International Mathematical Series • Volume 8

SOBOLEV SPACES IN MATHEMATICS I

Sobolev Type Inequalities

Vladimir Maz'ya
EDITOR

SOBOLEV SPACES IN MATHEMATICS I

SOBOLEV TYPE INEQUALITIES

INTERNATIONAL MATHEMATICAL SERIES

1. **Nonlinear Problems in Mathematical Physics and Related Topics I. In Honor of Professor O.A. Ladyzhenskaya** • M.Sh. Birman, S. Hildebrandt, V.A. Solonnikov, N.N. Uraltseva Eds. • 2002
2. **Nonlinear Problems in Mathematical Physics and Related Topics II. In Honor of Professor O.A. Ladyzhenskaya** • M.Sh. Birman, S. Hildebrandt, V.A. Solonnikov, N.N. Uraltseva Eds. • 2003
3. **Different Faces of Geometry** • S. Donaldson, Ya. Eliashberg, M. Gromov Eds. • 2004
4. **Mathematical Problems from Applied Logic I. Logics for the XXIst Century** • D. Gabbay, S. Goncharov, M. Zakharyaschev Eds. • 2006
5. **Mathematical Problems from Applied Logic II. Logics for the XXIst Century** • D. Gabbay, S. Goncharov, M. Zakharyaschev Eds. • 2007
6. **Instability in Models Connected with Fluid Flows. I** • C. Bardos, A. Fursikov Eds. • 2008
7. **Instability in Models Connected with Fluid Flows. II** • C. Bardos, A. Fursikov Eds. • 2008
8. **Sobolev Spaces in Mathematics I. Sobolev Type Inequalities** • V. Maz'ya Ed. • 2009
9. **Sobolev Spaces in Mathematics II. Applications in Analysis and Partial Differential Equations** • V. Maz'ya Ed. • 2009
10. **Sobolev Spaces in Mathematics III. Applications in Mathematical Physics** • V. Isakov Ed. • 2009

SOBOLEV SPACES IN MATHEMATICS I

Sobolev Type Inequalities

Editor: **Vladimir Maz'ya**

Ohio State University, USA
University of Liverpool, UK
Linköping University, SWEDEN

Springer
Tamara Rozhkovskaya Publisher

Editor

Prof. Vladimir Maz'ya
Ohio State University
Department of Mathematics
Columbus, USA

University of Liverpool
Department of Mathematical Sciences
Liverpool, UK

Linköping University
Department of Mathematics
Linköping, Sweden

This series was founded in 2002 and is a joint publication of Springer and "Tamara Rozhkovskaya Publisher." Each volume presents contributions from the Volume Editors and Authors exclusively invited by the Series Editor Tamara Rozhkovskaya who also prepares the Camera Ready Manuscript. This volume is distributed by "Tamara Rozhkovskaya Publisher" (tamara@mathbooks.ru) in Russia and by Springer over all the world.

ISBN 978-0-387-85647-6 e-ISBN 978-0-387-85648-3
ISBN 978-5-901873-24-3 (Tamara Rozhkovskaya Publisher)
ISSN 1571-5485

Library of Congress Control Number: 2008937494

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com

To the memory of
Sergey L'vovich Sobolev
on the occasion of his centenary

Main Topics

Sobolev's discoveries of the 1930's have a strong influence on development of the theory of partial differential equations, analysis, mathematical physics, differential geometry, and other fields of mathematics. The three-volume collection *Sobolev Spaces in Mathematics* presents the latest results in the theory of Sobolev spaces and applications from leading experts in these areas.

I. Sobolev Type Inequalities
In 1938, exactly 70 years ago, the original Sobolev inequality (an embedding theorem) was published in the celebrated paper by S.L. Sobolev "On a theorem of functional analysis." By now, the Sobolev inequality and its numerous versions continue to attract attention of researchers because of the central role played by such inequalities in the theory of partial differential equations, mathematical physics, and many various areas of analysis and differential geometry. The volume presents the recent study of different Sobolev type inequalities, in particular, inequalities on manifolds, Carnot–Carathéodory spaces, and metric measure spaces, trace inequalities, inequalities with weights, the sharpness of constants in inequalities, embedding theorems in domains with irregular boundaries, the behavior of maximal functions in Sobolev spaces, etc. Some unfamiliar settings of Sobolev type inequalities (for example, on graphs) are also discussed. The volume opens with the survey article "My Love Affair with the Sobolev Inequality" by David R. Adams.

II. Applications in Analysis and Partial Differential Equations
Sobolev spaces become the established language of the theory of partial differential equations and analysis. Among a huge variety of problems where Sobolev spaces are used, the following important topics are in the focus of this volume: boundary value problems in domains with singularities, higher order partial differential equations, nonlinear evolution equations, local polynomial approximations, regularity for the Poisson equation in cones, harmonic functions, inequalities in Sobolev–Lorentz spaces, properties of function spaces in cellular domains, the spectrum of a Schrödinger operator with negative potential, the spectrum of boundary value problems in domains with cylindrical and quasicylindrical outlets to infinity, criteria for the complete integrability of systems of differential equations with applications to differential geometry, some aspects of differential forms on Riemannian manifolds related to the Sobolev inequality, a Brownian motion on a Cartan–Hadamard manifold, etc. Two short biographical articles with unique archive photos of S.L. Sobolev are also included.

III. Applications in Mathematical Physics

The mathematical works of S.L. Sobolev were strongly motivated by particular problems coming from applications. The approach and ideas of his famous book "Applications of Functional Analysis in Mathematical Physics" of 1950 turned out to be very influential and are widely used in the study of various problems of mathematical physics. The topics of this volume concern mathematical problems, mainly from control theory and inverse problems, describing various processes in physics and mechanics, in particular, the stochastic Ginzburg–Landau model with white noise simulating the phenomenon of superconductivity in materials under low temperatures, spectral asymptotics for the magnetic Schrödinger operator, the theory of boundary controllability for models of Kirchhoff plate and the Euler–Bernoulli plate with various physically meaningful boundary controls, asymptotics for boundary value problems in perforated domains and bodies with different type defects, the Finsler metric in connection with the study of wave propagation, the electric impedance tomography problem, the dynamical Lamé system with residual stress, etc.

Contents

I. Sobolev Type Inequalities

Vladimir Maz'ya Ed.

II. Applications in Analysis and Partial Differential Equations

Vladimir Maz'ya Ed.

III. Applications in Mathematical Physics

Victor Isakov Ed.

Contributors

Editors

Vladimir Maz'ya

Ohio State University
Columbus, OH 43210
USA

University of Liverpool
Liverpool L69 7ZL
UK

Linköping University
Linköping SE-58183
SWEDEN

vlmaz@mai.liu.se
vlmaz@math.ohio-state.edu

Victor Isakov

Wichita State University
Wichita, KS 67206
USA

victor.isakov@wichita.edu

Contributors

Authors

Daniel Aalto
Institute of Mathematics
Helsinki University of Technology
P.O. Box 1100, FI-02015
FINLAND
e-mail: daniel.aalto@tkk.fi

David R. Adams
University of Kentucky
Lexington, KY 40506-0027
USA
e-mail: dave@ms.uky.edu

Hiroaki Aikawa
Hokkaido University
Sapporo 060-0810
JAPAN
e-mail: aik@math.sci.hokudai.ac.jp

Vasili Babich
Steklov Mathematical Institute
Russian Academy of Sciences
27 Fontanka Str., St.-Petersburg 191023
RUSSIA
e-mail: babich@pdmi.ras.ru

Mikhail Belishev
Steklov Mathematical Institute
Russian Academy of Sciences
27 Fontanka Str., St.-Petersburg 191023
RUSSIA
e-mail: belishev@pdmi.ras.ru

Sergey Bobkov
University of Minnesota
Minneapolis, MN 55455
USA
e-mail: bobkov@math.umn.edu

Yuri Brudnyi

Technion – Israel Institute of Technology
Haifa 32000
ISRAEL
e-mail: ybrudnyi@math.technion.ac.il

Victor Burenkov

Università degli Studi di Padova
63 Via Trieste, 35121 Padova
ITALY
e-mail: burenkov@math.unipd.it

Andrea Cianchi

Università di Firenze
Piazza Ghiberti 27, 50122 Firenze
ITALY
e-mail: cianchi@unifi.it

Serban Costea

McMaster University
1280 Main Street West
Hamilton, Ontario L8S 4K1
CANADA
e-mail: secostea@math.mcmaster.ca

Stephan Dahlke

Philipps–Universität Marburg
Fachbereich Mathematik und Informatik
Hans Meerwein Str., Lahnberge 35032 Marburg
GERMANY
e-mail: dahlke@mathematik.uni-marburg.de

Donatella Danielli

Purdue University
150 N. University Str.
West Lafayette, IN 47906
USA
e-mail: danielli@math.purdue.edu

David E. Edmunds

School of Mathematics Cardiff University
Senghennydd Road CARDIFF
Wales CF24 4AG
UK
e-mail: davideedmunds@aol.com

W. Desmond Evans

School of Mathematics Cardiff University
Senghennydd Road CARDIFF
Wales CF24 4AG
UK
e-mail: EvansWD@cf.ac.uk

Andrei Fursikov

Moscow State University
Vorob'evy Gory, Moscow 119992
RUSSIA

e-mail: fursikov@mtu-net.ru

Victor Galaktionov

University of Bath
Bath, BA2 7AY
UK

e-mail: vag@maths.bath.ac.uk

Nicola Garofalo

Purdue University
150 N. University Str.
West Lafayette, IN 47906
USA

e-mail: garofalo@math.purdue.edu

Friedrich Götze

Bielefeld University
Bielefeld 33501
GERMANY

e-mail: goetze@math.uni-bielefeld.de

Vladimir Gol'dshtein

Ben Gurion University of the Negev
P.O.B. 653, Beer Sheva 84105
ISRAEL

e-mail: vladimir@bgu.ac.il

Alexander Grigor'yan

Bielefeld University
Bielefeld 33501
GERMANY

e-mail: grigor@math.uni-bielefeld.de

Max Gunzburger

Florida State University
Tallahassee, FL 32306-4120
USA

e-mail: gunzburg@scs.fsu.edu

Piotr Hajłasz

University of Pittsburgh
301 Thackeray Hall, Pittsburgh, PA 15260
USA

e-mail: hajlasz@pitt.edu

Elton Hsu

Northwestern University
2033 Sheridan Road, Evanston, IL 60208-2730
USA

e-mail: ehsu@math.northwestern.edu

Victor Isakov

Wichita State University
Wichita, KS 67206
USA

e-mail: victor.isakov@wichita.edu

Victor Ivrii

University of Toronto
40 St.George Str., Toronto, Ontario M5S 2E4
CANADA

e-mail: ivrii@math.toronto.edu

Tünde Jakab

University of Virginia
Charlottesville, VA 22904
USA

e-mail: tj8y@virginia.edu

Nanhee Kim

Wichita State University
Wichita, KS 67206
USA

e-mail: kim@math.wichita.edu

Juha Kinnunen

Institute of Mathematics
Helsinki University of Technology
P.O. Box 1100, FI-02015
FINLAND

e-mail: juha.kinnunen@tkk.fi

Pier Domenico Lamberti

Universitá degli Studi di Padova
63 Via Trieste, 35121 Padova
ITALY

e-mail: lamberti@math.unipd.it

Irena Lasiecka

University of Virginia
Charlottesville, VA 22904
USA

e-mail: il2v@virginia.edu

Vladimir Maz'ya

Ohio State University
Columbus, OH 43210
USA

University of Liverpool
Liverpool L69 7ZL
UK
Linköping University
Linköping SE-58183
SWEDEN
e-mail: vlmaz@mai.liu.se
e-mail: vlmaz@math.ohio-state.edu

Enzo Mitidieri

Università di Trieste
Via Valerio 12/1, 34127 Trieste
ITALY
e-mail: mitidier@units.it

Irina Mitrea

University of Virginia
Charlottesville, VA 22904
USA
e-mail: im3p@virginia.edu

Marius Mitrea

University of Missouri
Columbia, MO
USA
e-mail: marius@math.missouri.edu

Alexander Movchan

University of Liverpool
Liverpool L69 3BX
UK
e-mail: abm@liverpool.ac.uk

Sergey Nazarov

Institute of Problems in Mechanical Engineering
Russian Academy of Sciences
61, Bolshoi pr., V.O., St.-Petersburg 199178
RUSSIA
e-mail: serna@snark.ipme.ru

Janet Peterson

Florida State University
Tallahassee FL 32306-4120
USA
e-mail: peterson@scs.fsu.edu

Nguyen Cong Phuc

Purdue University
150 N. University Str.
West Lafayette, IN 47906
USA
e-mail: pcnguyen@math.purdue.edu

Luboš Pick

Charles University
Sokolovská 83, 186 75 Praha 8
CZECH REPUBLIC

e-mail: pick@karlin.mff.cuni.cz

Yehuda Pinchover

Technion – Israel Institute of Technology
Haifa 32000
ISRAEL

e-mail: pincho@techunix.technion.ac.il

Stanislav Pokhozhaev

Steklov Mathematical Institute
Russian Academy of Sciences
8, Gubkina Str., Moscow 119991
RUSSIA

e-mail: pokhozhaev@mi.ras.ru

Yuri Reshetnyak

Sobolev Institute of Mathematics
Siberian Branch
Russian Academy of Sciences
4, Pr. Koptyuga, Novosibirsk 630090
RUSSIA

Novosibirsk State University
2, Pirogova Str., Novosibirsk 630090
RUSSIA

e-mail: Reshetnyak@math.nsc.ru

Grigori Rozenblum

University of Gothenburg
S-412 96, Gothenburg
SWEDEN

e-mail: grigori@math.chalmers.se

Laurent Saloff-Coste

Cornell University
Mallot Hall, Ithaca, NY 14853
USA

e-mail: lsc@math.cornell.edu

Nageswari Shanmugalingam

University of Cincinnati
Cincinnati, OH 45221-0025
USA

e-mail: nages@math.uc.edu

Tatyana Shaposhnikova

Ohio State University
Columbus, OH 43210
USA

Linköping University
Linköping SE-58183
SWEDEN
e-mail: tasha@mai.liu.se

Winfried Sickel

Friedrich-Schiller-Universität Jena
Mathematisches Institut
Ernst–Abbe–Platz 2, D-07740 Jena
GERMANY
e-mail: sickel@minet.uni-jena.de

Michael Solomyak

The Weizmann Institute of Science
Rehovot, 76100
ISRAEL
e-mail: michail.solomyak@weizmann.ac.il

Michael Taylor

University of North Carolina
Chapel Hill, NC 27599
USA
e-mail: met@email.unc.edu

Kyril Tintarev

Uppsala University
P.O. Box 480, SE-751 06 Uppsala
SWEDEN
e-mail: kyril.tintarev@math.uu.se

Hans Triebel

Mathematisches Institut
Friedrich-Schiller-Universität Jena
D-07737 Jena
GERMANY
e-mail: triebel@minet.uni-jena.de

Roberto Triggiani

University of Virginia
Charlottesville, VA 22904
USA
e-mail: rt7u@virginia.edu

Marc Troyanov

Institute of Geometry, Algebra, and Topology
École Polytechnique Fédérale de Lausanne
1015 Lausanne
SWITZERLAND
e-mail: marc.troyanov@epfl.ch

Sobolev Type Inequalities

Vladimir Maz'ya Ed.

Contents

My Love Affair with the Sobolev Inequality

David R. Adams

Abstract Reminiscence about different versions of the Sobolev inequality obtained by the author and others.

Due to the fact that the Sobolev Inequality is so central to much of mathematical analysis, especially to partial differential equations, it is not surprising that there are by now, 70 years after Sobolev's original paper, many different versions of the Sobolev Inequality and by many different authors. This paper is a tribute to S.L. Sobolev.

David R. Adams

On a cold December morning shortly after Christmas 1969, a small group of people were huddled against the cold near a limousine type bus parked in the middle of the main street of a very small South Dakota farming community in the central plains of the USA – my ancestral hometown. The bus, run by the Greyhound Company, was the "east-west connector bus" and the middle of Main Street was the usual passenger pick-up and drop-off spot in town. Here Main Street consisted of just one block of store front businesses and it

David R. Adams
University of Kentucky, Lexington, KY 40506-0027, USA, e-mail: dave@ms.uky.edu

V. Maz'ya (ed.), *Sobolev Spaces in Mathematics I, International Mathematical Series.*
doi: 10.1007/978-0-387-85648-3,

was wide enough for angular parking on each side and still with plenty of room to accommodate car traffic on either side of the parked bus and the waiting people. Most of these people were my relatives, gathered in town for Christmas and now taking the opportunity to see me off. This bus will meet a larger north-south bus at some highway crosswords out on the nearby prairie. I am expecting to tranfer to that bus for the next stage of my journey.

When I boarded the small bus, there was only one other passenger, an elderly woman. And as I sat down and waved a final farewell to my family, the woman turned to me and asked, “where are you going, young man?” I could have been very dramatic and responded, “into History!”, but I did not, nor did it even cross my mind to say such a thing. I just said, “to Rome, Italy”, which I am sure was dramatic enough. “My land”, she responded, “I am only going to see my sister in Minnesota, you are going a long ways.”

And I guess it was, both physically and psychologically for all concerned. And now I can confess that I was not very well prepared for my Italian sojourn. Though I did eventually adjust and adapt to life in Rome (January–August 1970), and even began to thrive there toward the end of my stay. However my budding Italian speech never broke away from my American-midwestern accent. Though if I kept my mouth shut, I eventually could pass for Italian at least in dress and demeanor. Once an American tourist stopped me on the street during my last days to ask, in English, where some place-street was located. I responded by telling her where it was and how to get there. “Wow, you speak good English!” she said. “Thank you” I replied and walked away, leaving her with the illusion.

Thus with my initial bus trip, I began my mathematical odyssey – first to the CNR in Rome as a Post Doc under the direction of Guido Stampacchia, later as an instructor at Rice University, an acting Assistant Professor at the University of California, San Diego, a visiting professor at Indiana University, and finally a Professor at the University of Kentucky – for the past 30+ years. When I left for Italy, I had just days earlier received my Ph.D. degree from the University of Minnesota under the direction of N.G. Meyers. And I began my Post Doc studies by tackling a question posed earlier by him and then at the CNR under the watchful eye of Stampacchia. As it turned out, it was this question of Meyers' that essentially started me down the road of looking at variations of the now classical Sobolev Inequality. For I did not consciously look to that direction, but as it all transpired, I have over the years returned again and again to this theme, eventually giving five or six versions of the Sobolev Inequality during my career. And due to the fact that the Sobolev Inequality is so central to much of mathematical analysis, especially to partial differential equations, it is not surprising that there are by now, 70 years after Sobolev's original paper, many different versions of the Sobolev Inequality and by many different authors. This paper is a tribute to S.L. Sobolev. I do not pretend to review all of this literature, only at best, part of my role in it. This is after all the story of my romance with the Sobolev Inequality.

The classical Sobolev Inequality that I refer to can take on one or two equivalent forms. For example, if $u(x)$ is a smooth function of compact support, then for $1 < p < n/m$, there is a constant c depending only on n, m and p such that

$$(*) \qquad ||u||_{L^q(I\mathbb{R}^n)} \leqslant c||D^m u||_{L^p(\mathbb{R}^n)},$$

where $1/q = 1/p - m/n$[1] Here $D^m u$ denotes the vector of all mth order derivatives of u. Or, if we use Riesz potentials, an equivalent form of $(*)$ is

$$(**) \qquad ||I_\alpha f||_{L^q(\mathbb{R}^n)} \leqslant c'||f||_{L^p(\mathbb{R}^n)}$$

with again $q = np/(n - \alpha p)$, where $\alpha = m$. Here I have written

$$I_\alpha f(x) = \int_{\mathbb{R}^n} |x-y|^{\alpha - n} f(y)\, dy,$$

$0 < \alpha < n$, $1 < p < n/\alpha$.

On the connection between the two, simply note $|u(x)| \leqslant cI_m\,(|D^m u|)(x)$ and singular integrals obtained by differentiating $I_m\, f$ m-times.

So now I begin the story of my involvement with the Sobolev Inequality – from my first struggle at the CNR to prove a trace inequality to my more recent work on a vanishing mean exponential integrability condition with R. Hurri-Syrjänen.

Some time lines are:

1971 – The trace inequality (Sect. 1)

1973 – An exponential trace inequality (Sect. 6)

1974 – A mixed norm inequality; with R. Bagby (Sect. 2)

1975 – A Morrey–Sobolev Inequality (Sect. 3)

1976 – A trace inequality with CSI, $q = p$ (Sect. 1)

1982 – A Morrey–Besov inequality; with J. Lewis (Sect. 4)

1988 – Exponential integrability (Sect. 5)

1998 – Estimates for $M_\alpha f$ (Sect. 7, (4))

2003 – Vanishing exponential integrability; with R. Hurri-Syrjänen (Sect. 6)

2004 – Trace estimates for Morrey–Sobolev functions; with J. Xiao (Sect. 3)

---

[1] Inequality $(*)$ also holds for $p = 1$ by the Gagliardo–Nirenberg estimates (see [55]). However, $(**)$ does not hold for $p = 1$.

1 The Trace Inequality

The struggle alluded to above was the question of finding necessary and sufficient conditions on a Borel measures μ defined on subsets of $\mathbb{R}^n$ that insures, in the language of N.G. Meyers, that μ has positive capacity. In the late 1960's, Meyers wrote a paper (unpublished, to this date) titled: *Capacities, extremal length and traces of strongly differentiable functions.* Here, to unify the ideas of capacity and extremal length, among other ideas, he defined a capacity of a set of measures on $\mathbb{R}^n$. The usual capacity of a standard subset $K \subset \mathbb{R}^n$ then reduced to taking the sets of Dirac measures $\{\delta_x\}_{x\in K}$. A simplified version of this might be

$$\mathbb{C}_{\alpha,p,q}(\mathcal{K}) = \inf\{||f||^p_{L^p(\mathbb{R}^n)} : ||I_\alpha f||_{L^q(\nu)} \geqslant 1 \ \forall \nu \in \mathcal{K} \text{ and } f \geqslant 0\},$$

where $\mathcal{K} \subset \mathcal{M}^+ =$ all Borel measures on $\mathbb{R}^n$. The question posed by Meyers was to characterize all measures μ with positive capacity $\mathbb{C}_{\alpha,p,q}(\{\mu\}) > 0$. This is clearly equivalent to the trace estimate

$$||I_\alpha f||_{L^q(\mu)} \leqslant c_1 ||f||_{L^p(\mathbb{R}^n)}. \tag{1.1}$$

What I eventually proved, in [1], was the simple and elegant necessary and sufficient condition

$$\mu(B(x,r)) \leqslant A r^d \tag{1.2}$$

for all $r > 0$ and all $x \in \mathbb{R}^n$. Here $B(x,r)$ is an open ball centered at x and of radius $r > 0$. The conditions of equivalency are: $0 < d \leqslant n$, $q = dp/(n-\alpha p)$, $1 < p < q < \infty$. The coefficient A is a constant independent of x and r. I worked this out in the Spring of 1970. At the time the only result I was aware of along these lines was the trace estimate of Il'in [37] which is (1.1) with $\mu =$ Lebesgue measure on a hyperplane. Later in [2], I found a much simpler proof of the equivalency of (1.1) and (1.2). There, the result followed easily from the weak type estimate

$$\mu([I_\alpha f > t]) \leqslant \left(\frac{c}{t}||f||_{L^p(\mathbb{R}^n)}\right)^q \tag{1.3}$$

for $f \geqslant 0$, followed by an application of the Marcinkiewicz Interpolation Theorem, since here $1/q < 1/p$ (see [11, Theorem 7.2.2] or [41, Theorem 1, p. 52] or even [55, Theorem 4.7.2]).

Of course, it easily becomes clear that (1.1) and (1.2) are no longer equivalent when $q = p$. Indeed, if $d = n - \alpha p$ in (1.2), then there is an $f \in L^p_+(\mathbb{R}^n)$ such that $I_\alpha f = +\infty$ on a set of positive finite μ measure, or to say it another way, $C_{\alpha,p}(K) \equiv \mathbb{C}_{\alpha,p}(\{\delta_x : x \in K\})$ can be zero for a compact set $K \subset \mathbb{R}^n$ with positive Hausdorff capacity (content) $H_\infty^{n-\alpha p}(K) > 0$. Here

$$H_\infty^d(K) = \inf\left\{\sum_i r_i^d : \begin{array}{l} K \text{ is covered by a countable} \\ \text{number of balls of radius } r_i > 0 \end{array}\right\}.$$

Left to right: V.P. Havin, V.G. Maz'ya, and D.R. Adams. Maz'ya lecturing on the finer points of Potential Theory on the wall of a building in Leningrad. Summer 1974.

All of my attempts to find a simple substitute for condition (1.2) when $q = p$ failed – until I made a pilgrimage to Leningrad (with L. Hedberg and J. Brennan) to meet with V. Maz'ya (and V. Havin) in the summer of 1974. Discussions with Maz'ya profoundly changed my view of the trace question. The limiting case $q = p$ needs the Maz'ya type capacity inequality

$$\int_0^\infty C_{\alpha,p}([I_\alpha f > t])\, dt^p \leqslant c||f||^p_{L^p(\mathbb{R}^n)} \tag{1.4}$$

for $f \geqslant 0$ and $1 < p < n/\alpha$. In [4], I dubbed (1.4) a *capacity strong type inequality* (or CSI) in analogy to the weak and strong type estimates for singular integrals in Harmonic Analysis; note that the weak type estimate

$$C_{\alpha,p}([I_\alpha f > t]) \leqslant t^{-p}||f||^p_{L^p(\mathbb{R}^n)}$$

is trivial. Maz'ya was the first to establish (1.4); the paper [39] treats the cases $\alpha = 1$ and 2. And then after my Leningrad meeting, I managed to prove (1.4) for α a positive integer (see [4]). Later, Dahlberg [26] noted how to get (1.4) in the remaining fractional α cases, and finally Hansson [34] gave a general argument that extended (1.4) considerably (see also [11, p. 187f]). Maz'ya [40] proved a CSI for Besov spaces also around this time. The connection of (1.4) to (1.1) is of course

$$\mu(K) \leqslant c\, C_{\alpha,p}(K) \tag{1.5}$$

for all compact sets $K \subset \mathbb{R}^n$. A similar condition: $\mu(K) \leqslant c\, C_{\alpha,p}(K)^{q/p}$ for all compact K works to give (1.1) when $1 < p < q < \infty$, but the advantage of checking (1.2) for only balls is enormous. But as we have observed above (1.1) does not hold simply by having (1.5) for $K = B(x,r)$ only. There is a ball condition in [38], but it is no longer so simple and a famous ball condition of Fefferman–Phong in [27], but that one is only sufficient.

Another interesting sufficient condition for (1.5) is the boundedness of the nonlinear potential $U^{\mu}_{\alpha,p}(x) = I_\alpha(I_\alpha \mu)^{p'-1}(x)$ or equivalently, for the Wolff potential, where $\alpha p < n$

$$W^{\mu}_{\alpha,p}(x) \equiv \int_0^\infty [r^{\alpha p - n}\mu(B(x,r))]^{p'-1}\frac{dt}{t}.$$

This last statement is a consequence of the Wolff inequality

$$||I_\alpha \mu||^{p'}_{L^{p'}(\mathbb{R}^n)} \leqslant c \int_{\mathbb{R}^n} W^{\mu}_{\alpha,p}(x)\, d\mu(x) \tag{1.8}$$

and

$$\mu(K) \leqslant \int I_\alpha f\, d\mu^K \leqslant ||f||_{L^p(\mathbb{R}^n)} ||I_\alpha \mu^K||_{L_{p'}(\mathbb{R}^n)}$$

$$\leqslant ||f||_{L^p(\mathbb{R}^n)}\; c||W^{\mu}_{\alpha,p}||^{1/p'}_{L^\infty}\, \mu(K)^{1/p'}$$

or

$$\mu(K) \leqslant c||W^{\mu}_{\alpha,p}||^{p-1}_{L^\infty}\, C_{\alpha,p}(K).$$

Here I have written $\mu^K = \mu \angle K$. Notice also that

$$||I_\alpha \mu||^{p'}_{L^{p'}(\mathbb{R}^n)} = \int U^{\mu}_{\alpha,p}\, d\mu$$

and so a lower bound matching (1.8) for p' norm of the potential $I_\alpha \mu$ holds – it is a simple estimate, whereas the Wolff inequality requires much heavier work (see [11, p. 109] or [6]).

Finally, I wish to say a few words about the case $q < p$, though I have had little to do with it. The result that I find most interesting here is due to Verbitsky et al [25]. Again, it involves the Wolff potential; the inequality (1.1) holds, $0 < q < p$, $1 < p < n/\alpha$ if and only if

$$W^{\mu}_{\alpha,p} \in L^{q(p-1)/(p-q)}(\mu). \tag{1.9}$$

My small contribution to this case is for $0 < p \leqslant 1$, where the $L^p(\mathbb{R}^n)$ spaces are now replaced by the real Hardy spaces $H^p(\mathbb{R}^n)$. Here is what I can prove for $p = 1$:

(1) (1.1) holds with $p = 1$ if and only if (1.2) holds,

$$q = d/(n-\alpha), \quad q > 1;$$

(2) (1.1) holds with $p = 1$ and $q = 1$ if and only if

$$\mu(K) \leqslant c H^{n-\alpha}_{\infty}(K) \tag{1.10}$$

holds for all compact set (K);

(3) (1.1) holds with $p = 1$ if and only if

$$\int (M_\alpha \mu)^{q/(1-q)}\, d\mu < \infty \tag{1.11}$$

for $0 < q < 1$.

The first result follows from the Semmes inequality

$$|I_\alpha f(x)| \leqslant c(s)[I_{\alpha s}(f^*)^s]^{1/s} \tag{1.12}$$

for $0 < s < 1$. Here f^* is the "grand maximal function" as used in H^p-theory (see [51] or [20, p. 217]). The estimate (1.11) follows from a covering argument; here

$$M_\alpha \mu(x) = \sup_{r>0} r^{\alpha-n} \int\limits_{|x-y|<r} d\mu(y)$$

the fractional maximal function of the measure μ. There are also analogues of (1)-(3) for the Hardy spaces H^p, $0 < p < 1$, replacing H^1, by similar methods.

2 A Mixed Norm Inequality

In the early 1970's, I had the idea to try to extend the L^p-capacity theory to the case of parabolic Riesz potentials. One distinctive feature of these potentials is that there is no real reason why the L^p-space for the space variable x should be the same as the L^p-space for the time variable t. But then I needed to consult the literature on mixed norm estimates, i.e., mixed norm Sobolev type inequalities. I knew only the papers [22] and [19] at the time. Only later did I find the two volume translation [21] that also treats mixed norm situations. Soon, I was working with R. Bagby, an expert on parabolic potentials. However, we discovered early on that the mixed norm estimates on (elliptic) Riesz potentials $I_\alpha f$ in [22] were far from complete, especially with regard to limiting cases.

The result I am referring to is:

$$||I_\alpha f||_{L_x^{q_1} L_t^{q_2}} \leqslant c||f||_{L_x^{p_1} L_t^{p_2}}, \tag{2.1}$$

where $\alpha = 1/p_1 + 1/p_2 - 1/q_1 - 1/q_2,\quad 0 < 1/q_i < 1/p_i < 1,\quad i = 1, 2$. Here the mixed norm is

$$\left[\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x,t)|^{p_1}\, dx \right)^{p_2/p_1} dt \right]^{1/p_2} = ||f||_{L_x^{p_1} L_t^{p_2}},$$

and I am confining myself to the situation $(x, t) \in R \times \mathbb{R}$ for simplicity – more general results can easily be established with the same methods (see [10, 18]). So Bagby and I dropped the parabolic considerations and sought to fill in the gaps we thought were left by [22]. As we saw it, (2.1) is a translation-dilation invariant estimate, so we wanted to prove (2.1) in as many cases as it remains valid, and when false, to find a translation – dilation inequality replacement. So here I list what we proved (using different notation) in [10] and in [18]. An important consideration in all of this turned out to be the mixed norm estimates of Fefferman–Stein [29] for the maximal function $M_0 f$. Here and throughout we are working with $(1/q_1, 1/q_2) \in$ rectangle $[0, 1/p_1] \times [0, 1/p_2],\quad 0 < 1/p_i < 1,\ i = 1, 2$.

(1) (2.1) holds for α as above and

$$0 \leqslant 1/q_1 \leqslant 1/p_1,\quad 0 < 1/q_2 < 1/p_2$$

or

$$1/q_2 = 1/p_2,\quad 0 < 1/q_1 < 1/p_1;$$

(2) when $1/q_2 = 0,\quad \alpha = 1/p_1 + 1/p_2 - 1/q_1$,

$$||I_\alpha f||_{L_x^{(q_1,p_2)} L_t^{\infty}} \leqslant c||f||_{L_x^{p_1} L_t^{p_2}}, \tag{2.2}$$

where the x-norm on the left-hand side is a Lorentz norm:

$$g \in L^{(p,q)} \text{ if and only if } \left[\int_0^\infty (t^p |\, [|g| > t]\, |)^{q/p} \frac{dt}{t} \right]^{1/q} = ||g||_{L^{(p,q)}} < \infty.$$

Here, it was interesting at the time that (2.2) was sharp, i.e., (2.1) is false when $1/p_2 < 1/q_1 < 1/p_1$. Later, we realized that this makes sense when one recalls that the trace of a Riesz potential on a hyperplane can be characterized as a Besov function - and there were the known estimates of Herz [35] (see also Sect. 4 below).

Next, consider the cases $(1/q_1, 1/q_2) = (1/p_1, 0), (0, 1/p_2)$ and $(0, 0)$. Here I will use the notation: $f \in L^{\Phi_p}(Q, \frac{dx}{|Q|})$ for the Orlicz space

$$\inf\left\{\lambda : \fint_Q \Phi_p\Big(\frac{f - f_Q}{\lambda}\Big)\, dx \leqslant 1\right\} < \infty, \tag{2.3}$$

where Q is a cube with sides parallel to the coordinate axes, the bared integral denotes integral average, and f_Q is the integral average of f over the cube Q. Also, $\Phi_p(t) = \exp(t^p) - 1$. With this it follows

(3) $\sup_Q ||I_\alpha f(\cdot, t) - I_\alpha f(\cdot, Q)||_{L_x^\infty L_t^{\Phi_{p_2'}}(Q, \frac{dt}{|Q|})} \leqslant c||f||_{L_x^{p_1} L_t^{p_2}}$,

$$1/q_1 = 1/q_2 = 0, \quad \alpha = 1/p_1 + 1/p_2;$$

(4) $\sup_Q ||I_\alpha f(\cdot, t) - I_\alpha f(\cdot, Q)||_{L_x^{p_1} L_t^{\Phi_{p_2'}}(Q, \frac{dt}{|Q|})} \leqslant c||f||_{L_x^{p_1} L_t^{p_2}}$,

$$1/q_2 = 0,\ 1/q_2 = 1/p_1,\ \alpha = 1/p_2;$$

(5) $\sup_Q ||I_\alpha f(\cdot, t) - I_\alpha f(Q, t)||_{L_t^{\Phi_{p_1'}}(Q, \frac{dt}{|Q|}) L_t^{p_2}} \leqslant c||f||_{L_x^{p_1} L_t^{p_2}}$,

$$1/q_1 = 0,\ 1/p_2 = 1/q_2,\ \alpha = 1/p_1.$$

I used the notation $I_\alpha f(\cdot, Q)$ to indicate that the integral average is taken with respect to the second variable over the cube (an interval in this case).

Estimates (2)–(5) are one way of replacing the false estimate (2.1) in these cases by translation-dilation invariant substitutes. Below (Sect. 5), the reader should note that these estimates are a motivation for the space BMO_p.

3 A Morrey–Sobolev Inequality

In 1938, C.B. Morrey introduced an L^p-growth condition on the gradient of a function that insures the function itself satisfies a Holder continuity condition. This became quite useful in proving the regularity of elliptic partial differential equations and systems. I think Morrey used it mainly in two dimensional situations. The condition, which now bears his name, is: f is a member of the Morrey space $L^{p;\lambda}$ if

$$\sup_{r>0,x} r^{\lambda-n} \int_{B(x,r)} |f(y)|^p \, dy \equiv ||f||^p_{L^{p;\lambda}} < \infty \tag{3.1}$$

for $1 \leqslant p < \infty$, $0 \leqslant \lambda \leqslant n$, and $B(x,r)$ a ball in $\mathbb{R}^n$ (or contained in some subdomain $\Omega \subset \mathbb{R}^n$). The Morrey estimate referred to above is: for $\lambda < p < \infty$, $1 < p < \infty$,

$$||u||_{C^{0,\gamma}} \leqslant c||\nabla u||_{L^{p;\lambda}}, \tag{3.2}$$

where $\gamma = 1 - \lambda/p$ (see [31, Chapt. 7]).

While working at the CNR in 1970, Stampacchia called my attention to his paper [52] where he treated the case $1 < p < \lambda$; i.e., he attempted to prove the corresponding Sobolev Inequality for functions u whose derivatives belong to $L^{p;\lambda}$. Since $L^{p;n} = L^p$, one would hope that such a result would simply become $(*)$, $m = 1$. But in his paper, he was only able to achieve the corresponding weak type estimate. And since there is no Marcinkiewicz Interpolation Theorem for the Morrey spaces, the result remained somewhat unsatisfying. I tried at the time to tackle this question, but was unsuccessful.

Left to right (front row): L.I. Hedberg, J. Brennan, D.R. Adams, and V.G. Maz'ya at Hedberg's party on the occasion of his 60th birthday. Summer 1996.

In 1972, L. Hedberg, in [33], while trying to verify some estimates of N.G. Meyers and myself (from an advanced copy of [16]), came up with a most remarkable proof of $(**)$. He showed that for $f \geqslant 0$,

$$I_\alpha f(x) \leqslant c[M_0 f(x)]^{1-\alpha p/n} ||f||^{\alpha p/n}_{L^p(\mathbb{R}^n)} \tag{3.3}$$

for $\alpha p < n$. Thus raising the left-hand side to the $p^* = np/(n - \alpha p)$ power and integrating, shows that the real work in proving $(**)$ is showing the L^p-maximal estimate

$$||M_0 f||_{L^p(\mathbb{R}^n)} \leqslant c||f||_{L^p(\mathbb{R}^n)} \tag{3.4}$$

(see [48] or [11]). Here $M_0 f$ is the standard Hardy–Littlewood maximal function of f.

So when I finally absorbed this idea, I began to wonder if a similar approach could be given for potentials of functions $f \in L^{p;\lambda}$. I eventually proved, for $f \geqslant 0$,

$$I_\alpha f(x) \leqslant c[M_{\lambda/p} f(x)]^{\alpha p/\lambda} [M_0 f(x)]^{1-\alpha p/\lambda} \tag{3.5}$$

for $\alpha p < \lambda$. Of course, $M_\alpha f$ is the fractional maximal function

$$\sup_{r>0} r^{\alpha - n} \int_{B(x,r)} |f(y)| \, dy.$$

Notice that the Morrey condition is: $M_\lambda f^p \in L^\infty$, so clearly (3.5) implies that under these conditions $I_\alpha f$ is at least locally $\widetilde{p}$-integrable, with $\widetilde{p} = \lambda p/(\lambda - \alpha)p),\ p > 1$. This then becomes the Sobolev exponent p^* when $\lambda = n$. The full Morrey–Sobolev Inequality now easily follows:

$$||I_\alpha f||_{L^{\widetilde{p};\lambda}} \leqslant c||f||_{L^{p;\lambda}} \tag{3.6}$$

$1 < p < \lambda/\alpha,\ 0 < \lambda \leqslant n$ (see [3]).

In a letter to me shortly before his death, Stampacchia mentioned again his attempts to prove (3.6) and that of others and lamented the fact that he had not thought to use the maximal function approach. I fully credit Hedberg for basically pointing me in the right direction with his clever proof of $(**)$. Stampacchia's passing was a personal tragedy for me and a great loss to the mathematical community.

One of my favorite anecdotes of Stampacchia is that one day I was arriving at the CNR just as Stampacchia was leaving, "towing a dignitary." We passed in the parking lot, and as we passed I asked him if he would be in his office later for I had some math questions about one of his papers. He said, "show me" and so on the hood of a car, we spread out the papers and books and he answered my questions there in the parking lot while his obviously annoyed companion stood by. He treated us Post Doc's very well and was always full of life – he obviously enjoyed his role at the CNR, especially when he could poke fun at some of his fellow rival mathematicians at the University of Rome (nearby). I think, in fact, he had a long standing feud with G. Fichera.

There are two other inequalities that I might add to my list. The first is a Morrey–Lorentz–Sobolev estimate

$$||I_\alpha f||_{L^{(\widetilde{p},r);\lambda}} \leqslant c||f||_{L^{(p,q);\lambda}}, \tag{3.7}$$

where $r = q\widetilde{p}/p,\ 0 < q < \infty$ and $\widetilde{p} = \lambda p/(\lambda - \alpha p),\ 1 < p < \lambda/\alpha$. In (3.7), the L^p-norm is replaced by the Lorentz norm $L^{(p,q)}$. This result follows from the usual Lorentz norm estimates for the maximal function.

A second inequality of the Morrey–Sobolev type is the estimate

$$||I_\alpha f||_{L_r^{\widetilde{p};\lambda}} \leqslant c||f||_{L_q^{p;\lambda}}, \tag{3.8}$$

where

$$||f||_{L_q^{p;\lambda}} = \left\{ \sup_x \int_0^\infty (r^{\lambda-n} \int_{B(x,r)} |f(y)|^p \, dy)^{q/p} \, \frac{dr}{r} \right\}^{1/q} < \infty \tag{3.9}$$

and $0 < \alpha < n,\ 0 < \lambda \leqslant n,\ 1 < p < \lambda/\alpha,\ 1 < q \leqslant \infty$ with $\widetilde{p} = \lambda p/(\lambda - \alpha p)$ and $r = \lambda q/(\lambda - \alpha p)$; this was first established in my Umeå Notes [5]. Here the usual Morrey space corresponds to $q = \infty$ in (3.9).

Trace results akin to those of Sect. 1 for Riesz potentials of Morrey functions can be obtained from capacity strong type estimates taken from the results of [17]. Here strong type estimates are found for certain Morrey–Sobolev capacities, then one resorts to measures μ that satisfy inequalities like (1.5) for these capacities. However, it should be noted that not all of these Morrey–Sobolev capacities satisfy such strong type estimates: only Type I capacities (see [17, Sect. 7]).

4 A Morrey–Besov Inequality

Early in the 1980's, I received a preprint from a former graduate student colleague of mine when we both attended the University of Minnesota: Jim Ross; [47]. He had shown an inequality he called a Morrey–Nilol'skii inequality: the function $u(x)$ satisfies the Morrey–Nikol'skii condition on cubes Q if

$$\int_Q |u(x+t) - u(x)|^p \, dx \leqslant c|t|^{\alpha p}|Q|^{1-\lambda/n} \tag{4.1}$$

whenever Q and $Q + t = \{x + t :\ x \in Q\}$ are parallel subcubes of a given fixed cube Q_0. Then Ross showed that $u \in L^s(Q_0)$ for all $s < \widetilde{p},\ 1 \leqslant p < \lambda/\alpha,\ 0 < \alpha \leqslant 1,\ 0 < \lambda \leqslant n$. The estimate (4.1) is to hold for all Q and all such t with C a constant independent of Q and t.

Now, since I recognized the exponent $\widetilde{p}$ and since I had recent success with the Morrey–Sobolev case, I was drawn to this and immediately began to think about what would happen if one replaced the Nikol'skii condition (4.1) with the more general Besov type condition: assume $u \in L^p(\mathbb{R}^n)$ and

$$\Big[\int_{\mathbb{R}^n} \Big(\int_Q |\Delta_t^k u(x)|^p \, dx \Big)^{q/p} |t|^{-(n+\alpha q)} \, dt \Big]^{1/q} \leqslant c|Q|^{(1-\lambda/n)/p} \tag{4.2}$$

with (4.1) the case $k = 1$ and $q = \infty$. Here $\Delta_t^k u$ is the k-th difference: $u(x+t) - u(x) = \Delta_t u(x)$ for $k = 1$ and $\Delta_t^k u(x) = \Delta_t(\Delta_t^{k-1} u)(x),\ k \geqslant 2$. In my original attempts at an $L^{\widetilde{p}}$-integrability result, I used balls rather than cubes and tried to use the maximal function on a potential representation given initially on $\mathbb{R}^{2n}$ and then restricted to $\mathbb{R}^n$ – since Besov functions can be represented in that way. I had only partial success. This idea to look at the restriction of a Riesz potential was useful, but now the maximal function considerations were not. At this point, John Lewis stepped in and found the nice argument that yielded: assume $u(x)$ satisfies (4.2), then it follows that the Lorentz norm condition

$$||u \cdot \chi_Q||_{L^{(\widetilde{p},r)}} \leqslant c|Q|^{(1-\lambda/n)/\widetilde{p}} \tag{4.3}$$

holds with $r = q\widetilde{p}/p$. Furthermore, (4.3) is sharp – no exponent smaller than r can replace r in (4.3) (see [15]). Notice that when $q = \infty$, (4.3) is also an improvement of the result of Ross, for now $s = \widetilde{p}$ is allowed.

At this point, I think a comment is in order about the cube vs. non-cube conditions (or ball vs. non-ball conditions) from this and the previous sections. Notice that in (3.7) when $\lambda = n$, we get the potential in the Lorentz space $L^{(p^*,p^*)},\ \ p^* = np/(n-\alpha p)$, whereas it is well known that when $f \in L^p$, the potential $I_\alpha f \in L^{(p^*,p)}$, a Lorentz space improvement of the classical Sobolev Inequality $(**)$; [23]. The same discontinuity occurs with (4.3), when $\lambda = n$, for the result of Herz [35] gives $u \in L^{(p^*,q)}$ in this case, a space strictly smaller than $L^{(p^*,r)},\ \ r = qp^*/p$. Thus the sharpness of (4.3), for $\lambda < n$ is perhaps a bit surprising. This type of discontinuity also occurs in the limiting case $\alpha p = \lambda$; for $\lambda < n$ vs. $\lambda = n :\ \exp(t)$ is improved to $\exp(t^{p'})$, as becomes evident from the consideration of Sect. 5. Again the example $u(x) = \log |x|,\ x \in R^n$, shows that these Orlicz functions are sharp, and that behavior is quite natural.

5 Exponential Integrability

In the Fall of 1972, I took a visiting position at the University of California-San Diego, where I was very fortunate to be on the faculty with the distinguished mathematician Adriano M. Garsia. He already enjoyed an outstanding reputation in analysis and he also had a slight reputation as a "simplifier" – one who has had some success in providing simpler proofs of some known results with rather original messy proofs. (In fact, I think his small book [30] testifies to this elegance of his ideas.) In my "Post Doc world" –

i.e., my first four years after my Ph.D. – I had three distinguished teachers and Garsia was one of these (the other two were Stampacchia in Rome and Frank Jones at Rice).

During my year at UCSD, I had many discussions with Garsia on topics in analysis. One of these concerned his recent visit to the Courant Institute in New York to lecture and meet with J. Moser – a trip in early '72, I believe. While there, Moser challenged Garsia to come up with a better proof of one of the two exponential estimates from [43]. That result is the now famous refinement of what is usually referred to as the Trudinger inequality, a limiting case of $(**)$, when $\alpha p = n$ (see also [54]).

Trudinger: Assume the support of $f \subset B$, a ball in $\mathbb{R}^n$, and $||f||_{L^p} \leqslant 1$, then there are constants, β and c, independent of f, such that

$$\fint_B \exp\left(\beta |I_\alpha f(x)|^{p'}\right) dx \leqslant c \tag{5.1}$$

for $p = n/\alpha > 1, \quad p' = p/(p-1)$.

Moser: Assume the support $u \subset B$ and $||\nabla u||_{L^n} \leqslant 1$, then

$$\fint_B \exp\left(\beta |u(x)|^{n'}\right) dx \leqslant c \tag{5.2}$$

for all $\beta \leqslant \beta_0 = n w_{n-1}^{1/(n-1)}$, w_{n-1} = area of the surface of the n-ball. Furthermore, (5.2) cannot hold for $\beta > \beta_0$ with the constant c independent of u.

Trudinger's proof relied on estimates for the L^q-norms of the potentials $I_\alpha f, \;\; \alpha p = n$, as $q \to \infty$, and then he summed the exponential series. Moser, relying on the fact that he was dealing only with the first derivative case, used decreasing rearrangements to reduce his estimate to a one dimensional calculus lemma; i.e., he used the crucial inequality

$$||\nabla u^*||_{L^p} \leqslant ||\nabla u||_{L^p}, \tag{5.3}$$

where u^* is the decreasing rearrangement of u. Then one needs only prove (5.2) for u^*.

So after Garsia returned to San Diego, he set to work, eventually devising a new clever proof of Moser's one dimensional calculus lemma used to prove (5.2). And it was in a letter dated March 29, 1972 to Moser that Garsia recorded this argument. I made a copy of that letter, dutifully studied it and then filed it away. I did not see the possibilities it contained. It took me 15 years to wake up! Finally, in the late 1980's, while working on a preliminary version of one of the chapters of [11], I got out Garsia's letter thinking that I could use his argument to redo some of the results we wanted to include, especially Trudinger's inequality. A bit earlier, L. Hedberg had shown that (5.1) held with any $\beta < n/w_{n-1}$, and was false for β larger than this value. Also, it was clear that if one could achieve (5.1) with $\beta = n/w_{n-1}$, then via the representation ($u(x)$ smooth compact support)

$$u(x) = \frac{1}{nw_{n-1}} \int_{\mathbb{R}^n} \frac{(x-y)\cdot \nabla u(y)}{|x-y|^n}\, dy$$

one could produce the Moser result! This seemed interesting, but out of my reach at the time.

Thus, as I worked with Garsia's argument, I fully realized what the stakes were. And hence it came as a complete surprise to me that the Garsia approach could, in fact, be adapted to give exactly the constant $\beta = n/w_{n-1}$ in (5.1). So this is when my confidence began to grow that I was on my way to producing a higher order Moser estimate; [7].

The amazing thing here is that one can just work with the potentials $I_\alpha f,\ f \geqslant 0$, to prove a sharp (5.2) and hence to get a higher order extension of Moser. Of course, one cannot use the idea of decreasing rearrangements u^* to treat the higher order case, for no inequality of the type (5.3) can hold for higher order derivatives. But, one can rearrange the convolution $I_\alpha f$, using a known rearrangement lemma of O'Neil [46] for convolutions, and by this, one can again reduce the result (5.1) to a one dimensional calculus inequality. This produces

Higher order Moser: If m is a positive integer less than n, then there are positive constants β_0 and c_0 depending only on m and n such that if u is a smooth function with support in the ball $B \subset R^n$, with $||\nabla^m u||_{L^p} \leqslant 1,\ \ p = n/m$, then

$$\fint_B \exp\left(\beta |u(x)|^{p'}\right) dx \leqslant c_0 \tag{5.4}$$

for all $\beta \leqslant \beta_0$. Here

$$\nabla^m = \begin{cases} \Delta^{m/2} u, & m = \text{ even}, \\ \nabla \Delta^{(m-1)/2} u, & m = \text{ odd}, \end{cases}$$

$\Delta =$ Laplacian $\nabla =$ gradient. Furthermore, (5.4) is false for $\beta > \beta_0$ in the same way that (5.2) fails.

The exact value of β_0 is given in [7] and its value has drawn some interest for higher order and higher dimensional problems (see in particular [36] where a nice article by A. Chang talks about the Moser–Trudinger inequality and its applications to conformal geometry). Also, I might add that the best value of c_0 in (5.4) is not known; a good estimate has been made for it in the Moser case [24], where they also discuss the existence of an extremal function that gives the best c_0. They use it to estimate the size of c_0.

A few years later, in the early 1990's, I got a call from Washington University in St. Louis, Mo, that a student of A. Bernstien, L. Fontana, had used my methods plus certain technical information about manifolds in $\mathbb{R}^n$, specifically about the sphere S^{n-1}, the boundary of the n-ball, to extend Moser's other exponential estimate from [43] for functions that live on such manifolds. The reader can find this very nice paper at [28].

Also, I might mention that [7] also contains a sharp exponential estimate for certain normalized Besov functions, an estimate similar to (5.4).

And, finally, I want to comment on the space that I will denote as $\mathrm{BMO}_p(\mathbb{R}^n)$. The standard space $\mathrm{BMO}(\mathbb{R}^n)$-functions of bounded mean oscillation – is a well-known space in Harmonic Analysis that often plays a pivotal role there together with its predual, the real Hardy space $H^1(\mathbb{R}^n)$, as a replacement for the spaces L^∞ and L^1, respectively. Using the now famous John–Nirenberg lemma, it is possible to give the functions of BMO an exponential characterization:

Definition: $f \in \mathrm{BMO}$ if for all subcubes Q of $\mathbb{R}^n$,

$$\sup_Q \fint_Q |f(x) - f_Q|\, dx = ||f||_{\mathrm{BMO}} < \infty,$$

where f_Q = average of f over Q.

