula

$$\int_{\mathbf{R}^n} \delta(\frac{1}{2}|y|^2 - \tilde{E})\,dy = \gamma_n \tilde{E}^{\frac{n}{2}-1} \tag{120}$$

(for $\tilde{E} \geq 0$), we obtain

$$\int_{R^N} \varphi(X) d\mu_N(X) = \frac{\int_{R^l} \varphi(x)(E - H_1(x))^{\frac{n}{2}-1} \delta_{\mathbf{R}^r}(f(x) - c)\ dx}{\int_{R^l} (E - H_1(x))^{\frac{n}{2}-1} \delta_{\mathbf{R}^r}(f(x) - c)\ dx}\,, \tag{121}$$

and the result follows from Lemma 1. ■

*Remark* One can interpret the calculation in the following way. The for the fixed number of variables $x_1, \ldots, x_l$ should be at equilibrium with the much larger system described by the variables $y_1, \ldots, y_n$. However, the total energy of the combined system is limited to $E$, and when this is uniformly distributed over all system, there is not much left for the subsystem $x_1, \ldots, x_l$. The system $y_1, \ldots, y_n$ acts effectively as a "refrigerator", removing as much energy out of the system $x_1, \ldots, x_l$ as is allowed by the requirement that $f$ be conserved.

Proposition 2 is "almost" applicable to the non-linear Schrödinger equation (Sect. 2.2) and the gKdV equation (Sect. 2.3), in the sense that although the Hamiltonian and conserved quantities there do not quite satisfy the assumptions of the proposition, they satisfy them after a small perturbation of the functionals. However, bridging the gap between Proposition 2 and the real situation would still require some effort if we wish to work with the exact constraints. It is an interesting problem which may not have been worked out completely in the existing mathematical literature. Things become easier if we are willing to relax the exact constraints somewhat, see below.

Let us now consider a similar approximation for the nonlinear wave equation from Sect. 2.1. We will consider a Fourier truncation to $N$ modes. We will consider the Fourier coefficients as real, having in mean the real form of the Fourier series. We fix a large natural number $l$. Our notation will be as follows.

| | |
|---|---|
| $V_0, V_1, \ldots, V_N$ | coefficients of the velocity field $v$ |
| $X_1, X_2, \ldots, X_N$ | coefficients of the derivative $u_x$ |
| $X_0$ | the "constant part" of $u$ |
| $x = (x_0, \ldots, x_l)$ | the first $l+1$ components of $X$, including $X_0$ |

We will now assume that the Hamiltonian is of the form

$$H = \frac{1}{2}V_0^2 + \frac{1}{2}V_1^2 + \cdots + \frac{1}{2}V_N^2 + \frac{1}{2}X_1^2 + \cdots + \frac{1}{2}X_N^2 + f(x)\,, \tag{122}$$

where $f$ is a truncation of $\int_{\mathbf{S}^1} \frac{\varkappa}{4}u^4$ to $l$ modes. This is the simplification we make in this model—we truncate the "interaction term" to $l$ modes, rather than $N$. As $X_1, \ldots, X_N$ are Fourier modes of the *derivative* $u_x$ and the interaction terms depends only on $u$, the error caused by this approximation can be assumed to be small, in a suitable sense.

The momentum function is of the form

$$P = X_1V_1 + X_2V_2 + \cdots + X_NV_N\,. \tag{123}$$

To diagonalize $P$, we will work with the variables

$$A_k = \frac{V_k + X_k}{\sqrt{2}}\,, \qquad B_k = \frac{V_k - X_k}{\sqrt{2}}\,. \tag{124}$$

and write

$$\begin{aligned} A = \quad (a,\tau) \quad &= (a_0, a_1, \ldots, a_l, \tau_1, \ldots, \tau_n)\,, \\ B = \quad (b,q) \quad &= (b_0, b_1, \ldots, b_l, q_1, \ldots, q_n)\,, \\ g = \tfrac{1}{2}V_0^2 + f(x) \; &= g(a,b)\,. \end{aligned} \tag{125}$$

We also let

$$\begin{aligned} \tilde{E} &= E - \left(\tfrac{1}{2}a_1^2 + \cdots + \tfrac{1}{2}a_l^2 + \tfrac{1}{2}b_1^2 + \cdots + \tfrac{1}{2}b_l^2 + g(a,b)\right)\,, \\ \tilde{p} &= p - \left(\tfrac{1}{2}a_1^2 + \cdots + \tfrac{1}{2}a_l^2 - \tfrac{1}{2}b_1^2 - \cdots - \tfrac{1}{2}b_l^2\right)\,. \end{aligned} \tag{126}$$

for the first $l+1$ components of $A$ and $B$ respectively. Both $\tilde{E}$ and $\tilde{p}$ can be considered as functions of $a, b$. We wish to evaluate, for a smooth test function $\varphi$ of $A, B$ which depends only on $a, b$, the integrals

$$\int_{\mathbf{R}^{N+1}\times\mathbf{R}^{N+1}} \varphi(A,B)d\mu_N(A,B) = \frac{\int_{\mathbf{R}^{N+1}\times\mathbf{R}^{N+1}} \varphi(a,b)\delta(H-E)\delta(P-p)\,dA\,dB}{\int_{\mathbf{R}^{N+1}\times\mathbf{R}^{N+1}} \delta(H-E)\delta(P-p)\,dA\,dB}\,, \tag{127}$$

assuming of course that $E$ and $p$ are such that the integral in the denominator does not vanish. Writing $\int dA\,dB$ as $\int da\,db \int d\tau dq$, the key point is again that the integral $\int d\tau\,dq$ can be evaluated explicitly.

$$\int_{\mathbf{R}^n\times\mathbf{R}^n} \delta(\tilde{E} - \frac{1}{2}|\tau|^2 - \frac{1}{2}|q|^2)\delta(\tilde{p} - \frac{1}{2}|\tau|^2 + \frac{1}{2}|q|^2)\,d\tau\,dq = \iota_n(\tilde{E}-\tilde{p})_+^{\frac{n}{2}-1}(\tilde{E}+\tilde{p})_+^{\frac{n}{2}-1}\,, \tag{128}$$

where $\iota_n$ is an immaterial constant which has no effect on the value of (127) and $(\cdot)_+$ denotes the positive part.

We see that when evaluating (127) for large $N$, we can again apply Lemma (1), and we only need to find the set where the expression

$$w = (\tilde{E} - \tilde{p})_+ (\tilde{E} + \tilde{p})_+ \tag{129}$$

attaints it maximum, when considered as a function of $a, b$. Letting

$$\alpha = \frac{1}{2}a_1^2 + \cdots + \frac{1}{2}a_l^2\,, \quad \beta = \frac{1}{2}b_1^2 + \cdots + \frac{1}{2}b_l^2\,, \quad \gamma = g(a,b) \tag{130}$$

We can take $l$ as large as we wish, as long as we keep it fixed when $N \to \infty$. It is easy to see that once $l$ is sufficiently large, than the parameters $\alpha, \beta, \gamma$ can be chosen independently of each other, as long as they are all positive. Hence we can consider $w$ in (129) as a function of $\alpha, \beta, \gamma$ (and the expression is quite simple), and maximize it over the set $\{\alpha \geq 0, \beta \geq 0, \gamma \geq 0\}$. An routine calculation shows that the maximum is attained at $\alpha = 0, \beta = 0, \gamma = 0$. We have shown

**Proposition 3** *In the situation above, the measures $\mu_N$ concentrate in the space $(X_E, w)$ defined by (14) at the point $(0,0)$ as $N \to \infty$.*

*Proof* See the calculation above. ■

We see that if the guess (10) is correct, then the generic solutions will indeed be escaping to higher and higher Fourier modes, leaving essentially nothing behind in the low modes. This would mean that for example the periodic solutions $U(t)$ which are constant in $x$ and which can have large amplitude would eventually completely "disintegrate". The times scales necessary for this effect (if it is real) might be enormous.

## 3.6 Diagonal Quadratic Forms in $\mathbf{C}^N$

Let us consider the situation when the phase space is $C^N$ and all the functionals involved are of the form

$$f(z_1, \ldots, z_N) = g(|z_1|^2, |z_2|^2, \ldots, |z_N|^2)\,, \tag{131}$$

where $g$ is a function on $\mathbf{R}^N$ (satisfying some mild technical assumptions which will be clear from the context). When $g$ is smooth and compactly supported, it is easy to see that

$$\int_{\mathbf{C}^N} g(|z_1|^2, \ldots, |z_N|^2)\, d\mathcal{L}_{2N}(z) = \pi^N \int_{\mathbf{R}^N_+} g(x_1, \ldots, x_N)\, dx\,, \tag{132}$$

where $\mathcal{L}_{2N}$ is the standard Lebesgue measure ion $\mathbf{C}^N$ and $\mathbf{R}_+^N$ are vectors in $\mathbf{R}^N$ with non-negative coordinates. (This change of variables has been used in [8].) If all the functionals involved in the definition of our measures are of the form (131), one can use the last formula to simplify the calculations in some cases. For example, in Sect. 3.3 the functionals $H, I$ on $\mathbf{C}^N$ can be replaced by the functionals

$$H(x) = x_1 + x_2 + \cdots + x_N\,, \tag{133}$$

$$I(x) = a_1 x_1 + a_2 x_2 + \cdots + a_N x_N\,. \tag{134}$$

The measure

$$\mu_{E,b,N} = Z^{-1}\delta(H - E)\,\delta(I - b)\mathcal{L}_N\,, \tag{135}$$

with

$$Z = \int_{\mathbf{R}_+^N} \delta(H - E)\,\delta(I - b)d\mathcal{L}_N(x) \tag{136}$$

is then a multiple of the $(N-2)$-dimensional Hausdorff measure on the convex set given by the linear equations $H = E, I = b$ in $\mathbf{R}_+^N$. As usual, we only consider the parameters for which $Z > 0$.

One advantage to working with linear constraints is that we do not have to worry about degenerate points of $f'$ when defining the measures $\delta(f(x) - b)$. When $f$ is linear, $f'$ is constant and there are no problems in this direction.