John Nirenberg: $f \in \mathrm{BMO}$ if there are constants c_1 and c_2 independent of f such that

$$|\{x \in Q : \ |f(x) - f_Q| > \lambda\}| \leqslant c_1 e^{-c_2\lambda}\, |Q| \tag{5.6}$$

for $||f||_{\mathrm{BMO}} \leqslant 1$.

Clearly (5.6) implies that BMO functions enjoy exponential integrability, a fact one can express as

$$\sup_Q \fint_Q \exp\left(\beta|f(x) - f_Q|\right) dx < \infty \tag{5.7}$$

for some constant β independent of f when $||f||_{\mathrm{BMO}} \leqslant 1$. Or in the previous notation

$$\sup_Q ||f - f_Q||_{L^{\Phi}(\mathbb{R}^n, \frac{dx}{|Q|})} < \infty,$$

where $\Phi(t) = e^t - 1$. Thus it seems quite reasonable to introduce the naturally occurring space, motivated by the exponential inequalities (5.1)–(5.4), $\mathrm{BMO}_p(R^n)$ as: $f \in \mathrm{BMO}_p$ if

$$\sup_Q ||f - f_Q||_{L^{\Phi_p}(\mathbb{R}^n, \frac{dx}{|Q|})} < \infty, \tag{5.8}$$

with now $\Phi_p(t) = \exp(t^p) - 1$. The connection to the Sobolev Inequality, in the limiting case $\alpha p = n$, is that the potential $I_\alpha f$, f of compact support, belongs to $\mathrm{BMO}_{p'}$, $1 < p < \infty$, when $||f||_{L^p} \leqslant 1$.

I introduce this space for two reasons:

(1) If we let $H^{1,p}(\mathbb{R}^n)$ denote the Hardy like space: $g \in H^{1,p}(R^n)$ if and only if $||g^*||_{L^{(1,p)}(\mathbb{R}^n)} = ||g||_{H^{1,p}} < \infty$, where g^* is the "grand maximal" function of g from Harmonic Analysis and the norm is the standard Lorentz norm, $p \geqslant 1$, then the Hardy space H^1 corresponds to $H^{1,1}$ (see [20, p. 201f]).

Indeed, one can show that there is a constant c such that

$$||I_{n/p} f||_{L^{p'}(\mathbb{R}^n)} \leqslant c||f||_{H^{1,p'}(\mathbb{R}^n)} \tag{5.9}$$

as a result of the extraordinary estimate of Semmes (1.12).

This strongly suggests that the predual of BMO_p should be $H^{1,p}$. For example the function

$$(\log |x|) \cdot |\log |x||^{1/p-1}$$

belongs to BMO_p in analogy to the $p = 1$ case and suggest the right order of growth for duality. Also, the operators that replace the Calderón–Zygmund singular integrals in the case $p = 1$, are now going to be translation invariant operators of smoothness $\alpha :\ T \in S_\alpha$ if and only if

$$T:\ \Lambda_\beta \to \Lambda_{\beta+\alpha},$$

where Λ_β denotes the space of Holder continuous functions of exponent $\beta \in (0, 1)$ (see [1]), for example, T can be taken to be the composition $I_\alpha \circ CZ$, I_α composed with a CZ-singular integral. $T \in S_{n/p}$ maps $H^{1,p'}$ into $L^{p'}$ via (5.9), and the limiting case $p \to \infty$ is then the standard: $\mathrm{CZ}:\ H^1 \to L^1$.

(2) This conjecture opens up a whole line of enquire in the analogy with the theory associated with the spaces BMO and VMO. In particular one notes that BMO_p can be defined asymptotically as:

$$\sup_Q \fint_Q |f - f_Q|^q \, dx \leqslant \lambda^q \Gamma(q/p + 1), \tag{5.10}$$

as $q \to \infty$. Here Γ is the standard gama function. A space VMO_p could be set as:

$$\lim_{|Q| \to 0} \fint_Q [\exp(|f - f_Q|^p) - 1] \, dx = 0$$

in analogy with Sarason's VMO. Now the analogous duality questions abound. It just appears that many of the standard questions can be posed here and that these spaces arise naturally and are ripe for further investigation.

Also, I should note that an extension of the trace estimates of Sect. 1 can be given in the case $\alpha p = n$. In particular one has the exponential estimate

$$\fint_B \exp\left(\beta(I_\alpha f|^{p'}\right) d\mu \leqslant c$$

for supp $f \subset$ ball, $||f||_{L^p} \leqslant 1$ and μ a measure such that

$$\mu(B(x_0, r)) \leqslant Ar^d$$

for all x_0 and $r > 0$ and some $d > 0$ (see [11], 7.6.4).

6 Vanishing Exponential Integrability

In my Umeå Notes [5], I showed that the exponential Lebesgue set for the Sobolev functions when $\alpha p = n$ could be expressed as:

$$\lim_{r \to 0} \fint_{B(x_0,r)} [\exp\left(\beta|I_\alpha f - [I_\alpha f]_{B(x_0,r)}|^{p'}\right) - 1]\, dx = 0 \tag{6.1}$$

for some constant β independent of f as long as supp f is compact and $||f||_{L^p} \leqslant 1$; this holds for $C_{\alpha,p}$-a.e. x_0. That is, (6.1) can be considered a refinement of the limiting case of the Sobolev Inequality.

When R. Hurri-Syrjänen came to visit the University of Kentucky from Finland, we looked around for something to work on together. In her recent work with D. Edmunds, they had a double exponential condition extending the Trudinger–Moser type estimate: now set $\Phi_\alpha(t) = t^n(\log(e+t))^\alpha, \quad t > 0, \quad 0 \leqslant \alpha \leqslant n-1$, and assume

$$\int_B \Phi_\alpha(|\nabla u|)\, dx \leqslant 1. \tag{6.2}$$

Then there are constants $c_i,\ i = 1, 2, 3$ such that

$$\fint_B \exp\left(c_1|u - u_B|^{n/(n-1-\alpha)}\right) dx \leqslant c_2 \tag{6.3}$$

when $\alpha < n-1$, and

$$\fint_B \exp\left(c_1 \exp\left(c_2|u - u_B|^{n/(n-1)}\right)\right) dx \leqslant c_3 \tag{6.4}$$

when $\alpha = n-1$.

Thus we set ourselves the challenge of finding the analogue of (6.1) for the estimates (6.3) and (6.4). Soon it became clear that the key to such improvements were the Maz'ya capacity type inequalities:

$$\int_0^\infty C_\Phi(\{x : |u(x)| > t\})\, d\Psi(t) \leqslant c \int \Phi(|\nabla u|)\, dx, \tag{6.5}$$

where C_Φ is the capacity defined by

$$C_\Phi(E) = \inf\left\{\int \Phi(|\nabla u|)\, dx : \ u \in C_0^\infty(G),\ u \geqslant 1 \text{ on } E\right\}$$

with $E \subset\subset G$ = bounded open set, usually taken to be a large ball. Here Φ is an Orlicz function as is Ψ, but the important observation here is that we cannot generally take Ψ to be equal to Φ, as is the case in (1.4). Here we want $\Phi = \Phi_\alpha$ and $\Psi = \Psi_\alpha(t) = t^n(\log(e + \frac{1}{t}))^{-\alpha}$, $0 \leqslant \alpha \leqslant n-1$. (6.5) is actually false if we try to replace Ψ_α with Φ_α, $\alpha > 0$.

So in [12], we were able to show the following extensions of (6.1):

$$\fint_{B(x_0,r)} [\exp(c_1\Psi_\alpha(|u(x) - u(x_0)|^{1/(n-1-\alpha)}) - 1]\, dx = o(1), \tag{6.6}$$

C_{Φ_α}-a.e. x_0 when $\alpha < n-1$ as $r \to 0$, and

$$\fint_{B(x_0,r)} [\exp(c_1 \exp(c_2\Psi_{n-1}(|u(x) - u(x_0)|^{1/(n-1)}) - e^{c_1}]\, dx = o(1), \tag{6.7}$$

$C_{\Phi_{n-1}}$-a.e. x_0 when $\alpha = n-1$, as $r \to 0$. The companion paper [13] deals solely with properties of the capacities C_Φ needed here.

R. Hurri-Syrjänen and I also looked at the same question with regard to Besov functions. We achieved much the same results, but there the description of the exceptional sets was a bit more complicated. In the 1970's, Neugebauer, in [44] and [45], developed a pointwise differentiation result for Besov functions that relied on a strange capacity set function – the melding of Hausdorff capacity (content) and the standard capacity natural to the Besov classes.

$$C(E) = \inf\{H(E_1) + A(E_2)\}, \tag{6.8}$$

where the infimum is over all disjoint partitions $E = E_1 \cup E_2$. Here H is a Hausdorff capacity and A is a Besov capacity. The problem is: there is often no known relationship between H and A to simplify (6.8) (see [8] for a study of the relations among the Besov capacities and Hausdorff capacities). This later work is contained in [14].

7 Concluding Remarks

In the 70 years since S.L. Sobolev published his now famous inequality, the subject has truely undergone a tremendous explosion, both in form and applications. Each new variation has been cleverly devised to treat a new and emerging area of research. My own efforts were only in part motivated by applications, but mostly by a curiosity to see how far these ideas can be extended. There is no doubt now that this area has become very rich and technical. The reader will notice that there are a lot of Sobolev type inequalities that I have not touched upon, even further estimates that I have worked on that I feel compelled to pass over for this note. Such topics include:

(1) *Sobolev–Poincaré inequalities*

$$||u - u_Q||_{L^{p^*}(Q)} \leqslant C||\nabla u||_{L^p(Q)} \tag{7.1}$$

and variations on such where the average of u over Q, u_Q, can be replaced by a polynomial depending on the cube Q with the right side of (7.1) replaced by the norm of certain higher order derivatives. Or u_Q can be replaced by a condition where u vanishes on a subset of Q and the constant on the right reflects the size of this set in terms of its capacity – such inequalities have been called Sobolev–Poincaré–Wirtinger type inequalities (see[2] [42] and [11, Chapt. 8]). Also, there is an estimate of this type for Besov function in [15].

(2) *Estimates of the Sobolev type for Parabolic Riesz potentials and their corresponding mixed norm*, for example,

$$P_\alpha f(x,t) = \int\limits_0^t \int\limits_{\mathbb{R}^n} \exp\left[-\frac{|x-y|^2}{t-s}\right] \cdot s^{(\alpha-n-2)/2} f(y,s)\, dy\, ds,$$

where x and $y \in \mathbb{R}^n$. Also generalizing this would include certain semi-groups of operators on f, especially integral representations of the potential type (see, for example, [3]).

[2] This result was established earlier (even for higher derivatives) in: Maz'ya, V.G.: The Dirichlet problem for elliptic equations of arbitrary order in unbounded regions. Dokl. Akad. Nauk SSSR **150**, 1221-1224 (1963); English transl.: Sov. Math., Dokl. **4**, (1963), 860-863 (1963) (see also Maz'ya, V.G.: On (p,l)-capacity, imbedding theorems and the spectrum of a selfadjoint elliptic operator. Izv. Akad. Nauk SSSR, Ser. Mat. **37**, 356-385 (1973); English transl.: Math. USSR-Izv. **7**, 357-387 (1973) and 357-387 and [41]). In the cited papers, it is also shown that the constant C admits two-sided estimates in terms of capacity if the capacity is small. In the opposite case, the constant C is equivalent to the capacitary interior diameter of $Q\backslash F$ (see Maz'ya, V.G.: On the connection between two kinds of capacity. Vestn. Leningr. Univ. Mat. Mekh. Astron. **7**, 33-40 (1974); English transl.: Vestn. Leningr. Univ. Math. **7**, 135-145 (1974).) The fractional variant of the same result can be found in Maz'ya, V.G., Otelbaev, M.: Imbedding theorems and the spectrum of a pseudodifferential operator. Sib. Mat. Zh. **18**, 1073-1087 (1977); English transl.: English translation: Sib. Math. J. **18**, 758-769 (1977). — *Note of Ed.*

(3) Recently, I have been looking at certain hyperbolic and/or retarded potentials of the type

$$\int_{\mathbb{R}^m} f(x-s, t-\Phi(s))\ |s|^{\alpha-m}\, ds,$$

$t > 0$. In this setting, solutions to the 3-dimensional wave equation can be treated when $\Phi(s) = |s|$ and $m = 3,\ \alpha = 2$. The corresponding Sobolev inequalities here are usually called *Strichartz inequalities*. And here $(s, \Phi(s))$ can be a more general function of finite type (see [49, Chapt. 11]).

(4) A whole set of Sobolev type estimates can be given with the fractional maximal function $M_\alpha f$ replacing $I_\alpha f$ throughout. Of course $M_\alpha f \leqslant c \cdot I_\alpha f$, for $f \geqslant 0$, so to make it interesting, one needs to "tighten up" such estimates for $M_\alpha f$ in L^q in terms of the L^p-norm of f. Such estimates can be found, for example, in [9].

(5) Several books have appeared recently on Sobolev Inequalities and their applications. I mention only: [53, 32, 50] and the references contained there in. All of these show some of the variety that has appeared in this area in recent years. Thus it seems clear that many authors feel compelled to write about and produce more and more variations on this Sobolev theme. Clearly, my love and devotion to this subject is not only a personal matter, but a general universal curiosity as well. I suspect it will continue for many more years to come.

Acknowledgement. I want to thank Mike Elery for transforming an unruly manuscript into something legible.

References

1. Adams, D.R.: Traces of potentials arising from translation invariant operators. Ann. Sc. Norm. Super. Pisa **25**, 203–217 (1971)
2. Adams, D.R.: A trace inequality for generalized potentials. Stud. Math. **48**, 99–105 (1973)
3. Adams, D.R.: A note on Riesz potentials. Duke Math. J. **42**, 765–778 (1975)
4. Adams, D.R.: On the existence of capacity strong type estimates on $\mathbb{R}^n$. Ark. Mat. **14**, 125–140 (1976)
5. Adams, D.R.: Lectures on L^p Potential Theory. Lect. Notes Umeå University (1981)
6. Adams, D.R.: Weighted nonlinear potential theory. Trans. Am. Math. Soc. **297**, 73–94 (1986)
7. Adams, D.R.: A sharp inequality of J. Moser's for higher order derivatives. Ann. Math. **128**, 385–398 (1988)
8. Adams, D.R.: The classification problem for the capacities associated with the Besov and Triebel–Lizorkin spaces. In: Approx. and Function Spaces. Banach Center pub. **22**, PWN Polish Sci. Pub., Warsaw (1989)

9. Adams, D.R.: Choquet integrals in Potential Theory. Pub. Math. **42**, 3–66 (1998)
10. Adams, D.R., Bagby, R.: Translation-dilation invarient estimates for Riesz potentials. Indiana Univ. Math. J. **23**, 1051–1067 (1974)
11. Adams, D.R., Hedberg, L.I.: Function Spaces and Potential Theory. Springer-Verlag (1999)
12. Adams, D.R., Hurri-Syrjänen, R.: Vanishing exponential integrability for functions whose gradients belong to $L^n(\log(e+L))^\alpha$, J. Funct. Anal. **197**, 162–178 (2003)
13. Adams, D.R., Hurri-Syrjänen, R.: Capacity estimates. Proc. Am. Math. Soc. **131**, 1159–1167 (2003)
14. Adams, D.R., Hurri-Syrjänen, R.: Besov functions and vanishing exponential integrability. Ill. J. Math. 47 (2003), 1137–1150.
15. Adams, D.R., Lewis, J.L.: On Morrey–Besov inequalities. Stud. Math. **74**, 169–182 (1982)
16. Adams, D.R., Meyers, N.G.: Thinness and Wiener criteria for nonlinear potentials. Indiana Univ. Math. J. **22**, 746–749 (1972)
17. Adams, D.R., Xiao, J.: Nonlinear potential analysis on Morrey spaces and their capacities. Indiana Univ. Math. J. **53**, 1629–1663 (2004)
18. Bagby, R.: An extended inequality for the maximal function. Proc. Am. Math. Soc. **48**, 419–422 (1975)
19. Benedek, M., Calderón, A.P., Panzone, M.: Convolution operators on Banach space valued functions. Proc. Nal. Acad. Sc. USA **48**, 356–365 (1962)
20. Bensoussan, A., Frehse, J.: Regularity Results for Nonlinear Elliptic Systems and Applications. Springer (2002)
21. Besov, O.V., Il'in, V.P., Nikol'skij, S.M.: Integral Representations of Functions and Embedding Theorems (Russian). Nauka, Moscow (1975); Second ed. (1996); English transl.: Wiley, New York (1978/79)
22. Benedek, M., Panzone, M.: The spaces with mixed norm. Duke Math. J. **281**, 301–324 (1961)
23. Bennett, C., Sharpley, R.: Interpolation of Operators. Academic Press, Boston (1988)
24. Carleson, L., Chang, S-YA.: On the existence of an extremal for an inequality of J. Moser. Bull. Sci. **110**, 113–127 (1986)
25. Cascante, C., Ortega, J., Verbitsky, I.: Trace inequalities of Sobolev type in upper triangle. Proc. London Math. Soc. **80**, 391–414 (2000)
26. Dahlberg, B.E.J.: Regularity properties of Riesz potentials. Indiana Univ. Math. J. **28**, 257–268 (1979)
27. Fefferman, C.: The uncertainty principle. Bull. Am. Math. Soc. **9**, 129–206 (1983)
28. Fontana, L.: Sharp borderline Sobolev Inequalities on compact Riemannian manifolds. Comment. Math. Helv. **68**, 415–454 (1993)
29. Fefferman, C., Stein, E.M.: Some maximal inequalities. Am. J. Math. **93**, 107–115 (1971)
30. Garsia, A.M.: Topics in Almost Everywhere Convergence. Markham Pub., Chicago (1970)
31. Gilbarg, D., Trudinger, N.S.: Elliptic Partial Differential Equations of Second Order. Springer (1977)
32. Hebey, E.: Nonlinear Analysis on Manifolds: Sobolev Spaces and Inequalities. Courant Lect. Notes **5**. Am. Math. Soc., Providence, RI (1999)
33. Hedberg, L.I.: On certain convolution inequalities. Proc. Am. Math. Soc. **36**, 505–510 (1972)
34. Hansson, K.: Imbedding theorems of Sobolev type in potential theory. Math. Scand. **45**, 77–102 (1979)
35. Herz, C.: Lipschitz spaces and Bernstein's theorem on absolutely convergent Fourier transforms. J. Math. Mech. **18**, 283–323 (1968)

36. Hardt, R., Wolf, M.: Nonlinear Partial Differential Questions in Differential Geometry. Am. Math. Soc., Providence, RI (1996)
37. Il'in, V.: Generalization of an integral inequality (Russian). Uspekhi Mat. Nauk. **2**, 131–138 (1956)
38. Kerman, R., Sawyer, E.T.: The trace inequality and eigenvalue estimate for Schödinger operators. Ann. Inst. Fourier (Grenoble) **36**, 207–228 (1986)
39. Maz'ya, V.G.: Certain integral inequalities for functions of many variables (Russian). Probl. Mat. Anal. **3**, 22–68 (1972); English transl.: J. Sov. Math **1**, 205–234 (1973)
40. Maz'ya, V.G.: On the capacitary strong type estimates for "fractional" norms (Russian). Zap. Nauchn. Semin. Leningr. Otd. Mat. Inst. Steklova **70**, 161–168 (1977)
41. Maz'ya, V.G.: Sobolev Spaces. Springer-Verlag, Berlin–Tokyo (1985)
42. Meyers, N.G.: Integral inequalities of Poincaré–Wirtinger types. Arch. Ration. Mech. Anal. **68**, 113–120 (1978)
43. Moser, J.: A sharp form of an inequality by N. Trudinger. Indiana Univ. Math. J. **20**, 1077–1092 (1971)
44. Neugbauer, C.: Lipschitz spaces and exponentially differential functions. Indiana Univ. Math. J. **23**, 103–106 (1973)
45. Neugbauer, C.: Strong differentiability of Lipschitz functions. Trans. Am. Math. Soc. **240**, 295–306 (1978)
46. O'Neil, R.: Convolution operators on $L(p, q)$ spaces. Duke Math. J. **30**, 129–142 (1963)
47. Ross, J.: A Morrey–Nikolski inequality. Proc. Am. Math. Soc. **78**, 97–102 (1980)
48. Stein, E.M.: Singular Integrals and Differentiability of Functions. Princeton Univ. Press, Princeton, NJ (1970)
49. Stein, E.M.: Harmonic Analysis. Princeton Univ. Press, Princeton, NJ (1993)
50. Saloff-Coste, L.: Aspects of Sobolev Type Inequalities. Cambridge Univ. Press, Cambridge (2002)
51. Semmes, S.: A primer on Hardy spaces and some remarks on a theorem of Evans and Miller. Commun. Partial Differ. Equ. **19**, 277–319 (1994)
52. Stampacchia, G.: The space $L^{p,\lambda}$ and $N^{p,\lambda}$ and interpolation. Ann. Sc. Norm. Super. Pisa **19**, 443–462 (1965)
53. Varopoulos, N.Th., Saloff-Coste, L., Coulhon, T.: Analysis and Geometry on Groups. Cambridge Univ. Press, Cambridge (1992)
54. Yudovich, V.I.: Some estimates connected with integral operators and with solutions of elliptic equations (Russian). Dokl. Akad. Nauk SSSR **138**, 805–808 (1961); English transl.: Sov. Math., Dokl. **2**, 746–749 (1961)
55. Ziemer, W.P.: Weakly Differentiable Functions. Sobolev Spaces and Functions of Bounded Variation. Springer, Berlin (1989)

Maximal Functions in Sobolev Spaces

Daniel Aalto and Juha Kinnunen

Abstract Applications of the Hardy–Littlewood maximal functions in the modern theory of partial differential equations are considered. In particular, we discuss the behavior of maximal functions in Sobolev spaces, Hardy inequalities, and approximation and pointwise behavior of Sobolev functions. We also study the corresponding questions on metric measure spaces.

1 Introduction

The centered Hardy–Littlewood maximal function $Mf : \mathbf{R}^n \to [0, \infty]$ of a locally integrable function $f : \mathbf{R}^n \to [-\infty, \infty]$ is defined by

$$Mf(x) = \sup \fint_{B(x,r)} |f(y)| \, dy,$$

where the supremum is taken over all radii $r > 0$. Here

$$\fint_{B(x,r)} |f(y)| \, dy = \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y)| \, dy$$

denotes the integral average and $|B(x,r)|$ is the volume of the ball $B(x,r)$. There are several variations of the definition in the literature, for example,

Daniel Aalto
Institute of Mathematics, P.O. Box 1100, FI-02015 Helsinki University of Technology, Finland, e-mail: daniel.aalto@tkk.fi

Juha Kinnunen
Institute of Mathematics, P.O. Box 1100, FI-02015 Helsinki University of Technology, Finland, e-mail: juha.kinnunen@tkk.fi

V. Maz'ya (ed.), *Sobolev Spaces in Mathematics I, International Mathematical Series.*
doi: 10.1007/978-0-387-85648-3, © Springer Science + Business Media, LLC 2009

depending on the requirement whether x is at the center of the ball or not. These definitions give maximal functions that are equivalent with two-sided estimates.

The maximal function theorem of Hardy, Littlewood, and Wiener asserts that the maximal operator is bounded in $L^p(\mathbf{R}^n)$ for $1 < p \leqslant \infty$,

$$\|Mf\|_p \leqslant c\|f\|_p, \tag{1.1}$$

where $c = c(n,p)$ is a constant. The case $p = \infty$ follows immediately from the definition of the maximal function. It can be shown that for the centered maximal function the constant depends only on p, but we do not need this fact here. For $p = 1$ we have the weak type estimate

$$|\{x \in \mathbf{R}^n : Mf(x) > \lambda\}| \leqslant c\lambda^{-1}\|f\|_1$$

for every $\lambda > 0$ with $c = c(n)$ (see [64]).

The maximal functions are classical tools in harmonic analysis. They are usually used to estimate absolute size, and their connections to regularity properties are often neglected. The purpose of this exposition is to focus on this issue. Indeed, applications to Sobolev functions and to partial differential equations indicate that it is useful to know how the maximal operator preserves the smoothness of functions.

There are two competing phenomena in the definition of the maximal function. The integral average is smoothing but the supremum seems to reduce the smoothness. The maximal function is always lower semicontinuous and preserves the continuity of the function provided that the maximal function is not identically infinity. In fact, if the maximal function is finite at one point, then it is finite almost everywhere. A result of Coifman and Rochberg states that the maximal function raised to a power which is strictly between zero and one is a Muckenhoupt weight. This is a clear evidence of the fact that the maximal operator may have somewhat unexpected smoothness properties.

It is easy to show that the maximal function of a Lipschitz function is again Lipschitz and hence, by the Rademacher theorem is differentiable almost everywhere. The question about differentiability in general is a more delicate one.

Simple one-dimensional examples show that the maximal function of a differentiable function is not differentiable in general. Nevertheless, certain weak differentiability properties are preserved under the maximal operator. Indeed, the Hardy–Littlewood maximal operator preserves the first order Sobolev spaces $W^{1,p}(\mathbf{R}^n)$ with $1 < p \leqslant \infty$, and hence it can be used as a test function in the theory of partial differential equations. More precisely, the maximal operator is bounded in the Sobolev space and for every $1 < p \leqslant \infty$ we have

$$\|Mu\|_{1,p} \leqslant c\|u\|_{1,p}$$

with $c = c(n,p)$. We discuss different aspects related to this result.

The maximal functions can also be used to study the smoothness of the original function. Indeed, there are pointwise estimates for the function in terms of the maximal function of the gradient. If $u \in W^{1,p}(\mathbf{R}^n)$, $1 \leqslant p \leqslant \infty$, then there is a set E of measure zero such that

$$|u(x) - u(y)| \leqslant c|x-y|\big(M|Du|(x) + M|Du|(y)\big)$$

for all $x, y \in \mathbf{R}^n \setminus E$. If $1 < p \leqslant \infty$, then the maximal function theorem implies that $M|Du| \in L^p(\mathbf{R}^n)$. This observation has fundamental consequences in the theory of partial differential equations. Roughly speaking, the oscillation of the function is small on the good set where the maximal function of the gradient is bounded. The size of the bad set can be estimated by the maximal function theorem. This can also be used to define Sobolev type spaces in a very general context of metric measure spaces. To show that our arguments are based on a general principle, we also consider the smoothness of the maximal function in this case. The results can be used to study the pointwise behavior of Sobolev functions.

2 Maximal Function Defined on the Whole Space

Recall that the Sobolev space $W^{1,p}(\mathbf{R}^n)$, $1 \leqslant p \leqslant \infty$, consists of functions $u \in L^p(\mathbf{R}^n)$ whose weak first order partial derivatives $D_i u$, $i = 1, 2, \ldots, n$, belong to $L^p(\mathbf{R}^n)$. We endow $W^{1,p}(\mathbf{R}^n)$ with the norm

$$\|u\|_{1,p} = \|u\|_p + \|Du\|_p,$$

where $Du = (D_1 u, D_2 u, \ldots, D_n u)$ is the weak gradient of u. Equivalently, if $1 \leqslant p < \infty$, the Sobolev space can be defined as the completion of smooth functions with respect to the norm above. For basic properties of Sobolev functions we refer to [17].

2.1 Boundedness in Sobolev spaces

Suppose that u is Lipschitz continuous with constant L, i.e.,

$$|u_h(y) - u(y)| = |u(y+h) - u(y)| \leqslant L|h|$$

for all $y, h \in \mathbf{R}^n$, where $u_h(y) = u(y+h)$. Since the maximal function commutes with translations and the maximal operator is sublinear, we have

$$|(Mu)_h(x) - Mu(x)| = |M(u_h)(x) - Mu(x)| \leqslant M(u_h - u)(x)$$
$$= \sup_{r>0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |u_h(y) - u(y)|\, dy \leqslant L|h|. \tag{2.1}$$

This means that the maximal function is Lipschitz continuous with the same constant as the original function provided that Mu is not identically infinity [19]. Observe that this proof applies to Hölder continuous functions as well [14].

It is shown in [33] that the Hardy–Littlewood maximal operator is bounded in the Sobolev space $W^{1,p}(\mathbf{R}^n)$ for $1 < p \leqslant \infty$ and hence, in that case, it has classical partial derivatives almost everywhere. Indeed, there is a simple proof based on the characterization of $W^{1,p}(\mathbf{R}^n)$ with $1 < p < \infty$ by integrated difference quotients according to which $u \in L^p(\mathbf{R}^n)$ belongs to $W^{1,p}(\mathbf{R}^n)$ if and only if there is a constant c for which

$$\|u_h - u\|_p \leqslant c\|Du\|_p|h|$$

for every $h \in \mathbf{R}^n$. As in (2.1), we have

$$|M(u_h) - Mu| \leqslant M(u_h - u)$$

and, by the Hardy–Littlewood–Wiener maximal function theorem, we conclude that

$$\begin{aligned}\|(Mu)_h - Mu\|_p = \|M(u_h) - Mu\|_p &\leqslant \|M(u_h - u)\|_p \\ &\leqslant c\|u_h - u\|_p \leqslant c\|Du\|_p|h|\end{aligned}$$

for every $h \in \mathbf{R}^n$, from which the claim follows. A more careful analysis gives even a pointwise estimate for the partial derivatives. The following simple proposition is used several times in the sequel. If $f_j \to f$ and $g_j \to g$ weakly in $L^p(\Omega)$ and $f_j(x) \leqslant g_j(x)$, $j = 1, 2, \ldots$, almost everywhere in Ω, then $f(x) \leqslant g(x)$ almost everywhere in Ω. Together with some basic properties of the first order Sobolev spaces, this implies that the maximal function semicommutes with weak derivatives. This is the content of the following result which was first proved in [33], but we recall the simple argument here (see also [40, 41]).

Theorem 2.2. *Let $1 < p < \infty$. If $u \in W^{1,p}(\mathbf{R}^n)$, then $Mu \in W^{1,p}(\mathbf{R}^n)$ and*

$$|D_i Mu| \leqslant MD_i u, \quad i = 1, 2, \ldots, n, \tag{2.3}$$

almost everywhere in $\mathbf{R}^n$.

Proof. If $\chi_{B(0,r)}$ is the characteristic function of $B(0,r)$ and

$$\chi_r = \frac{\chi_{B(0,r)}}{|B(0,r)|},$$

then

$$\frac{1}{|B(x,r)|}\int\limits_{B(x,r)} |u(y)|\,dy = |u| * \chi_r(x),$$

where $*$ denotes convolution. Now $|u| * \chi_r \in W^{1,p}(\mathbf{R}^n)$ and

$$D_i(|u| * \chi_r) = \chi_r * D_i|u|, \quad i = 1, 2, \ldots, n,$$

almost everywhere in $\mathbf{R}^n$.

Let r_j, $j = 1, 2, \ldots$, be an enumeration of the positive rational numbers. Since u is locally integrable, we may restrict ourselves to positive rational radii in the definition of the maximal function. Hence

$$Mu(x) = \sup_j (|u| * \chi_{r_j})(x).$$

We define functions $v_k : \mathbf{R}^n \to \mathbf{R}$, $k = 1, 2, \ldots$, by

$$v_k(x) = \max_{1\leqslant j\leqslant k} (|u| * \chi_{r_j})(x).$$

Now (v_k) is an increasing sequence of functions in $W^{1,p}(\mathbf{R}^n)$ which converges to Mu pointwise and

$$\begin{aligned} |D_i v_k| &\leqslant \max_{1\leqslant j\leqslant k} |D_i(|u| * \chi_{r_j})| = \max_{1\leqslant j\leqslant k} |\chi_{r_j} * D_i|u|| \\ &\leqslant MD_i|u| = MD_i u, \end{aligned}$$

$i = 1, 2, \ldots, n$, almost everywhere in $\mathbf{R}^n$. Here we also used the fact that $|D_i|u|| = |D_i u|$, $i = 1, 2, \ldots, n$, almost everywhere. Thus,

$$\|Dv_k\|_p \leqslant \sum_{i=1}^n \|D_i v_k\|_p \leqslant \sum_{i=1}^n \|MD_i u\|_p$$

and the maximal function theorem implies

$$\begin{aligned} \|v_k\|_{1,p} &\leqslant \|Mu\|_p + \sum_{i=1}^n \|MD_i u\|_p \\ &\leqslant c\|u\|_p + c\sum_{i=1}^n \|D_i u\|_p \leqslant c < \infty \end{aligned}$$

for every $k = 1, 2, \ldots$. Hence (v_k) is a bounded sequence in $W^{1,p}(\mathbf{R}^n)$ which converges to Mu pointwise. By the weak compactness of Sobolev spaces,

$Mu \in W^{1,p}(\mathbf{R}^n)$, v_k converges to Mu weakly in $L^p(\mathbf{R}^n)$, and $D_i v_k$ converges to $D_i Mu$ weakly in $L^p(\mathbf{R}^n)$. Since $|D_i v_k| \leqslant MD_i u$ almost everywhere, the weak convergence implies

$$|D_i Mu| \leqslant MD_i u, \quad i = 1, 2, \ldots, n,$$

almost everywhere in $\mathbf{R}^n$. □

Remark 2.4. (i) The case $p = 1$ is excluded in the theorem because our arguments fail in that case. However, Tanaka [66] proved, in the one-dimensional case, that if $u \in W^{1,1}(\mathbf{R})$, then the noncentered maximal function is differentiable almost everywhere and

$$\|DMu\|_1 \leqslant 2\|Du\|_1.$$

For extensions of Tanaka's result to functions of bounded variation in the one-dimensional case we refer to [3] and [4]. The question about the counterpart of Tanaka's result remains open in higher dimensions (see also discussion in [26]). Observe that

$$\|u\|_{n/n-1} \leqslant c\|Du\|_1$$

by the Sobolev embedding theorem and $Mu \in L^{n/(n-1)}(\mathbf{R}^n)$ by the maximal function theorem. However, the behavior of the derivatives is not well understood in this case.

(ii) The inequality (2.3) implies that

$$|DMu(x)| \leqslant M|Du|(x) \tag{2.5}$$

for almost all $x \in \mathbf{R}^n$. Fix a point at which the gradient $DMu(x)$ exists. If $|DMu(x)| = 0$, then the claim is obvious. Hence we may assume that $|DMu(x)| \neq 0$. Let

$$e = \frac{DMu(x)}{|DMu(x)|}.$$

Rotating the coordinates in the proof of the theorem so that e coincides with some of the coordinate directions, we get

$$|DMu(x)| = |D_e Mu(x)| \leqslant MD_h u(x) \leqslant M|Du|(x),$$

where $D_e u = Du \cdot e$ is the derivative to the direction of the unit vector e.

(iii) Using the maximal function theorem together with (2.3), we find

$$\begin{aligned} \|Mu\|_{1,p} &= \|Mu\|_p + \|DMu\|_p \\ &\leqslant c\|u\|_p + \|M|Du|\|_p \leqslant c\|u\|_{1,p}, \end{aligned} \tag{2.6}$$

where c is the constant in (1.1). Hence

$$M : W^{1,p}(\mathbf{R}^n) \to W^{1,p}(\mathbf{R}^n)$$

is a bounded operator, where $1 < p < \infty$.

(iv) If $u \in W^{1,\infty}(\mathbf{R}^n)$, then a slight modification of our proof shows that Mu belongs to $W^{1,\infty}(\mathbf{R}^n)$. Moreover,

$$\begin{aligned} \|Mu\|_{1,\infty} &= \|Mu\|_\infty + \|DMu\|_\infty \\ &\leqslant \|u\|_\infty + \|M|Du|\|_\infty \leqslant \|u\|_{1,\infty}. \end{aligned}$$

Hence, in this case, the maximal operator is bounded with constant one. Recall that, after a redefinition on a set of measure zero, $u \in W^{1,\infty}(\mathbf{R}^n)$ is a bounded and Lipschitz continuous function.

(v) A recent result of Luiro [53] shows that

$$M : W^{1,p}(\mathbf{R}^n) \to W^{1,p}(\mathbf{R}^n)$$

is a continuous operator. Observe that bounded nonlinear operators are not continuous in general. Luiro employs the structure of the maximal operator. He also obtained an interesting formula for the weak derivatives of the maximal function. Indeed, if $u \in W^{1,p}(\mathbf{R})$, $1 < p < \infty$, and $R(x)$ denotes the set of radii $r \geqslant 0$ for which

$$Mu(x) = \limsup_{r_i \to r} \fint_{B(x,r_i)} |u|\, dy$$

for some sequence (r_i) with $r_i > 0$, then for almost all $x \in \mathbf{R}^n$ we have

$$D_i Mu(x) = \fint_{B(x,r)} D_i|u|\, dy$$

for every strictly positive $r \in R(x)$ and

$$D_i Mu(x) = D_i|u|(x)$$

if $0 \in R(x)$. For this is a sharpening of (2.3) we refer to [53, Theorem 3.1] (see also [55]).

(vi) Let $0 \leqslant \alpha \leqslant n$. The fractional maximal function of a locally integrable function $f : \mathbf{R}^n \to [-\infty, \infty]$ is defined by

$$M_\alpha f(x) = \sup_{r>0} r^\alpha \fint_{B(x,r)} |f(y)|\, dy.$$

For $\alpha = 0$ we obtain the Hardy–Littlewood maximal function.

Theorem 2.2 can be easily extended to fractional maximal functions. Indeed, suppose that $1 < p < \infty$. Let $0 \leqslant \alpha < n/p$. If $u \in W^{1,p}(\mathbf{R}^n)$, then $M_\alpha u \in W^{1,q}(\mathbf{R}^n)$ with $q = np/(n - \alpha p)$ and

$$|D_i M_\alpha u| \leqslant M_\alpha D_i u, \qquad i = 1, 2, \dots, n,$$

almost everywhere in $\mathbf{R}^n$. Moreover, there is $c = c(n, p, \alpha)$ such that

$$\|M_\alpha u\|_{1,q} \leqslant c\, \|u\|_{1,p}.$$

The main result of [39] shows that the fractional maximal operator is smoothing in the sense that it maps L^p-spaces into certain first order Sobolev spaces.

2.2 A capacitary weak type estimate

As an application, we show that a weak type inequality for the Sobolev capacity follows immediately from Theorem 2.2. The standard proofs seem to depend, for example, on certain extension properties of Sobolev functions (see [17]). Let $1 < p < \infty$. The Sobolev p-capacity of the set $E \subset \mathbf{R}^n$ is defined by

$$\operatorname{cap}_p(E) = \inf_{u \in \mathcal{A}(E)} \int_{\mathbf{R}^n} \big(|u|^p + |Du|^p\big)\, dx,$$

where

$$\mathcal{A}(E) = \big\{u \in W^{1,p}(\mathbf{R}^n) : u \geqslant 1 \text{ on a neighborhood of } E\big\}.$$

If $\mathcal{A}(E) = \varnothing$, we set $\operatorname{cap}_p(E) = \infty$. The Sobolev p-capacity is a monotone and countably subadditive set function. Let $u \in W^{1,p}(\mathbf{R}^n)$. Suppose that $\lambda > 0$ and denote

$$E_\lambda = \{x \in \mathbf{R}^n : Mu(x) > \lambda\}.$$

Then E_λ is open and $Mu/\lambda \in \mathcal{A}(E_\lambda)$. Using (2.6), we get

$$\begin{aligned}\operatorname{cap}_p\big(E_\lambda\big) &\leqslant \frac{1}{\lambda^p} \int_{\mathbf{R}^n} \big(|Mu|^p + |DMu|^p\big)\, dx \\ &\leqslant \frac{c}{\lambda^p} \int_{\mathbf{R}^n} \big(|u|^p + |Du|^p\big)\, dx \leqslant \frac{c}{\lambda^p} \|u\|_{1,p}^p.\end{aligned}$$

This inequality can be used in the study of the pointwise behavior of Sobolev functions by standard methods. We recall that $x \in \mathbf{R}^n$ is a Lebesgue point for u if the limit

$$u^*(x) = \lim_{r\to 0} \fint_{B(x,r)} u\,dy$$

exists and

$$\lim_{r\to 0} \fint_{B(x,r)} |u(y) - u^*(x)|\,dy = 0.$$

The Lebesgue theorem states that almost all points of a $L^1_{\mathrm{loc}}(\mathbf{R}^n)$ function are Lebesgue points. If a function belongs to $W^{1,p}(\mathbf{R}^n)$, then, using the capacitary weak type estimate, we can prove that the complement of the set of Lebesgue points has zero p-capacity (see [17]).

3 Maximal Function Defined on a Subdomain

Let Ω be an open set in the Euclidean space $\mathbf{R}^n$. For a locally integrable function $f : \Omega \to [-\infty, \infty]$ we define the Hardy–Littlewood maximal function $M_\Omega f : \Omega \to [0, \infty]$ as

$$M_\Omega f(x) = \sup \fint_{B(x,r)} |f(y)|\,dy,$$

where the supremum is taken over all radii $0 < r < \delta(x)$, where

$$\delta(x) = \mathrm{dist}(x, \partial\Omega).$$

In this section, we make the standing assumption that $\Omega \neq \mathbf{R}^n$ so that $\delta(x)$ is finite. Observe that the maximal function depends on Ω. The maximal function theorem implies that the maximal operator is bounded in $L^p(\Omega)$ for $1 < p \leqslant \infty$, i.e.,

$$\|M_\Omega f\|_{p,\Omega} \leqslant c\|f\|_{p,\Omega}. \tag{3.1}$$

This follows directly from (1.1) by considering the zero extension to the complement. The Sobolev space $W^{1,p}(\Omega)$, $1 \leqslant p \leqslant \infty$, consists of those functions u which, together with their weak first order partial derivatives $Du = (D_1 u, \dots, D_n u)$, belong to $L^p(\Omega)$. When $1 \leqslant p < \infty$, we may define $W^{1,p}(\Omega)$ as the completion of smooth functions with respect to the Sobolev norm.

3.1 Boundedness in Sobolev spaces

We consider the counterpart of Theorem 2.2 for the maximal operator M_Ω. It turns out that the arguments in the previous section do not apply mainly because the maximal operator M_Ω does not commute with translations. The following result was proved in [35]. We also refer to [26] for an alternative approach.

Theorem 3.2. *Let $1 < p \leqslant \infty$. If $u \in W^{1,p}(\Omega)$, then $M_\Omega u \in W^{1,p}(\Omega)$ and*

$$|DM_\Omega u| \leqslant 2M_\Omega |Du|$$

almost everywhere in Ω.

Observe that the result holds for every open set and, in particular, we do not make any regularity assumption on the boundary. The functions $u_t : \Omega \to [-\infty, \infty]$, $0 < t < 1$, defined by

$$u_t(x) = \fint_{B(x,t\delta(x))} |u(y)|\, dy,$$

will play a crucial role in the proof of Theorem 3.2 because

$$M_\Omega u(x) = \sup_{0<t<1} u_t(x)$$

for every $x \in \Omega$. We begin with an auxiliary result which may be of independent interest.

Lemma 3.3. *Let Ω be an open set in $\mathbf{R}^n$, and let $1 < p \leqslant \infty$. Suppose that $u \in W^{1,p}(\Omega)$. Then for every $0 < t < 1$ we have $u_t \in W^{1,p}(\Omega)$ and*

$$|Du_t(x)| \leqslant 2M_\Omega |Du|(x) \tag{3.4}$$

for almost all $x \in \Omega$.

Proof. Since $|u| \in W^{1,p}(\Omega)$ and $|D|u|| = |Du|$ almost everywhere in Ω, we may assume that u is nonnegative. Suppose first that $u \in C^\infty(\Omega)$. Let t, $0 < t < 1$, be fixed. According to the Rademacher theorem, as a Lipschitz function δ is differentiable almost everywhere in Ω. Moreover, $|D\delta(x)| = 1$ for almost all $x \in \Omega$. The Leibniz rule gives

$$\begin{aligned} D_i u_t(x) =& D_i\Big(\frac{1}{\omega_n (t\delta(x))^n}\Big) \cdot \int_{B(x,t\delta(x))} u(y)\, dy \\ &+ \frac{1}{\omega_n (t\delta(x))^n} \cdot D_i \int_{B(x,t\delta(x))} u(y)\, dy \end{aligned}$$

for almost all $x \in \Omega$, and, by the chain rule,

$$D_i \int\limits_{B(x,t\delta(x))} u(y)\,dy = \int\limits_{B(x,t\delta(x))} D_i u(y)\,dy$$

$$+ t \int\limits_{\partial B(x,t\delta(x))} u(y)\,d\mathcal{H}^{n-1}(y) \cdot D_i\delta(x)$$

for almost all $x \in \Omega$. Here we also used the fact that

$$\frac{\partial}{\partial r} \int\limits_{B(x,r)} u(y)\,dy = \int\limits_{\partial B(x,r)} u(y)\,dy.$$

Collecting terms, we obtain

$$D_i u_t(x) = n \frac{D_i\delta(x)}{\delta(x)} \Big(\fint\limits_{\partial B(x,t\delta(x))} u(y)\,d\mathcal{H}^{n-1}(y)$$

$$- \fint\limits_{B(x,t\delta(x))} u(y)\,dy \Big) + \fint\limits_{B(x,t\delta(x))} D_i u(y)\,dy \tag{3.5}$$

for almost all $x \in \Omega$ and every $i = 1, 2, \ldots, n$.

In order to estimate the difference of the two integrals in the parentheses in (3.5), we have to take into account a cancellation effect. To this end, suppose that $B(x,R) \subset \Omega$. We use the first Green identity

$$\int\limits_{\partial B(x,R)} u(y) \frac{\partial v}{\partial \nu}(y)\,d\mathcal{H}^{n-1}(y)$$

$$= \int\limits_{B(x,R)} \big(u(y)\Delta v(y) + Du(y) \cdot Dv(y)\big)\,dy,$$

where $\nu(y) = (y-x)/R$ is the unit outer normal of $B(x,R)$, and we choose

$$v(y) = \frac{|y-x|^2}{2}.$$

With these choices the Green formula reads

$$\fint\limits_{\partial B(x,R)} u(y)\,d\mathcal{H}^{n-1}(y) - \fint\limits_{B(x,R)} u(y)\,dy$$

$$= \frac{1}{n} \fint_{B(x,R)} Du(y) \cdot (y - x)\, dy.$$

We estimate the right-hand side of the previous equality by

$$\begin{aligned} \Big| \fint_{B(x,R)} Du(y) \cdot (y-x)\, dy \Big| &\leqslant R \fint_{B(x,R)} |Du(y)|\, dy \\ &\leqslant R M_\Omega |Du|(x). \end{aligned}$$

Finally, we conclude that

$$\Big| \fint_{\partial B(x,R)} u(y)\, d\mathcal{H}^{n-1}(y) - \fint_{B(x,R)} u(y)\, dy \Big| \leqslant \frac{R}{n} M_\Omega |Du|(x). \qquad (3.6)$$

Let e be a unit vector. Using (3.5), (3.6) with $R = t\delta(x)$, and the Schwarz inequality, we find

$$\begin{aligned} &|Du_t(x) \cdot e| \\ &\qquad \leqslant n \frac{|e \cdot D\delta(x)|}{\delta(x)} \cdot \frac{t\delta(x)}{n} M|Du|(x) + \Big| \fint_{B(x,t\delta(x))} e \cdot Du(y)\, dy \Big| \\ &\qquad \leqslant tM|Du|(x) + \fint_{B(x,t\delta(x))} |Du(y)|\, dy \\ &\qquad \leqslant (t+1) M_\Omega |Du|(x) \end{aligned}$$

for almost all $x \in \Omega$. Since $t \leqslant 1$ and e is arbitrary, (3.4) is proved for nonnegative smooth functions.

The case $u \in W^{1,p}(\Omega)$ with $1 < p < \infty$ follows from an approximation argument. Indeed, suppose that $u \in W^{1,p}(\Omega)$ for some p with $1 < p < \infty$. Then there is a sequence (φ_j) of functions in $W^{1,p}(\Omega) \cap C^\infty(\Omega)$ such that $\varphi_j \to u$ in $W^{1,p}(\Omega)$ as $j \to \infty$.

Fix t with $0 < t < 1$. We see that

$$u_t(x) = \lim_{j \to \infty} (\varphi_j)_t(x)$$

if $x \in \Omega$. It is clear that

$$(\varphi_j)_t(x) = \fint_{B(x,t\delta(x))} |\varphi_j(y)|\, dy \leqslant M_\Omega \varphi_j(x)$$

for every $x \in \Omega$. By (3.4), for smooth functions we have

$$|D(\varphi_j)_t(x)| \leqslant 2M_\Omega|D\varphi_j|(x) \tag{3.7}$$

for almost all $x \in \Omega$ and every $j = 1, 2 \ldots$. These inequalities and the maximal function theorem imply that

$$\begin{aligned}\|(\varphi_j)_t\|_{1,p,\Omega} &= \|(\varphi_j)_t\|_{p,\Omega} + \|D(\varphi_j)_t\|_{p,\Omega} \\ &\leqslant c\big(\|\varphi_j\|_{p,\Omega} + \|D\varphi_j\|_{p,\Omega}\big) = c\|\varphi_j\|_{1,p,\Omega}.\end{aligned}$$

Thus, $((\varphi_j)_t)_{j=1}^\infty$ is a bounded sequence in $W^{1,p}(\Omega)$ and, since it converges to u_t pointwise, we conclude that the Sobolev derivative Du_t exists and $D(\varphi_j)_t \to Du_t$ weakly in $L^p(\Omega)$ as $j \to \infty$. This is a standard argument which gives the desired conclusion that u_t belongs to $W^{1,p}(\Omega)$. To establish the inequality (3.4), we want to proceed to the limit in (3.7) as $j \to \infty$. Using the sublinearity of the maximal operator and the maximal function theorem once more, we arrive at

$$\begin{aligned}\|M_\Omega|D\varphi_j| - M_\Omega|Du|\|_{p,\Omega} &\leqslant \|M_\Omega(|D\varphi_j| - |Du|)\|_{p,\Omega} \\ &\leqslant c\||D\varphi_j| - |Du|\|_{p,\Omega}.\end{aligned}$$

Hence $M_\Omega|D\varphi_j| \to M_\Omega|Du|$ in $L^p(\Omega)$ as $j \to \infty$. To complete the proof, we apply the proposition mentioned before Theorem 2.2 to (3.7).

Finally, we consider the case $p = \infty$. Slightly modifying the above proof, we see that $u_t \in W^{1,p}_{\mathrm{loc}}(\Omega)$ for every $1 < p < \infty$ and the estimate (2.3) holds for the gradient. The claim follows from the maximal function theorem. This completes the proof. □

The proof of Theorem 3.2 follows now easily since the hard work has been done in the proof of Lemma 3.3. Suppose that $u \in W^{1,p}(\Omega)$ for some $1 < p < \infty$. Then $|u| \in W^{1,p}(\Omega)$. Let t_j, $j = 1, 2, \ldots$, be an enumeration of the rational numbers between 0 and 1. Denote $u_j = u_{t_j}$. By the previous lemma, we see that $u_j \in W^{1,p}(\Omega)$ for every $j = 1, 2, \ldots$ and (3.4) gives us the estimate

$$|Du_j(x)| \leqslant 2M_\Omega|Du|(x)$$

for almost all $x \in \Omega$ and every $j = 1, 2, \ldots$. We define $v_k : \Omega \to [-\infty, \infty]$, $k = 1, 2, \ldots$, as

$$v_k(x) = \max_{1 \leqslant j \leqslant k} u_j(x).$$

Using the fact that the maximum of two Sobolev functions belongs to the Sobolev space, we see that (v_k) is an increasing sequence of functions in $W^{1,p}(\Omega)$ converging to $M_\Omega u$ pointwise and

$$|Dv_k(x)| = |D \max_{1 \leqslant j \leqslant k} u_j(x)| \leqslant \max_{1 \leqslant j \leqslant k} |Du_j(x)| \leqslant 2M_\Omega|Du|(x) \tag{3.8}$$

for almost all $x \in \Omega$ and every $j = 1, 2, \dots$. On the other hand,

$$v_k(x) \leqslant M_\Omega u(x)$$

for all $x \in \Omega$ and $k = 1, 2, \dots$. The rest of the proof goes along the lines of the final part of the proof of Theorem 2.2. By the maximal function theorem,

$$\begin{aligned}\|v_k\|_{1,p,\Omega} &= \|v_k\|_{p,\Omega} + \|Dv_k\|_{p,\Omega} \\ &\leqslant \|M_\Omega u\|_{p,\Omega} + 2\|M_\Omega |Du|\|_{p,\Omega} \leqslant c\|u\|_{1,p,\Omega}.\end{aligned}$$

Hence (v_k) is a bounded sequence in $W^{1,p}(\Omega)$ such that $v_k \to M_\Omega u$ everywhere in Ω as $k \to \infty$. A weak compactness argument shows that $M_\Omega u \in W^{1,p}(\Omega)$, $v_k \to M_\Omega u$, and $Dv_k \to DM_\Omega u$ weakly in $L^p(\Omega)$ as $k \to \infty$. Again, we may proceed to the weak limit in (3.8), using the proposition mentioned before Theorem 2.2.