**Connection to Free Bose Gas** Let us assume $a_1 \geq a_2 \geq \cdots > 0\,,\ a_k \to 0$ and set

$$y_k = a_k x_k\,, \quad E_k = \frac{1}{a_k}\,, \quad k = 1, 2, \ldots, N\,. \tag{137}$$

Then

$$H = E_1 y_1 + E_2 y_2 + \cdots + E_N b_N\,, \tag{138}$$

$$I = y_1 + y_2 + \cdots + y_N\,. \tag{139}$$

We can think of a quantum system with $N$ energy states with levels $E_1 \leq E_2 \leq \cdots \leq E_N$, respectively, with $E_N \nearrow \infty$ for $N \to \infty$, and interpret $y_k$ as the number of particles in the state $|k\rangle$. Then $I$ is the total number of particles and $H$ is the energy. One can consider the system under the constraints $H \sim E$ and $I = b$. (Note that the "hard constraint" $H = E$ is not useful here, as the achievable energy levels are discrete.) By $H \sim E$ we can mean for example that we only count states with energy in some well-chosen interval around $E$. This corresponds to "microcanonical ensemble". In physics one studies the question of what the "most likely states" are, although instead of the constraint $H \sim E$ one usually takes the "canonical

ensemble" (with respect to $H$) based on the Gibbs measure given by weighing a natural background measure by $Z^{-1}e^{-\beta H}$. (In the case at hand the background measure would be the counting measure.) A famous result by Bose and Einstein is that at very low temperature a non-negligible fraction of particles will be in the lowest-energy state. This is the Bose-Einstein condensation. See for example [10] for a mathematical treatment and a number of references. The concentration effects we consider here are in some sense a variant of this phenomenon when the variables are continuous and the temperature approaches zero.

We will study a slight generalization of the measures (135), when we can have several constraints of the type $I$. The results will be applicable to measures (101) and, more generally, to various other problems with diagonal quadratic constraints, which include the "ideal Schrödinger gas" discussed below (represented by the linear Schödinger equation) or the micro-canonical version of Kraichnan's well-known results about downward cascades in 2d flows.

On the space $\mathbf{R}^N_+$ we consider the function

$$H(x) = H^{(N)}(x) = x_1 + \cdots + x_N \,, \tag{140}$$

and a set of $r$ linear constraints given by a linear map $A = A^{(N)}\colon \mathbf{R}^N \to \mathbf{R}^r$, which will be identified with its matrix

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \dots & \dots & \dots & \dots \\ a_{r1} & a_{r2} & \dots & a_{rN} \,. \end{pmatrix} \tag{141}$$

For any fixed indices $k, l$ with $1 \le k \le r$ and $l \ge 1$ the coefficient $a_{kl}$ is independent of $N$. In other words, we can think of $A$ as a matrix with $r$ infinite rows, and of $A^{(N)}$ as its truncation to $N$ columns. We will often write $A$ instead of $A^{(N)}$ when the value of $N$ is clear from the context. We assume that

$$a_{kl} \to 0 \text{ as } l \to \infty \,, \quad k = 1, 2, \dots, r \,. \tag{142}$$

and that the rank of $A^{(N)}$ is $r$ once $N$ is sufficiently large. We will be interested in the measures

$$\mu_{E,b;N} = Z^{-1} \delta(H^{(N)}(x) - E)\, \delta_{\mathbf{R}^r}(A^{(N)}x - b)\, \mathcal{L}_N \,, \tag{143}$$

where

$$Z = \int_{\mathbf{R}^N} \delta(H^{(N)}(x) - E)\, \delta_{\mathbf{R}^r}(A^{(N)}x - b)\, d\mathcal{L}_N(x) \,, \tag{144}$$

and we assume that $E > 0$ and $b \in \mathbf{R}^r$ are such that $Z > 0$.

Another way to think about these measures is as follows. Assuming $Z > 0$, the set

$$\{x \in \mathbf{R}^N\,;\ x_i \geq 0,\ i = 1, 2, \ldots, N\,,\ H(x) = E\,,\ Ax = b\} \subset (\mathbf{R}_+)^N\,. \tag{145}$$

is a compact convex set of dimension $N - r - 1$ (we assume $N$ is large), and the measures (143) are just suitable multiples of the $(N - r - 1)$-dimensional Hausdorff measure on the set, normalized to total mass 1.

Strictly speaking, potentially there could be some degenerate cases when our definition in Sect. 3.1 would not be equivalent to this. To avoid this marginal issue, we will only consider the "non-degenerate case". The precise definition is as follows.

**Definition 1** We say that the measures $\mu_{E,b;N}$ are non-degenerate if the set $\{H^{(N)}(x) = E\,,\ A^{(N)}x = b\}$ has a non-empty intersection of dimension $N - r - 1$ with the interior of the cone $\mathbf{R}^N_+$.

Note that by our assumptions the set $\{H^{(N)}(x) = E\,,\ A^{(N)}x = b\}$ has dimension $N - r - 1$ once $N$ is sufficiently large.

The whole situation can be imbedded in the space $l^1(\mathbf{N})$ of absolutely summable sequences. Let

$$X_E = \{x \in l^1(\mathbf{N})\,,\ x_k \geq 0 \text{ for each } k \text{ and } x_1 + x_2 + \cdots \leq E\}\,. \tag{146}$$

It is natural to consider $X_E$ with the weak* topology (when the space $l^1$ is considered as the dual space of the space $c_0$ of infinite sequences converging to 0. We will denote this topology by $w^*$. It is well-known that $(X_E, w^*)$ is a compact metric space. The convergence in $(X_E, w^*)$ is the same as component-wise convergence. (Note that by its definition, the set $X_E$ is obviously bounded in norm in the space $l^1$.)

Our goal in this section is to prove the following result, which—as already indicated above—can be considered in some sense as a version of the Bose-Einstein condensation in the limit when particle numbers can be continuous and the total amount of energy is bounded. In contrast with the quantum case, in such situation the condensation cannot occur at any positive temperature as long as the total amount of energy of the system is bounded. In the quantum case the variables $x_j$ are natural numbers, and that restriction allows the condensation happen (at least partially) even at (small) positive temperature. See [10] for a precise mathematical treatment.

**Theorem 1** *Assume the parameters $E, b$ are such that the measures $\mu_{E,b;N}$ given by (143) are well defined for sufficiently large $N$. Then, as $N \to \infty$, the measures concentrate in the space $(X_E, w^*)$ on the set of minimizers of $H$ on the subset of $X_E$ given by the constraint $Ax = b$.*

*Proof* The proof is quite similar to the proof of Proposition of 2, although this time we will not be able to integrate over the "non-interactive" degrees of freedom explicitly, as the interaction does not have to have a definite cut-off. However, the

bulk of the interaction still comes from only finitely many degrees of freedom, and this will be sufficient in this the case considered in the theorem, as we can quite easily handle the perturbations caused by truncations to finitely many modes.

*Step 1.* (Understanding the minimizers)

Let $e_j$ be the $j$th vector of the canonical basis of $\mathbf{R}^N$ and set

$$a_j = A(Ee_j) \in \mathbf{R}^r \,. \tag{147}$$

This is the $j$th column of he matrix $EA$. The compatibility of the constraints $H = E$ and $Ax = b$ on $\mathbf{R}^N_+$ is easily seen to be equivalent to the condition that $b$ is in the convex hull of the vectors $a_1, a_2, \ldots, a_N$. The condition that $\mu_{E,b;N}$ be well-defined in the sense of Definition 1 means that $b$ is in fact the interior of the convex hull of $a_1, \ldots, a_N$ for large $N$. Let $C_N$ be the convex hull of the set $\{a_1, \ldots, a_N, 0\}$ and define

$$\bar{t} = \{\max t\,;\, tb \in C_N\} \,. \tag{148}$$

The point $\bar{t}b$ lies on the boundary of $C_N$ and since $a_j \to 0$ for $j \to \infty$, $\bar{t}b$ is easily seen to be independent of $N$ once $N$ is large enough. We can write

$$\bar{t}b = \bar{y}_1 a_1 + \bar{y}_2 a_2 + \cdots + \bar{y}_N a_N \,, \tag{149}$$

where $\bar{y}_j$ are positive and sum up to 1. In the "generic case" the point $\bar{t}b$ will lie on a face of the boundary which is an $(r-1)$-simplex and the coefficients $\bar{y}_j$ will be unique, independent of $N$ when $N$ is large, with exactly $r$ of them being strictly positive. In that case the point

$$\bar{x} = \frac{\bar{y}}{\bar{t}} \tag{150}$$

is the unique point where $H$ attains its minimum (which has value $E/\bar{t}$) on the set of constraints. The point $\bar{x}$ will be independent of $N$ for $N$ large enough and exactly $r$ of its coordinates will not vanish.

In the non-generic case the coefficients $\bar{y}$ may not be unique, but it is still clear that there exists some index $k_0$ such that $\bar{y}_k =$ for $k \geq k_0$ once $N$ is large enough. The set of the point of the form (150) will then be a compact convex polyhedron in the intersection of some affine subspace of $\mathbf{R}^N$ and $\mathbf{R}^N_+$.

An important point is that $\bar{t}$ and the point $\bar{b}$ are continuous with respect to small perturbations of $A$ and $b$. In other words, the map $(A, b) \to \bar{t} = \bar{t}(A, b)$ will be continuous with respect to $(A, b)$ as long as the point $b$ belongs to the interior of the convex hull of $a_1, \ldots, a_N$ for sufficiently large $N$. Note also that in the generic case the point $\bar{x}$ will also depend continuously on $(A, b)$ in some small neighborhood of the pair we started with.

Using these observations, it is not hard to prove the following statement.