Let us briefly consider the case $p = \infty$. Using the above argument, it is easy to see that $M_\Omega u \in W^{1,p}_{\mathrm{loc}}(\Omega)$ and the claim follows from the maximal function theorem.

Remark 3.9. Again, it follows immediately that

$$M_\Omega : W^{1,p}(\Omega) \to W^{1,p}(\Omega)$$

is a bounded operator. Luiro [54] shows that it is also a continuous operator for every open set Ω, with $1 < p \leqslant \infty$. In [55], he gives examples of natural maximal operators which are not continuous on Sobolev spaces.

3.2 Sobolev boundary values

We have shown that the local Hardy–Littlewood maximal operator preserves the Sobolev spaces $W^{1,p}(\Omega)$ provided that $1 < p \leqslant \infty$. Next we show that the maximal operator also preserves the boundary values in the Sobolev sense. Recall that the Sobolev space with zero boundary values, denoted by $W^{1,p}_0(\Omega)$ with $1 \leqslant p < \infty$, is defined as the completion of $C_0^\infty(\Omega)$ with respect to the Sobolev norm.

We begin with some useful condition which guarantees that a Sobolev function has zero boundary values in the Sobolev sense. The following result was proved in [36], but we present a very simple proof by Zhong [70, Theorem 1.9]. With a different argument this result also holds in metric measure spaces [32, Theorem 5.1].

Lemma 3.10. *Let $\Omega \neq \mathbf{R}^n$ be an open set. Suppose that $u \in W^{1,p}(\Omega)$. If*

$$\int_\Omega \left(\frac{|u(x)|}{\operatorname{dist}(x,\partial\Omega)}\right)^p dx < \infty,$$

then $u \in W_0^{1,p}(\Omega)$.

Proof. For $\lambda > 0$ we define $u_\lambda : \Omega \to [0,\infty]$ by

$$u_\lambda(x) = \min(|u(x)|, \lambda \operatorname{dist}(x,\partial\Omega)).$$

We see that $u_\lambda \in W_0^{1,p}(\Omega)$ for every $\lambda > 0$.

Then we show that (u_λ) is a uniformly bounded family of functions in $W_0^{1,p}(\Omega)$. Clearly, $u_\lambda \leqslant |u|$ and hence

$$\int_\Omega u_\lambda^p \, dx \leqslant \int_\Omega |u|^p \, dx.$$

For the gradient estimate we define

$$F_\lambda = \{x \in \Omega : |u(x)| > \lambda \operatorname{dist}(x,\partial\Omega)\},$$

where $\lambda > 0$. Then

$$\begin{aligned}\int_\Omega |Du_\lambda|^p \, dx &= \int_{\Omega\setminus F_\lambda} |Du|^p \, dx + \lambda^p \int_{F_\lambda} |D \operatorname{dist}(x,\partial\Omega)|^p \, dx \\ &\leqslant \int_\Omega |Du|^p \, dx + \lambda^p |F_\lambda|,\end{aligned}$$

where, by assumption,

$$\lambda^p |F_\lambda| \leqslant \int_\Omega \left(\frac{|u(x)|}{\operatorname{dist}(x,\partial\Omega)}\right)^p dx < \infty$$

for every $\lambda > 0$. Here we again used the fact that $|D \operatorname{dist}(x,\partial\Omega)| = 1$ for almost all $x \in \Omega$. This implies that (u_λ) is a uniformly bounded family of functions in $W_0^{1,p}(\Omega)$.

Since $|F_\lambda| \to 0$ as $\lambda \to \infty$ and $u_\lambda = |u|$ in $\Omega \setminus F_\lambda$, we have $u_\lambda \to |u|$ almost everywhere in Ω. A similar weak compactness argument that was used in the proofs of Theorems 2.2 and 3.2 shows that $|u| \in W_0^{1,p}(\Omega)$. □

Remark 3.11. The proof shows that, instead of $u/\delta \in L^p(\Omega)$, it is enough to assume that u/δ belongs to the weak $L^p(\Omega)$. Boundary behavior of the maximal function was studied in [37, 35]

Theorem 3.12. *Let* $\Omega \subset \mathbf{R}^n$ *be an open set. Suppose that* $u \in W^{1,p}(\Omega)$ *with* $p > 1$. *Then*

$$|u| - M_\Omega u \in W_0^{1,p}(\Omega).$$

Remark 3.13. In particular, if $u \in W_0^{1,p}(\Omega)$, then $M_\Omega u \in W_0^{1,p}(\Omega)$. Observe that this holds for every open subset Ω.

Proof. Fix $0 < t < 1$. A standard telescoping argument (see Lemma 4.1) gives

$$\begin{aligned}\big||u(x)| - u_t(x)\big| &= \left| |u(x)| - \fint_{B(x,t\delta(x))} |u(y)|\,dy \right| \\ &\leqslant ct \operatorname{dist}(x,\partial\Omega) M_\Omega |Du|(x).\end{aligned}$$

For every $x \in \Omega$ there is a sequence t_j, $j = 1, 2, \ldots$, such that

$$M_\Omega u(x) = \lim_{j\to\infty} u_{t_j}(x).$$

This implies that

$$\begin{aligned}\big||u(x)| - M_\Omega u(x)\big| &= \lim_{j\to\infty} \big||u(x)| - u_{t_j}(x)\big| \\ &\leqslant c \operatorname{dist}(x,\partial\Omega) M_\Omega |Du|(x).\end{aligned}$$

By the maximal function theorem, we conclude that

$$\begin{aligned}\int_\Omega \left(\frac{\big||u(x)| - M_\Omega u(x)\big|}{\operatorname{dist}(x,\partial\Omega)} \right)^p dx &\leqslant c \int_\Omega (M_\Omega |Du|(x))^p\,dx \\ &\leqslant c \int_\Omega |Du(x)|^p\,dx.\end{aligned}$$

This implies that

$$\frac{|u(x)| - M_\Omega u(x)}{\operatorname{dist}(x,\partial\Omega)} \in L^p(\Omega).$$

By Theorem 3.2, we have $M_\Omega u \in W^{1,p}(\Omega)$, and from Lemma 3.10 we conclude that $|u| - M_\Omega u \in W_0^{1,p}(\Omega)$. □

Remark 3.14. We observe that the maximal operator preserves nonnegative superharmonic functions; see [37]. (For superharmonic functions that change signs, we may consider the maximal function without absolute values.) Suppose that $u : \Omega \to [0, \infty]$ is a measurable function which is not identically ∞ on any component of Ω. Then it is easy to show that

$$M_\Omega u(x) = u(x)$$

for every $x \in \Omega$ if and only if u is superharmonic.

The least superharmonic majorant can be constructed by iterating the maximal function. For short we write

$$M_\Omega^{(k)} u(x) = M_\Omega \circ M_\Omega \circ \cdots \circ M_\Omega u(x), \qquad k = 1, 2, \dots .$$

Since $M_\Omega^{(k)} u$, $k = 1, 2, \dots$, are lower semicontinuous, we see that

$$M_\Omega^{(k)} u(x) \leqslant M_\Omega^{(k+1)} u(x), \qquad k = 1, 2, \dots,$$

for every $x \in \Omega$. Hence $(M_\Omega^{(k)} u(x))$ is an increasing sequence of functions and it converges for every $x \in \Omega$ (the limit may be ∞). We denote

$$M_\Omega^{(\infty)} u(x) = \lim_{k \to \infty} M_\Omega^{(k)} u(x)$$

for every $x \in \Omega$. If $M_\Omega^{(\infty)} u$ is not identically infinity on any component of Ω, then it is the smallest superharmonic function with the property that

$$M_\Omega^{(\infty)} u(x) \geqslant u(x)$$

for almost all $x \in \Omega$. If $u \in W^{1,p}(\Omega)$, then the obtained smallest superharmonic function has the same boundary values as u in the Sobolev sense by Theorem 3.12.

Fiorenza [18] observed that nonnegative functions of one or two variables cannot be invariant under the maximal operator unless they are constant. This is consistent with the fact that on the line there are no other concave functions and in the plane there are no other superharmonic functions but constants that are bounded from below (see also [42]).

4 Pointwise Inequalities

The following estimates are based on a well-known telescoping argument (see [28] and [16]). The proofs are based on a general principle and they apply in a metric measure space equipped with a doubling measure (see [25]). This fact will be useful below.

Let $0 < \beta < \infty$ and $R > 0$. The fractional sharp maximal function of a locally integrable function f is defined by

$$f^{\#}_{\beta,R}(x) = \sup_{0<r<R} r^{-\beta} \fint_{B(x,r)} |f - f_{B(x,r)}| \, dy,$$

If $R = \infty$ we simply write $f^{\#}_{\beta}(x)$.

Lemma 4.1. *Suppose that f is locally integrable. Let $0 < \beta < \infty$. Then there is a constant $c = c(\beta, n)$ and a set E with $|E| = 0$ such that*

$$|f(x) - f(y)| \leqslant c|x-y|^\beta \big(f^\#_{\beta,4|x-y|}(x) + f^\#_{\beta,4|x-y|}(y)\big) \tag{4.2}$$

for all $x, y \in \mathbf{R}^n \setminus E$.

Proof. Let E be the complement of the set of Lebesgue points of f. By the Lebesgue theorem, $|E| = 0$. Fix $x \in \mathbf{R}^n \setminus E$, $0 < r < \infty$ and denote $B_i = B(x, 2^{-i}r)$, $i = 0, 1, \dots$. Then

$$\begin{aligned}|f(x) - f_{B(x,r)}| &\leqslant \sum_{i=0}^\infty |f_{B_{i+1}} - f_{B_i}| \\ &\leqslant \sum_{i=0}^\infty \frac{\mu(B_i)}{\mu(B_{i+1})} \fint_{B_i} |f - f_{B_i}|\,dy \\ &\leqslant c\sum_{i=0}^\infty (2^{-i}r)^\beta (2^{-i}r)^{-\beta} \fint_{B_i} |f - f_{B_i}|\,dy \\ &\leqslant cr^\beta f^\#_{\beta,r}(x).\end{aligned}$$

Let $y \in B(x,r) \setminus E$. Then $B(x,r) \subset B(y,2r)$ and we obtain

$$\begin{aligned}|f(y) - f_{B(x,r)}| &\leqslant |f(y) - f_{B(y,2r)}| + |f_{B(y,2r)} - f_{B(x,r)}| \\ &\leqslant cr^\beta f^\#_{\beta,2r}(y) + \fint_{B(x,r)} |f - f_{B(y,2r)}|\,dz \\ &\leqslant cr^\beta f^\#_{\beta,2r}(y) + c\fint_{B(y,2r)} |f - f_{B(y,2r)}|\,dz \\ &\leqslant cr^\beta f^\#_{\beta,2r}(y).\end{aligned}$$

Let $x, y \in \mathbf{R}^n \setminus E$, $x \neq y$ and $r = 2|x-y|$. Then $x, y \in B(x,r)$ and hence

$$\begin{aligned}|f(x) - f(y)| &\leqslant |f(x) - f_{B(x,r)}| + |f(y) - f_{B(x,r)}| \\ &\leqslant c|x-y|^\beta \big(f^\#_{\beta,4|x-y|}(x) + f^\#_{\beta,4|x-y|}(y)\big).\end{aligned}$$

This completes the proof. □

Let $0 \leqslant \alpha < 1$ and $R > 0$. The fractional maximal function of a locally integrable function f is defined by

$$M_{\alpha,R}f(x) = \sup_{0<r<R} r^\alpha \fint_{B(x,r)} |f|\,dy,$$

For $R = \infty$, we write $M_{\alpha,\infty} = M_\alpha$. If $\alpha = 0$, we obtain the Hardy–Littlewood maximal function and write $M_0 = M$.

If $u \in W^{1,1}_{\mathrm{loc}}(\mathbf{R}^n)$, then, by the Poincaré inequality, there is a constant $c = c(n)$ such that

$$\fint_{B(x,r)} |u - u_{B(x,r)}|\,dy \leqslant cr \fint_{B(x,r)} |Du|\,dy$$

for every ball $B(x,r) \subset \mathbf{R}^n$. It follows that

$$r^{\alpha-1} \fint_{B(x,r)} |u - u_{B(x,r)}|\,dy \leqslant cr^\alpha \fint_{B(x,r)} |Du|\,dy$$

and, consequently,

$$u^\#_{1-\alpha,R}(x) \leqslant cM_{\alpha,R}|Du|(x)$$

for every $x \in \mathbf{R}^n$ and $R > 0$. Thus, we have proved the following useful inequality.

Corollary 4.3. *Let $u \in W^{1,1}_{\mathrm{loc}}(\mathbf{R}^n)$ and $0 \leqslant \alpha < 1$. Then there is a constant $c = c(n,\alpha)$ and a set $E \subset \mathbf{R}^n$ with $|E| = 0$ such that*

$$|u(x) - u(y)| \leqslant c|x-y|^{1-\alpha}\big(M_{\alpha,4|x-y|}|Du|(x) + M_{\alpha,4|x-y|}|Du|(y)\big)$$

for all $x, y \in \mathbf{R}^n \setminus E$.

If $u \in W^{1,p}(\mathbf{R}^n)$, $1 \leqslant p \leqslant \infty$, then

$$|u(x) - u(y)| \leqslant c|x-y|\big(M|Du|(x) + M|Du|(y)\big)$$

for all $x, y \in \mathbf{R}^n \setminus E$. If $1 < p \leqslant \infty$, then the maximal function theorem implies that $g = M|Du| \in L^p(\mathbf{R}^n)$ and, by the previous inequality, we have

$$|u(x) - u(y)| \leqslant c|x-y|\big(g(x) + g(y)\big)$$

for all $x, y \in \mathbf{R}^n \setminus E$ with $|E| = 0$. The following result shows that this gives a characterization of $W^{1,p}(\mathbf{R}^n)$ for $1 < p \leqslant \infty$. This characterization can be used as a definition of the first order Sobolev spaces on metric measure spaces (see [21, 24, 25]).

Theorem 4.4. *Let $1 < p \leqslant \infty$. Then the following four conditions are equivalent.*

(i) *$u \in W^{1,p}(\mathbf{R}^n)$.*

(ii) *$u \in L^p(\mathbf{R}^n)$ and there is $g \in L^p(\mathbf{R}^n)$, $g \geqslant 0$, such that*

$$|u(x) - u(y)| \leqslant |x - y|(g(x) + g(y))$$

for all $x, y \in \mathbf{R}^n \setminus E$ with $|E| = 0$.

(iii) *$u \in L^p(\mathbf{R}^n)$ and there is $g \in L^p(\mathbf{R}^n)$, $g \geqslant 0$, such that the Poincaré inequality holds,*

$$\fint_{B(x,r)} |u - u_{B(x,r)}|\,dy \leqslant c\,r \fint_{B(x,r)} g\,dy$$

for all $x \in \mathbf{R}^n$ and $r > 0$.

(iv) *$u \in L^p(\mathbf{R}^n)$ and $u_1^\# \in L^p(\mathbf{R}^n)$.*

Proof. We have already seen that (i) implies (ii). To prove that (ii) implies (iii), we integrate the pointwise inequality twice over the ball $B(x,r)$. After the first integration we obtain

$$\begin{aligned}|u(y) - u_{B(x,r)}| &= \left| u(y) - \fint_{B(x,r)} u(z)\,dz \right| \\ &\leqslant \fint_{B(x,r)} |u(y) - u(z)|\,dz \\ &\leqslant 2r \left(g(y) + \fint_{B(x,r)} g(z)\,dz \right),\end{aligned}$$

which implies

$$\begin{aligned}\fint_{B(x,r)} |u(y) - u_{B(x,r)}|\,dy &\leqslant 2r \left(\fint_{B(x,r)} g(y)\,dy + \fint_{B(x,r)} g(z)\,dz \right) \\ &\leqslant 4r \fint_{B(x,r)} g(y)\,dy.\end{aligned}$$

To show that (iii) implies (iv), we observe that

$$u_1^{\#}(x) = \sup_{r>0} \frac{1}{r} \fint_{B(x,r)} |u - u_{B(x,r)}| \, dy \leqslant c \sup_{r>0} \fint_{B(x,r)} g \, dy = cMg(x).$$

Then we show that (iv) implies (i). By Theorem 4.1,

$$|u(x) - u(y)| \leqslant c|x-y|(u_1^{\#}(x) + u_1^{\#}(y))$$

for all $x, y \in \mathbf{R}^n \setminus E$ with $|E| = 0$. If we denote $g = cu_1^{\#}$, then $g \in L^p(\mathbf{R}^n)$ and

$$|u(x) - u(y)| \leqslant |x-y|(g(x) + g(y))$$

for all $x, y \in \mathbf{R}^n \setminus E$ with $|E| = 0$. Then we use the characterization of Sobolev spaces $W^{1,p}(\mathbf{R}^n)$, $1 < p < \infty$, with integrated difference quotients. Let $h \in \mathbf{R}^n$. Then

$$|u_h(x) - u(x)| = |u(x+h) - u(x)| \leqslant |h|(g_h(x) + g(x)),$$

from which we conclude that

$$\|u_h - u\|_p \leqslant |h|(\|g_h\|_p + \|g\|_p) = 2|h|\|g\|_p,$$

which implies the claim. □

Remark 4.5. Hajłasz [22] showed that $u \in W^{1,1}(\mathbf{R}^n)$ if and only if $u \in L^1(\mathbf{R}^n)$ and there is a nonnegative function $g \in L^1(\mathbf{R}^n)$ and $\sigma \geqslant 1$ such that

$$|u(x) - u(y)| \leqslant |x-y|(M_{\sigma|x-y|}g(x) + M_{\sigma|x-y|}g(y))$$

for all $x, y \in \mathbf{R}^n \setminus E$ with $|E| = 0$. Moreover, if this inequality holds, then $|Du| \leqslant c(n, \sigma)g$ almost everywhere.

4.1 Lusin type approximation of Sobolev functions

Approximations of Sobolev functions were studied, for example, in [2, 10, 11, 13, 20, 25, 52, 56, 58, 60, 69].

Let $u \in W^{1,p}(\mathbf{R}^n)$ and $0 \leqslant \alpha < 1$. By Corollary (4.3),

$$|u(x) - u(y)| \leqslant c|x-y|^{1-\alpha}\big(M_\alpha|Du|(x) + M_\alpha|Du|(y)\big)$$

for all $x, y \in \mathbf{R}^n \setminus E$ with $|E| = 0$. For $p > n$ the Hölder inequality implies

$$M_{n/p}|Du|(x) \leqslant cM_n|Du|^p(x)^{1/p} \leqslant c\|Du\|_p$$

for every $x \in \mathbf{R}^n \setminus E$ with $c = c(n,p)$. Hence

$$|u(x) - u(y)| \leqslant c\|Du\|_p |x - y|^{1-n/p}$$

for all x, $y \in \mathbf{R}^n \setminus E$ and u is Hölder continuous with the exponent $1 - n/p$ after a possible redefinition on a set of measure zero. The same argument implies that if $M_\alpha|Du|$ is bounded, then $u \in C^{1-\alpha}(\mathbf{R}^n)$. Even if $M_\alpha|Du|$ is unbounded, then

$$|u(x) - u(y)| \leqslant c\lambda|x - y|^{1-\alpha}$$

for all $x, y \in \mathbf{R}^n \setminus E_\lambda$, where

$$E_\lambda = \{x \in \mathbf{R}^n : M_\alpha|Du|(x) > \lambda\}$$

for $\lambda > 0$. This means that the restriction of $u \in W^{1,p}(\mathbf{R}^n)$ to the set $\mathbf{R}^n \setminus E_\lambda$ is Hölder continuous after a redefinition on a set of measure zero.

Recall that the (spherical) Hausdorff s-content, $0 < s < \infty$, of $E \subset \mathbf{R}^n$ is defined by

$$\mathcal{H}^s_\infty(E) = \inf\Big\{ \sum_{i=1}^\infty r_i^s : E \subset \bigcup_{i=1}^\infty B(x_i, r_i)\Big\}.$$

The standard Vitali covering argument gives the following estimate for the size of the set $\mathbf{R}^n \setminus E_\lambda$. There is a constant $c = c(n, p, \alpha)$ such that

$$\mathcal{H}^{n-\alpha p}_\infty(E_\lambda) \leqslant c\lambda^{-p} \int_{\mathbf{R}^n} |Du|^p \, dx \tag{4.6}$$

for every $\lambda > 0$.

Theorem 4.7. *Let $u \in W^{1,p}(\mathbf{R}^n)$, and let $0 \leqslant \alpha < 1$. Then for every $\lambda > 0$ there is an open set E_λ and a function u_λ such that $u(x) = u_\lambda(x)$ for every $x \in \mathbf{R}^n \setminus E_\lambda$, $u_\lambda \in W^{1,p}(\mathbf{R}^n)$, u_λ is Hölder continuous with the exponent $1 - \alpha$, $\|u - u_\lambda\|_{W^{1,p}(\mathbf{R}^n)} \to 0$ as $\lambda \to \infty$, and $\mathcal{H}^{n-\alpha p}_\infty(E_\lambda) \to 0$ as $\lambda \to \infty$.*

Remark 4.8. (i) If $\alpha = 0$, then the theorem says that every function in the Sobolev space coincides with a Lipschitz function outside a set of arbitrarily small Lebesgue measure. The obtained Lipschitz function approximates the original Sobolev function also in the Sobolev norm.

(ii) Since

$$\operatorname{cap}_{\alpha p}(E_\lambda) \leqslant c\mathcal{H}^{n-\alpha p}_\infty(E_\lambda),$$

the size of the exceptional set can also be expressed in terms of capacity.

Proof. The set E_λ is open since M_α is lower semicontinuous. From (4.6) we conclude that

$$\mathcal{H}^{n-\alpha p}_\infty(E_\lambda) \leqslant c\lambda^{-p} \, \|Du\|_p^p$$

for every $\lambda > 0$ with $c = c(n, p, \alpha)$.

We already showed that $u|_{\mathbf{R}^n \setminus E_\lambda}$ is $(1 - \alpha)$-Hölder continuous with the constant $c(n)\lambda$.

Let Q_i, $i = 1, 2, \ldots$, be a Whitney decomposition of E_λ with the following properties: each Q_i is open, the cubes Q_i, $i = 1, 2, \ldots$, are disjoint, $E_\lambda = \bigcup_{i=1}^{\infty} \overline{Q}_i$, $4Q_i \subset E_\lambda$, $i = 1, 2, \ldots$,

$$\sum_{i=1}^{\infty} \chi_{2Q_i} \leqslant N < \infty,$$

and

$$c_1 \operatorname{dist}(Q_i, \mathbf{R}^n \setminus E_\lambda) \leqslant \operatorname{diam}(Q_i) \leqslant c_2 \operatorname{dist}(Q_i, \mathbf{R}^n \setminus E_\lambda)$$

for some constants c_1 and c_2.

Then we construct a partition of unity associated with the covering $2Q_i$, $i = 1, 2, \ldots$. This can be done in two steps. First, let $\widetilde{\varphi}_i \in C_0^\infty(2Q_i)$ be such that $0 \leqslant \widetilde{\varphi}_i \leqslant 1$, $\widetilde{\varphi}_i = 1$ in Q_i and

$$|D\widetilde{\varphi}_i| \leqslant \frac{c}{\operatorname{diam}(Q_i)}$$

for $i = 1, 2, \ldots$. Then we define

$$\varphi_i(x) = \frac{\widetilde{\varphi}_i(x)}{\sum\limits_{j=1}^{\infty} \widetilde{\varphi}_j(x)}$$

for every $i = 1, 2, \ldots$. Observe that the sum is taken over finitely many terms only since $\varphi_i \in C_0^\infty(2Q_i)$ and the cubes $2Q_i$, $i = 1, 2, \ldots$, are of bounded overlap. The functions φ_i have the property

$$\sum_{i=1}^{\infty} \varphi_i(x) = \chi_{E_\lambda}(x)$$

for every $x \in \mathbf{R}^n$.

Then we define the function u_λ by

$$u_\lambda(x) = \begin{cases} u(x), & x \in \mathbf{R}^n \setminus E_\lambda, \\ \sum\limits_{i=1}^{\infty} \varphi_i(x) u_{2Q_i}, & x \in E_\lambda. \end{cases}$$

The function u_λ is a Whitney type extension of $u|_{\mathbf{R}^n \setminus E_\lambda}$ to the set E_λ.

First we claim that

$$\|u_\lambda\|_{W^{1,p}(E_\lambda)} \leqslant c\|u\|_{W^{1,p}(E_\lambda)}. \tag{4.9}$$

Since the cubes $2Q_i$, $i = 1, 2, \ldots$, are of bounded overlap, we have

$$\int_{E_\lambda} |u_\lambda|^p\,dx = \int_{E_\lambda} \Big|\sum_{i=1}^\infty \varphi_i(x) u_{2Q_i}\Big|^p\,dx \leqslant c\sum_{i=1}^\infty \int_{2Q_i} |u_{2Q_i}|^p\,dx$$

$$\leqslant c\sum_{i=1}^\infty |2Q_i| \fint_{2Q_i} |u|^p\,dx \leqslant c\int_{E_\lambda} |u|^p\,dx.$$

Then we estimate the gradient. We recall that

$$\varPhi(x) = \sum_{i=1}^\infty \varphi_i(x) = 1$$

for every $x \in E_\lambda$. Since the cubes $2Q_i$, $i = 1, 2, \ldots$, are of bounded overlap, we see that $\varPhi \in C^\infty(E_\lambda)$ and

$$D_j\varPhi(x) = \sum_{i=1}^\infty D_j\varphi_i(x) = 0, \quad j = 1, 2, \ldots, n,$$

for every $x \in E_\lambda$. Hence we obtain

$$|D_j u_\lambda(x)| = \Big|\sum_{i=1}^\infty D_j\varphi_i(x) u_{2Q_i}\Big| = \Big|\sum_{i=1}^\infty D_j\varphi_i(x)(u(x) - u_{2Q_i})\Big|$$

$$\leqslant c\sum_{i=1}^\infty \operatorname{diam}(Q_i)^{-1}|u(x) - u_{2Q_i}|\chi_{2Q_i}(x)$$

and, consequently,

$$|D_j u_\lambda(x)| \leqslant c\sum_{i=1}^\infty \operatorname{diam}(Q_i)^{-p}|u(x) - u_{2Q_i}|^p\chi_{2Q_i}(x).$$

Here we again used the fact that the cubes $2Q_i$, $i = 1, 2, \ldots$, are of bounded overlap.

This implies that for every $j = 1, 2, \ldots, n$

$$\int_{E_\lambda} |D_j u_\lambda|\,dx \leqslant c\int_{E_\lambda} \Big(\sum_{i=1}^\infty \operatorname{diam}(Q_i)^{-p}|u - u_{2Q_i}|^p\chi_{2Q_i}\Big)\,dx$$

$$\leqslant \sum_{i=1}^\infty \int_{2Q_i} \operatorname{diam}(Q_i)^{-p}|u - u_{2Q_i}|^p\,dx$$

$$\leqslant c \sum_{i=1}^{\infty} \int_{2Q_i} |Du|^p \, dx \leqslant c \int_{E_\lambda} |Du|^p \, dx.$$

Then we show that $u_\lambda \in W^{1,p}(\mathbf{R}^n)$. We know that u_λ belongs to $W^{1,p}(E_\lambda)$ and is Hölder continuous in $\mathbf{R}^n$. Moreover, $u \in W^{1,p}(\mathbf{R}^n)$ and $u = u_\lambda$ in $\mathbf{R}^n \setminus E_\lambda$ by (i). This implies that $w = u - u_\lambda \in W^{1,p}(E_\lambda)$ and $w = 0$ in $\mathbf{R}^n \setminus E_\lambda$. By the ACL-property, u is absolutely continuous on almost every line segment parallel to the coordinate axes. Take any such a line. Now w is absolutely continuous on the part of the line segment which intersects E_λ. On the other hand, $w = 0$ in the complement of E_λ. Hence the continuity of w in the line segment implies that w is absolutely continuous on the whole line segment.

We have

$$\begin{aligned} \|u - u_\lambda\|_{W^{1,p}(\mathbf{R}^n)} &= \|u - u_\lambda\|_{W^{1,p}(E_\lambda)} \\ &\leqslant \|u\|_{W^{1,p}(E_\lambda)} + \|u_\lambda\|_{W^{1,p}(E_\lambda)} \leqslant c\|u\|_{W^{1,p}(E_\lambda)}. \end{aligned}$$

We leave it as an exercise for the interested reader to show that the function u_λ is Hölder continuous with the exponent $1 - \alpha$ (or see, for example, [27] for details). □

5 Hardy Inequality

In this section, we consider the Hardy inequality, which was originally studied by Hardy in the one-dimensional case. In the higher dimensional case, the Hardy inequality was studied, for example, in [5, 46, 51, 59, 67, 68]. Our approach is mainly based on more recent works [23, 36, 44, 47, 48, 61].

Suppose first that $p > n$, $n < q < p$, $0 \leqslant \alpha < q$, and $\Omega \neq \mathbf{R}^n$ is an open set. Let $u \in C_0^\infty(\Omega)$. Consider the zero extension to $\mathbf{R}^n \setminus \Omega$. Fix $x \in \Omega$ and take $x_0 \in \partial\Omega$ such that

$$|x - x_0| = \operatorname{dist}(x, \partial\Omega) = \delta(x) = R.$$

Denote $\chi = \chi_{B(x_0,2R)}$. By Corollary 4.3,

$$\begin{aligned} |u(x)| &= |u(x) - u(x_0)| \\ &\leqslant c|x - x_0|^{1-n/q} (M_{n/q}(|Du|\chi)(x) + M_{n/q}(|Du|\chi)(x_0)), \end{aligned}$$

where

$$M_{n/q}(|Du|\chi)(x) \leqslant c M_n(|Du|^q \chi)(x)^{1/q} \leqslant \|Du\chi\|_q,$$

and, by the same argument,

$$M_{n/q}(|Du|\chi)(x_0) \leqslant \|Du\chi\|_q.$$

This implies that

$$\begin{aligned}
|u(x)| &\leqslant c|x-x_0|^{1-n/q}\Bigg(\int_{B(x_0,2R)} |Du|^q\,dy\Bigg)^{1/q}\\
&\leqslant cR^{1-\alpha/q}\Bigg(R^{\alpha-n}\int_{B(x,4R)} |Du|^q\,dy\Bigg)^{1/q}\\
&\leqslant c\,\mathrm{dist}(x,\partial\Omega)^{1-\alpha/q}\big(M_{\alpha,4\delta(x)}|Du|^q(x)\big)^{1/q} \qquad (5.1)
\end{aligned}$$

for every $x \in \mathbf{R}^n$ with $c = c(n,q)$. This is a pointwise Hardy inequality. For $u \in W_0^{1,p}(\Omega)$ this inequality holds almost everywhere. Integrating (5.1) with $\alpha = 0$ over Ω and using the maximal function theorem, we arrive at

$$\begin{aligned}
\int_\Omega \left(\frac{|u(x)|}{\mathrm{dist}(x,\partial\Omega)}\right)^p dx &\leqslant c\int_\Omega \big(M|Du|^q(x)\big)^{p/q}\,dx\\
&\leqslant c\int_\Omega |Du(x)|^p\,dx \qquad (5.2)
\end{aligned}$$

for every $u \in W_0^{1,p}(\Omega)$ with $c = c(n,p,q)$. This is a version of the Hardy inequality which is valid for every open sets with nonempty complement if $n < p < \infty$. The case $1 < p \leqslant n$ is more involved since then extra conditions must be imposed on Ω (see [51, Theorem 3]). However, there is a sufficient condition in terms of capacity density of the complement.

A closed set $E \subset \mathbf{R}^n$ is uniformly $p-$fat, $1 < p < \infty$, if there is a constant $\gamma > 0$ such that

$$\mathrm{cap}_p\big(E \cap \overline{B}(x,r), B(x,2r)\big) \geqslant \gamma\,\mathrm{cap}_p\big(\overline{B}(x,r), B(x,2r)\big) \qquad (5.3)$$

for all $x \in E$ and $r > 0$. Here $\mathrm{cap}_p(K,\Omega)$ denotes the variational $p-$capacity

$$\mathrm{cap}_p(K,\Omega) = \inf \int_\Omega |Du(x)|^p\,dx,$$

where the infimum is taken over all $u \in C_0^\infty(\Omega)$ such that $u(x) \geqslant 1$ for every $x \in K$. Here Ω is an open subset of Ω and K is a compact subset of Ω. We recall that

$$\mathrm{cap}_p(\overline{B}(x,r), B(x,2r)) = cr^{n-p},$$

where $c = c(n,p)$.

If $p > n$, then all nonempty closed sets are uniformly p-fat. If there is a constant $\gamma > 0$ such that E satisfies the measure thickness condition

$$|B(x,r) \cap E| \geqslant \gamma |B(x,r)|$$

for all $x \in E$ and $r > 0$, then E is uniformly p-fat for every p with $1 < p < \infty$.

If E is uniformly p-fat for some p, then it is uniformly q-fat for every $q > p$. The fundamental property of uniformly fat sets is the following self improving result due to Lewis [51, Theorem 1]. For another proof see [61, Theorem 8.2].

Theorem 5.4. *Let $E \subset \mathbf{R}^n$ be a closed uniformly $p-$fat set. Then there is $1 < q < p$ such that E is uniformly $q-$fat.*

In the case where $\Omega \subset \mathbf{R}^n$ is an open set such that $\mathbf{R}^n \setminus \Omega$ is uniformly $p-$fat, Lewis [51, Theorem 2] proved that the Hardy inequality holds. We have already seen that the Hardy inequality follows from pointwise inequalities involving the Hardy–Littlewood maximal function if $p > n$. We show that this is also the case $1 < p \leqslant n$.

Theorem 5.5. *Let $1 < p \leqslant n$, $0 \leqslant \alpha < p$, and let $\Omega \subset \mathbf{R}^n$ be an open set such that $\mathbf{R}^n \setminus \Omega$ is uniformly $p-$fat. Suppose that $u \in C_0^\infty(\Omega)$. Then there are constants $c = c(n,p,\gamma)$ and $\sigma > 1$ such that*

$$|u(x)| \leqslant c \operatorname{dist}(x,\partial\Omega)^{1-\alpha/p} \big(M_{\alpha,\sigma\delta(x)} |Du|^p(x)\big)^{1/p} \tag{5.6}$$

for every $x \in \Omega$.

Proof. Let $x \in \Omega$. Choose $x_0 \in \partial\Omega$ such that

$$|x - x_0| = \operatorname{dist}(x,\partial\Omega) = \delta(x) = R.$$

Then

$$|u(x) - u_{B(x_0,2R)}| \leqslant cR^{1-\alpha/p}\big(M_{\alpha,R}|Du|^p(x)\big)^{1/p}$$

for every $x \in B(x_0,2R)$ with $c = c(n,p)$, and hence

$$\begin{aligned} |u(x)| &\leqslant |u(x) - u_{B(x_0,2R)}| + |u_{B(x_0,2R)}| \\ &\leqslant cR^{1-\alpha/p}\big(M_{\alpha,\delta(x)}|Du|^p(x)\big)^{1/p} + |u|_{B(x_0,2R)} \end{aligned}$$

for every $x \in B(x_0,2R)$. Denote $A = \{x \in \mathbf{R}^n : u(x) = 0\}$. Using a capacitary version of the Poincaré inequality, we arrive at

$$\frac{1}{|B(x_0,2R)|} \int\limits_{B(x_0,2R)} |u|\, dy$$

$$\leqslant c\left(\operatorname{cap}_p\left(A\cap\overline{B}(x_0,2R),B(x_0,4R)\right)^{-1}\int\limits_{B(x_0,4R)}|Du|^p\,dy\right)^{1/p}$$

$$\leqslant c\left(\operatorname{cap}_p\left((\mathbf{R}^n\setminus\Omega)\cap\overline{B}(x_0,2R),B(x_0,4R)\right)^{-1}\int\limits_{B(x_0,4R)}|Du|^p\,dy\right)^{1/p}$$

$$\leqslant c\left(R^{p-n}\int\limits_{B(x,8R)}|Du|^p\,dy\right)^{1/p}\leqslant cR^{1-\alpha/p}\left(M_{\alpha,8\delta(x)}|Du|^p(x)\right)^{1/p},$$

where $c=c(n,p,\gamma)$. □

If $\mathbf{R}^n\setminus\Omega$ is $p-$fat, then, by Theorem 5.4, it is $q-$fat for some $1<q<p\leqslant n$. Using (5.6) with $\alpha=0$, we get the pointwise q-Hardy inequality

$$|u(x)|\leqslant c\operatorname{dist}(x,\partial\Omega)\left(M_{\sigma\delta(x)}|Du|^q(x)\right)^{1/q}$$

for every $x\in\Omega$ with $c=c(n,q)$. Integrating and using the maximal function theorem exactly in the same way as in (5.2), we also prove the Hardy inequality in the case $1<p\leqslant n$. Again, a density argument shows that the Hardy inequality holds for every $u\in W_0^{1,p}(\Omega)$. Thus, we have proved the following assertion.

Corollary 5.7. *Let $1<p<\infty$. Suppose that $\Omega\subset\mathbf{R}^n$ is an open set such that $\mathbf{R}^n\setminus\Omega$ is uniformly $p-$fat. If $u\in W_0^{1,p}(\Omega)$, then there is a constant $c=c(n,p,\gamma)$ such that*

$$\int\limits_{\Omega}\left(\frac{|u(x)|}{\operatorname{dist}(x,\partial\Omega)}\right)^p dx\leqslant c\int\limits_{\Omega}|Du(x)|^p\,dx.$$

In particular, if $p>n$, then the inequality holds for every $\Omega\neq\mathbf{R}^n$.

Remark 5.8. The pointwise Hardy inequality is not equivalent to the Hardy inequality since there are open sets for which the Hardy inequality holds for some p, but the pointwise Hardy inequality fails. For example, the punctured ball $B(0,1)\setminus\{0\}$ satisfies the pointwise Hardy inequality only in the case $p>n$, but the usual Hardy inequality also holds when $1<p<n$. When $p=n$, the Hardy inequality fails for this set. This example also shows that the uniform fatness of the complement is not a necessary condition for an open set to satisfy the Hardy inequality since the complement of $B(0,1)\setminus\{0\}$ is not uniformly p-fat when $1<p<n$. If $p=n$, then the Hardy inequality is equivalent to the fact that $\mathbf{R}^n\setminus\Omega$ is uniformly p-fat (see [51, Theorem 3]).

A recent result of Lehrbäck [47] shows that the uniform fatness is not only sufficient, but also necessary condition for the pointwise Hardy inequality (see also [50, 48, 49]). When $n = p = 2$, Sugawa [65] proved that the Hardy inequality is also equivalent to the uniform perfectness of the complement of the domain. Recently this result was generalized in [43] for other values of p. The arguments of [43] are very general. It is also possible to study Hardy inequalities on metric measure spaces (see [9, 32, 43]).

Theorem 5.4 shows that the p-fatness is a self improving result. Next we give a proof of an elegant result of Koskela and Zhong [45] which states that the Hardy inequality is self improving.

Theorem 5.9. *Suppose that the Hardy inequality holds in Ω for some $1 < p < \infty$. Then there exists $\varepsilon > 0$ such that the Hardy inequality holds in Ω for every q with $p - \varepsilon < q \leqslant p$.*

Proof. Let u be a Lipschitz continuous function that vanishes in $\mathbf{R}^n \setminus \Omega$. For $\lambda > 0$ denote

$$F_\lambda = \{x \in \Omega : |u(x)| \leqslant \lambda \operatorname{dist}(x, \partial\Omega) \text{ and } M|Du|(x) \leqslant \lambda\}.$$

We claim that the restriction of u to $F_\lambda \cup (\mathbf{R}^n \setminus \Omega)$ is Lipschitz continuous with a constant $c\lambda$, where $c = c(n)$. If $x, y \in F_\lambda$, then

$$|u(x) - u(y)| \leqslant c|x - y|(M|Du|(x) + M|Du|(y)) \leqslant c\lambda|x - y|$$

by Corollary 4.3. If $x \in F_\lambda$ and $y \in \mathbf{R}^n \setminus \Omega$, then

$$|u(x) - u(y)| = |u(x)| \leqslant \lambda \operatorname{dist}(x, \partial\Omega) \leqslant \lambda|x - y|.$$

This implies that $u|_{F_\lambda \cup (\mathbf{R}^n \setminus \Omega)}$ is Lipschitz continuous with the constant $c\lambda$. We extend the function to the entire space $\mathbf{R}^n$, for example, with the classical McShane extension

$$v(x) = \inf\{u(y) + c\lambda|x - y| : y \in F_\lambda \cup (\mathbf{R}^n \setminus \Omega)\}.$$

The function v is Lipschitz continuous in $\mathbf{R}^n$ with the same constant $c\lambda$ as $u|_{F_\lambda \cup (\mathbf{R}^n \setminus \Omega)}$. Let

$$G_\lambda = \{x \in \Omega : |u(x)| \leqslant \lambda \operatorname{dist}(x, \partial\Omega)\}$$

and

$$E_\lambda = \{x \in \Omega : M|Du|(x) \leqslant \lambda\}.$$

Then $F_\lambda = G_\lambda \cap E_\lambda$, and we note that

$$|Dv(x)| \leqslant |Du(x)|\chi_{F_\lambda}(x) + c\lambda\chi_{\Omega \setminus F_\lambda}(x)$$

for almost all $x \in \mathbf{R}^n$. By the Hardy inequality,

$$\int_{F_\lambda} \left(\frac{|v(x)|}{\operatorname{dist}(x,\partial\Omega)}\right)^p dx \leqslant c \int_{F_\lambda} |Du(x)|^p\, dx + c\lambda^p|\Omega \setminus F_\lambda|$$

and, consequently,

$$\begin{aligned}
&\int_{G_\lambda} \left(\frac{|u(x)|}{\operatorname{dist}(x,\partial\Omega)}\right)^p dx \\
&\leqslant c \int_{F_\lambda} |Du(x)|^p\, dx + c\lambda^p|\Omega \setminus F_\lambda| + \int_{G_\lambda\setminus E_\lambda} \left(\frac{|u(x)|}{\operatorname{dist}(x,\partial\Omega)}\right)^p dx \\
&\leqslant c \int_{E_\lambda} |Du(x)|^p\, dx + c\lambda^p(|\Omega \setminus G_\lambda| + |\Omega \setminus E_\lambda|).
\end{aligned}$$

From this we obtain

$$\begin{aligned}
\frac{1}{\varepsilon}\int_{\Omega} \left(\frac{|u(x)|}{\operatorname{dist}(x,\partial\Omega)}\right)^{p-\varepsilon} dx &= \int_0^\infty \lambda^{-\varepsilon-1} \int_{G_\lambda} \left(\frac{|u(x)|}{\operatorname{dist}(x,\partial\Omega)}\right)^p dx\, d\lambda \\
&\leqslant c \int_0^\infty \lambda^{-\varepsilon-1} \int_{E_\lambda} |Du(x)|^p\, dx\, d\lambda \\
&+ c \int_0^\infty \lambda^{p-\varepsilon-1}|\Omega \setminus G_\lambda|\, d\lambda + c \int_0^\infty \lambda^{p-\varepsilon-1}|\Omega \setminus E_\lambda|\, d\lambda \\
&\leqslant \frac{c}{\varepsilon}\int_{\Omega} |Du(x)|^{p-\varepsilon}\, dx + \frac{1}{p-\varepsilon}\int_{\Omega} \left(\frac{|u(x)|}{\operatorname{dist}(x,\partial\Omega)}\right)^{p-\varepsilon} dx \\
&+ \frac{c}{p-\varepsilon}\int_{\Omega} (M|Du|(x))^{p-\varepsilon}\, dx.
\end{aligned}$$

The claim follows from this by using the maximal function theorem, choosing $\varepsilon > 0$ small enough, and absorbing the terms on the left-hand side. □

6 Maximal Functions on Metric Measure Spaces

In this section, we show that most of the results that we have discussed so far are based on a general principle and our arguments apply in the context of metric measure spaces.

6.1 Sobolev spaces on metric measure spaces

Let $X = (X, d, \mu)$ be a complete metric space endowed with a metric d and a Borel regular measure μ such that $0 < \mu(B(x,r)) < \infty$ for all open balls

$$B(x,r) = \{y \in X : d(y,x) < r\}$$

with $r > 0$.

The measure μ is said to be doubling if there exists a constant $c_\mu \geqslant 1$, called the doubling constant of μ, such that

$$\mu(B(x,2r)) \leqslant c_\mu \mu(B(x,r))$$

for all $x \in X$ and $r > 0$. Note that an iteration of the doubling property implies that, if $B(x,R)$ is a ball in X, $y \in B(x,R)$, and $0 < r \leqslant R < \infty$, then

$$\frac{\mu(B(y,r))}{\mu(B(x,R))} \geqslant c\Big(\frac{r}{R}\Big)^Q \tag{6.1}$$

for some $c = c(c_\mu)$ and $Q = \log c_\mu / \log 2$. The exponent Q serves as a counterpart of dimension related to the measure.

A nonnegative Borel function g on X is said to be an upper gradient of a function $u : X \to [-\infty, \infty]$ if for all rectifiable paths γ joining points x and y in X we have

$$|u(x) - u(y)| \leqslant \int\limits_\gamma g\, ds, \tag{6.2}$$

whenever both $u(x)$ and $u(y)$ are finite, and $\int\limits_\gamma g\, ds = \infty$ otherwise. The assumption that g is a Borel function is needed in the definition of the path integral. If g is merely a μ-measurable function and (6.2) holds for p-almost every path (i.e., it fails only for a path family with zero p-modulus), then g is said to be a p-weak upper gradient of u. If we redefine a p-weak upper gradient on a set of measure zero we obtain a p-weak upper gradient of the same function. In particular, this implies that, after a possible redefinition on a set of measure zero, we obtain a Borel function. If g is a p-weak upper gradient of u, then there is a sequence g_i, $i = 1, 2, \dots$, of the upper gradients of u such that

$$\int\limits_X |g_i - g|^p \, d\mu \to 0$$

as $i \to \infty$. Hence every p-weak upper gradient can be approximated by upper gradients in the $L^p(X)$-norm. If u has an upper gradient that belongs to $L^p(X)$, then it has a minimal p-weak upper gradient g_u in the sense that for every p-weak upper gradient g of u, $g_u \leqslant g$ μ-almost everywhere.

We define Sobolev spaces on the metric space X using the p-weak upper gradients. For $u \in L^p(X)$ we set

$$\|u\|_{N^{1,p}(X)} = \Big(\int_X |u|^p \, d\mu + \inf_g \int_X g^p \, d\mu \Big)^{1/p},$$

where the infimum is taken over all p-weak upper gradients of u. The Sobolev space (sometimes called the Newtonian space) on X is the quotient space

$$N^{1,p}(X) = \{u : \|u\|_{N^{1,p}(X)} < \infty\}/\sim,$$

where $u \sim v$ if and only if $\|u - v\|_{N^{1,p}(X)} = 0$. The notion of a p-weak upper gradient is used to prove that $N^{1,p}(X)$ is a Banach space. For properties of Sobolev spaces on metric measure spaces we refer to [30, 29, 62, 63, 6].

The p-capacity of a set $E \subset X$ is the number

$$\operatorname{cap}_p(E) = \inf \|u\|^p_{N^{1,p}(X)},$$

where the infimum is taken over all $u \in N^{1,p}(X)$ such that $u = 1$ on E [38]. We say that a property regarding points in X holds p-quasieverywhere (p-q.e.) if the set of points for which the property does not hold has capacity zero. If $u \in N^{1,p}(X)$, then $u \sim v$ if and only if $u = v$ p-q.e. Moreover, if $u, v \in N^{1,p}(X)$ and $u = v$ μ-a.e., then $u \sim v$. Hence the capacity is the correct gauge for distinguishing between two Newtonian functions (see [8]).

To be able to compare the boundary values of Sobolev functions, we need a Sobolev space with zero boundary values. Let E be a measurable subset of X. The Sobolev space with zero boundary values is the space

$$N^{1,p}_0(E) = \{u|_E : u \in N^{1,p}(X) \text{ and } u = 0 \text{ } p\text{-q.e. in } X \setminus E\}.$$

The space $N^{1,p}_0(E)$ equipped with the norm inherited from $N^{1,p}(X)$ is a Banach space.

We say that X supports a weak $(1,p)$-Poincaré inequality if there exist constants $c > 0$ and $\lambda \geqslant 1$ such that for all balls $B(x,r) \subset X$, all locally integrable functions u on X and for all p-weak upper gradients g of u,

$$\fint_{B(x,r)} |u - u_{B(x,r)}| \, d\mu \leqslant cr \Big(\fint_{B(x,\lambda r)} g^p \, d\mu \Big)^{1/p}, \tag{6.3}$$

where

$$u_{B(x,r)} = \fint_{B(x,r)} u \, d\mu = \frac{1}{\mu(B(x,r))} \int_{B(x,r)} u \, d\mu.$$

Since the p-weak upper gradients can be approximated by upper gradients in the $L^p(X)$-norm, we could require the Poincaré inequality for upper gradients as well.

By the Hölder inequality, it is easy to see that if X supports a weak $(1,p)$-Poincaré inequality, then it supports a weak $(1,q)$-Poincaré inequality for every $q > p$. If X is complete and μ doubling then it is shown in [31] that a weak $(1, p)$-Poincaré inequality implies a weak $(1, q)$-Poincaré inequality for some $q < p$. Hence the $(1, p)$-Poincaré inequality is a self improving condition. For simplicity, we assume throughout that X supports a weak $(1, 1)$-Poincaré inequality, although, by using the results of [31], it would be enough to assume that X supports a weak $(1, p)$-Poincaré inequality. We leave the extensions to the interested reader. In addition, we assume that X is complete and μ is doubling. This implies, for example, that Lipschitz functions are dense in $N^{1,p}(X)$ and the Sobolev embedding theorem holds.

6.2 Maximal function defined on the whole space

The standard centered Hardy–Littlewood maximal function on a metric measure space X is defined as

$$Mu(x) = \sup_{r>0} \fint_{B(x,r)} |u| \, d\mu.$$

By the Hardy–Littlewood maximal function theorem for doubling measures (see [15]), we see that the Hardy–Littlewood maximal operator is bounded on $L^p(X)$ when $1 < p \leqslant \infty$ and maps $L^1(X)$ into the weak $L^1(X)$. However, the standard Hardy–Littlewood maximal function does not seem to preserve the smoothness of the functions as examples by Buckley [12] clearly indicate. In order to have a maximal function which preserves, for example, the Sobolev spaces on metric measure spaces, we construct a maximal function based on a discrete convolution.

Let $r > 0$. We begin by constructing a family of balls which cover the space and are of bounded overlap. Indeed, there is a family of balls $B(x_i, r)$, $i = 1, 2, \ldots$, such that

$$X = \bigcup_{i=1}^{\infty} B(x_i, r)$$

and

$$\sum_{i=1}^{\infty} \chi_{B(x_i,6r)} \leqslant c < \infty.$$

This means that the dilated balls $B(x_i, 6r)$ are of bounded overlap. The constant c depends only on the doubling constant and, in particular, is independent of r. These balls play the role of Whitney cubes in a metric measure space.

Then we construct a partition of unity subordinate to the cover $B(x_i, r)$, $i = 1, 2, \ldots$, of X. Indeed, there is a family of functions φ_i, $i = 1, 2, \ldots$, such that $0 \leqslant \varphi_i \leqslant 1$, $\varphi_i = 0$ on $X \setminus B(x_i, 6r)$, $\varphi_i \geqslant c$ on $B(x_i, 3r)$, φ_i is Lipschitz with constant c/r_i with c depending only on the doubling constant, and

$$\sum_{i=1}^{\infty} \varphi_i = 1$$

in X. The partition of unity can be constructed by first choosing auxiliary cutoff functions $\widetilde{\varphi}_i$ so that $0 \leqslant \widetilde{\varphi}_i \leqslant 1$, $\widetilde{\varphi}_i = 0$ on $X \setminus B(x_i, 6r)$, $\widetilde{\varphi}_i = 1$ on $B(x_i, 3r)$ and each $\widetilde{\varphi}_i$ is Lipschitz with constant c/r. We can, for example, take

$$\widetilde{\varphi}_i(x) = \begin{cases} 1, & x \in B(x_i, 3r), \\ 2 - \dfrac{d(x, x_i)}{3r}, & x \in B(x_i, 6r) \setminus B(x_i, 3r), \\ 0, & x \in X \setminus B(x_i, 6r). \end{cases}$$

Then we can define the functions φ_i, $i = 1, 2, \dots$, in the partition of unity by

$$\varphi_i(x) = \frac{\widetilde{\varphi}_i(x)}{\sum\limits_{j=1}^{\infty} \widetilde{\varphi}_j(x)}.$$

It is not difficult to see that the defined functions satisfy the required properties.