**Lemma 2** *Assume that $m$ is sufficiently large. Then for each $\sigma > 0$ and each $A', b'$ sufficiently close to $(A, b)$ (in the sense that $A^{(m)}$ is close $A'^{(m)}$ and $b'$ is close to $b$) we have*

$$\frac{\int_{R^m_+} \delta(A'x - b')\mathbf{1}_{\{H(x)<\frac{E}{\tilde{I}(A,b)}+\sigma\}}\, dx}{\int_{R^m_+} \delta(A'x - b')\, \mathbf{1}_{\{H(x)<E\}}\, dx} \geq \varepsilon = \varepsilon(\tau, m, A, b) > 0\,. \tag{151}$$

*Step 2.* Postponing the proof of the lemma for the moment, let us complete the proof of the theorem. Let us write the points of $X = (X_1, \ldots, X_N)$ of $R^N_+$ as

$$X = (x, y)\,, \quad x = (x_1, \ldots, x_m) \in \mathbf{R}^m_+\,, \quad y = (y_1, \ldots, y_n) \in \mathbf{R}^n_+\,, \quad m+n = N\,. \tag{152}$$

Let us further write

$$y = (E - t)\eta \quad \eta \in \mathbf{R}^n_+\,, \quad \eta_1 + \eta_2 + \cdots + \eta_n = 1\,, \quad t \in [0, E]\,. \tag{153}$$

We will use the notation

$$\Delta^{n-1} = \{\eta \in \mathbf{R}^n_+\ ,\ \eta_1 + \eta_2 + \cdots + \eta_n = 1\} \tag{154}$$

Letting $H(x) = H^{(m)}(x) = x_1 + x_2 + \cdots + x_m$, it is easy to see that we ca write our integrals as

$$\int_{\mathbf{R}^N_+} f(X)\delta(H - E)dX = \tag{155}$$

$$c_{m,n} \int_{\Delta^{n-1}} \int_{R^m_+} f(x, (E - H(x))\eta)(E - H(x))^{n-1}\, dx\, d\mathcal{H}^{n-1}(\eta),$$

where $c_{m,n}$ as positive constant the exact value of which is immaterial for our calculations. (Evaluating $c_{m,n}$ explicitly is an interesting exercise.)

We need to evaluate the limit as $n \to \infty$ of integrals of the form

$$I_N = \frac{\int_{\mathbf{R}^N_+} \varphi(X)\, \delta(H(X) - E)\, \delta_{\mathbf{R}^r}(AX - b)\, dX}{\int_{\mathbf{R}^N_+}\ \delta(H(X) - E)\delta_{\mathbf{R}^r}(AX - b)\, dX}\,, \tag{156}$$

where we can assume without loss of generality that the function $\varphi$ depends only on the first $m$ variables $x = (x_1, \ldots, x_m)$. Using the change of variables (156), we can write

$$I_N = \frac{\int_{\Delta^{n-1}} d\eta \int_{\mathbf{R}^m_+}\ dx\ \varphi(x)\delta_{\mathbf{R}^r}(Ax + (E - H(x))A\eta - b)(E - H(x))^{n-1}_+}{\int_{\Delta^{n-1}} d\eta \int_{\mathbf{R}^m_+}\ dx\ \ \delta_{\mathbf{R}^r}(Ax + (E - H(x))A\eta - b)(E - H(x))^{n-1}_+}\,, \tag{157}$$

where $(\cdot)_+$ denotes the positive part. In related integrals in the proof of Proposition 2 we could integrate over $\eta$ explicitly, but here such an evaluation might be harder (although perhaps still not impossible). Therefore it seems to be easier to first integrate of $x$. We note that once $m$ is large the term $(E - H(x))A\eta$ will be small, due to assumption (142). The dominant term in (157) will again be $(E - H(x))^{n-1}$. Let us assume that $\varphi$ is supported away from the set where $(E - H(x))$ attains its maximum. We need to show that under this assumption $I_N \to 0$ as $N \to \infty$. Using Lemma 2, this can be done by essentially adjusting the proof of Lemma 1 to the case when there is an additional averaging involved, and instead of (109) we have

$$\frac{\int_{\Sigma_n} d\eta \int_X \varphi(x) w^n(x)\, d\mu(x,\eta)}{\int_{\Sigma_n} d\eta \int_X w^n(x)\, d\mu(x,\eta)}\,, \tag{158}$$

where $d\eta$ denotes some probability measure on the set $\Sigma_n$. (Note that the expression (158) is invariant under replacing $d\eta$ by $c\, d\eta,\ c > 0$.) The potential difficulty is that we do not have control which is sufficiently uniform in $\eta$. Lemma 2 and the upper semi-continuity of the set-valued map $A', b' \to Y(A', b', \frac{E}{7}$ mentioned in its proof address this issue. Using the notation from the proof of Lemma 1, the upper semi-continuity of our set-valued map implies that there exists $\tilde{M} > M_1$ such that

$$\max_{\operatorname{supp}\mu(\cdot,\eta)} w \geq \tilde{M}\,, \quad \text{for each } \eta \in \Sigma_n\,, \tag{159}$$

and Lemma 2 then shows that the expression (158) will still be bounded by

$$\frac{CM_1^n}{\epsilon \tilde{M}_2^n} \tag{160}$$

for some $\tilde{M}_2 > M_1$. The statement of the theorem then follows easily. ■

*Proof of Lemma 2:* During the proof we consider $m$ as fixed (and sufficiently large), and therefore we can write $A$ and $H$ instead of $A^{(m)}$ and $H^{(m)}$, respectively, and, in general, we do not have to indicate the dependence of $m$. We use the notation introduced in the proof of the theorem before Lemma 2. Let

$$Y(A, b, E) = \{x \in \mathbf{R}^m_+\,,\ Ax = b, H(x) = E\}. \tag{161}$$

In the non-degenerate case it is easily seen that

$$\mathcal{H}^{m-r-1}(Y(A', b', E) \geq \gamma > 0 \tag{162}$$

for $A', b'$, in a neighborhood of $A, b$ for some constant $\gamma$ which depends on $A, b, E, m$ and the size of the neighborhood. The set $Y(A', b', \frac{E}{7})$ consists of a single point $\bar{x}'$ in the generic case, when $A', b'$ is close to $A, b$. In the non-generic case the set $Y(A', b', \frac{E}{7})$ may be larger, and we just choose any $\bar{x}'$ in it. The dependence of $\bar{x}'$ on

the parameters may not be continuous in that case, but the important point is that $\bar{t}'$ is continuous.

Also, the set-valued map $A', b' \to Y(A', b', \frac{E}{\bar{t}'})$ is easily seen to be upper-semicontinuous, in the sense that for each open set $\mathcal{O}$ containing $Y(A, b, \frac{E}{\bar{t}})$ there is a neighborhood of $A, b$ such that $Y(A', b', \frac{E}{\bar{t}'}) \subset \mathcal{O}$ for $A', b'$ in that neighborhood.

Because $H$ is linear, we have

$$H((1-s)\bar{x}' + (s)Y(A', b', E)) = (1-s)\frac{E}{\bar{t}'} + sE\,. \tag{163}$$

Define $s_0$ by

$$(1-s_0)\frac{E}{\bar{t}} + s_0 E = \frac{E}{\bar{t}} + \frac{\sigma}{2},. \tag{164}$$

Note that the point $\bar{x}'$ is at distance at least

$$d' = \frac{1}{\sqrt{m}}\,(E - \frac{E}{\bar{t}'}) \tag{165}$$

from the plane $H = E$ in which $Y(A', b', E)$ is contained. Together with (162) this means the $\mathcal{H}^r$ measure of the set

$$\{(1-s)\bar{x}' + sy\,,\ s \in (0, s_0)\,,\ y \in Y(A', b', E)\} \tag{166}$$

is bounded below by $\frac{1}{r}\gamma s_0^{r-1} d'$. Recalling that $\bar{t}'$ is continuous in $A', b'$ the claim follows easily. ■

Let us now present two examples where Theorem 1 can be applied (in addition to the model problem with measures (101)).

*Example 1 (Kraichnan's Energy-Enstrophy Model for 2d Euler)* We consider the 2s Euler equation, see Sect. 2.5. We assume the conservation of the energy

$$H(\omega) = \frac{1}{2}\int_{\mathbf{T}^2} -\omega\psi\, dx = \frac{1}{2}\int_{\mathbf{T}^2} |u|^2\, dx \tag{167}$$

and enstrophy

$$I(\omega) = \int_{\mathbf{T}^2} \omega^2\, dx\,. \tag{168}$$

As we have seen, there are many other conserved quantities, but it turns out that the prediction based on $H$ and $I$ alone already captures the most important phenomenon, which is Kraichnan's downward cascade. Also, for direct Fourier truncation of Euler the quantities $\mathcal{E}$ and $I$ may could be the only conservation laws. (There are more sophisticated truncations, see for example [1, 58], but the direct Fourier truncation

probably remains the one which is used the most.) In the Fourier representation the Functionals $H$ and $I$ are quadratic and diagonal, and hence after the change of variables $|\hat{\omega}_k|^2 = x_k$ and using formula (132) Theorem 1 immediately applies. We obtain Kraichnan's conclusion that in the Fourier space the energy should concentrate in the lowest (non-zero) modes.

*Example 2 (Ideal "Schrödinger gas" in d Dimensions)* In the $d$-dimensional torus consider the classical Schrödinger equation

$$iu_t = -\Delta u + \kappa u\,. \tag{169}$$

The Noether-type conserved quantities (which survive a generic non-linear perturbation, assuming of course we adjust the Hamiltonian accordingly) are the Hamiltonian $H \sim \int_{\mathbf{T}^d} \nabla u \nabla \bar{u} + \kappa u \bar{u}$, the momentum $P \sim -\frac{i}{2}\int_{\mathbf{T}^d}(\bar{u}\nabla u - u\nabla\bar{u})$ and the mass $I = \int_{\mathbf{T}^d} u\bar{u}$. The value of $\kappa$ can be adjusted by the change of variables $u(x,t) \to u(x,t)e^{-iE_0 t}$. The linear equation of course does not have any interaction between the Fourier modes, but we can again think of a very small non-linear perturbation which will ensure ergodicity by will not significantly affect the Statistical Mechanics considerations, similarly to the considerations for the classical "ideal gas", where inter-molecular interaction energy is not considered although the system is assumed to be ergodic (for the purposes of Statistical Mechanics). All the quantities are quadratic diagonal in the Fourier variables, and hence Theorem 1 applies immediately after the change of variables $|\hat{u}_k|^2 = x_k$, at least when $\kappa > 0$. The general case follows easily either by directly checking the proof of the Theorem in the slightly more general situation arising due to the presence of modes with zero or negative energy, or by the change of variable indicated above. In any case the conclusion is that the measures representing the statistical equilibria will concentrate (in the weak topology of $H^1$) on the set of minimizers of the Hamiltonian subject to the constraints $P = p$ and $I = m$. The proof of Theorem 1 then shows how to find these minimizers. For example, when $\kappa = 0$ and $E, p, m$ are given, we can mix a particle of momentum $p$ by suitably using some of the closest "pure momentum states" in a way which minimizes the energy, and then put all the remaining mass into the zero mode. In particular, if $p = 0$, all the mass will be in the zero mode. The long time behavior of the (slightly non-linearly perturbed) equation should then be given by weak convergence of the solution to the set of these minimizers, with the excess energy going to high frequencies.

## *3.7 More General Functionals*

The method used in the proof of Theorem 1 can be used in more general situations, although one may have to modify the statements somewhat to avoid some natural complications. We will illustrate this on the simple example which can be thought of as a non-linear version of Proposition 1 considered in Sect. 3.2. Let us consider

$\mathbf{R}^N$ with coordinates $X = (X_1, \ldots, X_N)$. We will consider a situation with just one functional $H(X) = H^{(N)}(X)$. One can think of

$$H(u) = \int_{\mathbf{S}^1} \left( \frac{1}{2} u_x^2 + F(u) \right) dx \,, \tag{170}$$

with a smooth $F$ restricted to Fourier truncations, for example. Additional assumptions on $F$ will become clear as we proceed. One of them will be

$$F(u) \to \infty \text{ when } |u| \to \infty \,, \tag{171}$$

which in dimension $d = 1$ is sufficient to guarantee the existence of minimizers. For the finite-dimensional truncation we will again use the notation

$$X = (x, y) \,, \quad x = (x_1, \ldots, x_m) \,, \quad y = (y_1, \ldots, y_n) \,, \quad m + n = N \,, \tag{172}$$

where the meaning of $x$ is of course different than in (170). There will be no danger of misunderstanding from this slight abuse of notation. We will write

$$H(x, y) = H_m(x) + h_m(x, y) + \frac{1}{2}|y|^2 \,, \tag{173}$$

i.e. we think of the $y_j$ as Fourier coefficients of the derivative $u_x$ (possibly after some fixed finite shift in the indices). In terms of the original variable $u$ this may for example correspond to letting $u_N = u_m + v_m$, where $u_N$ is the Fourier truncation to $N$ modes and $u_m$ is the Fourier truncation to $m$ modes, with $v_m$ representing the error between the two truncations (whose dependence on $N$ is not indicated), and setting

$$H_m(u_N) = H(u_m) \qquad |y|^2 = \int_{\mathbf{S}^1} (v_m)_x^2 \, dx \,, \tag{174}$$

so that

$$h_m = \int_{\mathbf{S}^1} \left( F(u_m + v_m) - F(u_m) \right) \, dx \,. \tag{175}$$

Then

$$h_m = \int_0^1 ds \int_{\mathbf{S}^1} F'(u_m + s v_m) v_m \, dx \,. \tag{176}$$

The last expression is small in comparison to the $H^1$-norm of $v_m$ when $F$ is smooth, $u_N$ is controlled in $H^1$, and $v_m$ is only supported in high Fourier modes.

Similarly to our previous calculations, we define

$$\int_{\mathbf{R}^N} \varphi(X)\, d\mu_{E,N}(X) = \frac{\int_{\mathbf{R}^N} \varphi(x)\delta(H(x,y)-E)\,dx\,dy}{\int_{\mathbf{R}^N} \delta(H(x,y)-E)\,dx\,dy}\,. \tag{177}$$

and wish to show that the measures $\mu_{E,N}$ (assuming they are well-defined) concentrate (under suitable assumptions, and in the sense of the weak topology of $H^1$) on the set of minimizers of $H$. Let us first proceed with the calculation formally, without worrying about various issues which eventually will need to be addressed by a more careful analysis of our assumptions, or by slightly adjusting the statement.

We will change variables in the integrals as follows. We set

$$y = r\eta\,, \quad \eta \in \mathbf{S}^{n-1}\,, \quad r = |y|\,, \tag{178}$$

where $\mathbf{S}^{n-1}$ is the unit sphere in $\mathbf{R}^n$. Then

$$\int_{\mathbf{R}^N} g(X)\,dX = \int_{\mathbf{S}^{n-1}} d\eta \int_{\mathbf{R}^m} dx \int_0^\infty dr\; g(x, r\eta)\,, \tag{179}$$

for any $g$ (under appropriate assumptions), and hence we can write

$$\int_{\mathbf{R}^N} \varphi(X)\, d\,\mu_{E,N}(X) = \tag{180}$$
$$\frac{\int_{\mathbf{S}^{n-1}} d\eta \int_{\mathbf{R}^m} \varphi(x)dx \int_0^\infty r^{n-1}dr\;\delta(H_m(x)+h_m(x,r\eta)+\frac{1}{2}r^2-E\,)}{\int_{\mathbf{S}^{n-1}} d\eta \int_{\mathbf{R}^m} dx \int_0^\infty r^{n-1}dr\;\;\delta(H_m(x)+h_m(x,r\eta)+\frac{1}{2}r^2-E\,)\,.}$$

To evaluate the integral over $r$ in the last expression, let us consider the equation for $r$ given by

$$H_m(x) + h_m(x, r\eta) + \frac{1}{2}r^2 = E\,, \tag{181}$$

in which we consider $x$ and $\eta$ as parameters. Some of the issues we have to deal with surface already in this equation. We would like to consider (181) as a perturbation of the case $h_m = 0$, which can be solved explicitly. When $E - H_m(x)$ is not very small, we have no problem as long as the derivative

$$\frac{\partial}{\partial r} h_m(x, r\eta) = \eta \nabla_y h_m(x, r\eta)\,. \tag{182}$$

is sufficiently small, and standard implicit function theorem considerations give us unique solvability. We will denote the solution $R = R(x, \eta, E)$. However, when $H_m(x) - E$ is very small, the situation may be more complicated. Disregarding this difficulty for a moment, let us assume that (181) either has a unique solution $R(x, \eta, E)$ it has no solution, in which case we define $R(x, \eta, E) = 0$. Let $\theta\colon \mathbf{R} \to R$

be the usual Heaviside function, equal to one for non-negative numbers and to zero otherwise. Then

$$\int_0^\infty \theta(E - H_m(x) - h_m(x, r\eta) - \frac{1}{2}r^2) r^{n-1}\, dr = \frac{1}{n} R^n(x, \eta, E)\,, \tag{183}$$

Taking $\frac{\partial}{\partial E}$ of the last equation, we obtain

$$\int_0^\infty \delta(E - H_m(x) - h_m(x, r\eta) - \frac{1}{2}r^2) r^{n-1}\, dr = R^{n-1}(x, \eta, E) \frac{\partial R(x, \eta, E)}{\partial E}\,. \tag{184}$$

The derivative $\frac{\partial R}{\partial E}$ can be evaluated by taking the $\frac{\partial}{\partial E}$ derivative of (181).

$$(R + \eta \nabla_y h_m(x, R\eta)) \frac{\partial R}{\partial E} = 1 \ : \tag{185}$$

We again see the difficulties with small $R$ from this equation.

$$\frac{\int_{\mathbf{S}^{n-1}} d\eta \int_{\mathbf{R}^m} \varphi(x) R^{n-1}(x, \eta, E) \frac{\partial R(x,\eta,E)}{\partial E}\, dx}{\int_{\mathbf{S}^{n-1}} d\eta \int_{\mathbf{R}^m} R^{n-1}(x, \eta, E) \frac{\partial R(x,\eta,E)}{\partial E}\, dx}\,, \tag{186}$$

and we see that, modulo the difficulties near $R = 0$, the situation is similar as before: the dominant term in the integrals should be $\sim R^{n-2}$ and the measures $\mu_{E,N}$ should concentrate near the points where $R$ is maximal, taking into account the smallness of $h_m$ for large $m$, should be near the minimizers of $H$.

The difficulties for small $R$ may be genuine for certain functions $H$, but in many cases they are only caused by our choice of coordinates. Adjusting the coordinates to the function $H$ will no doubt solve these issues in at least in some cases, under appropriate assumptions on $H$ (some which are already needed for all the objects to be well-defined).

One can also deal with this difficulty by changing the setup slightly. We note that the difficulties just discussed come from small values of $R(x, \eta, E)$, which at the level reasoning based on real "physical states" should be harmless. Our difficulties come from working with the "hard-conditioned" microcanonical measures $Z^{-1}\delta(H - E)\, dX$. If we work with $\delta^{(\varepsilon)}$ as defined by (77) for some $\varepsilon > 0$, the difficulties largely disappear. Roughly speaking, denoting again by $\varepsilon > 0$ the "regularization parameter" in $\delta^{(\varepsilon)}$, instead of working with

$$\lim_{N\to\infty} \lim_{\varepsilon\to 0}\,, \tag{187}$$

we can work with

$$\lim_{\varepsilon\to 0} \lim_{N\to\infty}\,. \tag{188}$$

This approach is taken for example in [9] or in [18], although in these works the authors use different methods, based on the theory of large deviations. The methods used above can be also applied in those cases. The advantage of working with (188) is that one can handle quite "rough" perturbations from exactly calculable situations.