Now we are ready to define the approximation of u at the scale of $3r$ by setting

$$u_r(x) = \sum_{i=1}^{\infty} \varphi_i(x) u_{B(x_i, 3r)}$$

for every $x \in X$. The function u_r is called the discrete convolution of u. The partition of unity and the discrete convolution are standard tools in harmonic analysis on homogeneous spaces (see, for example, [15] and [57]).

Let r_j, $j = 1, 2, \dots$, be an enumeration of the positive rational numbers. For every radius r_j we choose balls $B(x_i, r_j)$, $i = 1, 2, \dots$, of X as above. Observe that for each radius there are many possible choices for the covering, but we simply take one of those. We define the discrete maximal function related to the coverings $B(x_i, r_j)$, $i, j = 1, 2, \dots$, by

$$M^* u(x) = \sup_j |u|_{r_j}(x)$$

for every $x \in X$. We emphasize the fact that the defined maximal operator depends on the chosen coverings. This is not a serious matter since we obtain estimates which are independent of the chosen coverings.

As the supremum of continuous functions, the discrete maximal function is lower semicontinuous and hence measurable. The following result shows that the discrete maximal function is equivalent with two-sided estimates to the standard Hardy–Littlewood maximal function.

Lemma 6.4. *There is a constant $c \geqslant 1$, which depends only on the doubling constant, such that*

$$c^{-1}Mu(x) \leqslant M^*u(x) \leqslant cMu(x)$$

for every $x \in X$.

Proof. We begin by proving the second inequality. Let $x \in X$, and let r_j be a positive rational number. Since $\varphi_i = 0$ on $X \setminus B(x_i, 6r_j)$ and $B(x_i, 3r_j) \subset B(x, 9r_j)$ for every $x \in B(x_i, 6r_j)$, we have, by the doubling condition,

$$\begin{aligned} |u|_{r_j}(x) &= \sum_{i=1}^{\infty} \varphi_i(x) |u|_{B(x_i,3r_j)} \\ &\leqslant \sum_{i=1}^{\infty} \varphi_i(x) \frac{\mu(B(x, 9r_j))}{\mu(B(x_i, 3r_j))} \fint_{B(x,9r_j)} |u| \, d\mu \leqslant cMu(x), \end{aligned}$$

where c depends only on the doubling constant c_μ. The second inequality follows by taking the supremum on the left-hand side.

To prove the first inequality, we observe that for each $x \in X$ there exists $i = i_x$ such that $x \in B(x_i, r_j)$. This implies that $B(x, r_j) \subset B(x_i, 2r_j)$ and hence

$$\begin{aligned} \fint_{B(x,r_j)} |u| \, d\mu \leqslant c \fint_{B(x_i,3r_j)} |u| \, d\mu \\ \leqslant c\varphi_i(x) \fint_{B(x_i,3r_j)} |u| \, d\mu \leqslant cM^*u(x). \end{aligned}$$

In the second inequality, we used the fact that $\varphi_i \geqslant c$ on $B(x_i, r_j)$. Again, the claim follows by taking the supremum on the left-hand side. □

Since the maximal operators are comparable, we conclude that the maximal function theorem holds for the discrete maximal operator as well. Our goal is to show that the operator M^* preserves the smoothness of the function in the sense that it is a bounded operator in $N^{1,p}(X)$. We begin by proving the corresponding result for the discrete convolution in a fixed scale.

Lemma 6.5. *Suppose that $u \in N^{1,p}(X)$ with $p > 1$. Let $r > 0$. Then $|u|_r \in N^{1,p}(X)$ and there is a constant c, which depends only on the doubling constant, such that cM^*g_u is a p-weak upper gradient of $|u|_r$ whenever g_u is a p-weak upper gradient of u.*

Proof. By Lemma 6.4, we have $|u|_r \leqslant cMu$. By the maximal function theorem with $p > 1$, we conclude that $|u|_r \in L^p(X)$.

Then we consider the upper gradient. We have

$$\begin{aligned} |u|_r(x) &= \sum_{i=1}^{\infty} \varphi_i(x)|u|_{B(x_i,3r)} \\ &= |u(x)| + \sum_{i=1}^{\infty} \varphi_i(x)\big(|u|_{B(x_i,3r)} - |u(x)|\big). \end{aligned}$$

Observe that, at each point, the sum is taken only over finitely many balls so that the convergence of the series is clear. Let $g_{|u|}$ be a p-weak upper gradient of $|u|$. Then

$$g_{|u|} + \sum_{i=1}^{\infty} g_{\varphi_i(|u|_{B(x_i,3r)}-|u|)}$$

is a p-weak upper gradient of $|u|_r$. On the other hand,

$$\Big(\frac{c}{r}\big||u| - |u|_{B(x_i,3r)}\big| + g_{|u|}\Big)\chi_{B(x_i,6r)}$$

is a p-weak upper gradient of $\varphi_i(|u|_{B(x_i,3r)} - |u|)$. Let

$$g_r = g_u + \sum_{i=1}^{\infty} \Big(\frac{c}{r}\big||u| - |u|_{B(x_i,3r)}\big| + g_u\Big)\chi_{B(x_i,6r)}.$$

Then g_r is a p-weak upper gradient of $|u|_r$. Here we used the fact that every p-weak upper gradient of u will do as a p-weak upper gradient of $|u|$ as well.

Then we show that $g_r \in L^p(X)$. Let $x \in B(x_i, 6r)$. Then $B(x_i, 3r) \subset B(x, 9r)$ and

$$\big||u(x)| - |u|_{B(x_i,3r)}\big| \leqslant \big||u(x)| - |u|_{B(x,9r)}\big| + \big||u|_{B(x,9r)} - |u|_{B(x_i,3r)}\big|.$$

We estimate the second term on the right-hand side by the Poincaré inequality and the doubling condition as

$$\begin{aligned} \big||u|_{B(x,9r)} - |u|_{B(x_i,3r)}\big| &\leqslant \fint_{B(x_i,3r)} \big||u| - |u|_{B(x,9r)}\big|\,d\mu \\ &\leqslant c \fint_{B(x,9r)} \big||u| - |u|_{B(x,9r)}\big|\,d\mu \leqslant cr \fint_{B(x,9r)} g_u\,d\mu. \end{aligned}$$

The first term on the right-hand side is estimated by a standard telescoping argument. Since μ-almost every point is a Lebesgue point for u, we have

$$\big||u(x)| - |u|_{B(x,9r)}\big| \leqslant \sum_{j=0}^{\infty} \big||u|_{B(x,3^{2-j}r)} - |u|_{B(x,3^{1-j}r)}\big|$$

$$\leqslant c\sum_{j=0}^{\infty} \fint_{B(x,3 2^{-j}r)} \big||u| - |u|_{B(x,3 2^{-j}r)}\big|\,d\mu$$

$$\leqslant c\sum_{j=0}^{\infty} 3 2^{-j} r \fint_{B(x,3 2^{-j}r)} g_u\,d\mu \leqslant crMg_u(x)$$

for μ-almost all $x \in X$. Here we used the Poincaré inequality and the doubling condition again. Hence we have

$$\big||u(x)| - |u|_{B(x_i,3r)}\big| \leqslant cr \fint_{B(x,9r)} g_u\,d\mu + crMg_u(x) \leqslant crMg_u(x)$$

for μ-almost all $x \in X$. From this we conclude that

$$g_r = g_u + \sum_{i=1}^{\infty}\Big(\frac{c}{r}\big||u| - |u|_{B(x_i,3r)}\big| + g_u\Big)\chi_{B(x_i,6r)} \leqslant cMg_u(x)$$

for μ-almost all $x \in X$. Here c depends only on the doubling constant. This implies that cMg_u is a p-weak upper gradient of u_r. The maximal function theorem shows that $g_r \in L^p(X)$ since $p > 1$. □

Now we are ready to conclude that the discrete maximal operator preserves Newtonian spaces. We use the following simple fact in the proof. Suppose that u_i, $i = 1, 2, \ldots$, are functions and g_i, $i = 1, 2, \ldots$, are p-weak upper gradients of u_i respectively. Let $u = \sup_i u_i$, and let $g = \sup_i g_i$. If $u < \infty$ $\mu-$almost everywhere, then g is a p-weak upper gradient of u. For the proof, we refer to [6]. The following result is a counterpart of Theorem 2.2 in metric measure spaces.

Theorem 6.6. *If $u \in N^{1,p}(X)$ with $p > 1$, then $M^*u \in N^{1,p}(X)$. In addition, the function cM^*g_u is a p-weak upper gradient of M^*u whenever g_u is a p-weak upper gradient of u. The constant c depends only on the doubling constant.*

Proof. By the maximal function theorem, we see that $M^*u \in L^p(X)$ and, in particular, $M^*u < \infty$ μ-almost everywhere. Since

$$M^*u(x) = \sup_j |u|_{r_j}(x)$$

and cM^*g_u is an upper gradient of $|u|_{r_j}$ for every j, we conclude that it is an upper gradient of M^*u as well. The claim follows from the maximal function theorem. □

Remark 6.7. (i) By Theorem 6.6 and the Hardy–Littlewood maximal theorem, we conclude that the discrete maximal operator M^* is bounded in $N^{1,p}(X)$.

(ii) The fact that the maximal operator is bounded in $N^{1,p}(X)$ can be used to prove a capacitary weak type estimate in metric spaces. This implies that $u \in N^{1,p}(X)$ has Lebesgue points outside a set of p-capacity zero (see [7] and [34]).

6.3 Maximal function defined on a subdomain

This subsection is based on [1]. We recall the following Whitney type covering theorem (see [15] and [57]).

Lemma 6.8. *Let $\Omega \subset X$ be an open set with a nonempty complement. Then for every $0 < t < 1$ there are balls $B(x_i, r_i) \subset \Omega$, $i = 1, 2, \dots$, such that*

$$\bigcup_{i=1}^{\infty} B(x_i, r_i) = \Omega,$$

for every $x \in B(x_i, 6r_i)$, $i = 1, 2, \dots$, we have

$$c_1 r_i \leqslant t \operatorname{dist}(x, X \setminus \Omega) \leqslant c_2 r_i$$

and the balls $B(x_i, 6r_i)$, $i = 1, 2, \dots$, are of bounded overlap. Here the constants c_1 and c_2 depend only on the doubling constant. In particular, the bound for the overlap is independent of the scale t.

Let $0 < t < 1$ be a rational number. We consider a Whitney type decomposition of Ω. We construct a partition of unity and discrete convolution related to the Whitney balls exactly in the same way as before. Let t_j, $j = 1, 2, \dots$, be an enumeration of the positive rational numbers. of the interval $(0, 1)$. For every scale t_j we choose a Whitney covering as in Lemma 6.8 and construct a discrete convolution $|u|_{t_j}$. Observe that for each scale there are many possible choices for the covering, but we simply take one of those. We define the discrete maximal function related to the discrete convolution $|u|_{t_j}$ by

$$M^*_\Omega u(x) = \sup\nolimits_j |u|_{t_j}(x)$$

for every $x \in X$. Again, the defined maximal operator depends on the chosen coverings, but this is not a serious matter for the same reason as above. It can be shown that there is a constant $c \geqslant 1$, depending only on the doubling constant, such that

$$M^*_\Omega u(x) \leqslant c M_\Omega u(x)$$

for every $x \in \Omega$.

Here,

$$M_\Omega u(x) = \sup \fint_{B(x,r)} |u| \, d\mu$$

is the standard maximal function related to the open subset $\Omega \subset X$ and the supremum is taken over all balls $B(x,r)$ contained in Ω. There is also an inequality to the reverse direction, but then we have to restrict ourselves in the definition of the maximal function to such balls that $B(x,\sigma r)$ is contained in Ω for some σ large enough. The pointwise inequality implies that the maximal function theorem holds for M^*_Ω as well.

Using a similar argument as above, we can show that, if the measure μ is doubling and the space supports a weak (1,1)-Poincaré inequality, then the maximal operator M^*_Ω preserves the Sobolev spaces $N^{1,p}(\Omega)$ for every open $\Omega \subset X$ when $p > 1$. Moreover,

$$M^*_\Omega : N^{1,p}(\Omega) \to N^{1,p}(\Omega)$$

is a bounded operator when $p > 1$. It is an interesting open question to study the continuity of the operator and the borderline case $p = 1$.

Then we consider the Sobolev boundary values. The following assertion is a counterpart of Theorem 3.12 in metric measure spaces.

Theorem 6.9. *Let $\Omega \subset X$ be an open set. Assume that $u \in N^{1,p}(\Omega)$ with $p > 1$. Then*

$$|u| - M^*_\Omega u \in N^{1,p}_0(\Omega).$$

Proof. Let $0 < t < 1$. Consider the discrete convolution $|u|_t$. Let $x \in \Omega$ with $x \in B(x_i, r_i)$. Using the same telescoping argument as in the proof of Lemma 4.1 and the properties of the Whitney balls we have

$$\big||u|_{B(x_i,3r_i)} - |u(x)|\big| \leqslant c r_i M_\Omega g_u(x) \leqslant ct \operatorname{dist}(x,\partial\Omega) M_\Omega g_u(x).$$

It follows that

$$\begin{aligned}\big||u|_t(x) - |u(x)|\big| &= \Big|\sum_{i=1}^{\infty} \psi_i(x)\big(|u|_{B(x_i,3r_i)} - |u(x)|\big)\Big| \\ &\leqslant \sum_{i=1}^{\infty} \psi_i(x)\big||u|_{B(x_i,3r_i)} - |u(x)|\big| \\ &\leqslant ct \operatorname{dist}(x,\partial\Omega) M_\Omega g_u(x).\end{aligned}$$

For every $x \in \Omega$ there is a sequence t_j, $j = 1, 2, \dots$, of scales such that

$$M^*_\Omega u(x) = \lim_{j\to\infty} |u|_{t_j}(x)$$

This implies that

$$\begin{aligned}\big||u(x)| - M^*_\Omega u(x)\big| &= \lim_{j\to\infty} \big||u(x) - |u|_{t_j}(x)\big| \\ &\leqslant c \operatorname{dist}(x,\partial\Omega) M_\Omega g_u(x),\end{aligned}$$

where we used the fact that $t_j \leqslant 1$. Hence, by the maximal function theorem, we conclude that

$$\int_\Omega \left(\frac{\big||u(x)| - M^*_\Omega u(x)\big|}{\operatorname{dist}(x,\partial\Omega)} \right)^p d\mu(x) \leqslant c \int_\Omega (M_\Omega g_u(x))^p \, d\mu(x) \leqslant c \int_\Omega |g_u(x)|^p \, d\mu(x).$$

This implies that

$$\frac{|u(x)| - M^*_\Omega u(x)}{\operatorname{dist}(x,\partial\Omega)} \in L^p(\Omega)$$

and from Theorem 5.1 in [32] we conclude that $|u| - M^*_\Omega u \in N^{1,p}_0(\Omega)$. □

6.4 Pointwise estimates and Lusin type approximation

Let u be a locally integrable function in X, let $0 \leqslant \alpha < 1$, and let $\beta = 1 - \alpha$. From the proof of Lemma 4.1 it follows that

$$|u(x) - u(y)| \leqslant c\, d(x,y)^\beta \big(u^\#_{\beta,4d(x,y)}(x) + u^\#_{\beta,4d(x,y)}(y) \big)$$

for every $x \neq y$. By the weak Poincaré inequality,

$$u^\#_{\beta,4d(x,y)}(x) \leqslant c M_{\alpha,4\lambda d(x,y)} g_u(x)$$

for every $x \in X$. Denote

$$E_\lambda = \{x \in X : M_\alpha g_u(x) > \lambda\},$$

where $\lambda > 0$. We see that $u|_{X \setminus E_\lambda}$ is Hölder continuous with the exponent β. We can extend this function to a Hölder continuous function on X by using a Whitney type extension. The Whitney type covering lemma (Lemma 6.8) enables us to construct a partition of unity as above. Let $B(x_i, r_i)$, $i = 1, 2, \dots$, be the Whitney covering of the open set E_λ. Then there are nonnegative functions φ_i, $i = 1, 2, \dots$, such that $\varphi_i = 0$ in $X \setminus B(x_i, 6r_i)$, $0 \leqslant \varphi_i(x) \leqslant 1$ for every $x \in X$, every φ_i is Lipschitz with the constant c/r_i and

$$\sum_{i=1}^\infty \varphi_i(x) = \chi_{E_\lambda}(x)$$

for every $x \in X$. We define the Whitney smoothing of u by

$$u_\lambda(x) = \begin{cases} u(x), & x \in X \setminus E_\lambda, \\ \sum_{i=1}^{\infty} \varphi_i(x) u_{B(x_i,3r_i)}, & x \in E_\lambda. \end{cases}$$

We obtain the following result by similar arguments as above. The exponent Q refers to the dimension given by (6.1).

Theorem 6.10. *Suppose that $u \in N^{1,p}(X)$, $1 < p \leqslant Q$. Let $0 \leqslant \alpha < 1$. Then for every $\lambda > 0$ there is a function u_λ and an open set E_λ such that $u = u_\lambda$ everywhere in $X \setminus E_\lambda$, $u_\lambda \in N^{1,p}(X)$, and u_λ is Hölder continuous with the exponent $1 - \alpha$ on every bounded set in X, $\|u - u_\lambda\|_{N^{1,p}(X)} \to 0$, and $\mathcal{H}^{n-\alpha p}_\infty(E_\lambda) \to 0$ as $\lambda \to \infty$.*

References

1. Aalto, D.: The discrete maximal function on a doubling metric space. Licetiate thesis, Helsinki University of Technology (2008)
2. Acerbi, E., Fusco, N.: Semicontinuity problems in the calculus of variations. Arch. Ration. Mech. Anal. **86**, no. 2, 125–145 (1984)
3. Aldaz, J.M., Pérez Lázaro, J.: Functions of bounded variation, the derivative of the one dimensional maximal function, and applications to inequalities. Trans. Am. Math. Soc. **359**, no. 5, 2443–2461 (2007)
4. Aldaz, J.M., Pérez Lázaro, J.: Boundedness and unboundedness results for some maximal operators on functions of bounded variation. `arXiv:math.CA/0605272`
5. Ancona, A.: On strong barriers and an inequality of Hardy for domains in $\mathbf{R}^n$. J. London Math. Soc. (2) **34**, 274–290 (1986)
6. Björn, A., Björn, J.: Nonlinear potential theory on metric spaces [In preparation]
7. Björn, A., Björn, J., Parviainen, M.: Convergence of superharmonic functions [In preparation]
8. Björn, A., Björn, J., Shanmugalingam, N.: Quasicontinuity of Newton–Sobolev Functions and Density of Lipschitz Functions on Metric Spaces. Preprint, Linköping (2006)
9. Björn, J., MacManus, P., Shanmugalingam, N.: Fat sets and pointwise boundary estimates for p-harmonic functions in metric spaces. J. Anal. Math. **85**, 339–369 (2001).
10. Bojarski, B., Hajłasz, P.: Pointwise inequalities for Sobolev functions and some applications. Stud. Math. **106**, 77–92 (1993)
11. Bojarski, B., Hajłasz, P., Strzelecki, P.: Improved $C^{k,\lambda}$ approximation of higher order Sobolev functions in norm and capacity. Indiana Univ. Math. J. **51**, no. 3, 507–540 (2002)
12. Buckley, S.M.: Is the maximal function of a Lipschitz function continuous? Ann. Acad. Sci. Fenn. Math. **24**, 519–528 (1999)
13. Calderón, A.P., Zygmund,A.: Local properties of solutions of elliptic partial differential equations. Stud. Math. **20**, 171–225 (1961)
14. F. Chiarenza, F., Frasca, M.: Morrey spaces and Hardy–Littlewood maximal function. Rend. Mat. Appl. (7) **7**, no. 3–4, 273–279 (1987)
15. Coifman, R.R., Weiss, G.: Analyse Harmonique Non-Commutative sur Certain Espaces Homogènes. Lect. Notes Math. **242**, Springer-Verlag (1971)
16. DeVore, R., Sharpley, R.C.: Maximal functions measuring smoothness. Mem. Am. Math. Soc. no. 293 (1984)

17. Evans, L.C., Gariepy, R.F.: Measure Theory and Fine Properties of Functions. CRC Press (1992)
18. Fiorenza, A.: A note on the spherical maximal function. Rend. Accad. Sci. Fis. Mat. Napoli (4) **54**, 77–83 (1987)
19. Fiorenza, A., Krbec, M.: On the domain and range of the maximal operator. Nagoya Math. J. **158**, 43–61 (2000)
20. Gutiérrez, S.: Lusin approximation of Sobolev functions by Hölder continuous functions. Bull. Inst. Math. Acad. Sinica **31**, no. 2, 95–116 (2003)
21. Hajłasz, P.: Boundary behavior of Sobolev mappings. Proc. Am. Math. Soc. **123**, 1145–1148 (1995)
22. Hajłasz, P.: A new characterization of the Sobolev space. (English summary) Dedicated to Professor Aleksander Pełczyński on the occasion of his 70th birthday (Polish). Stud. Math. **159**, no. 2, 263–275 (2003)
23. Hajłasz, P.: Pointwise Hardy inequalities. Proc. Am. Math. Soc. **127**, 417–423 (1999)
24. Hajłasz, P.: Sobolev spaces on an arbitrary metric space. Potential Anal. **5**, 403–415 (1996)
25. Hajłasz, P., Kinnunen, J.: Hölder quasicontinuity of Sobolev functions on metric spaces. Rev. Mat. Iberoam. **14**, 601–622 (1998)
26. Hajłasz, P., Onninen, J.: On boundedness of maximal functions in Sobolev spaces. Ann. Acad. Sci. Fenn. Math. **29**, 167–176 (2004)
27. Harjulehto, P., Kinnunen, J., Tuhkanen, K.: Hölder quasicontinuity in variable exponent Sobolev spaces. J. Inequal. Appl. (2007) doi:10.1155/2007/32324.
28. Hedberg, L.I.: On certain convolution inequalities. Proc. Am. Math. Soc. **36**, 505–510 (1972)
29. Heinonen, J., Kilpeläinen, T., Martio, O.: Nonlinear Potential Theory of Denegerate Elliptic Equations. Oxford Univ. Press, Oxford (1993)
30. Heinonen, J., Koskela, P.: Quasiconformal maps in metric spaces with controlled geometry. Acta Math. **181**, no. 1, 1–61 (1998)
31. Keith, S., Zhong, X.: The Poincaré inequality is an open ended condition. Ann. Math. [To appear]
32. Kilpeläinen, T., Kinnunen, J., Martio, O: Sobolev spaces with zero boundary values on metric spaces. Potential Anal. **12**, no. 3, 233–247 (2000)
33. Kinnunen, J.: The Hardy–Littlewood maximal function of a Sobolev function. Israel J. Math. **100**, 117–124 (1997)
34. Kinnunen, J., Latvala, V.: Lebesgue points for Sobolev functions on metric spaces. Rev. Mat. Iberoam. **18**, 685–700 (2002)
35. Kinnunen, J., Lindqvist, P.: The derivative of the maximal function. J. Reine Angew. Math. **503**, 161–167 (1998)
36. Kinnunen, J., Martio, O.: Hardy's inequalities for Sobolev functions. Math. Research Lett. **4**, 489–500 (1997)
37. Kinnunen, J., Martio, O.: Maximal operator and superharmonicity. In: Function Spaces, Differential Operators and Nonlinear Analysis (Pudasjärvi, 1999), pp. 157–169. Acad. Sci. Czech Repub., Prague (2000)
38. Kinnunen, J., Martio, O.: The Sobolev capacity on metric spaces. Ann. Acad. Sci. Fenn. Math. **21**, no. 2, 367–382 (1996)
39. Kinnunen, J., Saksman, E.: Regularity of the fractional maximal function. Bull. London Math. Soc. **35**, 529–535 (2003)
40. Korry, S.: A class of bounded operators on Sobolev spaces. Arch. Math. (Basel) **82**, no. 1, 40–50 (2004)
41. Korry, S.: Boundedness of Hardy–Littlewood maximal operator in the framework of Lizorkin–Triebel spaces. Rev. Mat. Comp. **15**, no. 2, 401–416 (2002)

42. Korry, S.: Fixed points of the Hardy–Littlewood maximal operator. Collect. Math. **52**, no. 3, 289–294 (2001)
43. Korte, R., Shanmugalingam, N.: Equivalence and Self-Improvement of p-Fatness and Hardy's Inequality, and Association with Uniform Perfectness. Preprint (2007)
44. Koskela, P., Lehrbäck, J.: Weighted Pointwise Hardy Inequalities. University of Jyväskylä, Preprint 332 (2007)
45. Koskela, P., Zhong, X.: Hardy's inequality and the boundary size. Proc. Am. Math. Soc. **131**, no. 4, 1151–1158 (2003)
46. Kufner, A.: Weighted Sobolev spaces. John Wiley & Sons, Inc. New York (1985)
47. Lehrbäck, J.: Pointwise Hardy Inequalities and Uniformly Fat Sets. University of Jyväskylä, Preprint 342 (2007)
48. Lehrbäck, J.: Self-Improving Properties of Weighted Hardy Inequalities. University of Jyväskylä, Preprint 348 (2007)
49. Lehrbäck, J.: Necessary Conditions for Weighted Pointwise Hardy Inequalities. Preprint (2008)
50. Lewis, J.: On very weak solutions of certain elliptic systems. Commun. Partial Differ. Equ. **18**, no. 9–10, 1515–1537 (1993)
51. Lewis, J.L.: Uniformly fat sets. Trans. Am. Math. Soc. **308**, 177–196 (1988)
52. Liu, F.C.: A Luzin type property of Sobolev functions. Indiana Univ. Math. J. **26**, no. 4, 645–651 (1977)
53. Luiro, H.: Continuity of maximal operator in Sobolev spaces. Proc. Am. Math. Soc. **135**, 243–251 (2007)
54. Luiro, H.: The Regularity of the Hardy–Littlewood Maximal Operator on Subdomains of $\mathbf{R}^n$. Preprint (2008)
55. Luiro, H.: On the Continuous and Discontinuous Maximal Operators. Preprint (2008)
56. Macías, R.A., Segovia, C.: Lipschitz functions on spaces of homogeneous type. Adv. Math. **33**, 257–270 (1979)
57. Macías, R.A., Segovia, C.: A decomposition into atoms of distributions on spaces of homogeneous type. Adv. Math. **33** 271–309 (1979)
58. Malý, J.: Hölder type quasicontinuity. Potential Anal. **2**, 249–254 (1993)
59. Maz'ja, V.G.: Sobolev Spaces. Springer-Verlag, Berlin (1985)
60. Michael, J.H., Ziemer, W.P.: A Lusin type approximation of Sobolev functions by smooth functions. In: Classical Real Analysis (Madison, Wis., 1982), pp. 135–167. Contemp. Math. **42**. Am. Math. Soc., Providence, RI (1985)
61. Mikkonen, P.: On the Wolff potential and quasilinear elliptic equations involving measures. Ann. Acad. Sci. Fenn. Ser. A I Math. Diss. **104** (1996)
62. Shanmugalingam, N.: Newtonian spaces: an extension of Sobolev spaces to metric measure spaces. Rev. Mat. Iberoam. **16**, no. 2, 243–279 (2000)
63. Shanmugalingam, N.: Harmonic functions on metric spaces. Ill. J. Math. **45**, 1021–1050 (2001)
64. Stein, E.M.: Singular Integrals and Differentiability Properties of Functions. Princeton Univ. Press, Princeton, NJ (1970)
65. Sugawa, T.: Uniformly perfect sets: analytic and geometric aspects. Sugaku Expositions **16**, no. 2, 225–242 (2003)
66. Tanaka, H.: A remark on the derivative of the one-dimensional Hardy–Littlewood maximal function. Bull. Austral. Math. Soc. **65**, 253–258 (2002)
67. Wannebo, A.: Hardy inequalities. Proc. Am. Math. Soc. **109**, no. 1, 85–95 (1990)
68. Wannebo, A.: Hardy inequalities and imbeddings in domains generalizing $C^{0,\lambda}$ domains. Proc. Am. Math. Soc. **122**, no. 4, 1181–1190 (1994)
69. Whitney, H.: Analytic extensions of differentiable functions defined in closed sets. Trans. Am. Math. Soc. **36**, no. 1, 63–89 (1934)
70. Zhong, X.: On homogeneous quasilinear elliptic equations. Ann. Acad. Sci. Fenn. Math. Diss. **117** (1998)

Hardy Type Inequalities via Riccati and Sturm–Liouville Equations

Sergey Bobkov and Friedrich Götze

Abstract We discuss integral estimates for domain of solutions to some canonical Riccati and Sturm–Liouville equations on the line. The approach is applied to Hardy and Poincaré type inequalities with weights.

1 Introduction

Given a function $V = V(t)$ in $t \geqslant 0$, consider the Riccati equation

$$y'(t) = y(t)^2 + V(t) \tag{1.1}$$

with initial condition

$$y(0) = 0. \tag{1.2}$$

A standard question about (1.1)–(1.2) is how to exactly determine or to estimate in terms of V the length of the maximal interval $[0, t_0)$, $t_0 > 0$, on which a (unique) solution y exists. Known results on estimates for t_0 usually treat more general Cauchy's problems, and being applied to the above special situation, they depend upon the growth of the maximum of $|V|$ on intervals $[0, t]$ with growing t. Throughout the paper, we assume that V is nonnegative, continuous, and is not identically zero. In this case, an important information can be derived by applying suitable comparison arguments, which lead to more sensitive integrable estimates. In particular, we prove the following

Sergey Bobkov
University of Minnesota, Minneapolis, MN 55455, USA, e-mail: bobkov@math.umn.edu

Friedrich Götze
Bielefeld University, Bielefeld 33501, Germany, e-mail: goetze@math.uni-bielefeld.de

V. Maz'ya (ed.), *Sobolev Spaces in Mathematics I, International Mathematical Series.*
doi: 10.1007/978-0-387-85648-3,

Theorem 1.1. *Define*

$$\overline{V}(t) = \int_0^t V(u)\,du, \quad t \geqslant 0.$$

The maximal value of t_0 *satisfies*

$$\frac{1}{4t_0} \leqslant \sup_{0<s<1} \big[(1-s)\overline{V}(t_0 s)\big] \leqslant \frac{1}{t_0}. \tag{1.3}$$

For example, for $V(t) = t^{\alpha-1}$, $\alpha \geqslant 1$, this gives

$$\frac{1}{4^{1/(\alpha+1)}} \frac{\alpha+1}{\alpha^{(\alpha-1)/(\alpha+1)}} \leqslant t_0 \leqslant \frac{\alpha+1}{\alpha^{(\alpha-1)/(\alpha+1)}}.$$

In particular, $t_0 \to 1$ as $\alpha \to +\infty$.

Transforming (1.1) to second order linear differential equations, one may give an equivalent formulation of Theorem 1.1 as a statement about the first eigenvalue λ_0 for the regular Sturm–Liouville equation

$$\frac{d}{dt}\left(q(t)\frac{d}{dt}z(t)\right) = \lambda p(t)z(t), \quad a \leqslant t \leqslant b, \tag{1.4}$$

with boundary conditions $z(a) = z'(b) = 0$. Introduce the quantity

$$A(p,q) = \sup_{a<x<b} \left[\int_a^x \frac{1}{q(t)}\,dt \int_x^b p(t)\,dt\right].$$

Theorem 1.2. *For all positive continuous functions* p *and* q *on* $[a,b]$

$$A(p,q) \leqslant \frac{1}{\lambda_0} \leqslant 4A(p,q). \tag{1.5}$$

We consider the estimates (1.3) and (1.5) as another approach, from differential equation point of view, to a result, obtained by Kac and Krein [7] in 1959 and later by Artola [1], Talenti [14], and Tomaselli [15], about Hardy type inequalities with weights. In these inequalities, one tries to determine or estimate the best constant $C = C(p,q)$ satisfying

$$\int_a^b f(x)^2 p(x)\,dx \leqslant C \int_a^b f'(x)^2 q(x)\,dx, \tag{1.6}$$

where f is an arbitrary absolutely continuous function on $[a,b)$ such that $f(a) = 0$. Their result, including the case $b = +\infty$ as well, asserts that

$$A(p,q) \leqslant C(p,q) \leqslant 4A(p,q) \tag{1.7}$$

(actually, they treated a more general L^α-norm in (1.6)). In 1972, Muckenhoupt [11] gave a complete account on this result and extended it to arbitrary positive measures in place of $p(x)dx$ and $q(x)dx$. In general (when the interval is unbounded), it might occur that $A(p,q)$ is infinite. The property $A(p,q) < +\infty$ is sometimes called the *Muckenhoupt condition*, although the two-sided inequality (1.7) is associated with it, as well. For more general results and references we refer the interested reader to the monograph [9]. The connection of (1.6) with (1.4) is as follows: in the regular case, the extremal functions in Hardy type inequalities exist and satisfy the boundary value problem of Theorem 1.2 with the smallest possible value of λ. In particular, $C(p,q) = 1/\lambda_0$. It should also be clear that, in (1.6) and (1.7), the regular case easily implies the general case, where p and q are defined on the half-axis $(a,+\infty)$.

In a more rigorous manner, we consider the corresponding variational problem in Sect. 4, where Theorem 1.2 is proved and is shown to imply (1.7). In Sect. 2, we prove Theorem 1.1 and some related statements. In Sect. 3, we consider a particular case of Theorem 1.2 with $q \equiv 1$. We finish the paper in Section 5, where we derive an analogue of (1.5) for the boundary conditions $z'(a) = z'(b) = 0$. These conditions turn out to be connected with another important family of inequalities of Poincaré type. The reader may find some results connecting Hardy type inequalities with weights with Poincaré and logarithmic Sobolev inequalities in [2], where Muckenhoupt's characterization was essentially used (see also [10] for discrete analogues).

Theorems 1.1 and 1.2 can easily be extended to more general equations such as $y'(t) = y(t)^\beta + V(t)$ and $(qz'(t)^\beta)' = -\lambda p(t)z^\beta$ respectively. We will not study these equations in order to make easy the presentation of main techniques in the basic case $\alpha = 2$. It should however be noted that these are precisely the equations which are needed for studying the Hardy type inequalities (1.6) with respect to the norms in general Lebesgue spaces (rather than in L^2).

2 Riccati Equations

At first, it is convenient to consider the Riccati equation (1.1) in the semi-open interval $[0,1)$ and assume that V is defined, is nonnegative and continuous on this interval (and is not identically zero). One is looking for some conditions, necessary and sufficient, which would guarantee the existence of a solution to (1.1) on the whole interval $[0,1)$. If it exists (and is thus unique), it should necessarily belong to the class $C^1[0,1)$ of all continuously differentiable functions on $[0,1)$. Introducing the integral operator

$$Af(t) = \int_0^t f(u)^2\,du + \overline{V}(t), \qquad 0 \leqslant t < 1,$$

we may reformulate our task as a problem on the existence of a solution $y = y(t)$ in $C^1[0,1)$ to the nonlinear integral equation

$$Ay = y \tag{2.1}$$

under the initial condition $y(0) = 0$.

A canonical way to construct a solution to the Cauchy problem $y'(t) = \Psi(t, y(t))$ and, in particular, to the problem (1.1), where $\Psi(t,y) = V(t) + y^2$, is to start from a function y_0, recursively defining the sequence

$$y_1 = Ay_0,\ y_2 = Ay_1,\ \ldots,\ y_{n+1} = Ay_n, \quad n \geqslant 0.$$

Certain conditions on V guarantee the convergence of Ay_n to a solution on some interval $[0, t_1)$. One general sufficient condition for convergence (see, for example, [5]) may be formulated as follows. Consider the maximum $M = \max_D \Psi$ on the rectangle $D = [0,\alpha] \times [0,\beta]$. Then one can take

$$t_1 = \min\left\{\alpha, \frac{\beta}{M}\right\} = \min\left\{\alpha, \frac{\beta}{\|V\|_{C[0,\alpha]} + \beta^2}\right\},$$

where $\|V\|_{C[0,\alpha]} = \max_{0 \leqslant t \leqslant \alpha} V(t)$. Optimizing over β so that to maximize t_1, we arrive at

$$t_1(V) = \sup_{0<\alpha<1} \min\left\{\alpha, \frac{1}{2}\ \|V\|_{C[0,\alpha]}^{-1/2}\right\}.$$

Although choosing some other domains D may improve this value t_1 for concrete V, we are in a typical situation where one has to require the boundedness of V on $[0,1)$ in order to reach the value $t_1 = 1$. In particular, the above formula gives $t_1(\lambda V) = 1$ only if $\lambda \leqslant 1/(4 \sup V)$.

Now let us look at the convergence of Ay_n by using some comparison arguments and first derive the following

Lemma 2.1. *A solution to* (2.1) *under the initial condition* $y(0) = 0$ *exists if and only if for some nonnegative measurable function* f *on* $[0,1)$ *for all* $t \in [0,1)$

$$f(t) \geqslant \overline{V}(t) \quad \text{and} \quad Af(t) \leqslant f(t). \tag{2.2}$$

Proof. Clearly, if y is a solution, then $f = y$ satisfies (2.2). To prove the converse, we assume that f satisfies (2.2). We start from $y_0 \equiv 0$ and define a sequence y_n as above. In particular, $y_1 = \overline{V}$. We set

$$y(t) = \sup_n y_n(t).$$

Note that the operator A is monotone: if $0 \leqslant g_1 \leqslant g_2$, then $0 \leqslant Ag_1 \leqslant Ag_2$. Since $\overline{V} \leqslant f$, we get $y_2 = A\overline{V} \leqslant Af \leqslant f$. Repeating the argument (i.e., by induction), we see that $y_n \leqslant f$, for all n. Therefore, the function y is finite, measurable, and satisfies $y \leqslant f$. In addition, for all n, $Ay \geqslant Ay_n = y_{n+1}$, so that $Ay \geqslant y$. On the other hand, the sequence y_n is nondecreasing: $y_2 = Ay_1 \geqslant Ay_0 = y_1$, $y_3 = Ay_2 \geqslant Ay_1 = y_2$, and so on. Thus, $y_n(t) \uparrow y(t)$ as $n \to \infty$. Hence, by the Tonneli monotone convergence theorem, $Ay_n(t) \uparrow Ay(t)$ for all $t \in [0, 1)$. Taking the limit in $Ay_n = y_{n+1} \leqslant y$, we conclude that $Ay \leqslant y$. The two estimates give $Ay = y$.

The function of the form Ay, as soon as it is finite, must be absolutely continuous. Hence y is absolutely continuous, and this implies that y is in $C^1[0, 1)$. □

Remark 2.2. According to the above proof, we may add to Lemma 2.1 another characterization. Consider a pointwise limit of the nondecreasing sequence

$$y_V(t) = \lim_{n\to\infty} [\underbrace{AA \dots A}_{n \text{ times}}\, y_0](t), \quad y_0(t) \equiv 0,$$

which might be finite or not. Then the existence on the interval [0,1) of a solution y to (2.1) under the initial condition (1.2) is equivalent to the property that $y_V(t) < +\infty$, for all $t \in [0, 1)$. In this case, $y = y_V$ provides the solution.

In particular, since from $0 \leqslant V \leqslant W$ it follows that $y_V \leqslant y_W$, the existence of a solution to (1.1) with a function W implies the existence of a solution to (1.1) with any (continuous) function $V \leqslant W$. Such a comparison property was given by Levin [8], who considered even more general situation, where V is not necessarily nonnegative, but still satisfies $|V| \leqslant W$.

The above reformulation also holds when we consider the Riccati equation on a larger interval or the whole half-axis $[0, +\infty)$. Then $t_0 = \sup\{t \geqslant 0 : y_V(t) < +\infty\}$.

We need the following assertion.

Lemma 2.3. *Any solution y to the Riccati equation* (1.1) *under the initial condition $y(0) = 0$ satisfies for all $t \in [0, 1)$*

$$y(t) < \frac{1}{1-t}.$$

Proof. Set $t_1 = \max\{t \in [0, 1) : y(t) = 0\}$. Since y must be nondecreasing and V is not identically zero, the point t_1 is well defined and lies in $[0, 1)$. For $t \in (t_1, 1)$ we have $y'(t) \geqslant y(t)^2 > 0$, which implies that the function $g(t) = \frac{1}{y(t)} + t$ decreases in $(t_1, 1)$. In particular, $g(t) > g(1-) \geqslant 1$. □

Now, we are ready to estimate the supremum $\lambda(V)$ of all $\lambda \geqslant 0$, for which there exists a solution $y = y(t)$ to the Riccati equation

$$y'(t) = y(t)^2 + \lambda\, V(t), \quad 0 \leqslant t < 1, \tag{2.3}$$

under the same initial condition $y(0) = 0$. As a consequence of the two lemmas, we derive the following statement closely related to Theorem 1.1.

Theorem 2.4. *We have*

$$\sup_{0<t<1} \left[(1-t)\overline{V}(t)\right] \leqslant \frac{1}{\lambda(V)} \leqslant 4 \sup_{0<t<1} \left[(1-t)\overline{V}(t)\right]. \tag{2.4}$$

In particular, $\lambda(V) > 0$ if and only if $\overline{V}(t) = O(\frac{1}{1-t})$ as $t \to 1$.

Proof. First, assume that y satisfies (2.3) with $y(0) = 0$. By Lemma 2.3, $\frac{1}{1-t} > y(t)$ in $[0,1)$. Since $y'(t) \geqslant \lambda V(t)$, we also have $y(t) \geqslant \lambda \overline{V}(t)$. Hence

$$\frac{1}{1-t} > \lambda \overline{V}(t).$$

This gives the first inequality in (2.4). To prove the second one, we use Lemma 2.1 with respect to the function λV. Take any $\lambda \geqslant 0$ such that

$$\frac{1}{\lambda} \geqslant 4 \sup_{0<t<1} (1-t)\overline{V}(t),$$

so that

$$\lambda \overline{V}(t) \leqslant \frac{1}{4(1-t)}, \qquad 0 \leqslant t < 1.$$

Then for the function $f(t) = \frac{1}{2(1-t)}$ we get

$$\begin{aligned} Af(t) \equiv \int_0^t f(u)^2\, du + \lambda \overline{V}(t) &= \frac{t}{4(1-t)} + \lambda \overline{V}(t) \\ &\leqslant \frac{t}{4(1-t)} + \frac{1}{4(1-t)} \leqslant \frac{1}{2(1-t)} = f(t). \end{aligned}$$

Thus, $Af(t) \leqslant f(t)$. On the other hand, $f(t) \geqslant \lambda \overline{V}(t)$, so the sufficient conditions of Lemma 2.1 are satisfied. Hence there is a solution y to (2.3) with $y(0) = 0$. This gives the second inequality in (2.4). Hence Theorem 2.4 is proved. □

Proof of Theorem 1.1. It remains to explain why Theorem 1.1 is an immediate consequence of Theorem 2.4. Considering (1.1) on a finite interval $[0, t_1)$ and introducing the functions $z(s) = t_1 y(t_1 s)$, $0 \leqslant s < 1$, we arrive at the Riccati equation on $[0,1)$

$$z'(s) = z(s)^2 + t_1^2 V(t_1 s) \tag{2.5}$$

under the same initial condition $z(0) = 0$. Now, apply Theorem 2.4 to $V_{t_1}(s) = V(t_1 s)$. The existence of a solution z to (2.5) implies that

$$\frac{1}{t_1^2} \geqslant \frac{1}{\lambda(V_{t_1})} \geqslant \sup_{0<s<1} \left[(1-s)\overline{V}_{t_1}(s)\right] = \frac{1}{t_1} \sup_{0<s<1} \left[(1-s)\overline{V}(t_1 s)\right].$$

This leads to the second inequality in (1.3) for any t_1 such that (1.1) has a solution on $[0, t_1)$ with the initial condition (1.2). Let t_0 denote the maximal value t_1 with this property. By the second inequality in (2.4), a solution z to the equation (2.5) exists on $[0, 1)$ if

$$\frac{1}{t_1^2} \geqslant 4 \sup_{0<s<1} [(1-s)\overline{V}_{t_1}(s)],$$

i.e., if

$$\frac{1}{t_1} \geqslant 4 \sup_{0<s<1} [(1-s)\overline{V}(t_1 s)]. \tag{2.6}$$

The right-hand side of (2.6) is nondecreasing and continuous in $t_1 > 0$, so there exists a unique point t_2 which turns this inequality into equality; moreover, for $t > t_2$, we have the converse inequality

$$\frac{1}{t} < 4 \sup_{0<s<1} \left[(1-s)\overline{V}(ts)\right].$$

Since $t_0 \geqslant t_2$, we thus obtain the left inequality in (1.3) and Theorem 1.1 follows. □

3 Transition to Sturm–Liouville Equations

One may equivalently reformulate Theorem 1.1 as a statement about the first zero of solutions to a second order differential equation. Here, we consider only the simplest equation

$$z''(t) = -V(t)z(t), \quad t \geqslant 0, \tag{3.1}$$

under the initial conditions

$$z(0) = 1, \quad z'(0) = 0 \tag{3.2}$$

(the condition $z(0) = 1$ has a matter of normalization, only). As in Theorem 1.1, assume that V is a nonnegative continuous function on $[0, +\infty)$, which is not identically zero. It is well known (see, for example, [12]) that any second order linear differential equation with continuous coefficients and given initial conditions has a unique nontrivial solution. Moreover, on every finite interval, the solution has a finite number of zeros. In the case of (3.1), (3.2) with $V \geqslant 0$

and $V \neq 0$, we may define

$$t_0 = \min\{t > 0 : z(t) = 0\},$$

and one would like to estimate t_0. Since $z(0) > 0$, the function z must be positive on $[0, t_0)$, and we may introduce a new function

$$y(t) = -\frac{z'(t)}{z(t)}, \quad 0 \leqslant t < t_0.$$

It satisfies the Riccati equation (1.1) with initial condition (1.2). Conversely, starting from a function y satisfying (1.1), (1.2) on $[0, t_0)$, one may define the function $z(t) = \exp\{-\int_0^t y(s)ds\}$, which will satisfy (3.1), (3.2) on the same interval. Thus, we may conclude:

Corollary 3.1. *The minimal zero t_0 of the solution z to the problem* (3.1), (3.2) *satisfies*

$$\frac{1}{4t_0} \leqslant \sup_{0<s<1} \left[(1-s)\overline{V}(t_0 s)\right] \leqslant \frac{1}{t_0}.$$

Similarly, we have an equivalent analogue of Theorem 2.4. Assume that V is now defined on $[0, 1)$, is continuous, nonnegative and is not identically zero. Consider in $[0, 1)$ the equation

$$z''(t) = -\lambda V(t) z(t). \tag{3.3}$$

Corollary 3.2. *Let $\lambda(V)$ be the supremum of all $\lambda \geqslant 0$, for which a solution z to the problem* (3.2), (3.3) *is positive in* $[0, 1)$. *Then*

$$\sup_{0<t<1} \left[(1-t)\overline{V}(t)\right] \leqslant \frac{1}{\lambda(V)} \leqslant 4 \sup_{0<t<1} \left[(1-t)\overline{V}(t)\right]. \tag{3.4}$$

If the limit $V(1-) = \lim_{t \to 1-0} V(t)$ exists and is finite, i.e., V is continuous on [0,1], the solutions z_λ to (3.2), (3.3) exist on the whole interval [0,1]. In particular, this is true for $\lambda = \lambda(V)$, and moreover, $z_{\lambda(V)}$ is still positive on $[0, 1)$. Indeed, z_λ depends continuously on λ, and in particular, for all $t \in [0, 1]$, $z_\lambda(t) \to z_{\lambda(V)}(t)$ as $\lambda \to \lambda(V)-$. But the functions z_λ are concave on $[0, 1]$ and satisfy $z_\lambda(0) = 1$, $z_\lambda(1) \geqslant 0$, so $z_{\lambda(V)}$ possesses the same properties. Thus, the supremum in Corollary 3.2 is actually the maximum, and a similar observation applies to Theorem 2.4.

In fact, $z_{\lambda(V)}(1) = 0$ since otherwise we would get, by continuity, that $z_\lambda(1) > 0$, for some $\lambda > \lambda(V)$, which contradicts to the maximality of $\lambda(V)$. Consequently, provided (3.2) holds, the following two conditions uniquely determine the value $\lambda = \lambda(V)$: z_λ is nonnegative and satisfies $z_\lambda(1) = 0$.

If V is additionally everywhere positive, one can further specify $\lambda(V)$ as the smallest eigenvalue λ_0 to the problem (3.2), (3.3) with boundary condition $z(1) = 0$. Indeed (see, for example, [3, 13]), in the regular case, the boundary value problem on [0,1]

$$z''(t) = -\lambda V(t)z(t), \qquad z'(0) = z(1) = 0, \tag{3.5}$$

has an infinite sequence $\lambda_0 < \lambda_1 < \dots$ of eigenvalues, and the corresponding eigenfunctions z_n have exactly n zeros in (0,1). Therefore, among these eigenfunctions and up to a constant, only z_0 does not vanish in (0,1). Getting rid of the normalization condition $z(0) = 1$, we may conclude the following:

Corollary 3.3. *Let V be continuous and positive on $[0,1]$. Then the value $\lambda(V)$ is the smallest eigenvalue λ_0 for the boundary value problem* (3.5)*. In particular, λ_0 admits the estimates* (3.4)*.*

4 Hardy Type Inequalities with Weights

As mentioned before, one may arrive at Sturm–Liouville equations starting from Hardy type inequalities with weights. Here, we show how to treat the constants in such inequalities using Corollary 3.3. To this end, consider the functional

$$J(f) = \frac{\int\limits_a^b f'(x)^2 q(x)\, dx}{\int\limits_a^b f(x)^2 p(x)\, dx},$$

where p and q are positive continuous functions on a finite interval $[a,b]$.

We denote by $W_1^2 = W_1^2[a,b]$ the Sobolev space of all absolutely continuous functions f on $[a,b]$ with square integrable (Radon–Nikodym) derivatives so that $J(f)$ is well defined for such functions provided that $f \neq 0$ (identically).

Lemma 4.1. *There exists a function f in W_1^2, $f \neq 0$, unique up to a constant, where the functional J attains its minimum within W_1^2 under the restriction $f(a) = 0$.*

The statement is well known (see, for example, [6] for related results). For the sake of completeness, we include a proof of the following assertion.

Theorem 4.2. *The quantity* $\min\{J(f) : f \in W_1^2, f \neq 0, f(a) = 0\}$ *represents the unique number $\lambda > 0$ such that the Sturm–Liouville equation*

$$(f'q)' = -\lambda f p \tag{4.1}$$

has a nontrivial nonnegative monotone solution on $[a, b]$ *with boundary conditions*

$$f(a) = f'(b) = 0. \tag{4.2}$$

Thus, it is equal to the smallest eigenvalue for this boundary value problem.

The argument consists of two parts.

Lemma 4.3. *Assume that a function* f *in* W_1^2, $f \neq 0$, *minimizes* J *on* W_1^2 *under the restriction* $f(a) = 0$. *Then the derivative* f' *may be modified on a set of Lebesgue measure zero such that the following properties are fulfilled:*

1) $f \in C^1[a, b]$;
2) f *is monotone, and moreover,* $f'(x) \neq 0$, *for all* $x \in [a, b)$;
3) $f'(b) = 0$;
4) $f'q \in C^1[a, b]$, *and Equation* (4.1) *holds.*

Proof. First note that, since $f(a) = 0$ and $f \neq 0$, we have

$$\int_a^b f(x)^2 p(x)\,dx > 0 \quad \text{and} \quad \int_a^b f'(x)^2 q(x)\,dx > 0.$$

Now, we take an arbitrary $h \in W_1^2$ with $h(a) = 0$ and consider for small ε the functions $f_\varepsilon = f + \varepsilon h$. By the Taylor expansion, as $\varepsilon \to 0$,

$$J(f_\varepsilon) = J(f)\left[1 + 2\varepsilon\left(\frac{\int_a^b f'(x)h'(x)q(x)\,dx}{\int_a^b f'(x)^2 q(x)\,dx} - \frac{\int_a^b f(x)h(x)p(x)\,dx}{\int_a^b f(x)^2 p(x)\,dx}\right) + O(\varepsilon^2)\right].$$

Since $J(f_\varepsilon) \leqslant J(f)$, the expression in the round brackets must be zero, i.e.,

$$\int_a^b (f'(x)q(x))\,h'(x)\,dx = \lambda \int_a^b (f(x)p(x))\,h(x)\,dx.$$

Using

$$h(x) = \int_a^x h'(t)\,dt, \quad a \leqslant x \leqslant b,$$

we rewrite the above expression as

$$\int_a^b (f'(x)q(x))\,h'(x)\,dx = \int_a^b \left(\lambda \int_x^b f(t)p(t)\,dt\right) h'(x)\,dx.$$

Since h' may be arbitrary in $L^2(a,b)$, we conclude that for almost all $x \in (a,b)$

$$f'(x)q(x) = \lambda \int_x^b f(t)p(t)\,dt. \tag{4.3}$$

This equality may be regarded as a definition of f'. Thus, we may assume that (4.3) holds for all $x \in [a,b]$, and as a result, immediately obtain properties 1), 3), and 4).

To get 2), we consider the function

$$g(x) = \int_a^x |f'(t)|\,dt.$$

Then $g(a) = 0$ and $g'(x) = |f'(x)|$ for almost all $x \in (a,b)$, so that

$$\int_a^b g'(x)^2 q(x)\,dx = \int_a^b f'(x)^2 q(x)\,dx.$$

We also have $g(x) \geqslant |f(x)|$ for all $[a,b]$, which implies

$$\int_a^b g(x)^2 p(x)\,dx \geqslant \int_a^b f(x)^2 p(x)\,dx$$

with equality possible only when $g(x) = |f(x)|$, for all $x \in [a,b]$ (since both g and f are continuous). This must be indeed the case since, otherwise, $J(g) < J(f)$ contradicts the basic assumption on f. Hence either $f' \geqslant 0$ almost everywhere or $f' \leqslant 0$ almost everywhere, and thus f is monotone.

Assume that $f' \geqslant 0$ almost everywhere, and thus $f' \geqslant 0$ everywhere by the continuity of f'. Since $f(a) = 0$ and $f \neq 0$ (identically), we get $f(b) > 0$. Hence f must be positive at least in a neighborhood of b, and this yields that the right-hand side of (4.3) is positive whenever $a \leqslant x < b$. Hence $f'(x) > 0$ on (a,b) according to (4.3). Lemma 4.3 follows. □

Lemma 4.4. *Given $\lambda > 0$, assume that the boundary value problem* (4.1), (4.2) *has a nontrivial monotone solution. Then for all $f \in W_1^2$, $f \neq 0$, we have $J(f) \geqslant \lambda$.*

The assumption about monotonicity is necessary. For example, in the case $p \equiv q \equiv 1$, on the interval $[0, 3\pi/2]$ there is solution $f(x) = \sin x$ to (4.1) with $\lambda = 1$, which satisfies the boundary conditions (4.2). However, the infimum of J is attained at $\psi(x) = \sin(x/3)$ and is equal to $1/9$.

Proof. The argument is not new; it was used, in particular, in [4]. Let ψ be a nontrivial nondecreasing solution. In particular, $\psi \in C^1[a,b]$, $\psi' q \in C^1[a,b]$, and ψ satisfies (4.1), (4.2). Integrating (4.1) over the interval (x,b) and using $\psi'(b) = 0$, we obtain (4.3) for ψ,

$$\psi'(x)q(x) = \lambda \int_x^b \psi(t)p(t)\,dt, \quad a \leqslant x \leqslant b. \tag{4.4}$$

Arguing as above, since $\psi(a) = 0$ and $\psi \neq 0$ (identically), we get $\psi(b) > 0$, and so ψ must be positive at least in a neighborhood of b. According to (4.4), we get $\psi'(x) > 0$ whenever $a \leqslant x < b$.

Now, we take an arbitrary f in W_1^2 with $f(a) = 0$ and write for $x \in (a,b)$

$$f(x) = \int_a^x f'(t)\,dt = \int_a^x \frac{f'(t)}{\sqrt{\psi'(t)}}\,\sqrt{\psi'(t)}\,dt,$$

so that, by the Schwarz inequality,

$$f(x)^2 \leqslant \int_a^x \frac{f'(t)^2}{\psi'(t)}\,dt \int_a^x \psi'(t)\,dt = \int_a^x \frac{f'(t)^2}{\psi'(t)}\,dt\ \psi(x).$$

Hence, by (4.4),

$$\begin{aligned}\lambda \int_a^b f(x)^2 p(x)\,dx &\leqslant \lambda \int_a^b \left(\int_a^x \frac{f'(t)^2}{\psi'(t)}\,dt\ \psi(x) \right) p(x)\,dx \\ &= \int_a^b \frac{f'(t)^2}{\psi'(t)} \left(\lambda \int_t^b \psi(x)p(x)\,dx \right) dt \ = \ \int_a^b f'(t)^2 q(t)\,dt.\end{aligned}$$

The proof is complete. □

Proof of Theorem 4.2. We combine Lemmas 4.3 and 4.4 (recalling an argument before Corollary 3.3 about zeros of eigenfunctions). □

Now, let us state a certain duality between Hardy type inequalities.

Lemma 4.5. *For every $c > 0$ the following two inequalities are equivalent*:

$$c \int_a^b f^2\, p\ dx \leqslant \int_a^b f'^2\, q\ dx, \quad \textit{for all } f \in W_1^2 \textit{ with } f(a) = 0; \tag{4.5}$$

$$c\int_a^b f^2/q\,dx \leqslant \int_a^b f'^2/p\,dx, \quad \textit{for all } f \in W_1^2 \textit{ with } f(b)=0. \tag{4.6}$$

In particular, the optimal constants c *in* (4.5) *and* (4.6) *coincide.*

Proof. It suffices to show that (4.5) implies (4.6). Denote by λ an optimal constant in (4.5) so that $c \leqslant \lambda$. By Theorem 4.2, there is a nonzero monotone function $\psi \in C^1[a,b]$ such that $\psi' q \in C^1[a,b]$, and ψ satisfies the equation $(\psi' q)' = -\lambda\psi p$ with boundary conditions $\psi(a) = \psi'(b) = 0$. In particular, the equality (4.4) holds. Define

$$y(x) = \int_x^b \psi(t)p(t)\,dt, \quad a \leqslant x \leqslant b.$$

This function is monotone, belongs to $C^1[a,b]$, and satisfies the boundary conditions $y(b)=0$ and $y'(a)=0$. In addition, $y'/p = -\psi$ belongs to $C^1[a,b]$. Moreover, the equality (4.4) can be rewritten in terms of y as

$$\Big(\frac{y'}{p}\Big)' = -\lambda\,\frac{y}{q},$$

which is again a Sturm–Liouville equation with respect to the functions $1/q$ and $1/p$ (in place of previous p and q). By Theorem 4.2, we conclude that (4.6) holds with constant λ in place c. ☐

As a consequence, we obtain Theorem 1.2 and the following assertion.

Corollary 4.6. *The smallest constant* C *such that the inequality*

$$\int_a^b f(x)^2 p(x)\,dx \leqslant C\int_a^b f'(x)^2 q(x)\,dx \tag{4.7}$$

holds for all f *in* W_1^2 *with* $f(a)=0$*, satisfies*

$$A(p,q) \leqslant C \leqslant 4A(p,q). \tag{4.8}$$

Recall that

$$A(p,q) = \sup_{a<x<b}\left[\int_a^x \frac{1}{q(t)}\,dt \int_x^b p(t)\,dt\right].$$

Proof of Theorem 1.2 *and Corollary* 4.6*.* We use Lemma 4.5. Without loss of generality, we assume that

$$\int_a^b p(x)\,dx = 1.$$

Introduce the distribution function

$$F(x) = \int_a^x p(t)\,dt$$

and its inverse $F^{-1} : [0,1] \to [a,b]$. Changing the variable $x = F^{-1}(t)$, we rewrite (4.5) as

$$c \int_0^1 f(F^{-1}(t))^2\,dt \leqslant \int_0^1 f'(F^{-1}(t))^2\,\frac{q(F^{-1}(t))}{p(F^{-1}(t))}\,dt.$$

In terms of $z(t) = f(F^{-1}(t))$, we again arrive at the Hardy type inequality on $[0,1]$

$$c \int_0^1 z(t)^2\,dt \leqslant \int_0^1 z'(t)^2\;p(F^{-1}(t))\,q(F^{-1}(t))\,dt$$

with boundary condition $z(0) = 0$. By Lemma 4.5, this is equivalent to

$$c \int_0^1 z(t)^2\,\frac{1}{p(F^{-1}(t))\,q(F^{-1}(t))}\,dt \;\leqslant\; \int_0^1 z'(t)^2\,dt \tag{4.9}$$

in the class of all $z \in W_2[0,1]$ such that $z(1) = 0$. Thus, the minimal constant $c = c(p,q)$ in (4.5) coincides with the optimal constant $c = c(V)$ in (4.9) on [0,1] under the restriction $z(1) = 0$ and with respect to the weight function

$$V(t) = \frac{1}{p(F^{-1}(t))\,q(F^{-1}(t))}.$$

On the other hand, by Theorem 4.2, $c(p,q)$ is the smallest eigenvalue $\lambda_0 = \lambda_0(p,q)$ for the boundary value problem (4.1), (4.2), while $c(V)$ is the smallest eigenvalue $\lambda(V)$ for the boundary value problem (3.5):

$$z'' = -\lambda V z, \quad z'(0) = z(1) = 0.$$

Hence $\lambda_0(p,q) = \lambda(V) = 1/C$, where C is the optimal constant in (4.7). By Corollary 3.3, all these quantities admit the estimates (3.4). However,

$$\sup_{0<t<1}\,(1-t)\overline{V}(t) = \sup_{0<t<1}\,(1-t)\int_0^t \frac{1}{p(F^{-1}(s))\,q(F^{-1}(s))}\,ds$$

$$= \sup_{0<t<1} (1-t) \int_a^{F^{-1}(t)} \frac{1}{q(x)}\,dx$$

$$= \sup_{a<r<b} (1-F(r)) \int_a^r \frac{1}{q(x)}\,dx$$

$$= \sup_{a<r<b} \int_r^b p(x)\,dx \int_a^r \frac{1}{q(x)}\,dx.$$

Therefore, (3.4) turns into (4.8). Consequently, Theorem 1.2 and Corollary 4.6 are proved. □

5 Poincaré Type Inequalities

Similarly to Theorem 1.2, we consider here the Sturm–Liouville equation

$$(f'q)' = -\lambda f p, \tag{5.1}$$

but with boundary conditions

$$f'(a) = f'(b) = 0. \tag{5.2}$$

As before, our case is regular, i.e., p and q are assumed to be positive continuous functions on a finite interval $[a,b]$. Denote by m the (unique) number in (a,b) such that $\int_a^m p(x)\,dx = \int_m^b p(x)\,dx$ and introduce the quantities

$$A_0 = \sup_{a<x<m} \int_x^m \frac{1}{q(t)}\,dt \int_a^x p(t)\,dt, \qquad A_1 = \sup_{m<x<b} \int_m^x \frac{1}{q(t)}\,dt \int_x^b p(t)\,dt.$$

Theorem 5.1. *The second smallest eigenvalue* λ_1 *for the boundary value problem* (5.1)-(5.2) *satisfies*

$$\frac{1}{2}\min(A_0, A_1) \leqslant \frac{1}{\lambda_1} \leqslant 4\min(A_0, A_1).$$

Recall that the smallest eigenvalue λ_0 is zero (and corresponds to the eigenfunction $f \equiv 1$). Often, λ_1 is called the first nontrivial eigenvalue.

Proof. As in Theorem 1.2, it is well known that λ_1 represents the best constant in the Poincaré type inequality

$$\lambda_1 \int_a^b f(x)^2 p(x)\,dx \leqslant \int_a^b f'(x)^2 q(x)\,dx, \tag{5.3}$$

where f is an arbitrary function in $W_1^2[a,b]$ such that

$$\int_a^b f(x)p(x)\,dx = 0. \tag{5.4}$$

We connect (5.3) and (5.4) to Hardy type inequalities and then apply Corollary 4.6. To this end, we observe that up to an absolute factor, in front of λ_1 in (5.3), the restriction (5.4) can be replaced by

$$f(m) = 0. \tag{5.5}$$

Indeed, without loss of generality, we may assume that

$$\int_a^b p(x)\,dx = 1$$

and denote by $\mu(dx)$ the measure $p(x)\,dx$ on $[a,b]$. Then (5.3) and (5.4) can be written as

$$\lambda_1 \mathrm{Var}_\mu(f) \equiv \lambda_1 \left[\int f^2\,d\mu - \left(\int f\,d\mu\right)^2\right] \leqslant \int_a^b f'(x)^2 q(x)\,dx, \tag{5.6}$$

which holds for all f in W_1^2 without any restrictions. Hence if

$$c \int f^2\,d\mu \leqslant \int_a^b f'(x)^2 q(x)\,dx \tag{5.7}$$

holds assuming (5.5), we obtain (5.6) with $\lambda_1 = c$ since $\mathrm{Var}_\mu(f) \leqslant \int f^2\,d\mu$.

Conversely, assume that (5.6) is fulfilled for a constant λ_1. Take any function f in W_1^2 such that $f = 0$ on $[a,m]$. Then, by the Cauchy inequality,

$$\left(\int f\,d\mu\right)^2 = \left(\int f\,\mathbf{1}_{[a,m]}\,d\mu\right)^2 \leqslant \frac{1}{2}\int f^2\,d\mu,$$

where $\mathbf{1}_{[a,m]}$ denotes the characteristic function of the interval $[a,m]$. Hence $\int f^2\,d\mu \leqslant \frac{1}{2}\,\mathrm{Var}_\mu(f)$ and, by (5.6), we obtain (5.7) for such a function f with $c = \lambda_1/2$. The same holds when $f = 0$ on $[m,b]$. At last, just assuming (5.5), we can apply (5.7) to $f_0 = f\,\mathbf{1}_{[a,m]}$ and $f_1 = f\,\mathbf{1}_{[m,b]}$ with $c = \lambda_1/2$. Adding the two corresponding inequalities, we arrive at (5.7) for f. Thus, the optimal constants in the Poincaré type inequality (5.6) and in the Hardy type inequality (5.7) (the latter being considered under (5.5)) are connected via

$$\frac{1}{2c} \leqslant \frac{1}{\lambda_1} \leqslant \frac{1}{c}. \tag{5.8}$$

It is obvious that $c = \min(c_0, c_1)$, where c_0 and c_1 are optimal in

$$c_0 \int_a^m f(x)^2 p(x)\,dx \leqslant \int_a^m f'(x)^2 q(x)\,dx,$$

$$c_1 \int_m^b f(x)^2 p(x)\,dx \leqslant \int_m^b f'(x)^2 q(x)\,dx$$

under the restriction (5.5). Therefore, by Corollary 4.6, we have $A_0 \leqslant \frac{1}{c_0} \leqslant 4A_0$ and $A_1 \leqslant \frac{1}{c_1} \leqslant 4A_1$. In view of (5.8), we arrive at the inequality of Theorem 5.1. □

Acknowledgement. We are grateful to V.G. Mazya for the reference to the work of I.S. Kac and M.G. Krein.

Research of the first author is supported in part by the NSF (grant no. DMS-0706866).

References

1. Artola, M.: Unpublished manuscript.
2. Bobkov, S.G., Götze, F.: Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. J. Funct. Anal. **163**, no. 1, 1–28 (1999)
3. Birkhoff, G., Rota, G-C.: Ordinary Differential Equations. Xerox College Publ. (1969)
4. Chen, L.H.Y., Lou, J.-H.: A Characterization of Probability Measures which Admit Poincaré Inequalities. Preprint (1991)
5. Ince, E.L.: Ordinary Differential Equations. Dover Publ. (1956)
6. Jost, J., Li-Jost, X.: Calculus of Variations. Cambridge Univ. Press, Cambridge (1998)
7. Kac, I.S., Krein, M.G.: Criteria for the discreteness of the spectrum of a singular string (Russian). Izv. VUZov. Matematika, no. 2(3), 136–153 (1958)
8. Levin, A.Yu.: A comparison principle for second-order differential equations (Russian). Dokl. Akad. Nauk SSSR 783–786 (1960); English transl.: Sov. Math., Dokl. **1**, 1313–1316 (1960)

9. Mazya, V.G.: Sobolev Spaces. Springer-Verlag, Berlin - New York (1985)
10. Miclo, L.: An example of application of discrete Hardy's inequalities. Markov Process. Relat. Fields **5**, no. 3, 319–330, (1999)
11. Muckenhoupt, B.: Hardy's inequality with weights. Stud. Math. **44**, 31–38 (1972)
12. Reid, W.T.: Ordinary Differential Equations. John Wiley & Sons (1971)
13. Ross, S.L.: Differential Equations. Xerox (1974)
14. Talenti, G.: Osservazioni sopra una classe di disuguaglianze. Rend. Sem. Mat. Fis. Milano **39**, 171–185 (1969)
15. Tomaselli, G.: A class of inequalities. Boll. Unione Mat. Ital. **21**, 627–631 (1969)

Quantitative Sobolev and Hardy Inequalities, and Related Symmetrization Principles

Andrea Cianchi

Abstract This survey paper deals with strengthened forms of classical Sobolev inequalities, involving remainder terms depending on the distance from the family of extremals, and with analogues for Hardy inequalities, where extremals do not exist, but can be replaced by "virtual" extremals. An account of the stability of isoperimetric and symmetrization inequalities, on which these Sobolev and Hardy inequalities rely, is provided as well.

1 Introduction

Sobolev spaces and pertinent embedding inequalities are fundamental tools in various branches of mathematical analysis, differential geometry and mathematical physics, and especially in the theory of partial differential equations and of the calculus of variations. In view of these applications, Sobolev type inequalities have been the object of a vast literature since their discovery, and a rich theory is now available on this topic, which includes a number of extensions, refinements, different approaches. We refer to the monographs [1, 2, 28, 15, 27, 51, 78, 93, 100] for accounts of some of the main contributions to this field.

In this survey paper, we focus on a basic form of the Sobolev inequality, which asserts that if $n \geqslant 2$ and $1 \leqslant p < n$, then there exists a constant $C = C(n, p)$ such that

$$C\|u\|_{L^{p^*}(\mathbb{R}^n)} \leqslant \|\nabla u\|_{L^p(\mathbb{R}^n)} \tag{1.1}$$

for every real-valued weakly differentiable function u in $\mathbb{R}^n$, decaying to 0 at infinity, and such that its gradient ∇u belongs to $L^p(\mathbb{R}^n)$. Here, $p^* = \frac{np}{n-p}$,

Andrea Cianchi
Università di Firenze, Piazza Ghiberti 27, 50122 Firenze, Italy, e-mail: cianchi@unifi.it

V. Maz'ya (ed.), *Sobolev Spaces in Mathematics I*, *International Mathematical Series*.
doi: 10.1007/978-0-387-85648-3,

the Sobolev conjugate of p, and $\|\nabla u\|_{L^p(\mathbb{R}^n)}$ is an abridged notation for the $L^p(\mathbb{R}^n)$ norm of the Euclidean length of ∇u.

The inequality (1.1) was established by Sobolev [86, 87] for $p \in (1, n)$ via Riesz potential techniques. Gagliardo [60] and Nirenberg [82] extended (1.1) to the case where $p = 1$, by a different approach based on a clever use of more elementary tools such as one-dimensional integration along lines, Fubini's theorem and Hölder's inequality. Suitable versions of (1.1) for $p > n$, with $L^{p^*}(\mathbb{R}^n)$ replaced by $L^\infty(\mathbb{R}^n)$, and even by spaces of Hölder continuous functions, are also well known and go back to Morrey [81].

A quite subtle issue concerning the inequality (1.1) is that of the optimal constant C. In the case where $p = 1$, this question was settled independently by Federer and Fleming [52] and Maz'ya [77] at the very beginning of the sixties of the last century, who pointed out the equivalence of the sharp version of (1.1) for $p = 1$ and of the isoperimetric theorem in $\mathbb{R}^n$ by De Giorgi [44]. In particular, extremals exist (and agree with characteristic functions of balls) in such a Sobolev inequality if the class of admissible functions is enlarged to include functions of bounded variation, and the right-hand side is modified accordingly. The strict connection of more general Sobolev inequalities with isoperimetric inequalities, and variants of them where perimeter is replaced by capacity, has been extensively investigated by Maz'ya (see the monograph [78]).

The best constant in the inequality (1.1) for $p \in (1, n)$ was found only about fifteen years later, again independently in two papers by Aubin [10] and Talenti [89], where a family of extremal functions is also exhibited for every $p \in (1, n)$. The approach of these papers makes use of a symmetrization argument which, in turn, relies upon the isoperimetric theorem. Owing to more recent refinements of the underlying symmetrization techniques, one can also show that the known extremals are, in fact, the only ones.

The purpose of the present paper is to report on some reinforcements of these results, which amount to quantitative versions of Sobolev inequalities with sharp constants, involving a remainder term depending on the distance from the family of extremals. The problem of inequalities of this kind was risen by Brezis and Lieb in [19], where remainder terms of a different nature for the Sobolev inequality on subdomains of $\mathbb{R}^n$ were considered (see also [49]). A positive answer to this question was given by Bianchi and Egnel [16] in the special case where $p = 2$. The general case has been attacked only recently in various papers, the main results of which will be described in what follows.

Besides Sobolev inequalities, Hardy type inequalities will be taken into account. In a broad sense, the expression Hardy inequality is usually referred to an inequality in the spirit of (1.1), where the left-hand side is replaced by a weighted norm of u (see, for example, the monographs [1, 73, 78, 83] and the survey papers [79, 98]). Customary weights depend on the distance from a (smooth) subset of $\mathbb{R}^n$. In particular, a prototypal version of the Hardy inequality, involving the distance from the origin, asserts that, if $1 \leqslant p < n$, then a constant $C = C(n,p)$ exists such that

$$C \int_{\mathbb{R}^n} \frac{|u(x)|^p}{|x|^p}\, dx \leqslant \int_{\mathbb{R}^n} |\nabla u|^p dx \tag{1.2}$$

for every function u as above (see, for example, [43]).

The optimal constant in (1.2), as well as in a modified limiting version which holds when $p = n$ (for functions with uniformly bounded supports), can be exhibited, for instance via symmetrization. However, unlike the Sobolev inequality, equality is never attained in these Hardy inequalities with sharp constants. This notwithstanding, a result will be presented showing that, when $p > 1$, it is still possible to add a remainder term depending on a suitable distance from a family of "virtual extremals," namely a family of functions which can be properly truncated in such a way to produce an optimizing sequence.

Results concerning the Sobolev inequality for $p = 1$, $1 < p < n$ and $p > n$ are contained in Sect. 3, whereas Hardy type inequalities for $1 < p < n$ and $p = n$ are discussed in Sect. 4. Their proofs are based on a common strategy, which consists in combining quantitative forms of classical isoperimetric inequalities or symmetrization principles with quantitative forms of the one-dimensional inequalities to which the Sobolev or Hardy inequalities are reduced after symmetrizing. However, although the general philosophy is similar, each inequality requires a different approach, and rests on a different quantitative symmetrization inequality. The new symmetrization inequalities coming into play are of independent interest, and are collected in Sect. 2, together with some basic definitions and properties concerning rearrangements of functions and related function spaces.

2 Symmetrization Inequalities

2.1 Rearrangements of functions and function spaces

Let Ω be a measurable subset of $\mathbb{R}^n$ with respect to the Lebesgue measure $\mathcal{L}^n$. A measurable function $u : \Omega \to \mathbb{R}$ will be called *admissible* if

$$\mathcal{L}^n\left(\{x \in \Omega : |u(x)| > t\}\right) < \infty \qquad \text{for } t > 0. \tag{2.1}$$

The *distribution function* of u is the function $\mu_u : [0,\infty) \to [0,\infty]$, defined by

$$\mu_u(t) = \mathcal{L}^n(\{x \in \mathbb{R}^n : |u(x)| > t\}) \qquad \text{for } t \geqslant 0\,.$$

Functions having the same distribution function are said to be *equidistributed* or *equimeasurable*. Equidistributed functions are called *rearrangements* of each other. The *decreasing rearrangement* of an admissible function u is the function $u^* : [0,\infty) \to [0,\infty]$ obeying

$$u^*(s) = \sup\{t \geqslant 0 : \mu_f(t) > s\} \qquad \text{for } s \geqslant 0.$$

By $u^{**} : (0,\infty) \to [0,\infty]$ we denote the function given by

$$u^{**}(s) = \frac{1}{s}\int_0^s u^*(r)\,dr$$

for $s > 0$. It is easy to see that u^{**} is also nonincreasing and satisfies

$$u^*(s) \leqslant u^{**}(s) \qquad \text{for } s > 0. \tag{2.2}$$

The *spherically symmetric rearrangement* $u^\bigstar : \mathbb{R}^n \to [0,\infty]$ of u is defined by

$$u^\bigstar(x) = u^*(\omega_n|x|^n) \qquad \text{for } x \in \mathbb{R}^n, \tag{2.3}$$

where $\omega_n = \pi^{n/2}/\Gamma(1+\frac{n}{2})$, the measure of the unit ball in $\mathbb{R}^n$. Notice that $u^*(s) = 0$ if $s > \mathcal{L}^n(\Omega)$, and $u^\bigstar(x) = 0$ if $x \notin \Omega^\bigstar$, the ball centered at the origin and having the same Lebesgue measure as Ω.

Clearly, u, u^*, and $u^\bigstar$ are equidistributed functions. Consequently,

$$\int_{\mathbb{R}^n} \Phi(u^\bigstar(x))dx = \int_0^\infty \Phi(u^*(s))ds = \int_\Omega \Phi(u(x))dx \tag{2.4}$$

for every nondecreasing function $\Phi : [0,\infty) \to [0,\infty)$. In particular,

$$\int_{\mathbb{R}^n} u^\bigstar(x)^p dx = \int_0^\infty u^*(s)^p ds = \int_\Omega |u(x)|^p dx \tag{2.5}$$

for every $p \geqslant 1$, and hence Lebesgue norms turn out to be invariant under the operations of decreasing rearrangement and of spherically symmetric rearrangement (and of any other rearrangement). An extension of this property leads to the notion of rearrangement invariant norms and spaces.

A *rearrangement invariant space* $X(\Omega)$ is a Banach function space (in the sense of Luxemburg) of real-valued measurable functions in Ω endowed with a norm $\|\cdot\|_{X(\Omega)}$ such that

$$\|u\|_{X(\Omega)} = \|v\|_{X(\Omega)} \qquad \text{if} \quad u^* = v^*. \tag{2.6}$$

For any rearrangement invariant space $X(\Omega)$ there exists a unique *representation space* $\overline{X}(0,|\Omega|)$ on the interval $(0,|\Omega|)$ fulfilling

$$\|u\|_{X(\Omega)} = \|u^*\|_{\overline{X}(0,|\Omega|)} \tag{2.7}$$

for every $u \in X(\Omega)$. Note that for customary spaces $X(\Omega)$ an expression for the norm in the representation space is immediately derived from the definition of the original norm, via elementary properties of rearrangements.

By (2.5), the *Lebesgue space* $L^p(\Omega)$, with $1 \leqslant p \leqslant \infty$, equipped with the standard norm, is a rearrangement invariant space. Both Lorentz spaces and Orlicz spaces provide generalizations of Lebesgue spaces in different directions. The *Lorentz space* $L^{p,q}(\Omega)$ is defined as the set of all measurable functions u in Ω for which the quantity

$$\|u\|_{L^{p,q}(\Omega)} = \begin{cases} \left(\int\limits_0^\infty u^*(s)^q d(s^{\frac{q}{p}})\right)^{\frac{1}{q}} & \text{if } 1 \leqslant p < \infty \text{ and } 1 \leqslant q < \infty, \\ \sup\limits_{0<s<\infty} s^{1/p} u^*(s) & \text{if } 1 \leqslant p \leqslant \infty \text{ and } q = \infty \end{cases} \tag{2.8}$$

is finite. In particular, one has $L^{p,p}(\Omega) = L^p(\Omega)$ for every $p \in [1,\infty]$ and

$$\|u\|_{L^{p,p}(\Omega)} = \|u\|_{L^p(\Omega)} \qquad \text{for } u \in L^{p,p}(\Omega)\ . \tag{2.9}$$

Lorentz spaces are monotone in the second exponent in the sense that, if $1 \leqslant q_1 < q_2$, then

$$L^{p,q_1}(\Omega) \subsetneqq L^{p,q_2}(\Omega) \tag{2.10}$$

and

$$\|u\|_{L^{p,q_2}(\Omega)} \leqslant \|u\|_{L^{p,q_1}(\Omega)} \tag{2.11}$$

for every $u \in L^{p,q_1}(\Omega)$. Lorentz spaces are in turn a special instance (up to equivalent norms) of the so–called *Lorentz–Zygmund spaces* $L^{p,q}(\operatorname{Log} L)^\gamma(\Omega)$, corresponding to $\gamma = 0$. Given $p, q \in [1,\infty]$, $\gamma \in \mathbb{R}$, and $C > 0$, such a space is defined as the set of all measurable functions in Ω for which the quantity

$$\|u\|_{L^{p,q}(\operatorname{Log} L)^\gamma(\Omega)} = \|s^{\frac{1}{p}-\frac{1}{q}}(C + \log(|\Omega|/s))^\gamma u^*(s)\|_{L^q(0,|\Omega|)} \tag{2.12}$$

is finite. Note that, replacing C by a different constant results in an equivalent expression (up to multiplicative constants). Note also that the quantities defined in (2.8) and (2.12) are only quasinorms in general since they fulfil the triangle inequality only up to a multiplicative constant. The quantity $\|\cdot\|_{L^{p,q}}$ is actually a rearrangement invariant norm if $1 \leqslant q \leqslant p$, and $\|\cdot\|_{L^{p,q}(\operatorname{Log} L)^\gamma(\Omega)}$ is a rearrangement invariant norm for those p, q, γ, and C such that the function $s \mapsto s^{\frac{1}{p}-\frac{1}{q}}(C + \log(|\Omega|/s))^\gamma$ is nonincreasing. However, for every $p > 1$ they are equivalent to rearrangement invariant norms obtained on replacing u^* by u^{**} in their definitions.

Given any Young function A, namely a convex function from $[0,\infty)$ into $[0,\infty]$ vanishing at 0, the *Orlicz space* $L^A(\Omega)$ is the rearrangement invariant space of those measurable functions u in Ω such that the Luxemburg norm

$$\|u\|_{L^A(\Omega)} = \inf\left\{\lambda > 0 : \int_\Omega A\Big(\frac{|u(x)|}{\lambda}\Big)\,dx \leqslant 1\right\}$$

is finite. Clearly, any Lebesgue space $L^p(\Omega)$, with $1 \leqslant p \leqslant \infty$, is reproduced as an Orlicz space $L^A(\Omega)$, with the choice $A(t) = t^p$ when $1 \leqslant p < \infty$ and $A(t) = \infty\chi_{(1,\infty)}(t)$ when $p = \infty$. Hereafter, χ_E stands for the characteristic function of the set E. The Orlicz spaces of exponential type, denoted by $\exp L^\sigma(\Omega)$ for $\sigma > 0$, are built upon the Young function given by $A(t) = e^{t^\sigma} - 1$ for $t \geqslant 0$.

In what follows, we work with a notion of Sobolev spaces slightly more general than the usual one. Given an open set $\Omega \subset \mathbb{R}^n$, we define the Sobolev type space $V^{1,p}(\Omega)$ by

$$\begin{aligned} V^{1,p}(\Omega) = \{u : u \text{ is a weakly differentiable function in } \Omega,\\ \text{fulfilling (2.1), and such that } |\nabla u| \in L^p(\Omega)\}. \end{aligned} \tag{2.13}$$

Accordingly, the subspace $V_0^{1,p}(\Omega)$ of those functions which vanish on $\partial\Omega$ is suitably defined by

$$\begin{aligned} V_0^{1,p}(\Omega) = \{u : u \text{ is a real-valued function in } \Omega \text{ such that}\\ \text{the continuation of } u \text{ by } 0 \text{ outside } \Omega\\ \text{belongs to } V^{1,p}(\mathbb{R}^n)\}. \end{aligned} \tag{2.14}$$

Functions of bounded variation will also come into play. We set

$$\begin{aligned} BV(\mathbb{R}^n) = \{u : u \text{ is a locally integrable function in } \Omega\\ \text{fulfilling (2.1), whose distributional gradient}\\ \text{is a vector-valued Random measure } Du \text{ with}\\ \text{finite total variation } \|Du\|(\Omega) \text{ in } \Omega\}. \end{aligned}$$

2.2 The Hardy–Littlewood inequality

A standard, but crucial, property of the decreasing rearrangement, playing a role in many applications, is the *Hardy–Littlewood inequality.* This very classical result, whose prototypal version goes back to [66], asserts that

$$\int_\Omega u(x)v(x)dx \leqslant \int_0^{\mathcal{L}^n(\Omega)} u^*(s)v^*(s)ds \tag{2.15}$$

for any nonnegative admissible functions u and v in a measurable set Ω. The inequality (2.15) was an object of a number of extensions, including those contained in [18, 22, 26, 47, 80]. Most of these contributions deal with rearrangement inequalities involving more general integrands than just the product of two functions. Here, we are concerned with an improvement of (2.15) in a different direction. Indeed, based on applications to Hardy type inequalities (see Sect. 4) a quantitative version of (2.15) with a remainder term depending on a distance from the family of extremals will be exhibited.

We first briefly discuss the cases of equality in (2.15). To begin with, let us mention that, fixed any admissible function v and a nonincreasing right-continuous function $\varphi : (0, \mathcal{L}^n(\Omega)) \to [0, \infty)$ fulfilling $\lim_{s \to \mathcal{L}^n(\Omega)^-} \varphi(s) = 0$, there always exists a nonnegative admissible function u attaining equality in (2.15) and satisfying $u^* = \varphi$. The existence of such a function u is proved, for example, in [14] when $\mathcal{L}^n(\Omega) < \infty$. The general case follows via a rather standard approximation argument.

As far as the identification of extremals in (2.15) is concerned, the following basic result [7, 30] tells us that, whenever equality holds in (2.15), the level sets of u and v are necessarily mutually nested.

Theorem 2.1 ([7, 30]). *Let Ω be a measurable subset of $\mathbb{R}^n$. Assume that equality holds in* (2.15) *for some nonnegative admissible functions u and v and that the common value of the two sides of* (2.15) *is finite. Then for every $t, \tau > 0$*

$$\text{either } \{u > t\} \subset \{v > \tau\} \text{ or } \{g > \tau\} \subset \{f > t\} \tag{2.16}$$

up to a set of Lebesgue measure zero.

Functions attaining equality in (2.15) need not be fully characterized by (2.16). In fact, fixed a nonnegative function v, extremal functions u in (2.15) are not uniquely determined by their rearrangement in general. In other words, given v, there may exist (infinitely) many functions u, with the same decreasing rearrangement, yielding equality in (2.15). As shown by simple examples, this always occurs in the case where the graph of v has a plateau, namely if a number $t > 0$ exists such that

$$\mathcal{L}^n\left(\{x \in \Omega : v(x) = t\}\right) > 0. \tag{2.17}$$

Indeed, if (2.17) holds for some $t > 0$, functions u achieving equality in (2.15) can be easily constructed in such a way that they are alterable in $\{x \in \Omega : v(x) = t\}$ while keeping u^* and $\int_\Omega u(x)v(x)dx$ constant.

The existence of some $t > 0$ at which (2.17) holds is equivalent to the existence of an interval in which v^* attains the constant value t, and hence to the nonstrict monotonicity of v^*. Thus, nonstrict monotonicity of v^* is responsible for nonuniqueness of extremal functions u in (2.15) with prescribed rearrangement. The next result ensures that lack of uniqueness can only occur in this case. Actually, it tells us that if v^* is strictly decreasing in $(0, \mathcal{L}^n(\Omega))$, then extremal functions u in (2.15) can be uniquely recovered from u^* and from v. A formula for the maximizing u is provided as well, which entails that each level set of such a maximizer agrees with the level set of v having the same measure.

Theorem 2.2 ([34]). *Let Ω be a measurable subset of $\mathbb{R}^n$, and let u and v be nonnegative admissible functions. Assume that v^* is strictly decreasing in $(0, \mathcal{L}^n(\Omega))$. Define $u_v : \Omega \to [0, \infty)$ by*

$$u_v(x) = u^*(\mu_v(v(x))) \qquad \text{for } x \in \Omega. \tag{2.18}$$

Then

$$\{x \in \Omega : u_v(x) > t\} = \{x \in \Omega : v(x) > v^*(\mu_u(t))\}, \tag{2.19}$$

up to a set of Lebesgue measure zero, for every $t > 0$, and

$$(u_v)^*(s) = u^*(s) \qquad \text{for } s \in (0, \mathcal{L}^n(\Omega)). \tag{2.20}$$

Moreover, equality holds in (2.15) *if and only if*

$$u(x) = u_v(x) \qquad \text{for a.e. } x \in \Omega. \tag{2.21}$$

The uniqueness of a function attaining equality in (2.15) in classes of functions with a prescribed rearrangement is proved in [28, Theorem 3] under the additional assumption that $\mathcal{L}^n(\Omega) < \infty$ and u and v belong to Lebesgue spaces which are duals of each other. Although no explicit representation formula like (2.18) is provided in that theorem, it could be used, in conjunction with suitable truncation arguments, as a starting point for the proof of Theorem 2.2. A direct argument relying on quite elementary properties of rearrangements and of measure preserving maps can be found in [34], where functions u and v defined on more general measure spaces are also considered.

A strengthened version of Theorem 2.2 is contained in [34] and ensures (in a quantitative form) that u is necessarily close to u_v if the two sides of (2.15) are almost equal. More precisely, it provides an estimate for a Lebesgue norm of $u - u_v$ in terms of the gap between the right-hand side and the left-hand side of (2.15). This result requires more stringent hypotheses involving both v^* and u^*.

Firstly, plain strict monotonicity of v^* has to be replaced by the stronger assumption that its derivative $v^{*\prime}$ is different from 0 a.e. in $(0, \mathcal{L}^n(\Omega))$ and that some negative power of $v^{*\prime}$ is locally integrable. Specifically, it is assumed that there exists $p \in [1, \infty)$ such that the (nonincreasing) function $\theta_p : [0, \mathcal{L}^n(\Omega)) \to [0, \infty]$, given by

$$\theta_p(s) = \begin{cases} \left(\int\limits_0^s \left(-v^{*\prime}(\sigma)\right)^{-\frac{1}{p-1}} d\sigma \right)^{1/p'} & \text{if } 1 \leqslant p < \infty, \\ \operatorname*{ess\,sup}_{\sigma \in [0,s)} \dfrac{1}{-v^{*\prime}(\sigma)} & \text{if } p = 1 \end{cases} \tag{2.22}$$

is finite for every $s \in [0, \mathcal{L}^n(\Omega))$. Note that, if $1 \leqslant p < \infty$, then θ_p is locally absolutely continuous in $[0, \mathcal{L}^n(\Omega))$ and $\theta_p(0) = 0$. Instead, θ_1 need not be absolutely continuous, nor continuous in $(0, \mathcal{L}^n(\Omega))$ – it is merely left-continuous in general – and one may have $\theta_1(0^+) > 0$, where

$$\theta_1(0^+) = \lim_{s \to 0^+} \theta_1(s).$$

Secondly, in a similar spirit, information on the size of u^* in the set where $-v^{*\prime}$ is small has to be retained. This is done by requiring that a number $q \in [1, \infty)$ exists such that the Lorentz type quasinorm given by

$$\|u\|_{\Lambda^q_p(\Omega)} = \begin{cases} \left(\displaystyle\int_0^{\mathcal{L}^n(\Omega)} u^*(s)^q \, \theta_p'(s) ds \right)^{1/q} & \text{if } 1 \leqslant p < \infty, \\ \left(\displaystyle\int_0^{\mathcal{L}^n(\Omega)} u^*(s)^q d\theta_1(s) + \theta_1(0^+) \|u\|^q_{L^\infty(\Omega)} \right)^{1/q} & \text{if } p = 1 \end{cases} \tag{2.23}$$

is finite. Here, the integral on the right-hand side of (2.23) is extended to the open interval $(0, \mathcal{L}^n(\Omega))$ when $p = 1$, and the second addend is missing if $\theta_1(0^+) = 0$. Moreover, a necessary condition for $\|u\|_{\Lambda^q_1}$ to be finite when $\theta_1(0^+) > 0$ is that $u \in L^\infty(\Omega)$.

Let us emphasize that no absolute continuity of v^* is needed: in (2.22), $v^{*\prime}$ just denotes the standard pointwise derivative of the monotone function v^*, which classically exists a.e. in $(0, \mathcal{L}^n(\Omega))$. Let us also note that the assumption $v^{*\prime} \neq 0$ a.e. in $(0, \mathcal{L}^n(\Omega))$ is equivalent to the absolute continuity of μ_v in $(0, \infty)$.

Theorem 2.3 ([34]). *Let Ω be a measurable subset of $\mathbb{R}^n$, and let u and v be nonnegative admissible functions. Assume that $p \in [1, \infty)$ and $q \in [1, \infty)$ exist such that $\theta_p(s) < \infty$ for every $s \in [0, \mathcal{L}^n(\Omega))$ and $\|u\|_{\Lambda^q_p(\Omega)} < \infty$. Set*

$$r = \frac{qp + 1}{p + 1}. \tag{2.24}$$

Then

$$\int_\Omega u(x)v(x)dx + \frac{1}{2^{p+1}\mathrm{e}q} \|u\|^{-qp}_{\Lambda^q_p(\Omega)} \|u - u_v\|^{1+qp}_{L^r(\Omega)} \leqslant \int_0^{\mathcal{L}^n(\Omega)} u^*(s)v^*(s)ds. \tag{2.25}$$

We conclude this section by specializing Theorems 2.2 and 2.3 in the customary instance when the function $v : \mathbb{R}^n \to [0, \infty)$ is radially strictly decreasing about 0. Hence, in particular,

$$v(x) = v^\bigstar(x) \qquad \text{for a.e. in } \mathbb{R}^n.$$

Since, in this case,

$$\mu_v(v(x)) = \omega_n |x|^n \qquad \text{for a.e. } x \in \mathbb{R}^n,$$

one has

$$u_v(x) = u^\bigstar(x) \qquad \text{for a.e. } x \in \mathbb{R}^n.$$

Thus, Theorem 2.2 recovers Theorem 3.4 in [74] and tells us that

$$\int_{\mathbb{R}^n} u^\bigstar(x)v(x)dx = \int_0^\infty u^*(s)v^*(s)ds$$

and that, if

$$\int_{\mathbb{R}^n} u(x)v(x)dx = \int_{\mathbb{R}^n} u^\bigstar(x)v(x)dx,$$

then

$$u(x) = u^\bigstar(x) \qquad \text{for a.e. } x \in \mathbb{R}^n.$$

Moreover, under the assumptions of Theorem 2.3, the inequality (2.25) reads

$$\int_{\mathbb{R}^n} u(x)v(x)dx + \frac{1}{2^{p+1}\mathrm{e}q} \|u\|^{-qp}_{\Lambda^q_p(\mathbb{R}^n)} \|u - u^\bigstar\|^{1+qp}_{L^r(\mathbb{R}^n)} \leqslant \int_{\mathbb{R}^n} u^\bigstar(x)v(x)dx\,. \quad (2.26)$$

2.3 The Pólya–Szegö inequality

A modern version of the classical *Pólya–Szegö principle* asserts that if $u \in V^{1,p}(\mathbb{R}^n)$, then u^* is locally absolutely continuous in $(0,\infty)$, $u^\bigstar \in V^{1,p}(\mathbb{R}^n)$, and

$$\int_{\mathbb{R}^n} |\nabla u^\bigstar|^p \, dx \leqslant \int_{\mathbb{R}^n} |\nabla u|^p \, dx \qquad (2.27)$$

(see, for example, [23, 67, 70, 88, 89]).

The inequality (2.27), together with its several variants (see, for example, [69]), is a powerful key to a number of variational problems of geometric and functional nature, concerning extremal properties of domains and functions. Besides optimal Sobolev embeddings, classical isoperimetric inequalities in mathematical physics and sharp eigenvalue inequalities fall within these results; a priori estimates for solutions to elliptic problems in sharp form are also a closely related topic. We refer to [72, 90, 94] for surveys on this matter.

Although alternative proofs of the Pólya–Szegö inequality are available in the literature (see, for example, [11, 24, 71]), the standard – and probably geometrically most transparent – approach relies upon the standard isoperimetric inequality in $\mathbb{R}^n$. Such an inequality tells us that

$$n\omega_n^{1/n}\mathcal{L}^n(E)^{1/n'} \leqslant P(E) \tag{2.28}$$

for every measurable set having finite measure and that equality holds in (2.28) if and only if E is (equivalent to) a ball [44]. Here, $P(E)$ stands for the perimeter of E defined according to geometric measure theory (see, for example, [9]) and $n' = n/(n-1)$, the Hölder conjugate of n.

Equality trivially holds in (2.27) when u equals a translate of $u^\bigstar$. However, the converse is not true. The problem of the description of the cases of equality in (2.27), already risen in [85], has been considered in the papers [70, 96, 23] appeared some twenty years ago, and has been recently the object of new contributions, including [24, 25, 38, 53, 54]. Minimal assumptions under which equality in (2.27) entails the radial symmetry of u were found by Brothers and Ziemer in [23], where the following result is established.

Theorem 2.4 ([23]). *Let $p > 1$, and let u be any function from $V^{1,p}(\mathbb{R}^n)$ attaining equality in* (2.27)*. Assume, in addition, that*

$$\mathcal{L}^n\big(\{\nabla u^\bigstar = 0\} \cap \{0 < u^\bigstar < \operatorname{ess\,sup} u\}\big) = 0\,. \tag{2.29}$$

Then

$$u = u^\bigstar \qquad \mathcal{L}^n\text{–a.e.} \quad (\textit{up to translations}). \tag{2.30}$$

The condition (2.29), as well as the hypothesis $p > 1$, is indispensable to conclude about the symmetry of u. Indeed, if $p = 1$, then, as a consequence of the coarea formula, any nonnegative function $u \in V^{1,1}(\mathbb{R}^n)$ having (non necessarily concentric) balls as level sets attains equality in (2.27). As far as (2.29) is concerned, if at least one plateau $\{u = t_0\}$ with $\mathcal{L}^n(\{u{=}t_0\}) > 0$ is allowed for some $t_0 > 0$, then equality holds in (2.27) for any function which is not necessarily globally symmetric, but which is separately symmetric in $\{0{<}u{<}t_0\}$ and in $\{t_0{<}u\}$. More subtle examples of (smooth) non symmetric minimal rearrangements u, not fulfilling (2.29), but yet with $\mathcal{L}^n(\{u = t\}) = 0$ for every $t > 0$, can also be worked out (see [23]).

In view of applications to quantitative Sobolev inequalities (see, for example, Sects. 3.2 and 3.3), the problem of strengthening the inequality (2.27) by relating the gap between $\int\limits_{\mathbb{R}^n} |\nabla u|^p dx$ and $\int\limits_{\mathbb{R}^n} |\nabla u^\bigstar|^p dx$ to the deviation of u from $u^\bigstar$ has been investigated. Two results, in a different spirit, will be presented here. The first one deals with arbitrary functions from $V^{1,p}(\mathbb{R}^n)$. In light of the picture sketched above, it is apparent that any remainder term, which accounts for the deficit between the right-hand side and the left-hand side of (2.27), has necessarily to depend not only on the deviation of u from $u^\bigstar$, but also on $u^\bigstar$ itself, and, specifically, on its gradient. Theorem 2.5 below ensures that, up to translations, the distance in $L^1(\mathbb{R}^n)$ of u from either $u^\bigstar$

or $-u^\bigstar$ can be estimated in terms of the (normalized) excess

$$E(u) = \frac{\int_{\mathbb{R}^n} |\nabla u|^p dx}{\int_{\mathbb{R}^n} |\nabla u^\bigstar|^p dx} - 1\,,$$

under the assumption that

$$\mathcal{L}^n\left(\{|u| > 0\}\right) < \infty\,. \tag{2.31}$$

The relevant estimate involves either the (normalized complementary) distribution function of $|\nabla u^\bigstar|$ defined by

$$M_{u^\bigstar}(\sigma) = \frac{\mathcal{L}^n\left(\{|\nabla u^\bigstar| \leqslant \sigma\} \cap \{0 < u^\bigstar < \operatorname{ess\,sup} u\}\right)}{\mathcal{L}^n(\{|u| > 0\})} \qquad \text{for } \sigma \geqslant 0 \tag{2.32}$$

or the function M_u, which is defined as in (2.32) on replacing $u^\bigstar$ by $|u|$. For simplicity of notation, we state our stability result for functions u normalized and rescaled in such a way that

$$\mathcal{L}^n\left(\{|u| > 0\}\right) = 1 \tag{2.33}$$

and

$$\int_{\mathbb{R}^n} |\nabla u^\bigstar|^p dx = 1\,. \tag{2.34}$$

Theorem 2.5 ([33]). *Let $p > 1$, and let $n \geqslant 2$. Then positive constants $r = r(n,p)$, $s = s(n,p)$, and $C = C(n,p)$, depending only on n and p, exist such that for every $u \in V^{1,p}(\mathbb{R}^n)$ satisfying* (2.33) *and* (2.34)

$$\min_{\pm} \inf_{x_0 \in \mathbb{R}^n} \int_{\mathbb{R}^n} |u(x) \pm u^\bigstar(x + x_0)|\, dx \leqslant C\left[M_{u^\bigstar}\left(E(u)^r\right) + E(u)\right]^s. \tag{2.35}$$

Moreover, the inequality (2.35) *continues to hold with $M_{u^\bigstar}$ replaced by M_u provided that $\int_{\mathbb{R}^n} |\nabla u^\bigstar|^p dx$ is replaced by $\int_{\mathbb{R}^n} |\nabla u|^p dx$ in* (2.34).

The conditions (2.33) and (2.34) can be easily removed via a rescaling and normalizing argument. An explicit form of the resulting estimate, in an even somewhat stronger version, is contained in the following result.

Theorem 2.6 ([33]). *Let $p > 1$, and let $n \geqslant 2$. Then positive constants $r_1 = r_1(n,p)$, $r_2 = r_2(n,p)$, $r_3 = r_3(n,p)$, and $C = C(n.p)$, depending only on n and p, exist such that for every $u \in V^{1,p}(\mathbb{R}^n)$ satisfying* (2.31)

$$\begin{aligned}\min_{\pm} \inf_{x_0 \in \mathbb{R}^n} \int\limits_{\mathbb{R}^n} &|u(x) \pm u^\bigstar(x+x_0)|\,dx \\ &\leqslant C\|\nabla u^\bigstar\|_{L^p(\mathbb{R}^n)} \mathcal{L}^n(\{|u|>0\})^{1+\frac{1}{n}-\frac{1}{p}} \\ &\quad \times \left[M_{u^\star}(\sigma) + E(u)^{r_1} + \frac{\|\nabla u^\bigstar\|_{L^p(\mathbb{R}^n)}}{\sigma \mathcal{L}^n(\{|u|>0\})^{\frac{1}{p}}} E(u)^{r_2} \right]^{r_3}, \end{aligned} \tag{2.36}$$

for $\sigma > 0$. *Moreover, the inequality* (2.36) *continues to hold with* $M_{u^\star}$ *and* $\|\nabla u^\bigstar\|_{L^p}$ *simultaneously replaced by* M_u *and* $\|\nabla u\|_{L^p}$.

Theorem 2.5 recovers, in particular, Theorem 2.4 corresponding to the case where $E(u) = 0$ and $M_{u^\star}(0) = 0$. Under the sole assumption that $E(u) = 0$, the inequality (2.35) enables us to estimate the distance in $L^1(\mathbb{R}^n)$ between u and a suitable translated of $u^\bigstar$ in terms of $M_{u^\star}(0)$, namely in terms of the left-hand side of (2.29). This reproduces (a special case of) [38]. Note that, in fact, Theorem 2.5 slightly improves these results in that, unlike [23] and [38], functions are allowed which need not be positive. Theorem 2.5 is also somehow related to a result of [25] which, however, only deals with qualitative issues.

That the whole function $M_{u^\star}$ comes into play in Theorem 2.5, instead of just $M_{u^\star}(0)$ is explained by the fact that large sets where $|\nabla u^\bigstar|$ and $|\nabla u|$ are small may allow u to be very asymmetric when $E(u) > 0$, in spite of $\int\limits_{\mathbb{R}^n} |\nabla u|^p dx$ being very close to $\int\limits_{\mathbb{R}^n} |\nabla u^\bigstar|^p dx$. Apropos examples can be easily exhibited, even when the dimension n equals one. For more details, we refer to [39] dealing with a parallel issue for Steiner symmetrization (see also [29, 37] for related results).