There can also consider an intermediate way between (187) and (188): we can take $\epsilon$ depending on $N$. In many cases this actually happens in some sense with the canonical ensemble. For readers not working in this area we recall this classical phenomenon for the simplest case of the Hamiltonian

$$Q(X) = \frac{1}{2}|X|^2 \tag{189}$$

in $\mathbf{R}^N$. The micro-canonical ensemble for energy level $E > 0$ is

$$\mu_{E,N} = Z^{-1}\delta(Q - E)\,dX\,, \qquad Z = \int_{\mathbf{R}^N} \delta(Q - E)\,dX\,. \tag{190}$$

This is just the surface measure on the sphere $\{|X|^2 = E\}$ normalized to total mass one. The corresponding canonical measure (the Gibbs measure) is

$$\gamma_{E,N} = Z^{-1}e^{-\beta|X|^2}\,dX\,, \qquad Z = \int_{\mathbf{R}^N} e^{-\beta|X|^2}\,dX\,, \tag{191}$$

where $\beta$ is chosen so that

$$\int_{\mathbf{R}^N} |X|^2\,d\nu_{E,N} = E\,. \tag{192}$$

This gives $\beta = \frac{N}{2E}$. A classical calculation shows that most of the mass of the measure $\gamma_{E,N}$ concentrates within distance $\sim \sqrt{\frac{E}{N}}$ of the sphere $\{|X| = E\}$ and therefore there is an analogy between $\gamma_{E,N}$ the measure

$$(Z^\varepsilon)^{-1}\delta^{(\varepsilon)}(Q - E)\,dX\,, \qquad Z^\varepsilon = \int_{\mathbf{R}^N} \delta^{(\varepsilon)}(Q - E)\,dX\,, \tag{193}$$

with

$$\varepsilon \sim \sqrt{\frac{E}{N}}\,, \tag{194}$$

and a suitable choice of the mollifying kernel $\phi$ in (76).

In this spirit we will consider measures $\nu_{N,E}$ defined for a sequence $\varepsilon_k \searrow 0$ by

$$d\,\nu_{E,N}(X) = Z_N^{-1}\delta^{(\varepsilon_N)}(H^{(N)}(X) - E)\,dX\,, \qquad Z_N = \int_{\mathbf{R}^N} \delta^{(\varepsilon_N)}(H^{(N)}(X) - E)\,dX\,. \tag{195}$$

We will write $H(X)$ instead of $H^{(N)}(X)$ in what follows if there is no danger of confusion. We will again use the notation (172), (173), and (178). Let us fix $E' > E$. We will assume that $H$ is Lipschitz on the subset $\{H \leq E'\}$ and, in addition,

$$\sup_{H(X)\leq E'} \left(|h_m(x, r\eta)| + |\eta \nabla_y h_m(x, r\eta)|\right) \to 0 \text{ as } m \to \infty. \tag{196}$$

These conditions are satisfied in many situations relevant for PDEs is lower dimensions. For example, in the context of functional (170) which we are considering in this section, we recall that we should think of $X$ as the Fourier coefficients of the derivative $u_x$ and hence the $X$-coordinates of high frequency part of $u$ will have a factor (high frequency)$^{-1}$ in front of them.

**Proposition 4** *Using the notation introduced above, assume condition (190) is satisfied. Let $\kappa \in (0, 1)$ and set $\varepsilon_k = \kappa^k$, $k = 1, 2, \ldots$. Then the measures $\nu_{E,N}$ defined by (189) concentrate (in the weak topology of $H^1$) on the set of minimizers of $H$ as $N \to \infty$.*

*Sketch of proof:* We use the change of variables (179) and write, similarly to (180),

$$\int_{\mathbf{R}^N} \varphi(X)\, d\nu_{E,N}(X) = \tag{197}$$
$$\frac{\int_{\mathbf{S}^{n-1}} d\eta \int_{\mathbf{R}^m} \varphi(x)dx \int_0^\infty r^{n-1}dr\ \delta^{(\varepsilon_N)}(H_m(x) + h_m(x, r\eta) + \frac{1}{2}r^2 - E)}{\int_{\mathbf{S}^{n-1}} d\eta \int_{\mathbf{R}^m} dx \int_0^\infty r^{n-1}dr\ \delta^{(\varepsilon_N)}(H_m(x) + h_m(x, r\eta) + \frac{1}{2}r^2 - E)}\,.$$

For a fixed $x, \eta$ and a fixed $m$ we set

$$\tilde{E} = E - H_m(x)\,, \qquad \tilde{h}(r) = h_m(x, r\eta)\,. \tag{198}$$

We will also temporarily write $\varepsilon = \varepsilon_N$. The key point is to have good estimates for the integral

$$I = \int_0^\infty \delta^{(\varepsilon)}\left(\frac{1}{2}r^2 + \tilde{h}(r) - \tilde{E}\right) r^{n-1}\, dr\,. \tag{199}$$

Let $\underline{E}$ be the minimum value of $H$. Assume

$$|\tilde{h}| \leq \sigma\,, \qquad |\tilde{h}'| \leq \tau\,. \tag{200}$$

Let $\alpha \in (\frac{9}{10}, 1)$ and choose $\tilde{E}_1$ so that

$$\alpha(E - \underline{E}) < \tilde{E}_1 < E - \underline{E}\,. \tag{201}$$

Also choose $\tilde{E}_0$ so that

$$\tilde{E}_0 = \frac{\kappa^2}{16}\tilde{E}_1 \,. \tag{202}$$

Once $\sigma, \tau, \epsilon$ are small enough, the largest values of $I$ will be for $\tilde{E}$ above $\tilde{E}_1$, and they will bounded below by

$$\sim (\tilde{E}_1)^{\frac{n}{2}-1} \,. \tag{203}$$

For $\tilde{E} > \tilde{E}_0$ and small enough $\sigma, \tau, \epsilon$ the value of $I$ will be given to a good approximation by

$$\sim (\tilde{E})^{\frac{n}{2}-1} \,. \tag{204}$$

In this respect the situation is quite similar to what we have seen in the proof of Theorem 1. For small $\tilde{E}$ the bound (205) will no longer be valid, but we can use the estimate

$$I \lesssim \frac{(2\sqrt{\tilde{E}_0})^{n-1}}{\epsilon} = \frac{(\frac{\kappa}{2})^{n-1}(\tilde{E}_1)^{\frac{n-1}{2}}}{\varepsilon} \,. \tag{205}$$

which will be true for sufficiently small $\sigma, \tau, \epsilon$. Now for $\varepsilon = \kappa^N$ the last expression of (205) is dominated by (203) as $N \to \infty$. The rest of the proof is similar to the proof of Theorem 1. One issue which may be worth mentioning is that we need to make sure that there will be enough mass (with respect to the measure $dx$, for any $\eta$) in the set where $I$ is bounded below by (203). At a heuristic level this will follow if we show that the solutions $R(x, \eta, E)$ of (181) have continuous dependence on $x$ for each fixed $m$. This is not hard to see by taking the $x_j$-derivatives of

$$H_m(x) + h_m(x, \eta R) + \frac{1}{2}R^2 = E \tag{206}$$

and using our assumptions on $H, h_m$ to estimates $\frac{\partial R}{\partial x_j}$ . ■

*Remark 1* The same idea can be used in the case when, in addition to $H$, we have several constraints $f(X)$ whose dependence of $X_n$ is becoming weak for large $n$, in the sense that

$$f(x, y) = f_m(x) + g_m(x, y) \,, \tag{207}$$

where $g_m(x, y)$ satisfies a condition of type (196). This case is planned to be discussed, together with other material, in [48].

*Remark 2* Since, in the non-linear case, finite dimensional truncations give typically only approximations of the original equation, taking $\delta^{(\varepsilon_N)}$ instead $\delta$ in our formulae, with $\varepsilon_N$ quickly converging actually seems quite appropriate.

## 4 On the Cauchy Problems for the Navier-Stokes Equations

In this last lecture we will discuss recent results in [27, 28] concerning the Cauchy problem for the Navier-Stokes equations. Our focus will be on the main ideas and heuristics, technical details can be found in the two papers just mentioned. The equations are

$$\left.\begin{aligned} u_t + u\nabla u + \nabla p - \Delta u &= 0 \\ \operatorname{div} u &= 0 \end{aligned}\right\} \quad \text{in } \mathbf{R}^3 \times (0,\infty)\,, \tag{208}$$

$$u|_{t=0} = u_0 \quad \text{in } \mathbf{R}^3\,, \tag{209}$$

There are two basic features of the equations: energy identity and the scaling symmetry. The energy identity describes the evolution of the kinetic energy in the fluid:

$$\int_{\mathbf{R}^3} \frac{1}{2}|u(x,t_2)|^2\,dx + \int_{t_1}^{t_2}\int_{\mathbf{R}^3} |\nabla u(x,t)|^2\,dx\,dt = \int_{\mathbf{R}^3} \frac{1}{2}|u(x,t_1)|^2\,dx\,, \tag{210}$$

for $0 \le t_1 \le t_2$. The scaling symmetry is

$$\begin{array}{lll} u(x,t) & u_\lambda(x,t) & \lambda u(\lambda x, \lambda^2 t)\,, \\ p(x,t) & \to p_\lambda(x,t) & = \lambda^2 p(\lambda x, \lambda^2 t)\,, \\ u_0(x) & u_{0\lambda}(x) & \lambda u_0(\lambda x)\,, \end{array} \tag{211}$$

where $\lambda$ is any number in $(0,\infty)$. It is immediate to verify that these transformations preserve the equations. We have

$$\int_{\mathbf{R}^3} \frac{1}{2}|u_{0\lambda}(x)|^2\,dx = \lambda^{-1}\int_{\mathbf{R}^3} \frac{1}{2}|u_0|^2\,dx\,, \tag{212}$$

which means that the magnitude of the kinetic energy is just a relative number from the point of view of the equation. We can scale it to any value by the symmetry (211). Moreover, we see that the energy becomes smaller if we "move" the solution to smaller scales. This is not good news for the regularity theory, as regularity is exactly about controlling what happens at small scales. At the time of this writing the energy

is the only coercive quantity which we control for general solutions. It means that the equation is "super-critical". This is a relative notion: if a new coercive estimate is discovered (however unlikely this may be), the status of the equation could be changed to "critical" or even "sub-critical".

## *4.1 Weak Solutions and the Problem of Their Local-In Time Uniqueness*

It is interesting to note that the scaling symmetry and the considerations around it do not play any role in the construction of the Leray-Hopf weak solution. That construction goes along the following lines. We first approximate the equation by a suitable regularization (in which it becomes sub-critical, for example), or a truncation to a finite-dimensional subspace, in such a way that the energy estimate is still valid for the approximation. That way we get for each level of approximation (which we can associate to a parameter $\varepsilon > 0$, for example) an approximate solution $u^{(\varepsilon)}$ and the remaining task is to prove that a suitably chosen subsequence of these approximations converges to the (weak) solution of the original equations. The details of this construction, which goes back to Leray's paper [38], can be found in any text on the weak solutions. One can go through the whole construction without ever needing to consider the scaling symmetry. The natural function space for the construction of the weak solutions, the space $L_t^\infty L_x^2 \cap L_t^2 \dot{H}_x^1$, is based exactly on the quantities controlled by the energy estimate. The use of this space and every other step in the construction seem to be exactly what is needed.

The main issue in the theory of the weak solutions is their uniqueness. Already in Leray's 1934 paper criteria for uniqueness are studied, and a number of works by other mathematicians have extended his results since. E. Hopf seemed to believe in uniqueness of the weak solutions (see [24]) whereas O. A. Ladyzhenskaya has believed that one cannot get uniqueness in this class, see [36]. This is discussed in more detail in the introduction of the paper [28].

Of course, the question of uniqueness is of fundamental importance. The Navier-Stokes equation is a newtonian model, and in the newtonian mechanics the state of the system at present should uniquely determine its state in the future. It was already known to Leray that, under natural assumptions, a weak solution of the Cauchy problem with $u_0$ in the space $L^p$ for some $p > 0$ is unique near time $t = 0$ and globally unique if it does not develop a singularity. Later this was extended by Kato to the case $p = 3$, see [29].

The construction of weak solutions works well for $u_0 \in L^2$, but the local-in time uniqueness of the weak solutions is open in that case.

## 4.2 Perturbation Theory

Another method for proving the existence of solutions (this time only local-in time) is perturbation theory. In this approach we treat the non-linear term in the equation as a perturbation, and construct a sequence of approximate solutions by solving the linear problems

$$\begin{aligned} {u^0}_t + \nabla p^{(0)} - \Delta u^0 &= 0 \\ \operatorname{div} u^0 &= 0 \\ u^0|_{t=0} &= u_0 \end{aligned} \tag{213}$$

and

$$\begin{aligned} {u^n}_t + \nabla p^{(n)} - \Delta u^n &= -\operatorname{div}\left(u^{n-1} \otimes u^{n-1}\right), \\ \operatorname{div} u^n &= 0, \\ u^n|_{t=0} &= 0, \end{aligned} \tag{214}$$

for $n = 1, 2, 3, \ldots$.

In the context of the Navier-Stokes equations the procedure (often called Picard iteration) goes back to Oseen's 1911 paper [47], and produces a sequence of solutions $u^n$ which with the right assumptions should converge to a solution $u$ of the original Cauchy problem. Leray essentially showed that it works at least for short times when $u_0 \in L^p$ for $p > 3$. Note that the space $L^p$ is "subcritical" when $p > 3$: the $L^p$-norm of the scaled function $u_{0\lambda}$ increases with increasing $\lambda$, so we cannot move the solution to small scales without having to pay a large penalty in terms of the $L_t^\infty L_x^p$ norm. The case $p = 3$ is "critical", a borderline between super-critical and sub-critical. The quantity $\int_{\mathbf{R}^3} |u_0|^3\, dx$ is invariant for the Navier-Stokes scaling (similarly to length in geometry being invariant under rotations or translations) and its magnitude has an intrinsic meaning from the point of view of the equation. Note that this means that if we can prove that the iteration above converges on the time interval $(0, T(\kappa))$ whenever $||u_0||_{L^3} \leq \kappa$, then the solution has to be global when $T(\kappa) > 0$ and $||u_0||_{L^3} \leq \kappa$. In this case we have global existence of smooth solutions based purely on the perturbation theory, without any help from the energy inequality, and hence that our proof should work for a more general class of equations.

The $L^p$ spaces in the above considerations can be replaced by many other spaces, such as various Besov spaces, Morrey spaces, etc. One of the best perturbation theory results is due to Koch-Tataru [33], with $u_0 \in (BMO)^{-1}$, which is again a critical space for the Navier-Stokes scaling.

There are various arguments which can help to elucidate the role the scaling symmetry and various function spaces in the perturbation theory arguments. We briefly describe one of them. Let us see when the non-linear term can be considered as a perturbation for data of the form

$$u_0(x) \sim |x|^{-\alpha} \tag{215}$$

near the origin. The function $u_0$ is of course vector-valued, the meaning of the notation 215 is that $u_0$ behaves as a $(-\alpha)$-homogeneous function. Then

$$\Delta u_0 \sim |x|^{-\alpha-2}\,, \qquad u_0 \nabla u_0 \sim |x|^{-2\alpha-1}\,. \tag{216}$$

We see that if $\alpha < 1$near the origin $\Delta u_0$ can be expected to be dominant (in some sense), and the non-linear term can be considered as a perturbation, which makes the perturbation theory feasible. This is the sub-critical case. On the other hand, for $\alpha > 1$ the perturbation theory is not feasible, as the non-linear term can dominate. This is the super-critical case. The case $\alpha = 1$ is critical. In that case one should look at

$$u_0 \sim a|x|^{-1}\,, \tag{217}$$

and one has

$$\Delta u_0 \sim a|x|^{-3}\,, \qquad u_0 \nabla u_0 \sim a^2|x|^{-3}\,. \tag{218}$$

Therefore when $a$ is small we should be in a perturbative regime, whereas when $a$ is large we are in the non-perturbative (or "large solutions") regime. This heuristic works very well, it captures more than what one might expect from the not-so-sophisticated argument.

The perturbation theory always needs that some small quantity. For example, even when we have a function $u_0 \in L^3$ which is not small, there is a "hidden" smallness quantity around: the integrals

$$\int_{B_r} |u|^3\,dx \tag{219}$$

(where $B_r$ represents balls of radius $r$) are small when $r$ is small. By rescaling $u$ to $ru_0(rx)$ we will have the same statement with $r = 1$ for the re-scaled function. This is essentially what makes it possible to prove the short-time existence for large $L^3$ data.

By contrast, the space $(BMO)^{-1}$, the Morrey space $M^{2,1}$ with the norm given by

$$\sup_{x,r} \frac{1}{r} \int_{B_{x,r}} |u_0(y)|^2\,dy\,, \tag{220}$$

or the Lorentz space $L^{3,\infty}$ (the weak $L^3$-space), are examples of spaces where functions do not behave in this way. The spaces $(BMO)^{-1}$ , $M^{1,2}$ and $L^{3,\infty}$ contain $(-1)$-homogeneous functions $u_0$ smooth away from the origin. For such functions we of course have

$$|u_0(x)| \sim a|x|^{-1} \tag{221}$$

and the scaling $u_0 \to u_{0\lambda}$ leaves the function invariant. There does not seem to be any "hidden" smallness condition which would be useful for the Navier-Stokes perturbation theory, unless the coefficient $a$ is already small. In that case one can indeed quite easily establish existence and uniqueness via the Picard iteration for functions $u_0$ which are $(-1)$-homogeneous and smooth away from the origin. Once $a$ is small, one can work in the spaces above, or use an even simpler space given by the norm

$$\sup_{x\in\mathbf{R}^3, t>0} |\sqrt{|x|^2+t}||u(x,t)|\,. \tag{222}$$

We see that there are essentially two type of critical spaces. One type is represented by $\dot{H}^{\frac{1}{2}}, L^3$ or certain Besov spaces. With any function in these spaces one can associate a small quantity ( related to a "uniform continuity condition") useful for the Navier-Stokes theory, one can prove local-in-time well-posedness results for any function in the space.

The other type are the space which contain $(-1)$-homogeneous functions, where the perturbation method works only for functions with a small norm.

We will argue that this not just a technical point, but it indeed reflects the behavior of the actual solutions of the Navier-Stokes equations.

## *4.3 Scale-Invariant Solutions for Large Data*

Let us take a $(-1)$-homogeneous div-free vector field $w_0(x)$ which is smooth away from the origin, and let us look at the Cauchy problem (208), (209) with

$$u_0 = u_0^{(\kappa)} = \kappa w_0(x)\,. \tag{223}$$

As we have just discussed, for small $\kappa$ we are in the range when the perturbation theory can be applied and we have no problem proving existence and uniqueness of the solutions to the Cauchy problem (in appropriate classes of functions). As the initial condition $u_0$ is scale invariant and we have uniqueness, the solution itself must be scale-invariant. This means that it is of the form

$$u(x,t) = \frac{1}{\sqrt{t}}\, U\left(\frac{x}{\sqrt{t}}\right)\,. \tag{224}$$

The profile function $U$ satisfies the following elliptic equations

$$\begin{aligned} -\Delta U - \tfrac{1}{2} x\nabla U - \tfrac{1}{2} U + U\nabla U + \nabla P = 0\,,\\ \text{div } U = 0 \end{aligned} \tag{225}$$

in the space $\mathbf{R}^3$, and the "boundary condition"

$$U(x) = u_0(x) + o(|x|^{-1}), \quad |x| \to \infty. \tag{226}$$

What happens as the parameter $\kappa$ is increased? A natural approach is to try to establish the existence of the solutions of elliptic problem (225) with the boundary conditions (226).

It is clear that to solve the elliptic problem for large data, we have to go beyond the perturbation theory. In the case of the usual steady Navier-Stokes in a bounded domain, where (226) is replaced by a Dirichlet boundary condition, the problem was solved by Leray by using the degree theory. The non-perturbative argument which is used in that approach is interesting even in the case of finite-dimensional equations. The following statement for $\mathbf{R}^n$ captures its main point.