The estimate (2.35) can be somewhat enhanced, on replacing the $L^1(\mathbb{R}^n$ norm by a stronger norm on the left-hand side. Indeed, on exploiting standard multiplicative Gagliardo–Nirenberg inequalities (see, for example, [78]), one can easily infer from (2.35) that

$$\min_{\pm} \inf_{x_0 \in \mathbb{R}^n} \|u(\cdot) \pm u^\bigstar(\cdot + x_0)\|_{L^q(\mathbb{R}^n)} \leqslant C\Big[M_{u^\star}\big(E(u)^r\big) + E(u)\Big]^s, \tag{2.37}$$

where $q \in [1, \frac{np}{n-p})$ if $1 < p < n$, q is any number $\geqslant 1$ if $p = n$, and $q = \infty$ if $p > n$ (and r, s, and C are suitable constants).

On the other hand, $M_{u^\star}(\sigma)$ and $M_u(\sigma)$ are unrelated in general for $\sigma \geqslant 0$, in the sense that no estimate between the two functions can hold in either direction. Actually, $M_u(0)$ cannot be controlled just by $M_{u^\star}(0)$ since functions u can be exhibited such that $M_u(0)$ is arbitrarily close to 1, but $M_{u^\star}(0) = 0$ (see, for example, [5]). Note also that, although the reverse estimate $M_{u^\star}(0) \leqslant M_u(0)$ holds for every $u \in W^{1,p}(\mathbb{R}^n)$, a bound of this kind fails for $\sigma > 0$. To see this, consider the one-dimensional function $u_k(x)$ given by $1 - (2k+1)|x|$ if $|x| < \frac{1}{2k+1}$, $k \in \mathbb{N}$, and extended periodically in $[-1, 1]$.

Obviously, $u_k^\bigstar(x) = 1 - |x|$ for $x \in [-1, 1]$. Therefore, $M_{u_k}(\sigma) = 0$ if $\sigma > 0$ and k is sufficiently large, whereas $M_{u_k^\bigstar}(\sigma) = 1$ if $\sigma \geqslant 1$ for every $k \in \mathbb{N}$. The next result shows that, nevertheless, $M_{u^\bigstar}$ and M_u can be estimated in terms of each other, if the extra information contained in $E(u)$ is exploited as well.

Theorem 2.7 ([33]). *Let $p > 1$, and let $n \geqslant 2$. Then there exist positive constants $r = r(n,p)$ and $C = C(n,p)$ such that, if u is any function as in Theorem 2.5, then*

$$M_u(\sigma) \leqslant C\left[M_{u^\bigstar}(4\sigma) + E(u)^r + \frac{E(u)^r}{\sigma}\right] \tag{2.38}$$

and

$$M_{u^\bigstar}(\sigma) \leqslant C\left[M_u(4\sigma) + E(u)^r + \frac{E(u)^r}{\sigma}\right] \tag{2.39}$$

for $\sigma > 0$.

A key ingredient in the proofs of Theorems 2.5–2.7, which are given in [33], is a quantitative version of (2.28) ensuring that a positive constant $C = C(n)$ exists such that

$$n\omega_n^{1/n}\mathcal{L}^n(E)^{1/n'}\left(1 + Ca(E)^2\right) \leqslant P(E) \tag{2.40}$$

for every set $E \subset \mathbb{R}^n$ of finite measure and perimeter. Here, $a(E)$ is the asimmetry of E measured as

$$a(E) = \inf_{B \text{ is a ball}, \mathcal{L}^n(B) = \mathcal{L}^n(E)} \frac{\mathcal{L}^n(E \triangle B)}{\mathcal{L}^n(E)}, \tag{2.41}$$

where $\triangle$ stands for symmetric difference of sets. A weaker form of (2.40), where the exponent 2 is replaced by 4, appears in [64]. The present version is the object of the recent paper [58] (see also [55] for an alternative proof including anisotropic notions of perimeter and sharpening a result of [50]). A refined version of (2.40) for convex sets in the plane is contained in [6].

The second quantitative version of the Pólya–Szegö inequality is stated in the next theorem and proved in [40]. It shows that, if functions u a priori enjoying certain partial symmetry properties are taken into account, then the distance of u from $u^\bigstar$ can actually be estimated in terms of the difference between the two sides of (2.27) without any extra information on the measure of the level sets of $|\nabla u|$ or $|\nabla u^\bigstar|$. Precisely, a quantitative Pólya–Szegö inequality holds for functions which are symmetric about n mutually orthogonal hyperplanes containing the origin.

Theorem 2.8 ([40]). *Let $n \geqslant 2$, and let $1 < p < n$. Set $q = \max\{p, 2\}$. Then a positive constant $C = C(n,p)$ exists such that*

$$\int_{\mathbb{R}^n} |u - u^\bigstar|^{p^*} dx \leqslant C\Bigg(\int_{\mathbb{R}^n} |u|^{p^*} dx\Bigg)^{p/n} \Bigg(\int_{\mathbb{R}^n} |\nabla u^\bigstar|^p dx\Bigg)^{1/q'}$$
$$\times \Bigg(\int_{\mathbb{R}^n} |\nabla u|^p dx - \int_{\mathbb{R}^n} |\nabla u^\bigstar|^p dx\Bigg)^{1/q} \qquad (2.42)$$

for every nonnegative function $u \in V^{1,p}(\mathbb{R}^n)$ symmetric about n orthogonal hyperplanes containing the origin.

Note that Theorem 2.8 continues to hold even if u is symmetric about n arbitrary orthogonal hyperplanes, not necessarily containing 0, provided that $u^\bigstar$ is replaced by a suitable translate.

3 Sobolev Inequalities

3.1 Functions of Bounded Variation

The sharp form of the Sobolev inequality for $p = 1$ goes back to [52, 77] and asserts that

$$n\omega_n^{1/n} \|u\|_{L^{n'}(\mathbb{R}^n)} \leqslant \| \nabla u \|_{L^1(\mathbb{R}^n)} \qquad (3.1)$$

for every $u \in V^{1,1}(\mathbb{R}^n)$. The constant $n\omega_n^{1/n}$ in (3.1) is the best possible, witness a suitable sequence of functions u converging to the characteristic function of a ball. Thus, in a sense, although equality is never attained in (3.1) unless u vanishes identically, (multiples of) characteristic functions of balls can be considered the virtual extremals. These functions turn into real extremals in an extended version of the inequality (3.1) in $BV(\mathbb{R}^n)$, the space of functions of bounded variation in $\mathbb{R}^n$. The relevant inequality states us that

$$n\omega_n^{1/n} \|u\|_{L^{n'}(\mathbb{R}^n)} \leqslant \|Du\|(\mathbb{R}^n) \qquad (3.2)$$

for every $u \in BV(\mathbb{R}^n)$. Actually, equality holds in (3.2) whenever

$$u = \lambda \chi_B \qquad (3.3)$$

for some $\lambda \in \mathbb{R}$ and for some ball $B \subset \mathbb{R}^n$. Moreover, functions given by (3.3) are the only extremals in (3.2).

Theorem 3.4 below provides us with a quantitative version of the inequalities (3.1)–(3.2). Loosely speaking, it tells us that the difference $\|Du\|(\mathbb{R}^n) - n\omega_n^{1/n}\|u\|_{L^{n'}(\mathbb{R}^n)}$ is not merely nonnegative for every $u \in BV(\mathbb{R}^n)$ (and vanishing if and only if u obeys (3.3)), but can also be estimated from below in terms of the distance in $L^{n'}(\mathbb{R}^n)$ between u and the (n+2)-parameter family of functions having the form (3.3). Precisely, on setting

$$d_{n'}(u) = \inf_{\lambda,B} \frac{\|u - \lambda\chi_B\|_{L^{n'}(\mathbb{R}^n)}}{\|u\|_{L^{n'}(\mathbb{R}^n)}} \tag{3.4}$$

if $u \neq 0$, and $d_{n'}(0) = 0$, one has the following result.

Theorem 3.1 ([31]). *Let $n \geqslant 2$. Then positive constants $\alpha = \alpha(n)$ and $C = C(n)$ exist such that*

$$n\omega_n^{1/n}\|u\|_{L^{n'}(\mathbb{R}^n)}\big(1 + Cd_{n'}(u)^\alpha\big) \leqslant \|Du\|(\mathbb{R}^n) \tag{3.5}$$

for every $u \in BV(\mathbb{R}^n)$.

Of course, the inequality (3.5) holds, in particular, for every $u \in V^{1,1}(\mathbb{R}^n)$, and, in this case, $\|Du\|(\mathbb{R}^n)$ agrees with $\|\nabla u\|_{L^1(\mathbb{R}^n)}$.

The inequality (3.2) (and (3.1)) is closely related and, in fact, equivalent, to the isoperimetric inequality (2.28). Indeed, the inequality (3.2) reduces to (2.28) when $u = \chi_E$, and, conversely, it quite easily follows from (2.28) applied to the level sets $\{|u| > t\}$, via the coarea formula.

Similarly, Theorem 3.1 relies upon the quantitative version of (2.28) given by (2.40). However, the derivation of (3.5) from an application of (2.40) to the level sets of u is not entirely straightforward. In fact, the sole content of (2.40) is not sufficient to deduce (3.5). This can be easily realized by noting that any function $u \in BV(\mathbb{R}^n)$, whose level sets are balls, satisfies $a(\{u > t\}) = 0$ for every $t > 0$, whereas $d_{n'}(u)$ can be very large. The key additional observation which, combined with (2.40), enables one to estimate $d_{n'}(u)$ in terms of $\|Du\|(\mathbb{R}^n) - n\omega_n^{1/n}\|u\|_{L^{n'}(\mathbb{R}^n)}$, is that, if the latter expression is small, then $\|u\|_{L^{n'}(\mathbb{R}^n)}$ and $\|u\|_{L^{n',1}(\mathbb{R}^n)}$ cannot differ too much. This is a consequence of an enhanced version of (3.2) in Lorentz spaces, stating that

$$n\omega_n^{1/n}\|u\|_{L^{n',q}(\mathbb{R}^n)} \leqslant \|Du\|(\mathbb{R}^n) \tag{3.6}$$

for every $q \in [1, n']$ and for every $u \in BV(\mathbb{R}^n)$. In view of (2.10) and (2.11), the inequality (3.6) improves (3.2), and, obviously, the constant $n\omega_n^{1/n}$ is the best possible also in (3.6). Moreover, if $1 < q \leqslant n'$, the extremals in (3.6) are exactly those given by (3.3).

In fact, a quantitative version of (3.6), extending (3.5), can be established for every $q \in (1, n']$. This is stated in Theorem 3.2 below, where the normalized distance defined by

$$d_{n',q}(u) = \inf_{\lambda,B} \frac{\|u - \lambda\chi_B\|_{L^{n',q}(\mathbb{R}^n)}}{\|u\|_{L^{n',q}(\mathbb{R}^n)}} \tag{3.7}$$

if $u \neq 0$, and $d_{n',q}(0) = 0$, comes into play. Note that the conclusion of Theorem 3.2 slightly strengthens (3.5) also in the case where $q = n'$, for it ensures that $d_{n'}(u)$ can be replaced by the stronger (normalized) distance $d_{n',1}(u)$.

Theorem 3.2 ([31]). *Let $n \geqslant 2$, and let $1 < q \leqslant n'$. Then positive constants $\alpha = \alpha(n)$ and $C = C(n)$ exist such that*

$$n\omega_n^{1/n}\|u\|_{L^{n',q}(\mathbb{R}^n)}\big(1 + C\big((q-1)d_{n',q}(u)\big)^{\alpha}\big) \leqslant \|Du\|(\mathbb{R}^n) \qquad (3.8)$$

for every $u \in BV(\mathbb{R}^n)$. Moreover, $d_{n',q}(u)$ can be replaced by $d_{n',1}(u)$ in (3.8).

Observe that the constant multiplying $d_{n',q}(u)$ in (3.8) approaches 0 as $q \to 1^+$. This is consistent with the fact that, since any function $u \in BV(\mathbb{R}^n)$ whose level sets are (not necessarily concentric) balls attains equality in (3.6) when $q = 1$, no estimate like (3.8) can hold in this case.

The quantitative Sobolev inequality, in the same spirit as (3.5), proved in [16] for $p = 2$, involves the distance between gradients in $L^2(\mathbb{R}^n)$. In view of this fact, the question might be risen of whether (3.5) and (3.8) can be augmented on replacing $\inf_{\lambda,B} \|u - \lambda\chi_B\|_{L^{n',q}(\mathbb{R}^n)}$ by $\inf_{\lambda,B} \|D(u - \lambda\chi_B)\|(\mathbb{R}^n)$ in the definition of $d_{n',q}(u)$. The answer turns out to be negative, as the choice of a sequence of characteristic functions of ellipsoids converging to a ball shows. A simple modification of this counterexample continues to work even for functions from $V^{1,1}(\mathbb{R}^n)$.

The outline of the proof of Theorem 3.2 is roughly as follows. Under the assumption that u is nonnegative, one can first prove that, if $\|Du\|(\mathbb{R}^n) - n\omega_n^{1/n}\|u\|_{L^{n',q}(\mathbb{R}^n)}$ is sufficiently small, then a level $T > 0$ exists enjoying the following properties:

(i) $a(\{u > T\})$ is small, i.e., $\{u > T\}$ is almost a ball, say B_T;
(ii) $|\{u > t\}|$ is nearly constant, in integral sense, for $t \in (0, T]$;
(iii) the contribution of $|\{u > t\}|$ to $\|u\|_{L^{n',q}(\mathbb{R}^n)}$ is small for $t > T$.

Next, all the level sets $\{u > t\}$ are shown to be close to B_T for $t \in (0, T]$. A combination of these facts then ensures that u does not differ much from $T\chi_{B_T}$ in the $L^{n',1}(\mathbb{R}^n)$ norm. Finally, the sign assumption on u is removed on splitting u as $u = u^+ - u^-$, where $u^+ = \max\{u, 0\}$ and $u^- = \max\{-u, 0\}$, and showing that either $\|u^+\|_{L^{n',1}(\mathbb{R}^n)}$ or $\|u^-\|_{L^{n',1}(\mathbb{R}^n)}$ can be bounded in terms of $\|Du\|(\mathbb{R}^n) - n\omega_n^{1/n}\|u\|_{L^{n',q}(\mathbb{R}^n)}$.

Detailed proofs of the results of the present section can be found in [31], where quantitative Sobolev inequalities involving non Euclidean norms of ∇u are discussed as well. A refinement of the inequality (3.5), where the optimal constant α is exhibited, is contained in [59].

3.2 The case $1 < p < n$

The sharp Sobolev inequality for $1 < p < n$ tells us that

$$S(p,n)\|u\|_{L^{p^*}(\mathbb{R}^n)} \leqslant \|\nabla u\|_{L^p(\mathbb{R}^n)} \tag{3.9}$$

for every $u \in V^{1,p}(\mathbb{R}^n)$, where

$$S(p,n) = \sqrt{\pi}\, n^{1/p} \left(\frac{n-p}{p-1}\right)^{(p-1)/p} \left(\frac{\Gamma(n/p)\Gamma(1+n-n/p)}{\Gamma(1+n/2)\Gamma(n)}\right)^{1/n},$$

and is the best possible constant [10, 89]. A family of extremals in (3.9) is given by the functions $v_{a,b,x_0} : \mathbb{R}^n \to [0,\infty)$ defined by

$$v_{a,b,x_0}(x) = \frac{a}{\left(1+b|x-x_0|^{p'}\right)^{(n-p)/p}} \qquad \text{for } x \in \mathbb{R}^n \tag{3.10}$$

for some $a \neq 0$, $b > 0$, $x_0 \in \mathbb{R}^n$. In fact, as recently pointed out in [41], functions having the form (3.10) are the only ones attaining equality in (3.9). Incidentally, note that, when $p = 2$, the classical result of [63], applied to the Euler equation of the functional $\|\nabla u\|_{L^2(\mathbb{R}^n)} / \|u\|_{L^{2^*}(\mathbb{R}^n)}$, can alternatively be used to derive this characterization of the extremals in (3.9).

In this section, we present a result which strengthens the inequality (3.9) by an additional term on the left-hand side which accounts for the deviation of u from a suitably chosen extremal. More precisely, Theorem 3.3 below yields a quantitative version of the inequality (3.9), with a remainder term depending on the (normalized) distance of u from the family of extremals (3.10) given by

$$d_{p^*}(u) = \inf_{a,\,b,\,x_0} \frac{\|u - v_{a,b,x_0}\|_{L^{p^*}(\mathbb{R}^n)}}{\|u\|_{L^{p^*}(\mathbb{R}^n)}} \tag{3.11}$$

if $u \neq 0$, and $d_{p^*}(0) = 0$.

Theorem 3.3 ([40]). *Let $n \geqslant 2$, and let $1 < p < n$. Then positive constants $\alpha = \alpha(n,p)$ and $C = C(n,p)$ exist such that*

$$S(p,n)\|u\|_{L^{p^*}(\mathbb{R}^n)}\left(1 + C d_{p^*}(u)^{\alpha}\right) \leqslant \|\nabla u\|_{L^p(\mathbb{R}^n)} \tag{3.12}$$

for every $u \in V^{1,p}(\mathbb{R}^n)$.

The inequality (3.12) gives a positive answer to a question raised by Brezis and Lieb in [19], which has been settled in [16] in the special case where $p = 2$ in the even stronger form with $\|u - v_{a,b,x_0}\|_{L^{2^*}(\mathbb{R}^n)}$ replaced by $\|\nabla u - \nabla v_{a,b,x_0}\|_{L^2(\mathbb{R}^n)}$ in (3.11) (see also [13] for the case of higher order derivatives). The method of [16] heavily rests upon the Hilbert space structure of $W^{1,2}(\mathbb{R}^n)$ and on eigenvalue properties of a weighted Laplacian in $\mathbb{R}^n$. Such an approach, which has been employed to deal with other related problems involving Sobolev spaces endowed with a Hilbert space structure [75], does not seem suitable for extensions to the general case where $p \neq 2$. Following the lines traced in [10] and [89], the proof of (3.12) given in [40] relies upon methods of geometric flavor, including isoperimetric inequalities and symmetrizations.

To be more specific, the proof of Theorem 3.3 basically consists of three steps, each step amounting to an extension of the inequality (3.12) to a broader class of functions. After starting with spherically symmetric functions, one proceeds with *n-symmetric functions*, namely functions which are symmetric about n orthogonal hyperplanes, and one eventually concludes with arbitrary Sobolev functions. This strategy can be clarified by the following considerations.

Since the operation of spherically symmetric rearrangement satisfies (2.5) and (2.27), one has

$$\|\nabla u^\star\|_{L^p(\mathbb{R}^n)} - S(p,n)\|u^\star\|_{L^p(\mathbb{R}^n)} \leqslant \|\nabla u\|_{L^p(\mathbb{R}^n)} - S(p,n)\|u\|_{L^{p^*}(\mathbb{R}^n)} \quad (3.13)$$

and

$$\|\nabla u\|_{L^p(\mathbb{R}^n)} - \|\nabla u^\star\|_{L^p(\mathbb{R}^n)} \leqslant \|\nabla u\|_{L^p(\mathbb{R}^n)} - S(p,n)\|u\|_{L^{p^*}(\mathbb{R}^n)} \quad (3.14)$$

for every $u \in V^{1,p}(\mathbb{R}^n)$. In view of (3.13) and (3.14), the idea in the proof of (3.12) is to split the problem: first, establish the inequality in the class of spherically symmetric functions; second, estimate the L^{p^*} distance of u from (a suitable translated of) $u^\star$ in terms of $\|\nabla u\|_{L^p(\mathbb{R}^n)} - \|\nabla u^\star\|_{L^p(\mathbb{R}^n)}$.

Even in the special class of spherically symmetric functions, the derivation of (3.12) is not straightforward. Actually, standard proofs of the one-dimensional Bliss inequality [17, 89] to which (3.9) reduces when restricted to spherically symmetric functions, do not seem suitable for modifications yielding stability results. A more flexible approach to the relevant one-dimensional inequality, which can be successfully augmented to provide a quantitative version, follows instead on specializing a mass transportation technique employed in [41] (see also [76]).

Major problems arise in the attempt at estimating the asymmetry of u in terms of the left-hand side of (3.14). Indeed, this is just impossible, without additional assumptions on u, as pointed out in Sect. 2.3. It is at this stage that the class of n-symmetric functions comes into play. Indeed, on the one hand, the distance of u from $u^\star$ can actually be estimated by $\|\nabla u\|_{L^p(\mathbb{R}^n)} - \|\nabla u^\star\|_{L^p(\mathbb{R}^n)}$ if u is a priori assumed to be n-symmetric (Theorem 2.8), thus enabling one to establish (3.12) in this class of functions. On the other hand, any function $u \in V^{1,p}(\mathbb{R}^n)$ can be replaced, through careful reflection arguments by a suitable n-symmetric function in such a way that $\|\nabla u\|_{L^p(\mathbb{R}^n)} - S(p,n)\|u\|_{L^{p^*}(\mathbb{R}^n)}$ and $d_{p^*}(u)$ do not increase and decrease, respectively, too much. This fact, combined with the former step, easily leads to the conclusion of Theorem 3.3.

We conclude this section by noting that the result of [16] leaves open the problem of whether a version of Theorem 3.3 holds with the distance of u from the family of extremals in $L^{p^*}(\mathbb{R}^n)$ replaced by the corresponding distance between gradients in $L^p(\mathbb{R}^n)$.

3.3 The case $p > n$

The present section is concerned with the Morrey–Sobolev embedding theorem stating that any function $u \in V^{1,p}(\mathbb{R}^n)$ for some $p > n$ is essentially bounded (and, in fact, locally Hölder continuous) in $\mathbb{R}^n$ (see, for example, [2, 78, 100]).

A special form of this embedding tells us that $\|u\|_{L^\infty(\mathbb{R}^n)}$ can be estimated in terms of $\|\nabla u\|_{L^p(\mathbb{R}^n)}$ under the assumption that

$$\mathcal{L}^n(\operatorname{sprt} u) < \infty\,, \tag{3.15}$$

where $\operatorname{sprt} u$ denotes the support of u. In particular, an optimal version of the relevant estimate yields

$$C_1(n,p)\mathcal{L}^n(\operatorname{sprt} u)^{\frac{1}{p}-\frac{1}{n}}\|u\|_{L^\infty(\mathbb{R}^n)} \leqslant \|\nabla u\|_{L^p(\mathbb{R}^n)} \tag{3.16}$$

for every $u \in V^{1,p}(\mathbb{R}^n)$ fulfilling (3.15) (see, for example, [91]). Here,

$$C_1(n,p) = n^{1/p}\omega_n^{1/n}\Big(\frac{p-n}{p-1}\Big)^{1/p'} \tag{3.17}$$

is the best possible constant since equality holds in (3.16) whenever u agrees with any of the functions $w_{a,b,x_0} : \mathbb{R}^n \to [0,\infty)$ given by

$$w_{a,b,x_0}(x) = \begin{cases} a\Big(b^{\frac{p-n}{p-1}} - |x-x_0|^{\frac{p-n}{p-1}}\Big) & \text{if } |x-x_0| \leqslant b, \\ 0 & \text{otherwise} \end{cases} \tag{3.18}$$

for some $a \in \mathbb{R}$, $b > 0$ and $x_0 \in \mathbb{R}^n$. When the assumption (3.15) is dropped, bounds for $\|u\|_{L^\infty(\mathbb{R}^n)}$ by $\|\nabla u\|_{L^p(\mathbb{R}^n)}$ are possible only in conjunction with some other norm $\|u\|_{L^q(\mathbb{R}^n)}$, where $q \in [1,\infty)$. A sharp form of these bounds is available in the endpoint case where $q = 1$. Actually, a result from [91] tells us that

$$C_2(n,p)\|u\|_{L^\infty(\mathbb{R}^n)} \leqslant \|u\|_{L^1(\mathbb{R}^n)}^{1-\eta}\|\nabla u\|_{L^p(\mathbb{R}^n)}^{\eta} \tag{3.19}$$

for every $u \in V^{1,p}(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$. Here,

$$\eta = \frac{np}{np+p-n} \tag{3.20}$$

and

$$C_2(n,p) = (n\omega_n^{1/n})^{\frac{np'}{n+p'}}\Big(\frac{1}{n}+\frac{1}{p'}\Big)^{-1}\Big(\frac{1}{n}-\frac{1}{p}\Big)^{\frac{n-(n-1)p'}{n+p'}}$$
$$\times\left(\frac{\Gamma(2+p'/n)}{\Gamma(1+p')\Gamma(1-p'/n')}\right)^{n/(n+p')}. \tag{3.21}$$

Furthermore, a family of extremals in (3.19) is given by

$$\overline{w}_{a,b,x_0}(x) = \begin{cases} a \displaystyle\int_{|x-x_0|}^{b} r^{\frac{1-n}{p-1}} (b^n - r^n)^{\frac{1}{p-1}}\, dr & \text{if } |x-x_0| \leqslant b, \\ 0 & \text{otherwise} \end{cases} \tag{3.22}$$

for $a \in \mathbb{R}$, $b > 0$, and $x_0 \in \mathbb{R}^n$.

The inequalities (3.16) and (3.19) rely upon the properties (2.7) and (2.27) of rearrangements. Moreover, the specification on (2.27) provided by Theorem 2.4 can be used to show that the $(n+2)$-parameter families of functions given by (3.18) and (3.22) yield, in fact, all the extremals in (3.16) and (3.19) respectively.

We present quantitative versions of (3.16) and (3.19), which strengthen the above results with a remainder term depending on the distance from extremals. Loosely speaking, the distance of any function u from the family of extremals w_{a,b,x_0} in terms of the gap between the two sides of the inequality (3.16) is estimated, and similarly for the inequality (3.19) and its extremals $\overline{w}_{a,b,x_0}$. Actually, on setting

$$d_\infty(u) = \inf_{a,\, b,\, x_0} \frac{\|u - w_{a,b,x_0}\|_{L^\infty(\mathbb{R}^n)}}{\|u\|_{L^\infty(\mathbb{R}^n)}}$$

if $u \neq 0$, and $d_\infty(0) = 0$, one has the following result.

Theorem 3.4 ([32]). *Let $p > n$. Then there exist positive constants $\alpha = \alpha(n,p)$ and $C_3 = C_3(n,p)$ such that*

$$C_1(n,p)\mathcal{L}^n(\operatorname{sprt} u)^{\frac{1}{p}-\frac{1}{n}} \|u\|_{L^\infty(\mathbb{R}^n)} \big[1 + C_3 d_\infty(u)^\alpha\big] \leqslant \|\nabla u\|_{L^p(\mathbb{R}^n)} \tag{3.23}$$

for every $u \in V^{1,p}(\mathbb{R}^n)$ satisfying (3.15).

The counterpart of Theorem 3.4 for the inequality (3.19) is contained in the next statement, where

$$\overline{d}_\infty(u) = \inf_{a,\, b,\, x_0} \frac{\|u - \overline{w}_{a,b,x_0}\|_{L^\infty(\mathbb{R}^n)}}{\|u\|_{L^\infty(\mathbb{R}^n)}}$$

if $u \neq 0$, and $\overline{d}_\infty(0) = 0$.

Theorem 3.5 ([32]). *Let $p > n$. Then there exist positive constants $\beta = \beta(n,p)$ and $C_4 = C_4(n,p)$ such that*

$$C_2(n,p)\|u\|_{L^\infty(\mathbb{R}^n)} \big[1 + C_4 \overline{d}_\infty(u)^\beta\big] \leqslant \|u\|_{L^1(\mathbb{R}^n)}^{1-\eta} \|\nabla u\|_{L^p(\mathbb{R}^n)}^{\eta} \tag{3.24}$$

for every $u \in V^{1,p}(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$.

The approach to Theorem 3.4, whose proof can be found in [32], consists of two steps. First, the inequality (3.23) is established for spherically symmetric functions. Second, the (normalized) distance in $L^\infty(\mathbb{R}^n)$ of any u from a suitable translated of its Schwarz symmetral $u^\star$ is estimated in terms of the gap between the two sides of (3.16). In particular, a key tool in this second step is the quantitative form of the inequality (2.27) given by Theorem 2.6. The outline of the proof of Theorem 3.5 is similar, although some complications arise, owing to the fact that it deals with a multiplicative inequality and that functions whose support need not have finite measure are involved.

4 Hardy Inequalities

4.1 The case $1 < p < n$

The basic Hardy inequality in $\mathbb{R}^n$ asserts that if $1 < p < n$, then

$$\left(\frac{n-p}{p}\right)^p \int_{\mathbb{R}^n} \frac{|u(x)|^p}{|x|^p}\, dx \leqslant \int_{\mathbb{R}^n} |\nabla u|^p dx \tag{4.1}$$

for every function $u \in V^{1,p}(\mathbb{R}^n)$. The constant $\left(\frac{n-p}{p}\right)^p$ is optimal in (4.1), as demonstrated by sequences obtained on truncating functions having the form

$$v_a(x) = a|x|^{\frac{p-n}{p}} \qquad \text{for } x \in \mathbb{R}^n, \tag{4.2}$$

with $a \in \mathbb{R} \setminus \{0\}$, at levels $1/k$ and k, and then letting $k \to \infty$. However, it is well known that equality is never achieved in (4.1), unless u is identically equal to 0. In fact, the natural candidates v_a to be extremals in (4.1) have a gradient which does not (even locally) belong to $L^p(\mathbb{R}^n)$.

The lack of extremals has inspired improved versions of (4.1) and of related inequalities, in the spirit of [78, Sect. 2.1.6], where $\mathbb{R}^n$ is replaced by any open bounded subset Ω containing 0 and u is assumed to belong to the Sobolev space $V_0^{1,p}(\Omega)$. Typically, these improvements of (4.1) amount to extra terms on the left-hand side that either involve integrals of $|u|^p$ with weights depending on $|x|$ which are less singular than $|x|^{-p}$ at 0 or integrals of $|\nabla u|^q$ with $q < p$ (see [3, 4, 8, 12, 20, 36, 45, 46, 56, 57, 61, 62, 68, 92, 97]).

This section is concerned with an analogue of the results of the preceding section for the inequality (4.1) in the whole of $\mathbb{R}^n$. Such an improved version contains a remainder term depending on a distance of u, in a suitable norm, from the family of those functions which have the form (4.2) and can be regarded as the virtual extremals in (4.1). In particular, this entails that any optimizing sequence in the inequality (4.1) must approach the family (4.2).

In order to give a precise statement, we begin by noting that, via spherically symmetric rearrangement, the inequality (4.1) is easily seen to be equivalent to the Lorentz-norm inequality

$$\omega_n^{1/n}\frac{n-p}{p}\|u\|_{L^{p^*,p}(\mathbb{R}^n)} \leqslant \|\nabla u\|_{L^p(\mathbb{R}^n)} \tag{4.3}$$

for $u \in V^{1,p}(\mathbb{R}^n)$. By (2.9) and (2.10), the inequality (4.3) improves the standard Sobolev inequality where $L^{p^*,p}(\mathbb{R}^n)$ is replaced by $L^{p^*}(\mathbb{R}^n)$ on the left-hand side (and $\omega_n^{1/n}\frac{n-p}{p}$ is replaced by a different constant).

In view of (4.3), the norm $\|\cdot\|_{L^{p^*,p}(\mathbb{R}^n)}$ could be considered the natural one to measure the distance of any u from the family (4.2) in terms of the gap between the two sides of (4.1). Unfortunately, this is not possible, since the functions v_a, which do not belong to $V^{1,p}(\mathbb{R}^n)$, neither belong to $L^{p^*,p}(\mathbb{R}^n)$. They do not even belong to the larger space $L^{p^*}(\mathbb{R}^n)$, appearing in the usual Sobolev inequality. In fact, the smallest rearrangement invariant space containing v_a is the Marcinkievicz space $L^{p^*,\infty}(\mathbb{R}^n)$, also called the weak-$L^{p^*}$ space. Thus, the $L^{p^*,\infty}(\mathbb{R}^n)$ norm appearing in the (normalized) distance given by

$$d_{p^*,\infty}(u) = \inf_{a\in\mathbb{R}} \frac{\|u-v_a\|_{L^{p^*,\infty}(\mathbb{R}^n)}}{\|u\|_{L^{p^*,p}(\mathbb{R}^n)}} \tag{4.4}$$

if $u \neq 0$, and $d_{p^*,\infty}(0) = 0$, which is employed in the following quantitative Hardy inequality, is actually the strongest possible in this setting.

Theorem 4.1 ([35]). *Let $n \geqslant 2$, and let $1 < p < n$. Then constants $\alpha = \alpha(n,p)$ and $C = C(n,p)$ exist such that*

$$\left(\frac{n-p}{p}\right)^p \int\limits_{\mathbb{R}^n} \frac{|u(x)|^p}{|x|^p} dx \left[1 + C d_{p^*,\infty}(u)^\alpha\right] \leqslant \int\limits_{\mathbb{R}^n} |\nabla u|^p dx \tag{4.5}$$

for every $u \in V^{1,p}(\mathbb{R}^n)$.

Note that, although the Hardy inequality (4.1) holds also for $p = 1$, the inequality (4.5) does not. Indeed, any spherically symmetric function can be shown to attain equality in (4.1) when $p = 1$.

Of course, Theorem 4.1 continues to hold if $\mathbb{R}^n$ is replaced by any open set Ω containing 0, provided that functions $u \in V_0^{1,p}(\Omega)$ are taken into account. Moreover, if $\mathcal{L}^n(\Omega) < \infty$ and $1 \leqslant q < p^*$, the space $L^{p^*,\infty}(\Omega)$ is continuously embedded into $L^q(\Omega)$, and hence an inequality analogous to (4.5), with $\|u - v_a\|_{L^{p^*,\infty}(\mathbb{R}^n)}$ replaced by $\|u - v_a\|_{L^q(\Omega)}$ in the definition of $d_{p^*,\infty}(u)$, follows from Theorem 4.1. In fact, minor changes in the proof yield a version of the inequality (4.5), where the functions v_a, which do not vanish on $\partial\Omega$, are replaced by the functions $\overline{v}_a : \Omega \to [0,\infty)$ given by

$$\overline{v}_a(x) = a(|x|^{\frac{p-n}{p}} - Q)_+ \qquad \text{for } x \in \Omega.$$

Here, Q is any positive number such that the support of v_a is contained in Ω; namely $Q > r_\Omega^{\frac{p-n}{n}}$, where

$$r_\Omega = \sup\{r > 0 : B_r(0) \subset \Omega\}\,,$$

and $B_r(0)$ denotes the ball centered at 0 and having radius r. Precisely, if $\mathcal{L}^n(\Omega) < \infty$ and $1 \leqslant q < p^*$, then on setting

$$d_q(u) = \inf_{a\in\mathbb{R}} \frac{\|u - \overline{v}_a\|_{L^q(\Omega)}}{\|u\|_{L^{p^*,p}(\Omega)}},$$

a constant $C = C(p, q, n, Q, \mathcal{L}^n(\Omega))$ exists such that

$$\left(\frac{n-p}{p}\right)^p \int_\Omega \frac{|u(x)|^p}{|x|^p}dx \left[1 + Cd_q(u)^{2p^*}\right] \leqslant \int_\Omega |\nabla u|^p dx$$

for every $u \in V_0^{1,p}(\Omega)$.

A quite simple proof of the inequality (4.1) relies upon symmetrization. Actually, the Hardy–Littlewood inequality (2.15) implies that

$$\int_{\mathbb{R}^n} \frac{|u(x)|^p}{|x|^p}dx \leqslant \int_{\mathbb{R}^n} \frac{u^\bigstar(x)^p}{|x|^p}dx \tag{4.6}$$

for every u as above. Owing to (4.6) and to the Pólya–Szegö principle (2.27), the inequality (4.1) is reduced to the well-known one-dimensional Hardy inequality

$$\left(\frac{1}{p^*}\right)^p \int_0^\infty \varphi(s)^p s^{-p/n} ds \leqslant \int_0^\infty (-\varphi'(s))^p s^{p/n'} ds \tag{4.7}$$

for every nonincreasing locally absolutely continuous function $\varphi : (0, \infty) \to [0, \infty)$ such that $\lim_{s\to\infty} \varphi(s) = 0$ (see, for example, [14, 83]).

Loosely speaking, the approach to Theorem 4.1 consists in proving the stability of this argument. To be more specific, reinforced versions of inequalities (4.6) and (4.7), containing quantitative information on the gap between their two sides, come into play. The former of these quantitative inequalities follows from (2.26), applied with $u(x)$ replaced by $|u(x)|^p$ and $v(x)$ replaced by $|x|^{-p}$, and enables one to show that, if the difference between the right-hand side and the left-hand side of (4.1) is small, then u is close to $u^\bigstar$. The latter is used to prove that, in the same circumstance, $u^\bigstar$ is close to some function having the form (4.2). The inequality (4.5) easily follows from these two pieces of information. The full proof of Theorem 4.1 is contained in [35].

4.2 The case $p = n$

The inequality (4.1) breaks down when $p = n$. In fact, no estimate like (4.1) (with $\left(\frac{n-p}{p}\right)^p$ replaced by any constant) can hold in this case since the weight $|x|^{-n}$ is not (even locally) integrable in $\mathbb{R}^n$. However, an inequality in the same spirit can be restored if $|x|^{-n}$ is replaced by a suitable less singular weight at 0, and $\mathbb{R}^n$ is replaced by any open bounded subset Ω. On defining

$$R_\Omega = \sup_{x \in \Omega} |x|, \tag{4.8}$$

the relevant inequality tells us that

$$\left(\frac{n-1}{n}\right)^n \int\limits_\Omega \frac{|u(x)|^n}{|x|^n \left(1 + \log \frac{D}{|x|}\right)^n} dx \leqslant \int\limits_\Omega |\nabla u|^n dx \tag{4.9}$$

for every $D \geqslant R_\Omega$ and $u \in V_0^{1,n}(\Omega)$. A similar phenomenon as in (4.1) occurs in (4.9) in the sense that the constant $\left(\frac{n-1}{n}\right)^n$ is the best possible for any bounded Ω containing 0, but it is not attained. Again, the optimality is shown by sequences of truncated (at levels k with $k \to \infty$) of a suitable family of functions, which in this case have the form

$$w_a(x) = a\left[\left(1 + \log \frac{D}{|x|}\right)^{1/n'} - Q\right]_+ \qquad \text{for } x \in \Omega \tag{4.10}$$

for some $a \in \mathbb{R} \setminus \{0\}$. Here, Q is any positive number fulfilling $Q > \left(1 + \log \frac{D}{r_\Omega}\right)^{1/n'}$, so that the support of w_a is contained in Ω.

A counterpart of Theorem 4.1 for the inequality (4.9) asserts that a remainder term can be added to the left-hand side of (4.9), which depends on the deviation of u from a suitable function of the form (4.10). Such a deviation can now be controlled by an exponential estimate. Precisely, recall that for each $D \geqslant R_\Omega$ the expression

$$\|u\|_{L^{n,\infty}(\operatorname{Log} L)^{-1}(\Omega),D} = \left(\int\limits_0^{|\Omega|} \frac{u^*(s)^n}{\left(n + \log \frac{\omega_n D^n}{s}\right)^n} \frac{ds}{s} \right)^{1/n}$$

defines a norm in the Lorentz–Zygmund space $L^{n,\infty}(\log L)^{-1}(\Omega)$, and set, for $C > 0$,

$$d_{C,D,Q}(u) = \inf_{a \in \mathbb{R}} \int\limits_\Omega \left(\exp\left(\frac{C|u(x) - w_a(x)|^{n'}}{\|u\|^{n'}_{L^{n,\infty}(\log L)^{-1}(\Omega),D}} \right) - 1 \right) dx \tag{4.11}$$

if $u \neq 0$, and $d_{C,D,Q}(0) = 0$. Then the following theorem holds.

Theorem 4.2 ([35]). *Let Ω be an open bounded subset of $\mathbb{R}^n$, $n \geqslant 2$, containing* 0. *Let $D > R_\Omega$, and let $Q > \big(1+\log \frac{D}{r_\Omega}\big)^{1/n'}$. Then positive constants $\alpha = \alpha(n)$ and $C = C(n, R_\Omega, D, Q)$ exist such that*

$$\Big(\frac{n-1}{n}\Big)^n \int\limits_\Omega \frac{|u(x)|^n}{|x|^n\big(1+\log\frac{D}{|x|}\big)^n} dx \left[1 + d_{C,D,Q}(u)^\alpha\right] \leqslant \int\limits_\Omega |\nabla u|^n dx \tag{4.12}$$

for every $u \in V_0^{1,n}(\Omega)$.

A few comments on Theorem 4.2 are in order. The presence of the norm $\|\cdot\|_{L^{n,\infty}(\mathrm{Log}\,L)^{-1}(\Omega)}$ in the definition of $d_{C,D,Q}(\cdot)$ is related to the fact that, in analogy with (4.1) and (4.3), the inequality (4.9) is equivalent to

$$\omega_n^{1/n}(n-1)\|u\|_{L^{n,\infty}(\mathrm{Log}\,L)^{-1}(\Omega),D} \leqslant \|\nabla u\|_{L^n(\Omega)} \tag{4.13}$$

for $u \in V_0^{1,n}(\Omega)$. The inequality (4.13) goes back (apart from the constant) to [21, 65, 78], and has recently been shown to be optimal as far as the norm on the left-hand side is concerned [42, 48]. On the other hand, the norm $\|\cdot\|_{L^{n,\infty}(\mathrm{Log}\,L)^{-1}(\Omega),D}$ cannot be used to measure the distance of u from the family (4.10) since $w_a \notin L^{n,\infty}(\mathrm{Log}\,L)^{-1}(\Omega)$. The exponential term in (4.11) serves as a replacement for this norm, in the same spirit as $\|\cdot\|_{L^{p^*,\infty}(\mathbb{R}^n)}$ replaces $\|\cdot\|_{L^{p^*,p}(\mathbb{R}^n)}$ in (4.4), and is related to the classical embedding theorem of [84, 95, 99] which states that

$$\int\limits_\Omega \left(\exp\left(\frac{C|u(x)|^{n'}}{\|\nabla u\|_{L^n(\Omega)}^{n'}}\right) - 1\right) dx \leqslant 1 \tag{4.14}$$

for some positive constant $C = C(n, \mathcal{L}^n(\Omega))$ and for every $u \in V_0^{1,n}(\Omega)$. Observe that (4.14) is equivalent to

$$\|u\|_{\exp L^{n'}(\Omega)} \leqslant C\|\nabla u\|_{L^n(\Omega)} \tag{4.15}$$

for some positive constant $C = C(n, \mathcal{L}^n(\Omega))$ and for every $u \in V_0^{1,n}(\Omega)$. The inequalities (4.14) and (4.15) are slightly weaker than (4.13) for

$$L^{n,\infty}(\mathrm{Log}\,L)^{-1}(\Omega) \subsetneq \exp L^{n'}(\Omega) \tag{4.16}$$

(with continuous embedding). However, the remainder $d_{C,D,Q}(u)$ appearing in (4.12) is again optimal, in that the function $e^{t^{n'}} - 1$ cannot be replaced by any other Young function growing essentially faster at infinity. Indeed, $\exp L^{n'}(\Omega)$, unlike $L^{p^*}(\Omega)$, agrees with its corresponding weak space and is the smallest rearrangement invariant space containing the family (4.10). The scheme of the proof of Theorem 4.2, which can be found in [35], is analogous to that of Theorem 4.1.

References

1. Adams, D.R., Hedberg, L.I.: Function Spaces and Potential Theory. Springer-Verlag, Berlin (1995)
2. Adams, R.A.: Sobolev Spaces. Academic Press, Orlando (1975)
3. Adimurthi, Chaudhuri, N., Ramaswamy, M.: An improved Hardy–Sobolev inequality and its applications. Proc. Am. Math. Soc. **130**, 489–505 (2002)
4. Adimurthi, Esteban, M.J.: An improved Hardy–Sobolev inequality in $W^{1,p}$ and its applications to Schrödinger operators. NoDEA Nonlinear Differ. Equ. Appl. **12**, 243–263 (2005)
5. Almgren, F.J.Jr., Lieb, E.H.: Symmetric rearrangement is sometimes continuous. J. Am. Math. Soc. **2**, 683–773 (1989)
6. Alvino, A., Ferone, V., Nitsch, C.: A Sharp Isoperimetric Inequality in the Plane. Preprint
7. Alvino, A., Lions, P.L., Trombetti, G.: A remark on comparison results via symmetrization. Proc. R. Soc. Edinb., Sect. A, Math. **102**, 37–48 (1986)
8. Alvino, A., Volpicelli, R., Volzone, B.: Best Constants in Hardy–Sobolev Inequalities with Remainder Terms. Preprint
9. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Clarendon Press, Oxford (2000)
10. Aubin, T.: Problèmes isopérimetriques et espaces de Sobolev. J. Differ. Geom. **11**, 573–598 (1976)
11. Baernstein, A.: A unified approach to symmetrization. In: Partial Differential Equations of Elliptic Type. Cambridge Univ. Press, Cambridge (1994)
12. Barbatis, G., Filippas, S., Tertikas, A.: A unified approach to improved L^p Hardy inequalities with best constants. Trans. Am. Math. Soc. **356**, 2169–2196 (2004)
13. Bartsch, T., Weth, T., Willem, M.: A Sobolev inequality with remainder term and critical equations on domains with topology for the polyharmonic operator. Calc. Var. Partial Differ. Equ. **18**, 57–75 (2003)
14. Bennett, C., Sharpley, R.: Interpolation of Operators. Academic Press, Boston (1988)
15. Besov, O.V., Il'in, V.P., Nikol'skij, S.M.: Integral Representations of Functions and Embedding Theorems (Russian). Nauka, Moscow (1975); Second ed. (1996); English transl.: Wiley, New York (1978/79)
16. Bianchi, G., Egnell, H.: A note on the Sobolev inequality. J. Funct. Anal. **100**, 18–24 (1991)
17. Bliss, G.A.: An integral inequality. J. London Math. Soc. **5**, 40–46 (1930)
18. Brascamp, H.J., Lieb, E.H., Luttinger, J.M.: A general rearrangement inequality for multiple integrals. J. Funct. Anal. **17**, 227–237 (1974)
19. Brezis, H., Lieb, E.: Sobolev inequalities with remainder terms. J. Funct. Anal. **62**, 73–86 (1985)
20. Brezis, H., Marcus, M., Shafrir, I.: Extremal functions for Hardy's inequality with weights. J. Funct. Anal. **171**, 177–191 (2000)
21. Brezis, H., Wainger, S.: A note on limiting cases of Sobolev embeddings and convolution inequalities. Commun. Partial Differ. Equ. **5**, 773–789 (1980)
22. Brock, F.: A general rearrangement inequality á la Hardy–Littlewood. J. Inequal. Appl. **5**, 309–320 (2000)
23. Brothers, J.E., Ziemer, W.P.: Minimal rearrangements of Sobolev functions. J. Reine Angew. Math **384**, 153–179 (1988)
24. Burchard, A.: Steiner symmetrization is continuous in $W^{1,p}$. Geom. Funct. Anal. **7**, 823–860 (1997)
25. Burchard, A., Guo, Y.: Compactness via symmetrization. J. Funct. Anal. **214**, 40–73 (2004)

26. Burchard, A., Hajaiej, H.: Rearrangement inequalities for functionals with monotone integrands. J. Funct. Anal. **233**, 561–582 (2006)
27. Burenkov, V.: Sobolev Spaces on Domains. Teubner, Stuttgart (1998)
28. Burton, G.R.: Rearrangements of functions, maximization of convex functionals, and vortex rings. Math. Ann. **276**, 225–253 (1987)
29. Chlebik, M., Cianchi, A., Fusco, N.: The perimeter inequality under Steiner symmetrization: cases of equality. Ann. Math. **162**, 525–555 (2005)
30. Cianchi, A.: On the L^q norm of functions having equidistributed gradients. Nonlinear Anal. **26**, 2007–2021 (1996)
31. Cianchi, A.: A quantitative Sobolev inequality in BV. J. Funct. Anal. **237**, 466–481 (2006)
32. Cianchi, A.: Sharp Sobolev-Morrey inequalities and the distance from extremals. Trans. Am. Math. Soc. **360**, 4335–4347 (2008)
33. Cianchi, A., Esposito, L., Fusco, N., Trombetti, C.: A quantitative Pólya–Szegö principle. J. Reine Angew. Math. **614** (2008)
34. Cianchi, A., Ferone, A.: A strenghtened version of the Hardy–Littlewood inequality. J. London Math. Soc. **77**, 581–592 (2008)
35. Cianchi, A., Ferone, A.: Hardy inequalities with non standard remainder terms. Ann. Inst. H. Poincaré. Anal. Nonlinéaire [To appear]
36. Cianchi, A., Ferone, A.: Best Remainder Norms in Sobolev–Hardy Inequalities. Indiana Univ. Math. J. [To appear]
37. Cianchi, A., Fusco, N.: Steiner symmetric extremals in Pólya–Szegö type inequalities. Adv. Math. **203**, 673–728 (2006)
38. Cianchi, A., Fusco, N.: Minimal rearrangements, strict convexity and critical points. Appl. Analysis **85**, 67–85 (2006)
39. Cianchi, A., Fusco, N.: Dirichlet integrals and Steiner asymmetry. Bull. Sci. Math. **130**, 675–696 (2006)
40. Cianchi, A., Fusco, N., Maggi, F., Pratelli, A.: The Sharp Sobolev Inequality in Quantitative Form. Preprint
41. Cordero-Erausquin, D., Nazaret, B., Villani, C.: A mass transportation approach to sharp Sobolev and Gagliardo–Nirenberg inequalities. Adv. Math. **182**, 307–332 (2004)
42. Cwikel, M., Pustylnik, E.: Sobolev type embeddings in the limiting case. J. Fourier Anal. Appl. **4**, 433–446 (1998)
43. Davies, E.B.: A review of Hardy inequalities. Oper. Theory, Adv. Appl. **110**, 55–67 (1999)
44. De Giorgi, E.: 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) **5**, 33–44 (1958)
45. Detalla, A., Horiuchi, T., Ando, H.: Missing terms in Hardy–Sobolev inequalities and its applications. Far East J. Math. Sci. **14**, 333–359 (2004)
46. Dolbeault, J., Esteban, M.J., Loss. M., Vega, L.: An analytical proof of Hardy-like inequalities related to the Dirac operator. J. Funct. Anal. **216**, 1–21 (2004)
47. Draghici, C., Rearrangement inequalities with application to ratios of heat kernels. Potential Anal. **22**, 351–374 (2005)
48. Edmunds, D.E., Kerman, R.A., Pick, L.: Optimal Sobolev imbeddings involving rearrangement invariant quasi-norms. J. Funct. Anal. **170**, 307–355 (2000)
49. Egnell, H., Pacella, F., Tricarico, M.: Some remarks on Sobolev inequalities. Nonlinear Anal., Theory Methods Appl. **13**, 671–681 (1989)
50. Esposito, L., Fusco, N., Trombetti, C.: A quantitative version of the isoperimetric inequality: the anisotropic case. Ann. Sc. Norm. Super. Pisa **4**, 619–651 (2005)
51. Evans, L.C., Gariepy, R.F.: Measure Theory and Fine Properties of Functions. CRC Press, Boca Raton (1992)

52. Federer, H., Fleming, W.: Normal and integral currents. Ann. Math. **72**, 458–520 (1960)
53. Ferone, A., Volpicelli, R.: Minimal rearrangements of Sobolev functions: a new proof. Ann. Inst. H.Poincaré, Anal. Nonlinéaire **20**, 333–339 (2003)
54. Ferone, A., Volpicelli, R.: Convex rearrangement: equality cases in the Pólya–Szegö inequality. Calc. Var. Partial Differ. Equ. **21**, no. 3, 259–272 (2004)
55. Figalli, A., Maggi, E., Pratelli, A.: A Mass Transportation Approach to Quantitative Isoperimetric Inequalities. Preprint
56. Filippas, S., Maz'ya, V.G., Tertikas, A.: On a question of Brezis and Marcus. Calc. Var. Partial Differ. Equ. **25**, 491–501 (2006)
57. Filippas, S., Maz'ya, V.G., Tertikas, A.: Critical Hardy–Sobolev inequalities. J. Math. Pures Appl. **87**, 37–56 (2007)
58. Fusco, N., Maggi, F., Pratelli, A.: The sharp quantitative isoperimetric inequality. Ann. Math. [To appear]
59. Fusco, N., Maggi, F., Pratelli, A.: The sharp quantitative Sobolev inequality for functions of bounded variation. J. Funct. Anal. **244**, 315–341 (2007)
60. Gagliardo, E., Proprietà di alcune classi di funzioni di più variabili. Ric. Mat. **7**, 102–137 (1958)
61. Gazzola, F., Grunau, H.C., Mitidieri, E.: Hardy inequalities with optimal constants and remainder terms. Trans. Am. Math. Soc. **356**, 2149–2168 (2004)
62. Ghoussoub, N., Moradifam, A.: On the Best Possible Remaining Term in the Hardy Inequality. Preprint
63. Gidas, B., Ni, W.M., Nirenberg, L.: Symmetry and related properties via the maximum principle. Comm. Math. Phys. **68**, 209–243 (1979)
64. Hall, R.R.: A quantitative isoperimetric inequality in n-dimensional space. J. Reine Angew. Math. **428**, 161–196 (1992)
65. Hansson, K.: Imbedding theorems of Sobolev type in potential theory. Math. Scand. **45**, 77–102 (1979)
66. Hardy, G.H., Littlewood, J.E., Pólya, G.: Inequalities. Cambridge Univ. Press, Cambridge (1934)
67. Hilden, K.: Symmetrization of functions in Sobolev spaces and the isoperimetric inequality. Manus. Math. **18**, 215–235 (1976)
68. Hoffman-Ostenhof, M., Hoffman-Ostenhof, T., Laptev, A.: A geometrical version of Hardy's inequality. J. Funct. Anal. **189**, 539–548 (2002)
69. Kawohl, B.: Rearrangements and convexity of level sets in PDE. Lect. Notes Math. **1150**, Springer-Verlag, Berlin (1985)
70. Kawohl, B.: On the isoperimetric nature of a rearrangement inequality and its consequences for some variational problems. Arch. Ration. Mech. Anal. **94**, 227–243 (1986)
71. Kesavan, S.: Comparison theorems via symmetrization: revisited. Boll. Unione Mat. Ital., VII, Ser. A **11**, no. 1, 163–172 (1997)
72. Kesavan, S.: Symmetrization and Applications. World Scientific, Hackensack, NJ (2006)
73. Kufner, A.: Weighted Sobolev Spaces. Teubner, Leipzig (1980)
74. Lieb, E.H., Loss, M.: Analysis. Am. Math. Soc., Providence, RI (2001)
75. Loiudice, A.: Improved Sobolev inequalities on the Heisenberg group. Nonlinear Anal., Theory Methods Appl. **62**, 953–962 (2005)
76. Lutwak, E., Yang, D., Zhang, G.: Sharp affine L_p Sobolev inequalities. J. Differ. Geom. **62**, 17–38 (2002)
77. Maz'ya, V.: Classes of domains and imbedding theorems for function spaces (Russian). Dokl. Akad. Nauk SSSR **133**, 527–530 (1960); English transl.: Sov. Math., Dokl. **1**, 882–885 (1960)
78. Maz'ya, V.G.: Sobolev Spaces. Springer-Verlag, Berlin-Tokyo (1985)

79. Maz'ya, V.G.: Analytic criteria in the qualitative spectral analysis of the Schrödinger operator. In: Proceedings of Symposia in Pure Mathematics, Vol. 76, pp. 257–288 (2007)
80. Morpurgo, C.: Sharp inequalities for functional integrals and traces of conformally invariant operators. Duke Math. J. **114**, 477–553 (2002)
81. Morrey, C.B.: Functions of several variables and absolute continuity II. Duke Math. J. **6**, 187–215 (1940)
82. Nirenberg, L.: On elliptic partial differential equations. Ann. Sc. Norm. Super. Pisa **13**, 1–48 (1959)
83. Opic, B., Kufner, A.: Hardy-Type Inequalities. Longman Scientific & Technical, Harlow (1990) 279–317.
84. Pokhozhaev, S.I.: On the imbedding Sobolev theorem for $pl = n$ (Russian). In: Dokl. Conference, Sec. Math. Moscow Power Inst., pp. 158–170 (1965)
85. Pólya, G., Szegö, G.: Isoperimetric Inequalities in Mathematical Physics. Princeton Univ. Press, Princeton, NJ (1951)
86. Sobolev, S.L.: On some estimates relating to families of functions having derivatives that are square integrable (Russian, French). Dokl. Akad. Nauk SSSR **1**, 267–282 (1936)
87. Sobolev, S.L.: On a theorem of functional analysis (Russian). Mat. Sb. **46**, 471–497 (1938); English transl.: Am. Math. Soc., Transl., II. Ser. **34**, 39–68 (1963)
88. Sperner, E.: Symmetrisierung für Funktionen mehrerer reeller Variablen. Manus. Math. **11**, 159–170 (1974)
89. Talenti, G.: Best constant in Sobolev inequality. Ann. Mat. Pura Appl. **110**, 353–372 (1976)
90. Talenti, G.: On isoperimetric theorems in mathematical physics. In: Handbook of Convex Geometry. North-Holland, Amsterdam (1993)
91. Talenti, G.: Inequalities in rearrangement invariant function spaces. In: Nonlinear Analysis, Function Spaces and Applications. Vol. 5. Prometheus Publ. House, Prague (1994)
92. Tidblom, J.: A Hardy inequality in the half-space. J. Funct. Anal. **221**, 482–495 (2005)
93. Triebel, H.: Theory of Function Spaces. Birkhäuser, Basel (1983)
94. Trombetti, G.: Symmetrization methods for partial differential equations. Boll. Unione Mat. Ital. **3**, 601–634 (2000)
95. Trudinger, N.S.: On imbeddings into Orlicz spaces and some applications. J. Math. Mech. **17**, 473–483 (1967)
96. Uribe, A.: Minima of the Dirichlet Norm and Toeplitz Operators. Preprint (1985)
97. Vazquez, J.L., Zuazua, E.: The Hardy inequality and the asympthotic behavior of the heat equation with an inverse-square potential. J. Funct. Anal. **173**, 103–153 (2000)
98. Verbitsky, I.G.: Superlinear equations, potential theory and weighted norm inequalities. In: Nonlinear Analysis, Function Spaces and Applications, Vol. 6, Acad. Sci. Czech Repub., Prague, pp. 223–269 (1999)
99. Yudovich, V.I.: Some estimates connected with integral operators and with solutions of elliptic equations (Russian). Dokl. Akad. Nauk SSSR **138**, 805–808 (1961); English transl.: Sov. Math., Dokl. **2**, 746–749 (1961)
100. Ziemer, W.P.: Weakly Differentiable Functions. Spriger-Verlag, New York (1989)

Inequalities of Hardy–Sobolev Type in Carnot–Carathéodory Spaces

Donatella Danielli, Nicola Garofalo, and Nguyen Cong Phuc

Abstract We consider various types of Hardy–Sobolev inequalities on a Carnot–Carathéodory space (Ω, d) associated to a system of smooth vector fields $X = \{X_1, X_2, \ldots, X_m\}$ on $\mathbb{R}^n$ satisfying the Hörmander finite rank condition $rank\ Lie[X_1, \ldots, X_m] \equiv n$. One of our main concerns is the trace inequality

$$\int\limits_{\Omega} |\varphi(x)|^p V(x) dx \leqslant C \int\limits_{\Omega} |X\varphi|^p dx, \qquad \varphi \in C_0^\infty(\Omega),$$

where V is a general weight, i.e., a nonnegative locally integrable function on Ω, and $1 < p < +\infty$. Under sharp geometric assumptions on the domain $\Omega \subset \mathbb{R}^n$ that can be measured equivalently in terms of subelliptic capacities or Hausdorff contents, we establish various forms of Hardy–Sobolev type inequalities.

1 Introduction

A celebrated inequality of S.L. Sobolev [49] states that for any $1 < p < n$ there exists a constant $S(n,p) > 0$ such that for every function $\varphi \in C_0^\infty(\mathbb{R}^n)$

Donatella Danielli
Purdue University, 150 N. University Str., West Lafayette, IN 47906, USA,
e-mail: danielli@math.purdue.edu

Nicola Garofalo
Purdue University, 150 N. University Str., West Lafayette, IN 47906, USA,
e-mail: garofalo@math.purdue.edu

Nguyen Cong Phuc
Purdue University, 150 N. University Str., West Lafayette, IN 47906, USA,
e-mail: pcnguyen@math.purdue.edu

V. Maz'ya (ed.), *Sobolev Spaces in Mathematics I, International Mathematical Series.*
doi: 10.1007/978-0-387-85648-3,

$$\left(\int_{\mathbb{R}^n} |\varphi|^{\frac{np}{n-p}} dx \right)^{\frac{n-p}{np}} \leqslant S(n,p) \left(\int_{\mathbb{R}^n} |D\varphi|^p dx \right)^{\frac{1}{p}}. \tag{1.1}$$

Such an inequality admits the following extension (see [8]). For $0 \leqslant s \leqslant p$ define the critical exponent relative to s as follows:

$$p^*(s) = p \, \frac{n-s}{n-p}.$$

Then for every $\varphi \in C_0^\infty(\mathbb{R}^n)$ one has

$$\left(\int_{\mathbb{R}^n} \frac{|\varphi|^{p^*(s)}}{|x|^s} dx \right)^{\frac{1}{p^*(s)}} \leqslant \left(\frac{p}{n-p} \right)^{\frac{s}{p^*(s)}} S(n,p)^{\frac{n(p-s)}{p(n-s)}} \left(\int_{\mathbb{R}^n} |D\varphi|^p dx \right)^{\frac{1}{p}}. \tag{1.2}$$

In particular, when $s = 0$, then (1.2) is just the Sobolev embedding (1.1), whereas for $s = p$ we obtain the Hardy inequality

$$\int_{\mathbb{R}^n} \frac{|\varphi|^p}{|x|^p} dx \leqslant \left(\frac{p}{n-p} \right)^p \int_{\mathbb{R}^n} |D\varphi|^p dx. \tag{1.3}$$

The constant $\left(\frac{p}{n-p} \right)^p$ on the right-hand side of (1.3) is sharp. If one is not interested in the best constant, then (1.2), and hence (1.3), follows immediately by combining the generalized Hölder inequality for weak L^p spaces in [32] with the Sobolev embedding (1.1), after having observed that $| \cdot |^{-s} \in L^{\frac{n}{s},\infty}(\mathbb{R}^n)$ (the weak $L^{\frac{n}{s}}$ space).

Inequalities of Hardy–Sobolev type play a fundamental role in analysis, geometry, and mathematical physics, and there exists a vast literature concerning them. Recently, there has been a growing interest in such inequalities in connection with the study of linear and nonlinear partial differential equations of subelliptic type and related problems in CR and sub-Riemannian geometry. In this context, it is also of interest to study the situation in which the whole space is replaced by a bounded domain Ω and, instead of a one point singularity such as in (1.2), (1.3), one has the distance from a lower dimensional set. We will be particularly interested in the case in which such a set is the boundary $\partial\Omega$ of the ground domain.

In this paper, we consider various types of Hardy–Sobolev inequalities on a Carnot–Carathéodory space (Ω, d) associated to a system of smooth vector fields $X = \{X_1, X_2, \dots, X_m\}$ on $\mathbb{R}^n$ satisfying the Hörmander finite rank condition [31]

$$rank \; Lie[X_1, \dots, X_m] \equiv n. \tag{1.4}$$

Here, Ω is a connected, (Euclidean) bounded open set in $\mathbb{R}^n$, and d is the Carnot–Carathéodory (CC hereafter) metric generated by X. For instance, a situation of special geometric interest is that when the ambient manifold is a nilpotent Lie group whose Lie algebra admits a stratification of finite step $r \geqslant 1$ (see [18, 20, 53]. These groups are called Carnot groups of step r. When $r > 1$ such groups are non-Abelian, whereas when $r = 1$ one essentially has Euclidean $\mathbb{R}^n$ with its standard translations and dilations.

For a function $\varphi \in C^1(\Omega)$ we indicate with $X\varphi = (X_1\varphi, \dots, X_m\varphi)$ its "gradient" with respect to the system X. One of our main concerns is the trace inequality

$$\int_\Omega |\varphi(x)|^p V(x)dx \leqslant C \int_\Omega |X\varphi|^p dx, \qquad \varphi \in C_0^\infty(\Omega), \tag{1.5}$$

where V is a general weight, i.e., a nonnegative locally integrable function on Ω, and $1 < p < +\infty$. This includes Hardy inequalities of the form

$$\int_\Omega \frac{|\varphi(x)|^p}{\delta(x)^p} dx \leqslant C \int_\Omega |X\varphi|^p dx, \tag{1.6}$$

and

$$\int_\Omega \frac{|\varphi(x)|^p}{d(x,x_0)^p} dx \leqslant C \int_\Omega |X\varphi|^p dx, \tag{1.7}$$

as well as the mixed form

$$\int_\Omega \frac{|\varphi(x)|^p}{\delta(x)^{p-\gamma} d(x,x_0)^\gamma} dx \leqslant C \int_\Omega |X\varphi|^p dx. \tag{1.8}$$

In (1.6), we denote by $\delta(x) = \inf\{d(x,y) : y \in \partial\Omega\}$ the CC distance of x from the boundary of Ω. In (1.7), we denote by x_0 a fixed point in Ω, whereas in (1.8) we have $0 \leqslant \gamma \leqslant p$.

Our approach to the inequalities (1.6)-(1.8) is based on results on subelliptic capacitary and Fefferman–Phong type inequalities in [13], Whitney decompositions, and the so-called pointwise Hardy inequality

$$|\varphi(x)| \leqslant C\delta(x) \Big(\sup_{0<r\leqslant 4\delta(x)} \frac{1}{|B(x,r)|} \int_{B(x,r)} |X\varphi|^q dy \Big)^{\frac{1}{q}}, \tag{1.9}$$

where $1 < q < p$. In (1.9), $B(x,r)$ denotes the CC ball centered at x of radius r.

We use the ideas in [25] and [37] to show that (1.9) is essentially equivalent to several conditions on the geometry of the boundary of Ω, one of which is the uniform (X,p)-fatness of $\mathbb{R}^n \setminus \Omega$, a generalization of that of uniform p-fatness introduced in [38] in the Euclidean setting (see Definition 3.2

below). The inequality (1.9) is also equivalent to other thickness conditions of $\mathbb{R}^n \setminus \Omega$ measured in terms of a certain Hausdorff content which is introduced in Definition 3.5. For the precise statement of these results we refer to Theorem 3.9.

We stress here that the class of uniformly (X,p)-fat domains is quite rich. For instance, when $\mathbf{G}$ is a Carnot group of step $r = 2$, then every (Euclidean) $C^{1,1}$ domain is uniformly (X,p)-fat for every $p > 1$ (see [7, 43]). On the other hand, one would think that the Carnot–Carathéodory balls should share this property, but it was shown in [7] that this is not the case since even in the simplest setting of the Heisenberg group these sets fail to be regular for the Dirichlet problem for the relevant sub-Laplacian.

We now discuss our results concerning the trace inequality (1.5). In the Euclidean setting, a necessary and sufficient condition on V was found by Maz'ya [40] in 1962 (see also [41, Theorem 2.5.2]), i.e., the inequality (1.5) with the standard Euclidean metric induced by $X = \{\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\}$ holds if and only if

$$\sup_{\substack{K \subset \Omega \\ K \text{ compact}}} \frac{\int_K V(x)dx}{\operatorname{cap}_p(K,\Omega)} < +\infty, \tag{1.10}$$

where $\operatorname{cap}_p(K,\Omega)$ is the (X,p)-capacity K defined by

$$\operatorname{cap}_p(K,\Omega) = \inf \left\{ \int_\Omega |Xu|^p dx : u \in C_0^\infty(\Omega), u \geqslant 1 \text{ on } K \right\}.$$

Maz'ya's result was generalized to the subelliptic setting by the first named author in [13]. However, although Corollary 5.9 in [13] implies that $V \in L^{\frac{Q}{p},\infty}(\Omega)$ is sufficient for (1.5), which is the case of an isolated singularity as in (1.7), the Hardy inequality (1.6) could not be deduced directly from it since $\delta(\cdot)^{-p} \notin L^{\frac{Q}{p},\infty}(\Omega)$. Here, $1 < p < Q$, where Q is the local homogeneous dimension of Ω (see Sect. 2). On the other hand, in the Euclidean setting the Hardy inequality (1.6) was established in [1], [38] and [51] (see also [42] and [3] for other settings) under the assumption that $\mathbb{R}^n \setminus \Omega$ is uniformly p-fat.

In this paper, we combine a "localized" version of (1.10) and the uniform (X,p)-fatness of $\mathbb{R}^n \setminus \Omega$ to allow the treatment of weights V with singularities which are distributed both inside and on the boundary of Ω. More specifically, we show that if $\mathbb{R}^n \setminus \Omega$ is uniformly (X,p)-fat, then the inequality (1.5) holds if and only if

$$\sup_{B \in \mathcal{W}} \sup_{\substack{K \subset 2B \\ K \text{ compact}}} \frac{\int_K V(x)dx}{\operatorname{cap}_p(K,\Omega)} < +\infty,$$

where $\mathcal{W} = \{B_j\}$ is a Whitney decomposition of Ω as in Lemma 4.2 below (see Theorem 4.3). In the Euclidean setting, this idea was introduced in [26]. Moreover, a localized version of Fefferman–Phong condition

$$\sup_{B\in\mathcal{W}} \sup_{\substack{x\in 2B \\ 0<r<\operatorname{diam}(B)}} \int_{B(x,r)} V(y)^s dy \leqslant C\,\frac{|B(x,r)|}{r^{sp}}$$

for some $s > 1$, is also shown to be sufficient for (1.5) (see Theorem 4.5).

With these general results in hands, in Corollaries 4.6 and 4.7 we deduce the Hardy type inequalities (1.6), (1.7), and (1.8) for domains Ω whose complements are uniformly (X,p)-fat. Note that in (1.7) and (1.8) one has to restrict the range of p to $1 < p < Q(x_0)$, where $Q(x_0)$ is the homogeneous dimension at x_0 with respect to the system X (see Sect. 2). It is worth mentioning that in the Euclidean setting inequalities of the form (1.8) were obtained in [16], but only for more regular domains, say, $C^{1,\alpha}$ domains or domains that satisfy a uniform exterior sphere condition. In closing we mention that our results are of a purely metrical character and that, similarly to [13], they can be easily generalized to the case in which the vector fields are merely Lipschitz continuous and they satisfy the conditions in [23].

2 Preliminaries

Let $X = \{X_1,\ldots,X_m\}$ be a system of C^∞ vector fields in $\mathbb{R}^n$, $n \geqslant 3$, satisfying the Hörmander finite rank condition (1.4). For any two points $x, y \in \mathbb{R}^n$ a piecewise C^1 curve $\gamma(t) : [0,T] \to \mathbb{R}^n$ is said to be *sub-unitary* with respect to the system of vector fields X if for every $\xi \in \mathbb{R}^n$ and $t \in (0,T)$ for which $\gamma'(t)$ exists one has

$$(\gamma'(t)\cdot\xi)^2 \leqslant \sum_{i=1}^m (X_i(\gamma(t))\cdot\xi)^2.$$

We note explicitly that the above inequality forces $\gamma'(t)$ to belong to the span of $\{X_1(\gamma(t)),\ldots, X_m(\gamma(t))\}$. The sub-unit length of γ is by definition $l_s(\gamma) = T$. Given $x, y \in \mathbb{R}^n$, denote by $\mathcal{S}_\Omega(x,y)$ the collection of all sub-unitary $\gamma : [0,T] \to \Omega$ which join x to y. The accessibility theorem of Chow and Rashevsky (see [46] and [9]) states that, given a connected open set $\Omega \subset \mathbb{R}^n$, for every $x, y \in \Omega$ there exists $\gamma \in \mathcal{S}_\Omega(x,y)$. As a consequence, if we pose

$$d_\Omega(x,y) \;=\; \inf\{l_s(\gamma) \mid \gamma \in \mathcal{S}_\Omega(x,y)\},$$

we obtain a distance on Ω, called the *Carnot–Carathéodory* (CC) *distance on* Ω, associated with the system X. When $\Omega = \mathbb{R}^n$, we write $d(x,y)$ instead of $d_{\mathbb{R}^n}(x,y)$. It is clear that $d(x,y) \leqslant d_\Omega(x,y)$, $x, y \in \Omega$, for every connected open set $\Omega \subset \mathbb{R}^n$. In [44], it was proved that for every connected $\Omega \subset\subset \mathbb{R}^n$ there exist $C, \varepsilon > 0$ such that

$$C\ |x-y|\ \leqslant d_\Omega(x,y)\ \leqslant C^{-1}\ |x-y|^\varepsilon, \qquad x,y \in \Omega. \tag{2.1}$$

This gives $d(x,y)\ \leqslant C^{-1}|x-y|^\varepsilon$, $x, y \in \Omega$, and therefore

$$i : (\mathbb{R}^n, |\cdot|) \to (\mathbb{R}^n, d) \qquad \textit{is continuous.}$$

It is easy to see that also the continuity of the opposite inclusion holds [23], hence the metric and the Euclidean topology are compatible. In particular, the compact sets with respect to either topology are the same.

For $x \in \mathbb{R}^n$ and $r > 0$ we let $B(x,r) = \{y \in \mathbb{R}^n \mid d(x,y) < r\}$. The basic properties of these balls were established by Nagel, Stein and Wainger in their seminal paper [44]. Denote by $Y_1, \dots, Y_l$ the collection of the X_j's and of those commutators which are needed to generate $\mathbb{R}^n$. A formal "degree" is assigned to each Y_i, namely the corresponding order of the commutator. If $I = (i_1, \dots, i_n), 1 \leqslant i_j \leqslant l$, is an n-tuple of integers, following [44] we let $d(I) = \sum_{j=1}^{n} deg(Y_{i_j})$, and $a_I(x) = \det\ (Y_{i_1}, \dots, Y_{i_n})$. The *Nagel–Stein–Wainger polynomial* is defined by

$$\Lambda(x,r)\ =\ \sum_I\ |a_I(x)|\ r^{d(I)}, \qquad r > 0. \tag{2.2}$$

For a given compact set $K \subset \mathbb{R}^n$ we denote by

$$Q = \sup\{d(I) : |a_I(x)| \neq 0,\ x \in K\} \tag{2.3}$$

the *local homogeneous dimension* of K with respect to the system X and by

$$Q(x) = \inf\{d(I) : |a_I(x)| \neq 0\} \tag{2.4}$$

the *homogeneous dimension* at x with respect to X. It is obvious that $3 \leqslant n \leqslant Q(x) \leqslant Q$. It is immediate that for every $x \in K$, and every $r > 0$, one has

$$t^Q \Lambda(x,r) \leqslant \Lambda(x,tr) \leqslant t^{Q(x)} \Lambda(x,r) \tag{2.5}$$

for any $0 \leqslant t \leqslant 1$, and thus

$$Q(x) \leqslant \frac{r\Lambda'(x,r)}{\Lambda(x,r)} \leqslant Q. \tag{2.6}$$

For a simple example consider in $\mathbb{R}^3$ the system

$$X\ =\ \{X_1, X_2, X_3\}\ =\ \left\{\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, x_1\frac{\partial}{\partial x_3}\right\}.$$

It is easy to see that $l = 4$ and

$$\{Y_1, Y_2, Y_3, Y_4\} = \{X_1, X_2, X_3, [X_1, X_3]\}.$$

Moreover, $Q(x) = 3$ for all $x \neq 0$, whereas for any compact set K containing the origin $Q(0) = Q = 4$.

The following fundamental result is due to Nagel, Stein, and Wainger [44]: *For every compact set* $K \subset \mathbb{R}^n$ *there exist constants* $C, R_0 > 0$ *such that for any* $x \in K$ *and* $0 < r \leqslant R_0$ *one has*

$$C\Lambda(x,r) \leqslant |B(x,r)| \leqslant C^{-1}\Lambda(x,r). \tag{2.7}$$

As a consequence, there exists C_0 such that for any $x \in K$ and $0 < r < s \leqslant R_0$ we have

$$C_0 \left(\frac{r}{s}\right)^Q \leqslant \frac{|B(x,r)|}{|B(x,s)|}. \tag{2.8}$$

Henceforth, the numbers C_0 and R_0 above will be referred to as the *local parameters* of K with respect to the system X. If E is any (Euclidean) bounded set in $\mathbb{R}^n$, then the local parameters of E are defined as those of $\overline{E}$. We mention explicitly that the number R_0 is always chosen in such a way that the closed metric balls $\overline{B}(x,R)$, with $x \in K$ and $0 < R \leqslant R_0$, are compact (see [23, 24]). This choice is motivated by the fact that in a CC space the closed metric balls of large radii are not necessarily compact. For instance, if one considers the Hörmander vector field on $\mathbb{R}$ given by $X_1 = (1+x^2)\frac{d}{dx}$, then for any $R \geqslant \pi/2$ one has $B(0,R) = \mathbb{R}$ (see [23]).

Given an open set $\Omega \subset \mathbb{R}^n$, and $1 \leqslant p \leqslant \infty$, we denote by $S^{1,p}(\Omega)$, the subelliptic Sobolev space associated with the system X is defined by

$$S^{1,p}(\Omega) = \{u \in L^p(\Omega) : X_i u \in L^p(\Omega), i = 1, \dots, m\},$$

where $X_i u$ is understood in the distributional sense, i.e.,

$$\langle X_i u, \varphi \rangle = \int\limits_{\Omega} u X_i^* \varphi dx$$

for every $\varphi \in C_0^\infty(\Omega)$. Here, X_i^* denotes the formal adjoint of X_i. Endowed with the norm

$$\|u\|_{S^{1,p}(\Omega)} = \left(\int\limits_{\Omega} (|u|^p + |Xu|^p) dx \right)^{\frac{1}{p}}, \tag{2.9}$$

$S^{1,p}(\Omega)$ is a Banach space which admits $C^\infty(\Omega) \cap S^{1,p}(\Omega)$ as a dense subset (see [23, 21]). The local version of $S^{1,p}(\Omega)$ is denoted by $S^{1,p}_{\mathrm{loc}}(\Omega)$, whereas the completion of $C_0^\infty(\Omega)$ under the norm in (2.9) is denoted by $S_0^{1,p}(\Omega)$.

A fundamental result in [47] shows that for any bounded open set $\Omega \subset \mathbb{R}^n$ the space $S_0^{1,p}(\Omega)$ embeds into a standard fractional Sobolev space $W_0^{s,p}(\Omega)$, where $s = 1/r$ and r is the largest number of commutators which are needed to generate the Lie algebra over $\overline{\Omega}$. Since, on the other hand, we have classically $W_0^{s,p}(\Omega) \subset L^p(\Omega)$, we obtain the following Poincaré inequality:

$$\int_\Omega |\varphi|^p \, dx \ \leqslant \ C(\Omega) \int_\Omega |X\varphi|^p \, dx \ , \quad \varphi \in S_0^{1,p}(\Omega). \tag{2.10}$$

Another fundamental result which plays a pervasive role in this paper is the following global Poincaré inequality on metric balls due to Jerison [33]. Henceforth, given a measurable set $E \subset \mathbb{R}^n$, the notation φ_E indicates the average of φ over E with respect to Lebesgue measure.

Theorem 2.1. *Let $K \subset \mathbb{R}^n$ be a compact set with local parameters C_0 and R_0. For any $1 \leqslant p < \infty$ there exists $C = C(C_0, p) > 0$ such that for any $x \in K$ and every $0 < r \leqslant R_0$ one has for all $\varphi \in S^{1,p}(B(x,r))$*

$$\int_{B(x,r)} |\varphi - \varphi_{B(x,r)}|^p dy \ \leqslant \ C \ r^p \int_{B(x,r)} |X\varphi|^p dy. \tag{2.11}$$

We also need the following basic result on the existence of cut-off functions in metric balls (see [24] and also [21]). Given a set $\Omega \subset \mathbb{R}^n$, we indicate with $C_d^{0,1}(\Omega)$ the collection of functions $\varphi \in C(\Omega)$ for which there exists $L \geqslant 0$ such that

$$|\varphi(x) - \varphi(y)| \ \leqslant \ L \ d(x,y), \quad x, y \in \Omega.$$

We recall that, thanks to the Rademacher–Stepanov type theorem proved in [24, 21], if Ω is metrically bounded, then any function in $C_d^{0,1}(\Omega)$ belongs to the space $S^{1,\infty}(\Omega)$. This is true, in particular, when Ω is a metric ball.

Theorem 2.2. *Let $K \subset \mathbb{R}^n$ be a compact set with local parameters C_0 and R_0. For every $0 < s < t < R_0$ there exists $\varphi \in C_d^{0,1}(\mathbb{R}^n)$, $0 \leqslant \varphi \leqslant 1$, such that*

(i) *$\varphi \equiv 1$ on $B(x,s)$ and $\varphi \equiv 0$ outside $B(x,t)$,*

(ii) *$|X\varphi| \leqslant \frac{C}{t-s}$ for a.e. $x \in \mathbb{R}^n$,*

for some $C > 0$ depending on C_0. Furthermore, we have $\varphi \in S^{1,p}(\mathbb{R}^n)$ for every $1 \leqslant p < \infty$.

A *condenser* is a couple (K, Ω), where Ω is open and $K \subset \Omega$ is compact. The *subelliptic p-capacity* of (K, Ω) is defined by

$$\mathrm{cap}_p(K, \Omega) = \inf \left\{ \int_\Omega |X\varphi|^p dx : \varphi \in C_d^{0,1}(\mathbb{R}^n), \mathrm{supp} \ \ \varphi \subset \Omega, \varphi \geqslant 1 \text{ on } K \right\}.$$

As usual, it can be extended to arbitrary sets $E \subset \Omega$ by letting

$$\operatorname{cap}_p(E,\Omega) = \inf_{\substack{G\subset\Omega \text{ open} \\ E\subset G}} \ \sup_{\substack{K\subset G \\ K \text{ compact}}} \operatorname{cap}_p(K,\Omega).$$

It was proved in [12] that the subelliptic p-capacity of a metric “annular” condenser has the following two-sided estimate which will be used extensively in the paper. Given a compact set $K \subset \mathbb{R}^n$ with local parameters C_0 and R_0, and homogeneous dimension Q, for any $1 < p < \infty$ there exist $C_1, C_2 > 0$ depending only on C_0 and p such that

$$C_1 \frac{|B(x,r)|}{r^p} \leqslant \operatorname{cap}_p(B(x,r), B(x,2r)) \leqslant C_2 \frac{|B(x,r)|}{r^p} \tag{2.12}$$

for all $x \in K$, and $0 < r \leqslant R_0/2$.

The subelliptic p-Laplacian associated to the system X is the quasilinear operator defined by

$$\mathcal{L}_p[u] = -\sum_{i=1}^m X_i^*(|Xu|^{p-2}X_i u).$$

A weak solution $u \in S^{1,p}_{\text{loc}}(\Omega)$ to the equation $\mathcal{L}_p[u] = 0$ is said to be $\mathcal{L}_p$-*harmonic* in Ω. It is well-known that every $\mathcal{L}_p$-harmonic function in Ω has a Hölder continuous representative (see [4]). This means that, if C_0 and R_0 are the local parameters of Ω, then there exist $0 < \alpha < 1$ and $C > 0$, depending on C_0 and p, such that for every $0 < R \leqslant R_0$ for which $B_{4R}(x_0) \subset \Omega$ one has

$$|u(x) - u(y)| \leqslant C \left(\frac{d(x,y)}{R}\right)^{\alpha} \left(\frac{1}{|B_{2R}(x_0)|} \int\limits_{B_{2R}(x_0)} |u|^p dx\right)^{1/p}. \tag{2.13}$$

Given a bounded open set $\Omega \subset \mathbb{R}^n$ and $1 < p < \infty$, the Dirichlet problem for Ω and $\mathcal{L}_p$ consists in finding, for every given $\varphi \in S^{1,p}(\Omega) \cap C(\overline{\Omega})$, a function $u \in S^{1,p}(\Omega)$ such that

$$\mathcal{L}_p[u] = 0 \quad \text{in } \Omega, \qquad u - \varphi \in S^{1,p}_0(\Omega). \tag{2.14}$$

Such a problem admits a unique solution (see [12]). A point $x_0 \in \partial\Omega$ is called *regular* if for every $\varphi \in S^{1,p}(\Omega) \cap C(\overline{\Omega})$,one has $\lim\limits_{x\to x_0} u(x) = \varphi(x_0)$. If every $x_0 \in \partial\Omega$ is regular, then we say that Ω is *regular*. We need the following basic Wiener type estimate proved in [12].

Theorem 2.3. *Given a bounded open set $\Omega \subset \mathbb{R}^n$ with local parameters C_0 and R_0, let $\varphi \in S^{1,p}(\Omega) \cap C(\overline{\Omega})$. Consider the (unique) solution u to the Dirichlet problem* (2.14)*. There exists $C = C(p, C_0) > 0$ such that for given $x_0 \in \partial\Omega$ and $0 < r < R \leqslant R_0/3$ one has with $\Omega^c = \mathbb{R}^n \setminus \Omega$*

$$\begin{aligned}\operatorname{osc}\{u,\Omega\cap B(x_0,r)\} &\leqslant \operatorname{osc}\{\varphi,\partial\Omega\cap\overline{B}(x_0,2R)\}\\ &+\operatorname{osc}(\varphi,\partial\Omega)\ \exp\left\{-\ C\int_r^R\left[\frac{\operatorname{cap}_p\ (\Omega^c\cap\overline{B}(x_0,t),B(x_0,2t))}{\operatorname{cap}_p\ (\overline{B}(x_0,t),B(x_0,2t))}\right]\ \frac{dt}{t}\right\}.\end{aligned}$$

Remark 2.4. It is clear from Theorem 2.3 that if Ω is *thin at* $x_0\in\partial\Omega$, i.e., if one has

$$\liminf_{t\to 0^+}\frac{\operatorname{cap}_p\ (\Omega^c\cap\overline{B}(x_0,t),B(x_0,2t))}{\operatorname{cap}_p\ (\overline{B}(x_0,t),B(x_0,2t))}\ >\ 0\ ,$$

then x_0 is regular for the Dirichlet problem (2.14).

A lower semicontinuous function $u:\Omega\to(-\infty,\infty]$ such that $u\not\equiv+\infty$ is called $\mathcal{L}_p$-*superharmonic* in Ω if for all open sets D such that $\overline{D}\subset\Omega$, and all $\mathcal{L}_p$-harmonic functions $h\in C(\overline{D})$ the inequality $h\leqslant u$ on ∂D implies $h\leqslant u$ in D. Similarly to what is done in the classical case in [30], one can associate with each $\mathcal{L}_p$-superharmonic function u in Ω a nonnegative (not necessarily finite) Radon measure $\mu[u]$ such that $-\mathcal{L}_p[u]=\mu[u]$. This means that

$$\int_\Omega |Xu|^{p-2}Xu\cdot X\varphi\ dx\ =\ \int_\Omega \varphi\ d\mu[u]$$

for all $\varphi\in C_0^\infty(\Omega)$. Here, Xu is defined a.e. by

$$Xu=\lim_{k\to\infty}X(\min\{u,k\}).$$

It is known that if either $u\in L^\infty(\Omega)$ or $u\in S^{1,r}_{\rm loc}(\Omega)$ for some $r\geqslant 1$, then Xu coincides with the regular distributional derivatives. In general, we have $Xu\in L^s_{\rm loc}(\Omega)$ for $0<s<\frac{Q(p-1)}{Q-1}$ (see, for example, [50] and [30]).

We need the following basic pointwise estimates for $\mathcal{L}_p$-superharmonic functions. This result was first established by Kilpeläinen and Malý [35] in the elliptic case and extended to the setting of CC metrics by Trudinger and Wang [50]. For a generalization to more general metric spaces we refer the reader to [3]. We recall that for given $1<p<\infty$ the *p-Wolff's potential* of a Radon measure μ on a metric ball $B(x,R)$ is defined by

$$\mathbf{W}^R_p\mu(x)\ =\ \int_0^R\left[\frac{\mu(B(x,t))}{t^{-p}|B(x,t)|}\right]^{\frac{1}{p-1}}\frac{dt}{t}. \tag{2.15}$$

Theorem 2.5. *Let $K\subset\mathbb{R}^n$ be a compact set with relative local parameters C_0 and R_0. If $x\in K$ and $R\leqslant R_0/2$, let $u\geqslant 0$ be $\mathcal{L}_p$-superharmonic in $B(x,2R)$ with associated measure $\mu=-\mathcal{L}_p[u]$. There exist positive constants C_1 and C_2 depending only on p and C_0 such that*

$$C_1 \mathbf{W}_p^R \mu(x) \leqslant u(x) \leqslant C_2 \left\{ \mathbf{W}_p^{2R} \mu(x) + \inf_{B(x,R)} u \right\}.$$

3 Pointwise Hardy Inequalities

We begin this section by generalizing a Sobolev type inequality that, in the Euclidean setting, was found by Maz'ya [41, Chapt. 10].

Lemma 3.1. *Let $K \subset \mathbb{R}^n$ be a compact set with local parameters C_0 and R_0. For $x \in K$ and $r \leqslant R_0/2$ we set $B = B(x,r)$. Given $1 \leqslant q < \infty$, there exists a constant $C > 0$ depending only on C_0 and q such that for all $\varphi \in C^\infty(2B)$*

$$|\varphi_B| \leqslant C \left(\frac{1}{\operatorname{cap}_q(\{\varphi = 0\} \cap \overline{B}, 2B)} \int_{2B} |X\varphi|^q dx \right)^{\frac{1}{q}}. \tag{3.1}$$

Proof. We may assume that $\varphi_B \neq 0$; otherwise, there is nothing to prove. Let $\eta \in C_d^{0,1}(\mathbb{R}^n)$, $0 \leqslant \eta \leqslant 1$, supp $\eta \subset 2B$, $\eta = 1$ on $\overline{B}$ and $|X\eta| \leqslant \frac{C}{r}$ be a cut-off function as in Theorem 2.2. Define $\phi = \eta(\varphi_B - \varphi)/\varphi_B$. Then $\phi \in C_d^{0,1}(\mathbb{R}^n)$, supp $\phi \subset 2B$, and $\phi = 1$ on $\{\varphi = 0\} \cap \overline{B}$. It thus follows that

$$\begin{aligned} \operatorname{cap}_q(\{\varphi = 0\} \cap \overline{B}, 2B) &\leqslant \int_{2B} |X\phi|^q dx \\ &\leqslant |\varphi_B|^{-q} \int_{2B} |X\eta|^q |\varphi - \varphi_B|^q dx + |\varphi_B|^{-q} \int_{2B} |X\varphi|^q dx \\ &\leqslant C|\varphi_B|^{-q} r^{-q} \int_{2B} |\varphi - \varphi_B|^q dx + |\varphi_B|^{-q} \int_{2B} |X\varphi|^q dx. \end{aligned} \tag{3.2}$$

On the other hand, by Theorem 2.1 and (2.8), we infer

$$\begin{aligned} \int_{2B} |\varphi - \varphi_B|^q dx &\leqslant C \int_{2B} |\varphi - \varphi_{2B}|^q dx + C \int_{2B} |\varphi_B - \varphi_{2B}|^q dx \\ &\leqslant Cr^q \int_{2B} |X\varphi|^q dx + C \int_{2B} |\varphi - \varphi_{2B}|^q dx \\ &\leqslant Cr^q \int_{2B} |X\varphi|^q dx. \end{aligned}$$

Inserting the latter inequality in (3.2), we find

$$\text{cap}_q(\{\varphi = 0\} \cap \overline{B}, 2B) \leqslant C|\varphi_B|^{-q} \int_{2B} |X\varphi|^q dx,$$

which gives the desired inequality (3.1). □

We now introduce the notion of uniform (X,p)-fatness. As Theorem 3.9 below proves, such a notion turns out to be equivalent to a pointwise Hardy inequality and to a uniform thickness property expressed in terms of the Hausdorff content.

Definition 3.2. We say that a set $E \subset \mathbb{R}^n$ is *uniformly (X,p)-fat* with constants $c_0, r_0 > 0$ if

$$\text{cap}_p(E \cap \overline{B}(x,r), B(x,2r)) \;\geqslant\; c_0 \;\text{cap}_p(\overline{B}(x,r), B(x,2r))$$

for all $x \in \partial E$ and for all $0 < r \leqslant r_0$.

The potential theoretic relevance of Definition 3.2 is underscored in Remark 2.4. From the latter it follows that if $\mathbb{R}^n \setminus \Omega$ is uniformly (X,p)-fat, then for every $x_0 \in \partial\Omega$ one has for every $\varphi \in S^{1,p}(\Omega) \cap C(\overline{\Omega})$

$$\text{osc}\,\{u, \Omega \cap B(x_0,r)\} \;\leqslant \text{osc}\,\{\varphi, \partial\Omega \cap \overline{B}(x_0,2R)\}$$

and, therefore, Ω is regular for the Dirichlet problem for the subelliptic p-Laplacian $\mathcal{L}_p$.

Uniformly (X,p)-fat sets enjoy the following self-improvement property which was discovered in [38] in the Euclidean setting. Such a property holds also in the setting of weighted Sobolev spaces and degenerate elliptic equations [42]. The proof in [42] uses the Wolff potential and works also in the general setting of metric spaces [3]. For the sake of completeness, we include its details here.

Theorem 3.3. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with local parameters C_0 and R_0. There exists a constant $0 < r_0 \leqslant R_0/100$ such that whenever $\mathbb{R}^n \setminus \Omega$ is uniformly (X,p)-fat with constants c_0 and r_0, then it is also uniformly (X,q)-fat for some $q < p$ with constants c_1 and r_0.*

Proof. Let $\text{dist}(x,\Omega) = \inf\{d(x,y) : y \in \Omega\}$. Denote by $U \subset \mathbb{R}^n$ the compact set

$$U = \{x \in \mathbb{R}^n : \text{dist}(x,\Omega) \leqslant R_0\},$$

with local parameters C_1, R_1. We show that if $\mathbb{R}^n \setminus \Omega$ is uniformly (X,p)-fat with constants c_0 and $r_0 = \min\{R_0, R_1\}/100$, then it is also uniformly (X,q)-fat for some $q < p$ with constants c_1 and r_0. To this end, we fix $x_0 \in \partial\Omega$ and $0 < R \leqslant r_0$. Following [38], we first claim that there e xists a compact set $K \subset (\mathbb{R}^n \setminus \Omega) \cap \overline{B}(x_0,R)$ containing x_0 such that K is uniformly (X,p)-fat with constants $c_1 > 0$ and R. Indeed, let $E_1 = (\mathbb{R}^n \setminus \Omega) \cap B(x_0, \frac{R}{2})$, and inductively let

$$E_k = (\mathbb{R}^n \setminus \Omega) \cap \left(\bigcup_{x \in E_{k-1}} B(x, \tfrac{R}{2^k}) \right), \quad k \in \mathbb{N}.$$

Then it is easy to see that K can be taken as the closure of $\cup_k E_k$.

Let now $B = B(x_0, R)$. Denote by $\widehat{P}_K$ the potential of K in $2B$, i.e., $\widehat{P}_K$ is the lower semicontinuous regularization

$$\widehat{P}_K(x) = \lim_{r \to 0} \inf_{B_r(x)} P_K,$$

where P_K is defined by

$$P_K = \inf\{u : u \text{ is } \mathcal{L}_p\text{-superharmonic in } 2B, \text{ and } u \geqslant \chi_K\}.$$

Let $\mu = -\mathcal{L}_p[\widehat{P}_K]$. Then $\operatorname{supp} \mu \subset \partial K$ and

$$\mu(K) = \operatorname{cap}_p(K, 2B). \tag{3.3}$$

Moreover, $\widehat{P}_K = P_K$ except for a set of zero capacity $\operatorname{cap}_p(\cdot, 2B)$ (see [50]). Hence $\widehat{P}_K$ is the unique solution in $S_0^{1,p}(2B)$ to the Dirichlet problem

$$\mathcal{L}_p[u] = 0 \quad \text{in} \quad 2B \setminus K, \qquad u - f \in S_0^{1,p}(2B \setminus K)$$

for any $f \in C_0^\infty(2B)$ such that $f \equiv 1$ on K. Thus, by Theorem 2.3 and the (X, p)-fatness of K, there are constants $C > 0$ and $\alpha > 0$ independent of R such that

$$\operatorname{osc}(\widehat{P}_K, B(x, r)) \leqslant C R^{-\alpha} r^\alpha \tag{3.4}$$

for all $x \in \partial K$ and $0 < r \leqslant R/2$. From the lower Wolff potential estimate in Theorem 2.5 we have

$$\begin{aligned}\left[\frac{\mu(B(x,r))}{r^{-p}|B(x,r)|}\right]^{\frac{1}{p-1}} &\leqslant C\,\mathbf{W}_p^{2r}\mu(x) \leqslant C\left(\widehat{P}_K(x) - \inf_{B(x,4r)} \widehat{P}_K\right) \\ &\leqslant C \operatorname{osc}(\widehat{P}_K, B(x, 4r)).\end{aligned}$$

Thus, from (3.4) it follows that

$$\mu(B(x, r)) \leqslant C R^{-\alpha(p-1)} r^{\alpha(p-1)-p} |B(x, r)| \tag{3.5}$$

for all $x \in \partial K$ and $0 < r \leqslant R/8$. Moreover, since $\operatorname{supp} \mu \subset \partial K$, we see from the doubling property (2.8) that (3.5) holds also for all $x \in B(x_0, 2R)$ and $0 < r \leqslant R/16$. In fact, it then holds for all $R/16 < r \leqslant 3R$ as well since, again by (2.8), the ball $B(x, r)$ can be covered by a fixed finite number of balls of radius $R/16$.

We next pick $q \in \mathbb{R}$ such that $p - \alpha(p - 1) < q < p$ and define a measure $\nu = R^{p-q}\mu$. From (3.5) it follows that for all $x \in B(x_0, 2R)$,

$$\mathbf{W}_q^{3R}\nu(x) \leqslant CR^{\frac{p-q-\alpha(p-1)}{q-1}} \int_0^{3R} r^{\frac{q-p+\alpha(p-1)}{q-1}} \frac{dr}{r} \leqslant M, \tag{3.6}$$

where M is independent of R. Thus, by [2, Lemma 3.3], ν belongs to the dual space of $S_0^{1,q}(2B)$ and there is a unique solution $v \in S_0^{1,q}(2B)$ to the problem

$$\begin{aligned} -\mathcal{L}_q[v] = \nu \quad \text{in} \quad 2B \\ v = 0 \quad \text{on} \quad \partial(2B). \end{aligned} \tag{3.7}$$

We now claim that

$$v(x) \leqslant c \tag{3.8}$$

for all $x \in 2B$ and for a constant c independent of R. To this end, it is enough to show (3.8) only for $x \in \overline{B}$ since v is $\mathcal{L}_q$-harmonic in $2B \setminus \overline{B}$ and $v = 0$ on $\partial(2B)$. Fix now $x \in \overline{B}$. By Theorem 2.5, we have

$$v(x) \leqslant C \left\{ \mathbf{W}_q^{3R}\nu(x) + \inf_{B(x,R/4)} v \right\}. \tag{3.9}$$

To bound the term $\inf_{B(x,R/4)} v$ in (3.9), we first use $\min\{v,k\}$, $k > 0$, as a test function in (3.7) to obtain

$$\begin{aligned} \int_{2B} |X(\min\{v,k\})|^q dx &= \int_{2B} |Xv|^{q-2} Xv \cdot X(\min\{v,k\}) dx \\ &= \int_{2B} \min\{v,k\} d\nu \ \leqslant \ k\ \nu(K). \end{aligned} \tag{3.10}$$

Consequently,

$$\operatorname{cap}_q(\{v \geqslant k\}, 2B) \leqslant \int_{2B} |X(\min\{v,k\}/k)|^q dx \leqslant k^{1-q}\nu(K) \tag{3.11}$$

for any $k > 0$. The inequality (3.11) with $k = \inf_{B(x,R/4)} v$ then gives

$$\begin{aligned} R^{-q}|B(x,R)| &\leqslant C \operatorname{cap}_q(B(x,R/4), B(x,4R)) \\ &\leqslant C \operatorname{cap}_q(\{v \geqslant k\}, 2B) \\ &\leqslant Ck^{1-q}\nu(K), \end{aligned}$$

which yields the estimate

$$\inf_{B(x,R/4)} v \leqslant C \left(\frac{\nu(K)}{R^{-q}|B(x,R)|} \right)^{\frac{1}{q-1}}. \tag{3.12}$$

Combining (3.6), (3.9), and (3.12). we obtain (3.8), thus proving the claim. Note that for any $\varphi \in C_0^\infty(2B)$ such that $\varphi \geqslant \chi_K$, by the Hölder inequality and by applying (3.10) with $k = c$, we have

$$\begin{aligned}\nu(K) &\leqslant \int\limits_{2B} \varphi d\nu = \int\limits_{\Omega} |Xv|^{q-2} Xv \cdot X\varphi dx \\ &\leqslant \Bigg(\int\limits_{2B} |Xv|^q dx\Bigg)^{\frac{q-1}{q}} \Bigg(\int\limits_{2B} |X\varphi|^q dx\Bigg)^{\frac{1}{q}} \\ &\leqslant [c\,\nu(K)]^{\frac{q-1}{q}} \Bigg(\int\limits_{2B} |X\varphi|^q dx\Bigg)^{\frac{1}{q}}.\end{aligned}$$

Thus, minimizing over such functions φ, we obtain

$$\nu(K) \leqslant c^{q-1} \operatorname{cap}_q(K, 2B).$$

The latter inequality and (2.12) give

$$\begin{aligned}\operatorname{cap}_q((\mathbb{R}^n \setminus \Omega) \cap \overline{B}, 2B) &\geqslant \operatorname{cap}_q(K, 2B) \geqslant C\,\nu(K) = CR^{p-q}\mu(K) \\ &= CR^{p-q}\operatorname{cap}_p(K, 2B) \geqslant CR^{p-q}\operatorname{cap}_p(\overline{B}, 2B) \\ &\geqslant CR^{-q}|B| \geqslant C\operatorname{cap}_q(\overline{B}, 2B)\end{aligned}$$

by (3.3) and the uniform (X,p)-fatness of K. This proves that $\mathbb{R}^n \setminus \Omega$ is uniformly (X,q)-fat, thus completing the proof of the theorem. □

In what follows, given $f \in L^1_{\text{loc}}(\mathbb{R}^n)$, we denote by $\mathcal{M}_R$, $0 < R < \infty$, the truncated centered Hardy–Littlewood maximal function of f defined by

$$\mathcal{M}_R(f)(x) = \sup_{0<r\leqslant R} \frac{1}{|B(x,r)|} \int\limits_{B(x,r)} |f(y)| dy, \qquad x \in \mathbb{R}^n.$$

We note explicitly that if $R_1 < R_2$, then $\mathcal{M}_{R_1}(f)(x) \leqslant \mathcal{M}_{R_2}(f)(x)$. The first consequence of the self-improvement property of uniformly (X,p)-fat set is the following pointwise Hardy inequality which generalizes a result originally found by Hajłasz [25] in the Euclidean setting.

Theorem 3.4. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with local parameters C_0 and R_0. Suppose that $\mathbb{R}^n \setminus \Omega$ is uniformly (X,p)-fat with constants c_0 and r_0, where $0 < r_0 \leqslant R_0/100$ is as in Theorem 3.3. There exist $1 < q < p$ and a constant $C > 0$, both depending on C_0 and p, such that the inequality*

$$|u(x)| \leqslant C\delta(x)\Big(\mathcal{M}_{4\delta(x)}(|\nabla u|^q)(x)\Big)^{\frac{1}{q}} \tag{3.13}$$

holds for all $x \in \Omega$ *with* $\delta(x) < r_0$ *and all compactly supported* $u \in C_d^{0,1}(\Omega)$.

Proof. For $x \in \Omega$ with $\delta(x) < r_0$ we let $B = B(\overline{x}, \delta(x))$, where $\overline{x} \in \partial\Omega$ is chosen so that $|x - \overline{x}| = \delta(x)$. By the fatness assumption and Theorem 3.3, there exists $1 < q < p$ such that

$$\mathrm{cap}_{1,q}(\overline{B} \cap (\mathbb{R}^n \setminus \Omega), 2B) \geqslant C|B|\delta(x)^{-q}.$$

Thus, by Lemma 3.1 above and Theorem 1.1 in [6],

$$\begin{aligned} u(x) &\leqslant |u(x) - u_B| + |u_B| \\ &\leqslant C\int_{2B} |Xu(y)|\frac{d(x,y)}{|B(x,d(x,y))|}dy + C\Big(\frac{\int_{2B}|Xu|^q dx}{|B|\delta(x)^{-q}}\Big)^{\frac{1}{q}}. \end{aligned} \tag{3.14}$$

Note that by the doubling property (2.8),

$$\begin{aligned} &\int_{2B} |Xu(y)|\frac{d(x,y)}{|B(x,d(x,y))|}dy \\ &\leqslant \int_{B(x,4\delta(x))} |Xu(y)|\frac{d(x,y)}{|B(x,d(x,y))|}dy \\ &= \sum_{k=0}^{\infty} \int_{B(x,2^{-k}4\delta(x))\setminus B(x,2^{-k-1}4\delta(x))} |Xu(y)|\frac{d(x,y)}{|B(x,d(x,y))|}dy \\ &\leqslant C\sum_{k=0}^{\infty}\frac{2^{-k}4\delta(x)}{|B(x,2^{-k}4\delta(x))|}\int_{B(x,2^{-k}4\delta(x))} |Xu(y)|dy \\ &\leqslant C\delta(x)\mathcal{M}_{4\delta(x)}(|Xu|)(x). \end{aligned} \tag{3.15}$$

Also,

$$\begin{aligned} \Big(\frac{\int_{2B}|Xu|^q dx}{|B|\delta(x)^{-q}}\Big)^{\frac{1}{q}} &\leqslant C\delta(x)\Big(\frac{\int_{B(x,4\delta(x))}|Xu|^q dx}{|B(x,4\delta(x))|}\Big)^{\frac{1}{q}} \\ &\leqslant C\delta(x)\Big(\mathcal{M}_{4\delta(x)}(|Xu|^q)(x)\Big)^{\frac{1}{q}}. \end{aligned} \tag{3.16}$$

From (3.14), (3.15), (3.16) and the Hölder inequality we now obtain

$$u(x) \leqslant C\delta(x)\Big(\mathcal{M}_{4\delta(x)}(|Xu|^q)(x)\Big)^{\frac{1}{q}},$$

which completes the proof of the theorem. □

As it turns out, the pointwise Hardy inequality (3.13) is in fact equivalent to certain geometric conditions on the boundary of Ω that can be measured in terms of a Hausdorff content. We introduce the relevant definition.