**Lemma 1 (Leray's Argument for Steady Solutions, Finite-Dimensional Version)** *Let $f\colon \mathbf{R}^n \to \mathbf{R}^n$ be a continuous function such that*

$$(x, f(x)) = 0, \quad x \in \mathbf{R}^n, \tag{227}$$

*where $(\cdot, \cdot)$ denotes the usual scalar product. Let $y \in \mathbf{R}^n$. Then the equation*

$$x + f(x) = y \tag{228}$$

*has at least one solution in $\mathbf{R}^n$ and, moreover, every solution $x$ of the equation satisfies*

$$|x| \le |y|. \tag{229}$$

The statement is interesting even for $n = 2$, in which case its validity should be heuristically clear from suitable pictures. The proof of the general case (for any dimension) requires some non-trivial topological argument. One use for example Browder's degree or, alternatively, Sard's theorem (after approximating $f$ by a smooth function) applied to the function $F\colon \mathbf{R}^n \times \mathbf{R} \to \mathbf{R}$

$$F(x, \lambda) = x + \lambda f(x). \tag{230}$$

A more general version of the lemma can be obtained by considering a function $F(x, \lambda)$ from $\mathbf{R}^n \times [0, 1]$ such that $F(x, 0) = x$. If the equation

$$F(x, \lambda) = y \tag{231}$$

has no solution with $|x| \ge r > |y|$ for any $\lambda \in [0, 1]$, then the equation

$$F(x, 1) = y \tag{232}$$

has a solution $x$ with $|x| \leq r$. From Sard's theorem it is not hard to see that when $F$ is smooth, then—in the above situation—there will be a smooth curve joining the solution $x^{(0)}$ of $F(x, 0) = y$ and our solution $x^{(1)}$ of (231).

However, the curve of solutions can potentially "turn back" a few times before it reaches the a point with $\lambda = 0$. In other words, the curve may not be a graph of a function $\lambda \to x(\lambda)$.

It turns out one can use Leray's strategy to solve (225), (226). The most difficult step is to obtain good a-priori estimates, analogous to the requirement that the Eq. (231) has no solutions with $|x| \geq r$.

This strategy is implemented in [28], where the following results is proved.

**Theorem 1 ([28])** *For any $(-1)$-homogeneous divergence free vector field $u_0$ which is locally Hölder continuous away from the origin, the Cauchy problem (208), (209) has at least one scale-invariant solution u.*

The main point of course is that there are no smallness assumptions.

## 4.4 Possible Non-uniqueness of Leray-Hopf Solutions

If the elliptic problem (225), (226) has more than one solution $U$ the Cauchy problem (208), (209) will also have more than one solution. Notice that the non-uniqueness arises immediately—the two solutions will be different at any positive time, although their limit as $t \to 0$ will be the same. At the spectral level the non-uniqueness will be detectable as follows. We again consider the family $u_0^{(\kappa)}$ given by (223). For small $\kappa$ we have a unique curve of solutions $U^{(\kappa)}$ of (225), (226) (where we take $u_0 = u_0^{(\kappa)}$). Let $L_\kappa$ be the linearization at $U^\kappa$ of the Eq. (225), with zero boundary conditions. In the finite-dimensional setting, the "turning" of the curve $U_{(\kappa)}$ (when is just stops being a graph of a function of $\kappa$) is characterized (in the generic case) by an eigenvalue of $L_\kappa$ crossing the imaginary axis at zero. (For small $\kappa$ the spectrum of $L_\kappa$ is on the left of the imaginary axis.) If we extent our class of solutions to a larger class of *discretely self-similar solutions* (where the invariance under scaling is only required for a discrete set $\lambda_0^n$, $n = 1, 2, \ldots$, rather than all $\lambda > 0$), any crossing of the imaginary axis (even when it is away from zero) will cause (under some natural technical assumptions) non-uniqueness of the solutions of the Cauchy problem.

The non-uniqueness in the class of invariant or discretely invariant solutions is not directly applicable to Leray-Hopf solutions, due to the slow decay of the fields when $|x| \to \infty$. However, it turns out that one can perform a truncation procedure and—assuming the spectrum will cross the imaginary line as $\kappa$ increases—achieve the same phenomenon starting from compactly supported data with one-point singularity near the origin where the datum grows as $\sim |x|^{-1}$.

The proofs of these statement are technically non-trivial and can be found in [28]. We conjecture that the spectral condition will be satisfied for certain solutions, and

therefore we believe that we have non-uniqueness. If that is the case, then, quite remarkably, the simple heuristic arguments and analogies in Sect. 4.2 would capture the real behavior of the equation, and the Navier-Stokes equations would only be well-posed in the spaces where the classical perturbation theory works.

**Acknowledgements** The author thanks Sergei Kuksin, Geordie Richards and Ofer Zeitouni for very helpful discussions. The research of was supported in part by grants DMS 1159376 and DMS 1362467 from the National Science Foundation.

## References

1. R.V. Abramov, G. Kovacic, A.J. Majda, Hamiltonian structure and statistically relevant conserved quantities for the truncated Burgers-Hopf equation. Commun. Pure Appl. Math. **56**(1), 1–46 (2003)
2. V.I. Arnold, B.A. Khesin, *Topological Methods in Hydrodynamics*. Applied Mathematical Sciences, vol. 125 (Springer, New York, 1998)
3. J. Bedrossian, N. Masmoudi, Inviscid damping and the asymptotic stability of planar shear flows in the 2D Euler equations. Publ. Math. Inst. Hautes Etudes Sci. **122**, 195–300 (2015)
4. G. Benettin, A. Ponno, Time-scales to equipartition in the Fermi-Pasta-Ulam problem: finite-size effects and thermodynamic limit. J. Stat. Phys. **144**(4), 793–812 (2011)
5. F. Bouchet, A. Venaille, Statistical mechanics of two-dimensional and geophysical flows. Phys. Rep. **515**(5), 227–295 (2012)
6. J. Bourgain, *Global Solutions of Nonlinear Schrödinger Equations*. American Mathematical Society Colloquium Publications, vol. 46 (American Mathematical Society, Providence, 1999)
7. J. Bourgain, Problems in Hamiltonian PDE's. GAFA 2000 (Tel Aviv, 1999). Geom. Funct. Anal. Special Volume, Part I (Birkhäuser, Basel, 2000), pp. 32–56
8. A. Biryuk, On invariant measures of the 2D Euler equation. J. Stat. Phys. **122**(4), 597–616 (2006)
9. S. Chatterjee, Invariant measures and the soliton resolution conjecture. Commun. Pure Appl. Math. **67**(11), 1737–1842 (2014)
10. S. Chatterjee, P. Diaconis, Fluctuations of the Bose-Einstein condensate. J. Phys. A **47**(8), 085201, 23 p. (2014)
11. L. Chierchia, J. You, KAM tori for 1D nonlinear wave equations with periodic boundary conditions. Commun. Math. Phys. **211**(2), 497–525 (2000)
12. A. Choffrut, V. Sverak, Local structure of the set of steady-state solutions to the 2D incompressible Euler equations. Geom. Funct. Anal. **22**(1), 136–201 (2012)
13. J. Colliander, M. Keel, G. Staffilani, H. Takaoka, T. Tao, Almost conservation laws and global rough solutions to a nonlinear Schrödinger equation. Math. Res. Lett. **9**(5–6), 659–682 (2002)
14. J. Colliander, M. Keel, G. Staffilani, H. Takaoka, T. Tao, Transfer of energy to high frequencies in the cubic defocusing nonlinear Schrödinger equation. Invent. Math. **181**(1), 39–113 (2010)
15. A. Dembo, O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd edn. Applications of Mathematics (New York), vol. 38 (Springer, New York, 1998)
16. P. Diaconis, D. Freedman, A dozen de Finetti-style results in search of a theory. Ann. Inst. H. Poincar Probab. Stat. **23**(2, suppl.), 397–423 (1987)
17. R.S. Ellis, *The Theory of Large Deviations and Applications to Statistical Mechanics*. Long-Range Interacting Systems, vol. 13 (Oxford University Press, Oxford, 2010), pp. 228–277
18. R.S. Ellis, R. Jordan, P. Otto, B. Turkington, A statistical approach to the asymptotic behavior of a class of generalized nonlinear Schrödinger equations. Commun. Math. Phys. **244**(1), 187–208 (2004)

19. G.L. Eyink, H. Spohn, Negative-temperature states and large-scale, long-lived vortices in two-dimensional turbulence. J. Stat. Phys. **70**(3–4), 833–886 (1993)
20. L.D. Faddeev, L. Takhtajan, *Hamiltonian Methods in the Theory of Solitons* (Springer, Berlin, 1987)
21. H. Federer, *Geometric Measure Theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153 (Springer, New York, 1969)
22. E. Fermi, J. Pasta, S. Ulam, Studies of non linear problems, Los-Alamos internal report, Document LA-1940 (1955), in: Enrico Fermi Collected Papers, vol. II (The University of Chicago Press/Accademia Nazionale dei Lincei, Chicago/Roma, 1965), pp. 977–988
23. Z. Hani, B. Pausader, N. Tzvetkov, N. Visciglia, Modified scattering for the cubic Schrödinger equation on product spaces and applications. Forum Math. Pi **3**, e4, 63 pp. (2015)
24. E. Hopf, Über die Anfangswertaufgabe für die hydrodynamischen Grundgleichungen. Math. Nachr. **4**, 213–231 (1951)
25. L. Hörmander, *The Analysis of Linear Partial Differential Operators. I*. Distribution Theory and Fourier Analysis (Springer, Berlin, 1990)
26. A. Izosimov, B. Khesin, Characterization of steady solutions to the 2D Euler equation, arXiv:1511.05623
27. H. Jia, V. Sverak, Local-in-space estimates near initial time for weak solutions of the Navier-Stokes equations and forward self-similar solutions. Invent. Math. **196**(1), 233–265 (2014)
28. H. Jia, V. Sverak, Are the incompressible 3d Navier-Stokes equations locally ill-posed in the natural energy space? J. Funct. Anal. **268**(12), 3734–3766 (2015)
29. T. Kato, Strong $L^p$-solutions of the Navier-Stokes equation in Rm, with applications to weak solutions. Math. Z. **187**(4), 471–480 (1984)
30. C.E. Kenig, G. Ponce, L. Vega, Quadratic forms for the 1-D semilinear Schrödinger equation. Trans. Am. Math. Soc. **348**(8), 3323–3353 (1996)
31. A. Kiselev, V. Sverak, Small scale creation for solutions of the incompressible two-dimensional Euler equation. Ann. Math. (2) **180**(3), 1205–1220 (2014)
32. R. Killip, M. Vişan, *Nonlinear Schrödinger Equations at Critical Regularity*. Evolution Equations. Clay Mathematics Proceedings, vol. 17 (American Mathematical Society, Providence, 2013), pp. 325–437
33. H. Koch, D. Tataru, Well-posedness for the Navier-Stokes equations. Adv. Math. **157**(1), 22–35 (2001)
34. R.H. Kraichnan, Inertial ranges in two dimensional turbulence. Phys. Fluids **10**(7), 1417–1423 (1967)
35. S.B. Kuksin, *Nearly Integrable Infinite-Dimensional Hamiltonian Systems*. Lecture Notes in Mathematics (Springer, Berlin, 1556)
36. O.A. Ladyzhenskaya, Example of non-uniqueness in the Hopf class of weak solutions for the Navier Stokes equations. Izv. Ross. Akad. Nauk Ser. Mat. **33**(1), 229–236 (1969)
37. J.L. Lebowitz, H.A. Rose, E.R. Speer, Statistical mechanics of the nonlinear Schrödinger equation. J. Stat. Phys. **50**(3–4), 657–687 (1988)
38. J. Leray, Sur le mouvement d'un liquide visqueux emplissant l'espace. Acta Math. **63**, 193–248 (1934)
39. B.V. Lidskij, E.I. Shulman, Periodic solutions of the equation $u_{tt} - u_{xx} + u^3 = 0$. Funct. Anal. Appl. **22**, 332–333 (1988)
40. A.J. Majda, A.L. Bertozzi, *Vorticity and Incompressible Flow*. Cambridge Texts in Applied Mathematics, vol. 27 (Cambridge University Press, Cambridge, 2002)
41. A.J. Majda, X. Wang, *Non-linear Dynamics and Statistical Theories for Basic Geophysical Flows* (Cambridge University Press, Cambridge, 2006)
42. J. Marsden, A. Weinstein, Coadjoint orbits, vortices, and Clebsch variables for incompressible fluids. Order in chaos (Los Alamos, N.M., 1982). Phys. D **7**(1–3), 305–323 (1983)
43. J. Miller, Statistical mechanics of Euler equations in two dimensions. Phys. Rev. Lett. **65**(17), 2137–2140 (1990)
44. C. Mouhot, C. Villani, On Landau damping. Acta Math. **207**(1), 29–201 (2011)

45. S.P. Novikov, The Hamiltonian formalism and a multivalued analogue of Morse theory. Uspekhi Mat. Nauk **37**(5(227)), 3–49, 248 (1982)
46. L. Onsager, Statistical hydrodynamics. Nuovo Cimento (9) **6**(Supplemento, 2) (Convegno Internazionale di Meccanica Statistica), 279–287 (1949)
47. C.W. Oseen, Sur let formules de Green généralisées qui se présentent dans l'hydrodynamique et sur quelques-unes de leus application. Acta Math. **34**(1), 205–284 (1911)
48. G. Richards, V. Sverak, O. Zeitouni, in preparation.
49. R. Robert, J. Sommeria, Statistical equilibrium states for two-dimensional flows. J. Fluid Mech. **229**, 291–310 (1991)
50. A.I. Shnirelman, Lattice theory and flows of ideal incompressible fluid. Russ. J. Math. Phys. **1**(1), 105–114 (1993)
51. V. Sverak, Selected Topic in Fluid Mechanics. Online course notes, http://www.math.umn.edu/~sverak/course-notes2011.pdf
52. T. Tao's notes on NLW well-posedness, http://www.math.ucla.edu/~tao/Dispersive/wave.html
53. B. Turkington, Statistical equilibrium measures and coherent states in two-dimensional turbulence. Commun. Pure Appl. Math. **52**(7), 781–809 (1999)
54. S.R.S. Varadhan, Online course notes on large deviations, https://www.math.nyu.edu/faculty/varadhan/LDP.html
55. X. Yuan, Quasi-periodic solutions of completely resonant nonlinear wave equations. J. Differ. Equ. **230**(1), 213–274 (2006)
56. V.I. Yudovich, Non-stationary flows of an ideal incompressible fluid. Ž. Vyčisl. Mat. i Mat. Fiz. **3**, 1032–1066 (1963)
57. V.E. Zakharov, V.S. L'vov, G. Falkovich, *Kolmogorov Spectra of Turbulence I* (Springer, Berlin, 1992)
58. V. Zeitlin, Finite-mode analogs of 2D ideal hydrodynamics: coadjoint orbits and local canonical structure. Phys. D **49**(3), 353–362 (1991)

*LECTURE NOTES IN MATHEMATICS* Springer

Editors in Chief: J.-M. Morel, B. Teissier;

**Editorial Policy**

1. Lecture Notes aim to report new developments in all areas of mathematics and their applications – quickly, informally and at a high level. Mathematical texts analysing new developments in modelling and numerical simulation are welcome.

   Manuscripts should be reasonably self-contained and rounded off. Thus they may, and often will, present not only results of the author but also related work by other people. They may be based on specialised lecture courses. Furthermore, the manuscripts should provide sufficient motivation, examples and applications. This clearly distinguishes Lecture Notes from journal articles or technical reports which normally are very concise. Articles intended for a journal but too long to be accepted by most journals, usually do not have this "lecture notes" character. For similar reasons it is unusual for doctoral theses to be accepted for the Lecture Notes series, though habilitation theses may be appropriate.

2. Besides monographs, multi-author manuscripts resulting from SUMMER SCHOOLS or similar INTENSIVE COURSES are welcome, provided their objective was held to present an active mathematical topic to an audience at the beginning or intermediate graduate level (a list of participants should be provided).

   The resulting manuscript should not be just a collection of course notes, but should require advance planning and coordination among the main lecturers. The subject matter should dictate the structure of the book. This structure should be motivated and explained in a scientific introduction, and the notation, references, index and formulation of results should be, if possible, unified by the editors. Each contribution should have an abstract and an introduction referring to the other contributions. In other words, more preparatory work must go into a multi-authored volume than simply assembling a disparate collection of papers, communicated at the event.

3. Manuscripts should be submitted either online at www.editorialmanager.com/lnm to Springer's mathematics editorial in Heidelberg, or electronically to one of the series editors. Authors should be aware that incomplete or insufficiently close-to-final manuscripts almost always result in longer refereeing times and nevertheless unclear referees' recommendations, making further refereeing of a final draft necessary. The strict minimum amount of material that will be considered should include a detailed outline describing the planned contents of each chapter, a bibliography and several sample chapters. Parallel submission of a manuscript to another publisher while under consideration for LNM is not acceptable and can lead to rejection.

4. In general, **monographs** will be sent out to at least 2 external referees for evaluation.

   A final decision to publish can be made only on the basis of the complete manuscript, however a refereeing process leading to a preliminary decision can be based on a pre-final or incomplete manuscript.

   Volume Editors of **multi-author works** are expected to arrange for the refereeing, to the usual scientific standards, of the individual contributions. If the resulting reports can be

forwarded to the LNM Editorial Board, this is very helpful. If no reports are forwarded or if other questions remain unclear in respect of homogeneity etc, the series editors may wish to consult external referees for an overall evaluation of the volume.

5. Manuscripts should in general be submitted in English. Final manuscripts should contain at least 100 pages of mathematical text and should always include
   - a table of contents;
   - an informative introduction, with adequate motivation and perhaps some historical remarks: it should be accessible to a reader not intimately familiar with the topic treated;
   - a subject index: as a rule this is genuinely helpful for the reader.
   - For evaluation purposes, manuscripts should be submitted as pdf files.

6. Careful preparation of the manuscripts will help keep production time short besides ensuring satisfactory appearance of the finished book in print and online. After acceptance of the manuscript authors will be asked to prepare the final LaTeX source files (see LaTeX templates online: https://www.springer.com/gb/authors-editors/book-authors-editors/manuscriptpreparation/5636) plus the corresponding pdf- or zipped ps-file. The LaTeX source files are essential for producing the full-text online version of the book, see http://link.springer.com/bookseries/304 for the existing online volumes of LNM). The technical production of a Lecture Notes volume takes approximately 12 weeks. Additional instructions, if necessary, are available on request from lnm@springer.com.

7. Authors receive a total of 30 free copies of their volume and free access to their book on SpringerLink, but no royalties. They are entitled to a discount of 33.3 % on the price of Springer books purchased for their personal use, if ordering directly from Springer.

8. Commitment to publish is made by a *Publishing Agreement*; contributing authors of multiauthor books are requested to sign a *Consent to Publish form*. Springer-Verlag registers the copyright for each volume. Authors are free to reuse material contained in their LNM volumes in later publications: a brief written (or e-mail) request for formal permission is sufficient.

**Addresses:**
Professor Jean-Michel Morel, CMLA, École Normale Supérieure de Cachan, France
E-mail: moreljeanmichel@gmail.com

Professor Bernard Teissier, Equipe Géométrie et Dynamique,
Institut de Mathématiques de Jussieu – Paris Rive Gauche, Paris, France
E-mail: bernard.teissier@imj-prg.fr

Springer: Ute McCrory, Mathematics, Heidelberg, Germany,
E-mail: lnm@springer.com