Definition 3.5. Let $s \in \mathbb{R}$, $r > 0$ and $E \subset \mathbb{R}^n$. The (X, s, r)-*Hausdorff content* of E is the number

$$\widetilde{\mathcal{H}}^s_r(E) = \inf \sum_j r_j^s |B_j|,$$

where the infimum is taken over all coverings of E by balls $B_j = B(x_j, r_j)$ such that $x_j \in E$ and $r_j \leqslant r$.

We next follow the idea in [37] to prove the following important consequence of the pointwise Hardy inequality (3.13).

Theorem 3.6. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with local parameters C_0 and R_0. Suppose that there exist $r_0 \leqslant R_0/100$, $q > 0$, and a constant $C > 0$ such that the inequality*

$$|u(x)| \leqslant C\delta(x)\Big(\mathcal{M}_{4\delta(x)}(|\nabla u|^q)(x)\Big)^{\frac{1}{q}} \tag{3.17}$$

holds for all $x \in \Omega$ with $\delta(x) < r_0$ and all compactly supported $u \in C^{0,1}_d(\Omega)$. There exists $C_1 > 0$ such that the inequality

$$\widetilde{\mathcal{H}}^{-q}_{\delta(x)}(\overline{B}(x, 2\delta(x)) \cap \partial\Omega) \geqslant C_1 \delta(x)^{-q} |B(x, \delta(x))| \tag{3.18}$$

holds for all $x \in \Omega$ with $\delta(x) < r_0$.

Proof. We argue by contradiction and suppose that (3.18) fails. We can thus find a sequence $\{x_k\}_{k=1}^\infty \subset \Omega$ with $\delta(x_k) < r_0$ such that

$$\widetilde{\mathcal{H}}^{-q}_{\delta(x)/4}(\overline{B}(x_k, 5\delta(x_k)) \cap \partial\Omega) < k^{-1}\delta(x_k)^{-q}|B(x_k, \delta(x_k))|.$$

Here, we used the fact that, by the continuity of the distance function δ and the doubling property (2.8), the inequality (3.18), which holds for all $x \in \Omega$ with $\delta(x) < r_0$, is equivalent to the validity of

$$\widetilde{\mathcal{H}}^{-q}_{\delta(x)/4}(\overline{B}(x, 5\delta(x)) \cap \partial\Omega) \geqslant C_2 \delta(x)^{-q}|B(x, \delta(x))|$$

for all $x \in \Omega$ with $\delta(x) < r_0$ and for a constant $C_2 > 0$. By compactness, we can now find a finite covering $\{B_i\}_{i=1}^N$, $B_i = B(z_i, r_i)$ with $z_i \in \overline{B}(x_k, 5\delta(x_k)) \cap \partial\Omega$ and $0 < r_i < \delta(x_k)/4$, such that

$$\overline{B}(x_k, 5\delta(x_k)) \cap \partial\Omega \ \subset \ \bigcup_{i=1}^N B_i \tag{3.19}$$

and

$$\sum_{i=1}^{N} r_i^{-q}|B_i| < k^{-1}\delta(x_k)^{-q}|B(x_k,\delta(x_k))|. \tag{3.20}$$

Next, for each $k \in \mathbf{N}$ we define a function φ_k by

$$\varphi_k(x) = \min\{1, \min_{1\leqslant i\leqslant N} r_i^{-1}\mathrm{dist}(x, 2B_i)\}$$

and let $\phi_k \in C_d^{0,1}(B(x_k, 5\delta(x_k)))$ be such that $0 \leqslant \phi_k \leqslant 1$ and $\phi_k \equiv 1$ on $B(x_k, 4\delta(x_k))$. Clearly, the function $u_k = \phi_k\varphi_k$ belongs to $C_d^{0,1}(\Omega)$ and, in view of (3.19), it has compact support. Moreover, $u_k(x_k) = 1$ since from the fact that $z_i \in \partial\Omega$ we have

$$d(x_k, z_i) \geqslant \delta(x_k) > 4r_i \tag{3.21}$$

for all $1 \leqslant i \leqslant N$. Also, since $\varphi_k(x) = 1$ for $x \notin \bigcup_{i=1}^{N} 3\overline{B}_i$ and $\varphi_k(x) = 0$ for $x \in \bigcup_{i=1}^{N} 2\overline{B}_i$, it is easy to see that

$$\mathrm{supp}\ \ (|Xu_k|) \cap B(x_k, 4\delta(x_k)) \subset \bigcup_{i=1}^{N}(3\overline{B}_i \setminus 2B_i)$$

and that for a.e. $y \in B(x_k, 4\delta(x_k))$ we have

$$|Xu_k(y)|^q \leqslant \sum_{i=1}^{N} r_i^{-q}\chi_{3\overline{B}_i\setminus 2B_i}(y). \tag{3.22}$$

Hence, using (3.21) and (3.22), we can calculate

$$\begin{aligned}
&\mathcal{M}_{4\delta(x_k)}(|Xu_k|^q)(x_k)\\
&\leqslant C \sup_{\frac{1}{4}\delta(x_k)\leqslant r\leqslant 4\delta(x_k)} \frac{1}{|B(x_k,r)|}\int_{B(x_k,r)} |Xu_k(y)|^q dy\\
&\leqslant C\frac{1}{|B(x_k,\delta(x_k))|}\int_{B(x_k,4\delta(x_k))} |Xu_k(y)|^q dy\\
&\leqslant C\frac{1}{|B(x_k,\delta(x_k))|}\sum_{i=1}^{N}|3\overline{B_i} \setminus 2B_i| r_i^{-q}
\end{aligned}$$

$$\leqslant C \frac{1}{|B(x_k, \delta(x_k))|} \sum_{i=1}^{N} |B_i| r_i^{-q}. \tag{3.23}$$

From (3.20) and (3.23) we obtain

$$\delta(x_k)^q \mathcal{M}_{4\delta(x_k)}(|Xu_k|^q)(x_k) \leqslant Ck^{-1}.$$

Since $u_k = 1$ for any k, this implies that the pointwise Hardy inequality (3.17) fails to hold with a uniform constant for all compactly supported $u \in C_d^{0,1}(\Omega)$. This contradiction completes the proof of the theorem. □

As in [37], from (3.18) we can also obtain the following thickness condition on $\mathbb{R}^n \setminus \Omega$.

Theorem 3.7. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with local parameters C_0 and R_0. Suppose that there exist $r_0 \leqslant R_0/100$, $q > 0$, and a constant $C > 0$ such that the inequality*

$$\widetilde{\mathcal{H}}^{-q}_{\delta(x)}(\overline{B}(x, 2\delta(x)) \cap \partial\Omega) \geqslant C\delta(x)^{-q}|B(x, \delta(x))| \tag{3.24}$$

holds for all $x \in \Omega$ with $\delta(x) < r_0$. Then there exists $C_1 > 0$ such that

$$\widetilde{\mathcal{H}}^{-q}_r(B(w, r) \cap (\mathbb{R}^n \setminus \Omega)) \geqslant C_1 r^{-q}|B(w, r)| \tag{3.25}$$

for all $w \in \partial\Omega$ and $0 < r < r_0$.

Proof. Let $w \in \partial\Omega$ and $0 < r < r_0$. If

$$|B(w, \tfrac{r}{2}) \cap (\mathbb{R}^n \setminus \Omega)| \geqslant \tfrac{1}{2}|B(w, \tfrac{r}{2})|,$$

then it is easy to see that (3.25) holds with $C_1 = 2^{-Q}C_0/2$. Thus, we may assume that

$$|B(w, \tfrac{r}{2}) \cap \Omega| \geqslant \tfrac{1}{2}|B(w, \tfrac{r}{2})|,$$

which, by (2.8), gives

$$|B(w, \tfrac{r}{2}) \cap \Omega| \ \geqslant \ 2^{-Q}C_0 \ |B(w, r)|/2. \tag{3.26}$$

Now, to prove (3.25), it is enough to show that

$$\widetilde{\mathcal{H}}^{-q}_r(B(w, r) \cap \partial\Omega) \geqslant C_1 r^{-q}|B(w, r)|. \tag{3.27}$$

To this end, let $\{B_i\}_{i=1}^{\infty}$, $B_i = B(z_i, r_i)$ with $z_i \in \partial\Omega$ and $0 < r_i \leqslant r$ be a covering of $B(w, r) \cap \partial\Omega$. Then if

$$\sum_i |B_i| \geqslant (2^{-Q}C_0)^2|B(w, r)|/4,$$

it follows that (3.27) holds with $C_1 = \frac{1}{4}(2^{-Q}C_0)^2$. Hence we are left with considering only the case

$$\sum_i |B_i| < (2^{-Q}C_0)^2|B(w,r)|/4. \tag{3.28}$$

Using (2.8), (3.26), and (3.28), we can now estimate

$$\begin{aligned}|(B(w,\tfrac{r}{2})\cap\Omega)\setminus\bigcup_i 2B_i| &\geqslant |B(w,\tfrac{r}{2})\cap\Omega| - 2^Q C_0^{-1}\sum_i |B_i| \\ &\geqslant 2^{-Q}C_0|B(w,r)|/2 - 2^{-Q}C_0|B(w,r)|/4 \\ &= 2^{-Q}C_0|B(w,r)|/4.\end{aligned}$$

Thus, by a covering lemma (see [52, p. 9]), we can find a sequence of pairwise disjoint balls $B(x_k, 6\delta(x_k))$ with $x_k \in (B(w,\frac{r}{2})\cap\Omega)\setminus\bigcup_i 2B_i$ such that

$$|B(w,r)| \leqslant C|(B(w,\tfrac{r}{2})\cap\Omega)\setminus\bigcup_i 2B_i| \leqslant C\sum_k |B(x_k, 30\delta(x_k))|.$$

This, together with (2.8) and (3.24), gives

$$\begin{aligned}|B(w,r)|r^{-q} &\leqslant C\sum_k |B(x_k,\delta(x_k))|\delta(x_k)^{-q} \\ &\leqslant C\sum_k \widetilde{\mathcal{H}}^{-q}_{\delta(x_k)}(\overline{B}(x_k, 2\delta(x_k))\cap\partial\Omega)\end{aligned} \tag{3.29}$$

since $\delta(x_k) < \frac{r}{2}$ for all k.

We next observe that we can further assume that

$$\delta(x) < \tfrac{r}{4} \text{ for all } x \in B(w,\tfrac{r}{2})\cap\Omega. \tag{3.30}$$

In fact, if there exits $x \in B(w,\frac{r}{2})\cap\Omega$ such that $\delta(x) \geqslant \frac{r}{4}$, then there exists $x_0 \in B(w,\frac{r}{2})\cap\Omega$ such that $\delta(x_0) = \frac{r}{4}$ by the continuity of δ. Thus, $B(x_0, 2\delta(x_0)) \subset B(w,r)$ and, in view of the assumption (3.24), we obtain

$$\begin{aligned}\widetilde{\mathcal{H}}^{-q}_r(B(w,r)\cap\partial\Omega) &\geqslant C\widetilde{\mathcal{H}}^{-q}_{\delta(x_0)}(B(x_0, 2\delta(x_0)\cap\partial\Omega)) \\ &\geqslant C\delta(x_0)^{-q}|B(x_0,\delta(x_0))| \geqslant Cr^{-q}|B(w,r)|\ ,\end{aligned}$$

which gives (3.27). Now, the inequality (3.30), in particular, implies that

$$\overline{B}(x_k, 2\delta(x_k))\cap\partial\Omega \subset B(w,r)\cap\partial\Omega \subset \bigcup_i B_i,$$

and hence for every k one has

$$\widetilde{\mathcal{H}}^{-q}_{2\delta(x_k)}(\overline{B}(x_k, 2\delta(x_k) \cap \partial\Omega)) \leqslant \sum_{\{i\in\mathbb{N} | B_i \cap \overline{B}(x_k, 2\delta(x_k)) \neq \varnothing\}} |B_i| r_i^{-q}. \tag{3.31}$$

Here, we used the fact that $r_i < 2\delta(x_k)$ since $x_k \notin 2B_i$. From (3.29) and (3.31), after changing the order of summation, we obtain

$$\begin{aligned} |B(w,r)| r^{-q} &\leqslant C \sum_i \sum_{\{k\in\mathbb{N} | B_i \cap \overline{B}(x_k, 2\delta(x_k)) \neq \varnothing\}} |B_i| r_i^{-q} \\ &\leqslant C \sum_i C(i) |B_i| r_i^{-q}, \end{aligned} \tag{3.32}$$

where $C(i)$ is the number of balls $\overline{B}(x_k, 2\delta(x_k))$ that intersect B_i. Note that if $B_i \cap \overline{B}(x_k, 2\delta(x_k)) \neq \varnothing$, then, since $r_i < 2\delta(x_k)$, we see that $B_i \subset B(x_k, 6\delta(x_k))$. Hence $C(i) \leqslant 1$ for all i since, by our choice, the balls $B(x_k, 6\delta(x_k))$ are pairwise disjoint. This and (3.32) give

$$|B(w,r)| r^{-q} \leqslant C \sum_i |B_i| r_i^{-q},$$

and the inequality (3.27) follows as the coverings $\{B_i\}_i$ of $B(w,r) \cap \partial\Omega$ are arbitrary. This completes the proof of the theorem. □

The thickness condition (3.25) that involves the Hausdorff content will now be shown to imply the uniform (X,p)-fatness of $\mathbb{R}^n \setminus \Omega$. To achieve this we borrow an idea from [29].

Theorem 3.8. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with local parameters C_0 and R_0. Suppose that there exist $r_0 \leqslant R_0/100$, $1 < q < p$, and a constant $C > 0$ such that the inequality*

$$\widetilde{\mathcal{H}}^{-q}_r(B(w,r) \cap (\mathbb{R}^n \setminus \Omega)) \geqslant C r^{-q} |B(w,r)| \tag{3.33}$$

holds for all $w \in \partial\Omega$ and $0 < r < r_0$. Then there exists $C_1 > 0$ such that the $\mathbb{R}^n \setminus \Omega$ is uniformly (X,p)-fat with constants C_1 and r_0.

Proof. Let $z \in \partial\Omega$, and let $0 < r < r_0$. We need to find a constant $C_1 > 0$ independent of z and r such that

$$\operatorname{cap}_p(K, B(z,2r)) \geqslant C_1 r^{-p} |B(z,r)|, \tag{3.34}$$

where $K = (\mathbb{R}^n \setminus \Omega) \cap \overline{B}(z,r)$. From (3.33) we have

$$\widetilde{\mathcal{H}}^{-q}_r(K) \geqslant C r^{-q} |B(z,r)|. \tag{3.35}$$

Let $\varphi \in C_0^\infty(B(z,2r))$ be such that $\varphi \geqslant 1$ on K. If there is $x_0 \in K$ such that

$$|\varphi(x_0) - \varphi_{B(x_0,4r)}| \leqslant 1/2,$$

then

$$1 \leqslant \varphi(x_0) \leqslant |\varphi(x_0) - \varphi_{B(x_0,4r)}| + |\varphi_{B(x_0,4r)}| \leqslant 1/2 + |\varphi_{B(x_0,4r)}|.$$

By Lemma 3.1, the doubling property (2.8), and (2.12), we obtain

$$1/2 \leqslant |\varphi_{B(x_0,4r)}| \leqslant C\Big(r^p |B(z,r)|^{-1} \int\limits_{B(z,2r)} |X\varphi|^p dx\Big)^{\frac{1}{p}},$$

which gives (3.34). Thus, we may assume that

$$1/2 < |\varphi(x) - \varphi_{B(x,4r)}| \quad \text{for all } x \in K.$$

Under such an assumption, using the covering argument in Theorem 5.9 in [29], the inequality (3.34) follows from (3.35) and Theorem 2.1. □

Finally, we summarize in one single theorem the results obtained in Theorems 3.4, 3.6, 3.7, and 3.8.

Theorem 3.9. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with local parameters C_0 and R_0 and let $1 < p < \infty$. There exists $0 < r_0 \leqslant R_0/100$ such that the following statements are equivalent:*

(i) *The set $\mathbb{R}^n \setminus \Omega$ is uniformly (X,p)-fat with constants c_0 and r_0 for some $c_0 > 0$, i.e.,*

$$\operatorname{cap}_p((\mathbb{R}^n \setminus \Omega) \cap \overline{B}(w,r), B(w,2r)) \geqslant c_0 r^{-p} |B(w,r)|$$

for all $w \in \partial\Omega$ and $0 < r < r_0$.

(ii) *There exists $1 < q < p$ and a constant $C > 0$ such that*

$$|u(x)| \leqslant C\delta(x)\Big(\mathcal{M}_{4\delta(x)}(|\nabla u|^q)(x)\Big)^{\frac{1}{q}}$$

for all $x \in \Omega$ with $\delta(x) < r_0$ and all compactly supported $u \in C^{0,1}_d(\Omega)$.

(iii) *There exists $1 < q < p$ and a constant $C > 0$ such that*

$$\widetilde{\mathcal{H}}^{-q}_{\delta(x)}(\overline{B}(x, 2\delta(x)) \cap \partial\Omega) \geqslant C\delta(x)^{-q} |B(x,\delta(x))|$$

for all $x \in \Omega$ with $\delta(x) < r_0$.

(iv) *There exists $1 < q < p$ and a constant $C > 0$ such that*

$$\widetilde{\mathcal{H}}^{-q}_{r}(B(w,r) \cap (\mathbb{R}^n \setminus \Omega)) \geqslant Cr^{-q} |B(w,r)|$$

for all $w \in \partial\Omega$ and $0 < r < r_0$.

Remark 3.10. As an example in [37] shows, we cannot replace the set $\mathbb{R}^n \setminus \Omega$ in statement (iv) in Theorem 3.9 with the smaller set $\partial\Omega$.

4 Hardy Inequalities on Bounded Domains

Our first result in this section is the following Hardy inequality which is a consequence of Theorem 3.4 and the L^s boundedness of the Hardy–Littlewood maximal function for $s > 1$. We remark that no assumption on the smallness of the diameter of the domain is required, as opposed to the Poicaré inequality (2.11) and Sobolev inequalities established in [23].

Theorem 4.1. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with local parameters C_0 and R_0. Suppose that $\mathbb{R}^n \setminus \Omega$ is uniformly (X,p)-fat with constants $c_0 > 0$ and $0 < r_0 \leqslant R_0/100$. There is a constant $C > 0$ such that for all $\varphi \in C_0^\infty(\Omega)$*

$$\int_\Omega \frac{|\varphi(x)|^p}{\delta(x)^p}\, dx \;\leqslant\; C \int_\Omega |X\varphi|^p\, dx. \tag{4.1}$$

Proof. Let $\Omega_{r_0} = \{x \in \Omega : \delta(x) \geqslant r_0\}$, and let $\varphi \in C_0^\infty(\Omega)$. By Theorem 3.4, we can find $1 < q < p$ such that

$$\begin{aligned}
\int_\Omega |\varphi(x)|^p \delta(x)^{-p} dx &= \int_{\Omega_{r_0}} |\varphi(x)|^p \delta(x)^{-p} dx + \int_{\Omega\setminus\Omega_{r_0}} |\varphi(x)|^p \delta(x)^{-p} dx \\
&\leqslant r_0^{-p} \int_\Omega |\varphi(x)|^p dx + C \int_\Omega \Big(\mathcal{M}_{4r_0}(|X\varphi|^q)(x)\Big)^{\frac{p}{q}} dx \\
&\leqslant C \int_\Omega |X\varphi(x)|^p dx.
\end{aligned}$$

In the last inequality above, we used the Poincaré inequality (2.10) and the boundedness property of $\mathcal{M}_{4r_0}$ on $L^s(\Omega)$, $s > 1$ (see [52]). The proof of Theorem 4.1 is then complete. □

To state Theorems 4.3 and 4.5 below, we need to fix a Whitney decomposition of Ω into balls as in the following lemma, whose construction can be found, for example, in [33] or [18].

Lemma 4.2. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with local parameters C_0 and R_0. There exists a family of balls $\mathcal{W} = \{B_j\}$ with $B_j = B(x_j, r_j)$ and a constant $M > 0$ such that*

(a) $\Omega \subset \cup_j B_j$,
(b) $B(x_j, \frac{r_j}{4}) \cap B(x_k, \frac{r_k}{4}) \neq \varnothing$ for $j \neq k$,
(c) $r_j = 10^{-3} \min\{R_0/\operatorname{diam}(\Omega), 1\} \operatorname{dist}(B_j, \partial\Omega)$,
(d) $\sum_j \chi_{4B_j}(x) \leqslant M\chi_\Omega(x)$.

In (c),

$$\operatorname{diam}(\Omega) = \sup_{x,y\in\Omega} d(x,y)$$

is the diameter of Ω *with respect to the* CC *metric. In particular, we have* $r_j \leqslant 10^{-3}R_0$.

We can now go further in characterizing weight functions V on Ω for which the embedding

$$\int_\Omega |\varphi(x)|^p V(x)dx \leqslant C \int_\Omega |X\varphi|^p dx$$

holds for all $\varphi \in C_0^\infty(\Omega)$. Here, the condition on V is formulated in terms of a localized capacitary condition adapted to a Whitney decomposition of Ω. Such a condition can be simplified further in the setting of Carnot groups as we point out in Remark 4.4 below. In the Euclidean setting, it was used in [26] to characterize the solvability of multi-dimensional Riccati equations on bounded domains.

Theorem 4.3. *Let* $\Omega \subset \mathbb{R}^n$ *be a bounded domain with local parameters* C_0 *and* R_0. *Let* $V \geqslant 0$ *be in* $L^1_{\mathrm{loc}}(\Omega)$. *Suppose that* $\mathbb{R}^n \setminus \Omega$ *is uniformly* (X,p)*-fat with* $1 < p < Q$. *Then the embedding*

$$\int_\Omega |\varphi(x)|^p V(x)dx \leqslant C \int_\Omega |X\varphi|^p dx, \qquad \varphi \in C_0^\infty(\Omega), \tag{4.2}$$

holds if and only if

$$\sup_{B\in\mathcal{W}} \sup_{\substack{K\subset 2B \\ K \text{ compact}}} \frac{\int_K V(x)dx}{{}' \bigcap limits_p(K,\Omega)} \leqslant C, \tag{4.3}$$

where $\mathcal{W} = \{B_j\}$ *is a Whitney decomposition of* Ω *as in Lemma* 4.2.

Remark 4.4. In the setting of a Carnot group $\mathbf{G}$ with homogeneous dimension Q, we can replace $\operatorname{cap}_p(K,\Omega)$ by $\operatorname{cap}_p(K,\mathbf{G})$ in (4.3) since, if $B \in \mathcal{W}$ and K is a compact set in $2B$, we have

$$c \operatorname{cap}_p(K,\Omega) \leqslant \operatorname{cap}_p(K,\mathbf{G}) \leqslant \operatorname{cap}_p(K,\Omega). \tag{4.4}$$

The second inequality in (4.4) is obvious. To see the first one, let $\varphi \in C_0^\infty(\mathbf{G})$, $\varphi \geqslant 1$ on K, and choose a cut-off function $\eta \in C_0^\infty(4B)$ such that $0 \leqslant \eta \leqslant 1$, $\eta \equiv 1$ on $2B$ and $|X\eta| \leqslant \frac{C}{r_B}$, where r_B is the radius of B. Since $\varphi\eta \in C_0^\infty(\Omega)$, $\varphi\eta \geqslant 1$ on K, we have

$$\begin{aligned}
\operatorname{cap}_p(K,\Omega) &\leqslant \int_\Omega |X(\varphi\eta)|^p dg \\
&\leqslant \int_{\mathbf{G}} |X\varphi|^p dg + C \int_{4B\setminus 2B} \frac{|\varphi|^p}{r_B^p} dg \\
&\leqslant \int_{\mathbf{G}} |X\varphi|^p dg + C \int_{\mathbf{G}} \frac{|\varphi|^p}{\rho(g,g_0)^p} dg,
\end{aligned}$$

where g_0 is the center of B, and we denoted by $\rho(g,g_0)$ the pseudo-distance induced on $\mathbf{G}$ by the anisotropic Folland–Stein gauge (see [18, 20]). To bound the third integral on the right-hand side of the latter inequality, we use the following Hardy type inequality:

$$\int_{\mathbf{G}} \frac{\varphi^p}{\rho(g,g_0)^p}\, dg \;\leqslant\; C \int_{\mathbf{G}} |X\varphi|^p\, dg, \qquad \varphi \in C_0^\infty(\mathbf{G}), \tag{4.5}$$

which is easily proved as follows. Recall the Folland-Stein Sobolev embedding (see [20])

$$\left(\int_{\mathbf{G}} |\varphi|^{\frac{pQ}{Q-p}}\, dg\right)^{\frac{Q-p}{pQ}} \leqslant S_p \left(\int_{\mathbf{G}} |X\varphi|^p\, dg\right)^{\frac{1}{p}}, \quad \varphi \in C_0^\infty(\mathbf{G}). \tag{4.6}$$

Observing that for every $g_0 \in \mathbf{G}$ one has $g \to \dfrac{1}{\rho(g,g_0)^p} \in L^{Q/p,\infty}(\mathbf{G})$, from the generalized Hölder inequality for weak L^p spaces due to Hunt [32] one obtains with an absolute constant $B > 0$

$$\begin{aligned}
\int_{\mathbf{G}} \frac{\varphi^p}{\rho(g,g_0)^p}\, dg \;&\leqslant\; B \left(\int_{\mathbf{G}} |\varphi|^{\frac{pQ}{Q-p}}\, dg\right)^{\frac{Q-p}{Q}} ||\rho(\cdot,g_0)^{-p}||_{L^{Q/p,\infty}(\mathbf{G})} \\
&\leqslant\; C \int_{\mathbf{G}} |X\varphi|^p\, dg,
\end{aligned}$$

where in the last inequality we used (4.6). This proves (4.5). In conclusion, we find

$$\operatorname{cap}_p(K,\Omega) \leqslant C \int_{\mathbf{G}} |X\varphi|^p dg,$$

which gives the first inequality in (4.4).

Proof of Theorem 4.3. That the emdedding (4.2) implies the capacitary condition (4.3) is clear. To prove the converse, let $\{\varphi_j\}$ be a Lipschitz partition of unity associated with the Whitney decomposition $\mathcal{W} = \{B_j\}$ (see [24]), i.e., $0 \leqslant \varphi_j \leqslant 1$ is Lipschitz with respect to the CC metric, supp $\varphi_j \Subset 2B_j$, $|X\varphi_j| \leqslant C/\mathrm{diam}(B_j)$, and

$$\sum_j \varphi_j(x) = \chi_\Omega(x).$$

Moreover, by property (d) in Lemma 4.2, there is a constant $C(p)$ such that

$$\left(\sum_j \varphi_j(x)\right)^p = C(p)\sum_j \varphi_j(x)^p.$$

Then for any $\varphi \in C_0^\infty(\Omega)$ we have

$$\begin{aligned}\int_\Omega |\varphi(x)|^p\, V(x)dx &\leqslant C\, \sum_j \int_\Omega |\varphi_j\varphi(x)|^p\, V(x)dx \\ &\leqslant C\, \sum_j \int_{4B_j} |X(\varphi_j\varphi)|^p dx\end{aligned}$$

by (4.3) and [13, Theorem 5.3]. Thus, from Theorem 4.1 and Lemma 4.2 we obtain

$$\begin{aligned}&\int_\Omega |\varphi(x)|^p\, V(x)dx \\ &\leqslant C\, \sum_j \int_{4B_j} |X\varphi|^p dx + C\sum_j [\mathrm{diam}(B_j)]^{-p} \int_{4B_j} |\varphi|^p dx \\ &\leqslant C\int_\Omega |X\varphi|^p dx + C\int_\Omega |\varphi|^p \delta^{-p}(x)dx \\ &\leqslant C\int_\Omega |X\varphi|^p dx.\end{aligned}$$

This completes the proof of the theorem. □

In view of [13, Theorem 1.6], the above proof also gives the following Fefferman–Phong type sufficiency result [17].

Theorem 4.5. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with local parameters C_0 and R_0. Let $V \geqslant 0$ be in $L^1_{\mathrm{loc}}(\Omega)$. Suppose that $\mathbb{R}^n \setminus \Omega$ is uniformly (X,p)-fat* *with* $1 < p < Q$*. Then the embedding*

$$\int_{\Omega} |\varphi(x)|^p V(x)dx \leqslant C \int_{\Omega} |X\varphi|^p dx, \qquad \varphi \in C_0^\infty(\Omega), \tag{4.7}$$

holds if for some $s > 1$*,* V *satisfies the following localized Fefferman–Phong type condition adapted to* Ω*:*

$$\sup_{B\in\mathcal{W}} \sup_{\substack{x\in 2B \\ 0<r<\operatorname{diam}(B)}} \int_{B(x,r)} V(y)^s dy \;\leqslant\; C\,\frac{|B(x,r)|}{r^{sp}} \tag{4.8}$$

where $\mathcal{W} = \{B_j\}$ *is a Whitney decomposition of* Ω *as in Lemma* 4.2*.*

Let $L^{s,\infty}(\Omega)$, $0 < s < \infty$, denote the weak L^s space on Ω, i.e.,

$$L^{s,\infty}(\Omega) = \Big\{ f : \|f\|_{L^{s,\infty}(\Omega)} < \infty \Big\},$$

where

$$\|f\|_{L^{s,\infty}(\Omega)} \;=\; \sup_{t>0}\; t\,|\{x\in\Omega : |f(x)| > t\}|^{\frac{1}{s}}.$$

Equivalently, one can take

$$\|f\|_{L^{s,\infty}(\Omega)} = \sup_{E\subset\Omega:\,|E|>0} |E|^{\frac{1}{s}-\frac{1}{r}} \left(\int_E |f|^r dx \right)^{\frac{1}{r}}$$

for any $0 < r < s$. For $s = \infty$ we define

$$L^{\infty,\infty}(\Omega) = L^\infty(\Omega).$$

From Theorem 4.8 we obtain the following corollary, which improves a similar result in [16, Remark 3.7] in the sense that not only does it cover the subelliptic case, but also require a milder assumption on the boundary.

Corollary 4.6. *Let* $\Omega \subset \mathbb{R}^n$ *be a bounded domain with local parameters* C_0 *and* R_0*. Suppose that* $\mathbb{R}^n \setminus \Omega$ *is uniformly* (X, p)*-fat for* $1 < p < Q$*, where* Q *is the homogeneous dimension of* Ω*. If* $v \in L^{\frac{Q}{\gamma},\infty}(\Omega)$ *for some* $0 \leqslant \gamma \leqslant p$*, then the embedding* (4.7) *holds for the weight* $V(x) = \delta(x)^{-p+\gamma} v(x)$*.*

Proof. Let $\mathcal{W} = \{B_j\}$ is a Whitney decompositon of Ω as in Lemma 4.2. For $x \in 2B$, $B \in \mathcal{W}$, $0 < r < \operatorname{diam}(B)$, and $1 < s < \frac{Q}{\gamma}$ we have

$$\int_{B(x,r)} V(y)^s dy \leqslant Cr^{-sp+s\gamma} \int_{B(x,r)} v(y)^s dy.$$

It is then easily seen from the Hölder inequality and the doubling property (2.8) that

$$\int\limits_{B(x,r)} V(y)^s dy \leqslant Cr^{-sp}|B(x,r)|\,\|v\|^s_{L^{\frac{Q}{\gamma},\infty}(\Omega)} \left(\frac{r}{|B(x,r)|^{\frac{1}{Q}}}\right)^{s\gamma}$$
$$\leqslant Cr^{-sp}|B(x,r)|\,\|v\|_{L^{\frac{Q}{\gamma},\infty}(\Omega)}.$$

By Theorem 4.5, we obtain the corollary. □

The results obtained in Corollary 4.6 do not in general cover the case in which $v(x)$ has a point singularity in Ω, such as $V(x) = \delta(x)^{-p+\gamma}d(x,x_0)^{-\gamma}$, with $0 \leqslant \gamma \leqslant p$ and $1 < p < Q(x_0)$ for some $x_0 \in \Omega$, where $Q(x_0)$ is the homogeneous dimension at x_0. The reason is that it may happen that $Q(x_0) < Q$ and hence $d(\cdot, x_0)^{-\gamma} \notin L^{\frac{Q}{\gamma},\infty}(\Omega)$. However, by the upper estimate in (2.5), we still can obtain the inequality (4.7) for such weights as follows.

Corollary 4.7. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with local parameters C_0 and R_0. Given $x_0 \in \Omega$, suppose that $\mathbb{R}^n \setminus \Omega$ is uniformly (X,p)-fat for $1 < p < Q(x_0)$. Then for any $0 \leqslant \gamma \leqslant p$ the embedding* (4.7) *holds for the weight*

$$V(x) = \delta(x)^{-p+\gamma}d(x,x_0)^{-\gamma}.$$

Proof. Let $\mathcal{W} = \{B_j\}$ be a Whitney decomposition of Ω as in Lemma 4.2. For $x \in 2B$, $B \in \mathcal{W}$, $0 < r < \operatorname{diam}(B)$, and $1 < s < \frac{Q(x_0)}{\gamma}$ we have

$$\int\limits_{B(x,r)} V(y)^s dy \leqslant C\, r^{-sp+s\gamma} \int\limits_{B(x,r)} d(y,x_0)^{-\gamma s} dy. \tag{4.9}$$

Thus, if $x \notin B(x_0, 2r)$, then

$$\int\limits_{B(x,r)} V(y)^s dy \leqslant C\, \frac{|B(x,r)|}{r^{sp}}$$

since for such x we have $d(y,x_0) \geqslant r$ for every $y \in B(x,r)$. On the other hand, if $x \in B(x_0,2r)$ then from (4.9) we find

$$\int\limits_{B(x,r)} V(y)^s dy \leqslant C\, r^{s\gamma-sp} \int\limits_{B(x_0,3r)} d(y,x_0)^{-\gamma s} dy$$

$$= C\, r^{s\gamma-sp} \sum_{k=0}^{\infty} \int\limits_{\frac{3r}{2^{k+1}} \leqslant d(y,x_0) < \frac{3r}{2^k}} d(y,x_0)^{-\gamma s} dy$$

$$\leqslant C\ r^{s\gamma-sp}\sum_{k=0}^{\infty}\left(\frac{r}{2^k}\right)^{-\gamma s}\left|B(x_0,\tfrac{3r}{2^k})\right|.$$

Thus, in view of (2.5) and the doubling property (2.8), we obtain

$$\begin{aligned}\int\limits_{B(x,r)} V(y)^s dy &\leqslant C\ r^{s\gamma-sp}\sum_{k=0}^{\infty}\left(\frac{r}{2^k}\right)^{-\gamma s}\left(\frac{1}{2^k}\right)^{Q(x_0)}|B(x_0,3r)|\\ &\leqslant C\ \frac{|B(x_0,3r)|}{r^{sp}}\sum_{k=0}^{\infty}\left(\frac{1}{2^k}\right)^{Q(x_0)-\gamma s}\\ &\leqslant C(x_0)\ \frac{|B(x,r)|}{r^{sp}}.\end{aligned}$$

Thus, by Theorem 4.5, we obtain the corollary. □

Remark 4.8. If we have $\gamma = p$ in Corollary 4.7, then we do not need to assume $\mathbb{R}^n \setminus \Omega$ to be uniformly (X,p)-fat. In fact, to obtain the embedding (4.7) in this case, we use [13, Theorem 1.6], the Poincaré inequality (2.10), and a finite partition of unity for Ω.

5 Hardy Inequalities with Sharp Constants

In this section, we collect, without proofs, for illustrative purposes some theorems from the forthcoming article [15]. The relevant results pertain certain Hardy–Sobolev inequalities on bounded and unbounded domains with a point singularity which are included in them.

We begin by recalling that when $X = \{X_1, \dots, X_m\}$ constitutes an orthonormal basis of bracket generating vector fields in a Carnot group $\mathbf{G}$, then a fundamental solution Γ_p for $-\mathcal{L}_p$ in all of $\mathbf{G}$ was constructed in [14]. For any bounded open set $\Omega \subset \mathbb{R}^n$ one can construct a positive fundamental solution with generalized zero boundary values, i.e., a Green function, in the more general situation of a Carnot–Carathéodory space. Henceforth, for a fixed $x \in \Omega$ we denote by $\Gamma_p(x,\cdot)$ such a fundamental solution with singularity at some fixed $x \in \Omega$. This means that $\Gamma_p(x,\cdot)$ satisfies the equation

$$\int\limits_{\Omega} |X\Gamma_p(x,y)|^{p-2} < X\Gamma_p(x,y), X\varphi(y) > dy \ = \ \varphi(x) \tag{5.1}$$

for every $\varphi \in C_0^\infty(\Omega)$.

We recall the following fundamental estimate, which is Theorem 7.2 in [5]. Let $K \subset \Omega \subset \mathbb{R}^n$ be a compact set with local parameters C_0 and R_0. Given $x \in K$, and $1 < p < Q(x)$, there exists a positive constant C depending on

C_0 and p such that for any $0 < r \leqslant R_0/2$, and $y \in B(x,r)$ one has

$$C \left(\frac{d(x,y)^p}{\Lambda(x,d(x,y))}\right)^{\frac{1}{p-1}} \leqslant \Gamma_p(x,y) \leqslant C^{-1} \left(\frac{d(x,y)^p}{\Lambda(x,d(x,y))}\right)^{\frac{1}{p-1}} . \tag{5.2}$$

The estimate (5.2) generalizes that obtained by Nagel, Stein, and Wainger [44] and independently by Sanchez-Calle [48] in the case $p = 2$.

For any given $x \in K$ we fix a number $p = p(x)$ such that $1 < p < Q(x)$ and introduce the function

$$E(x,r) \stackrel{def}{=} \left(\frac{\Lambda(x,r)}{r^p}\right)^{\frac{1}{p-1}} . \tag{5.3}$$

Because of the constraint imposed on $p = p(x)$, we see that for every fixed $x \in K$ the function $r \to E(x,r)$ is strictly increasing, and thereby invertible. We denote by $F(x,\cdot) = E(x,\cdot)^{-1}$, the inverse function of $E(x,\cdot)$, so that

$$F(x,E(x,r)) = E(x,F(x,r)) = r.$$

We now define for every $x \in K$

$$\rho_x(y) = F\left(x, \frac{1}{\Gamma(x,y)}\right). \tag{5.4}$$

We emphasize that, in a Carnot group $\mathbf{G}$, one has for every $x \in \mathbf{G}$, $Q(x) \equiv Q$ the homogeneous dimension of the group, and therefore the Nagel–Stein–Wainger polynomial is, in fact, just a monomial, i.e., $\Lambda(x,r) \equiv C(\mathbf{G}) r^Q$. It follows that there exists a constant $\omega(\mathbf{G}) > 0$ such that

$$E(x,r) \equiv \omega(\mathbf{G})\, r^{(Q-p)/(p-1)}. \tag{5.5}$$

Using the function $E(x,r)$ in (5.3), it should be clear that we can recast the estimate (5.2) in the following more suggestive form:

$$\frac{C}{E(x,d(x,y))} \leqslant \Gamma_p(x,y) \leqslant \frac{C^{-1}}{E(x,d(x,y))}. \tag{5.6}$$

As a consequence of (5.6) and (5.4), we obtain the following estimate: *there exist positive constants C and R_0 depending on $X_1,\dots,X_m$ and K such that for every $x \in K$ and every $0 < r \leqslant R_0$ one has for $y \in B(x,r)$*

$$C\, d(x,y) \leqslant \rho_x(y) \leqslant C^{-1}\, d(x,y). \tag{5.7}$$

We can thus think of the function ρ_x as a *regularized pseudo-distance* adapted to the nonlinear operator $\mathcal{L}_p$. We denote by

$$B_X(x,r) = \{y \in \mathbb{R}^n \mid \rho_x(y) < r\},$$

the ball centered at x with radius r with respect to the pseudo-distance ρ_x. Because of (5.7), it is clear that

$$B(x, Cr) \subset B_X(x, r) \subset B(x, C^{-1}r).$$

Our main assumption is that for any $p > 1$ the fundamental solution of the operator $\mathcal{L}_p$ satisfy the following

Hypothesis. *For any compact set $K \subset \Omega \subset \mathbb{R}^n$ there exist $C > 0$ and $R_0 > 0$ depending on K and $X_1, \dots, X_m$ such that for every $x \in \Omega$, $0 < R < R_0$ for which $B_X(x, 4R) \subset \Omega$, and a.e. $y \in B(x, R) \setminus \{x\}$ one has*

$$|X\Gamma_p(x, y)| \leqslant C^{-1} \left(\frac{d(x, y)}{\Lambda(x, d(x, y))}\right)^{\frac{1}{p-1}}. \tag{5.8}$$

We mention explicitly that, as a consequence of the results in [44] and [48], the assumption (5.8) is fulfilled when $p = 2$. For $p \neq 2$ it is also satisfied in any Carnot group of Heisenberg type $\mathbf{G}$. This follows from the results in [5], where for every $1 < p < \infty$ the following explicit fundamental solution of $-\mathcal{L}_p$ was found:

$$-\Gamma_p(g) = \begin{cases} \frac{p-1}{Q-p}\sigma_p^{-\frac{1}{p-1}} N(g)^{-\frac{Q-p}{p-1}}, & p \neq Q, \\ \sigma_Q^{-\frac{1}{Q-1}} \log N(g), & p = Q, \end{cases} \tag{5.9}$$

where we denoted by $N(g) = (|x(g)|^4 + 16|y(g)|^2)^{\frac{1}{4}}$ the Kaplan gauge on $\mathbf{G}$ (see [34]), and we set $\sigma_p = Q\omega_p$ with

$$\omega_p = \int\limits_{\{g \in \mathbf{G} | N(g) < 1\}} |XN(g)|^p \, dg.$$

We note that the case $p = 2$ of (5.9) was first discovered by Folland [19] for the Heisenberg group and subsequently generalized by Kaplan [34] to groups of Heisenberg type. The conformal case $p = Q$ was also found in [28].

We stress that the hypothesis (5.8) is not the weakest one that could be made, and that to the expenses of additional technicalities, we could have chosen substantially weaker hypothesis.

We now recall the classical one-dimensional Hardy inequality [27]: let $1 < p < \infty$, $u(t) \geqslant 0$, and $\varphi(t) = \int\limits_0^t u(s) ds$. Then

$$\int\limits_0^\infty \left(\frac{\varphi(t)}{t}\right)^p dt \leqslant \left(\frac{p}{p-1}\right)^p \int\limits_0^\infty \varphi'(t)^p dt.$$

Here is our main result.

Theorem 5.1. *Given a compact set $K \subset \Omega \subset \mathbb{R}^n$, let $x \in K$ and $1 < p < Q(x)$. For any $0 < R < R_0$ such that $B_X(x, 4R) \subset \Omega$ one has for $\varphi \in S_0^{1,p}(B_X(x,R))$*

$$\int_{B_X(x,R)} |\varphi|^p \left\{ \frac{E'(x,\rho_x)}{E(x,\rho_x)} \right\}^p |X\rho_x|^p \, dy \leqslant \left(\frac{p}{p-1} \right)^p \int_{B_X(x,R)} |X\varphi|^p \, dy.$$

When $\Lambda(x,r)$ is a monomial (thus, for example, in the case of a Carnot group) the constant on the right-hand side of the above inequality is best possible.

We do not present here the proof of Theorem 5.1, but refer the reader to the forthcoming article [15]. Some comments are in order. First of all, concerning the factor $|X\rho_x|^p$ on the left-hand side of the inequality in Theorem 5.1, we emphasize that the hypothesis (5.8) implies that $X\rho_x \in L^\infty_{\mathrm{loc}}$. Secondly, as is shown in [15], one has

$$\left(\frac{Q(x)-p}{p-1} \right)^p \frac{1}{\rho_x^p} \leqslant \left\{ \frac{E'(x,\rho_x)}{E(x,\rho_x)} \right\}^p \leqslant \left(\frac{Q-p}{p-1} \right)^p \frac{1}{\rho_x^p}. \tag{5.10}$$

As a consequence of Theorem 5.1 and (5.10) we thus obtain the following

Corollary 5.2. *Under the same assumptions of Theorem 5.1, one has for $\varphi \in S_0^{1,p}(B_X(x,R))$*

$$\int_{B_X(x,R)} \frac{|\varphi|^p}{\rho_x^p} |X\rho_x|^p \, dy \leqslant \left(\frac{p}{Q(x)-p} \right)^p \int_{B_X(x,R)} |X\varphi|^p \, dy. \tag{5.11}$$

Thirdly, it is worth observing that, with the optimal constants, neither Theorem 5.1 nor Corollary 5.2 can be obtained from Corollary 4.7.

We mention in closing that for the Heisenberg group $\mathbf{H}^n$ with $p = 2$ Corollary 5.2 was first proved in [22]. The inequality (5.11) was extended to the nonlinear case $p \neq 2$ in [45]. For Carnot groups of Heisenberg type and also for some operators of Baouendi–Grushin type the inequality (5.11) was obtained in [11]. In the case $p = 2$, various weighted Hardy inequalities with optimal constants in groups of Heisenberg type were also independently established in [36]. An interesting generalization of the results in [45], along with an extension to nilpotent Lie groups with polynomial growth, was recently obtained in [39]. In this latter setting, an interesting form of the uncertainty principle connected to the case $p = 2$ of the Hardy type inequality (5.11) was established in [10]. These latter two references are not concerned however with the problem of finding the sharp constants.

Acknowledgement. The first author was supported in part by NSF CAREER Grant, DMS-0239771. The second author was supported in part by NSF Grant DMS-0701001.

References

1. Ancona, A.: On strong barriers and an inequality of Hardy for domains in $\mathbb{R}^n$. J. London Math. Soc. **34**, 274–290 (1986)
2. Biroli, M.: Schrödinger type and relaxed Dirichlet problems for the subelliptic p-Laplacian. Potential Anal. **15**, 1–16 (2001)
3. Björn, J., MacManus, P., Shanmugalingam, N.: Fat sets and pointwise boundary estimates for p-harmonic functions in metric spaces. J. Anal. Math. **85**, 339–369 (2001)
4. Capogna, L., Danielli, D., Garofalo, N.: An embedding theorem and the Harnack inequality for nonlinear subelliptic equations. Commun. Partial Differ. Equ. **18**, no. 9-10, 1765–1794 (1993)
5. Capogna, L., Danielli, D., Garofalo, N.: Capacitary estimates and the local behavior of solutions of nonlinear subelliptic equations. Am. J. Math. **118**, no.6, 1153–1196 (1996)
6. Capogna, L., Danielli, D., Garofalo, N.: Subelliptic mollifiers and a basic pointwise estimate of Poincaré type. Math. Z. **226**, 147–154 (1997)
7. Capogna, L., Garofalo, N.: Boundary behavior of nonnegative solutions of subelliptic equations in NTA domains for Carnot–Carathéodory metrics. J. Fourier Anal. Appl. **4**, no. 4-5, 403–432 (1998)
8. Caffarelli, L., Kohn, R., Nirenberg, L.: First order interpolation inequalities with weights. Compos. Math. **53**, 259–275 (1984)
9. Chow, W.L.: *Über* systeme von linearen partiellen Differentialgleichungen erster Ordnung. Math. Ann. **117**, 98–105 (1939)
10. Ciatti, P., Ricci, F., Sundari, M.: Heisenberg–Pauli–Weyl uncertainty inequalities and polynomial volume growth. Adv. Math. **215**, 616–625 (2007)
11. D'Ambrosio, L.: Hardy-type inequalities related to degenerate elliptic differential operators. Ann. Sc. Norm. Super. Pisa, Cl. Sci. (5) **IV**, 451–486 (2005)
12. Danielli, D.: Regularity at the boundary for solutions of nonlinear subelliptic equations. Indiana Univ. Math. J. **44**, 269–285 (1995)
13. Danielli, D.: A Fefferman–Phong type inequality and applications to quasilinear subelliptic equations. Potential Anal. **11**, 387–413 (1999)
14. Danielli, D., Garofalo, N.: Green Functions in Nonlinear Potential Theory in Carnot Groups and the Geometry of their Level Sets. Preprint (2003)
15. Danielli, D., Garofalo, N., Phuc, N.C.: Hardy–Sobolev Inequalities with Sharp Constants in Carnot–Carathéodory Spaces. Preprint (2008)
16. Duc, D.M., Phuc, N.C., Nguyen, T.V.: Weighted Sobolev's inequalities for bounded domains and singular elliptic equations. Indiana Univ. Math. J. **56**, 615–642 (2007)
17. Fefferman, C.: The uncertainty principle. Bull. Am. Math. Soc. **9**, 129–206 (1983)
18. Folland, G.B., Stein,E.M.: Hardy Spaces on Homogeneous Groups. Princeton Univ. Press, Princeton, NJ (1982)
19. Folland, G.B.: A fundamental solution for a subelliptic operator. Bull. Am. Math. Soc. **79**, 373–376 (1973)
20. Folland, G.B.: Subelliptic estimates and function spaces on nilpotent Lie groups. Ark. Mat. **13**, 161–207 (1975)

21. Franchi, B., Serapioni, R., Serra Cassano, F.: Approximation and imbedding theorems for weighted Sobolev spaces associated with Lipschitz continuous vector fields. Boll. Unione Mat. Ital., VII, Ser. B **11**, no. 1, 83–117 (1997)
22. Garofalo, N., Lanconelli, E.: Frequency functions on the Heisenberg group, the uncertainty principle and unique continuation. Ann. Inst. Fourier (Grenoble) **40**, no. 2, 313–356 (1990)
23. Garofalo, N., Nhieu, D.M.: Isoperimetric and Sobolev inequalities for Carnot–Carathéodory spaces and the existence of minimal surfaces. Commun. Pure Appl. Math. **49**, 1081–1144 (1996)
24. Garofalo, N., Nhieu, D.M.: Lipschitz continuity, global smooth approximations and extension theorems for Sobolev functions in Carnot–Carathéodory spaces. J. Anal. Math. **74**, 67–97 (1998)
25. Hajłasz, P.: Pointwise Hardy inequalities, Proc. Am. Math. Soc. **127**, 417–423 (1999)
26. Hansson, K., Maz'ya, V.G., Verbitsky, I.E.: Criteria of solvability for multidimensional Riccati equations. Ark. Mat. **37**, 87–120 (1999)
27. Hardy, G.: Note on a theorem of Hilbert. Math. Z. **6**, 314–317 (1920)
28. Heinonen, J., Holopainen, I.: Quasiregular maps on Carnot groups. J. Geom. Anal. **7**, no. 1, 109–148 (1997)
29. Heinonen, J., Koskela, P.: Quasiconformal maps in metric spaces with controlled geometry. Acta Math. **181**, 1–61 (1998)
30. Heinonen, J., Kilpeläinen, T., Martio, O.: Nonlinear Potential Theory of Degenerate Elliptic Equations. Oxford Univ. Press, Oxford (1993)
31. Hörmander, L.: Hypoelliptic second order differential equations. Acta Math. **119**, 147–171 (1967)
32. Hunt, R.: On $L^{p,q}$ spaces. Eins. Math. **12**, 249–276 (1966)
33. Jerison, D.: The Poincaré inequality for vector fields satisfying Hörmander's condition. Duke Math. J. **53**, 503–523 (1986)
34. Kaplan, A.: Fundamental solutions for a class of hypoelliptic PDE generated by composition of quadratic forms. Trans. Am. Math. Soc. **258**, 147–153 (1980)
35. Kilpeläinen, T., Malý, J.: The Wiener test and potential estimates for quasilinear elliptic equations. Acta Math. **172**, 137–161 (1994)
36. Kombe, I.: Sharp Weighted Hardy Type Inequalities and Uncertainty Principle Inequality on Carnot Groups. Preprint (2005)
37. Lehrbäck, J.: Pointwise Hardy inequalities and uniformly fat sets. Proc. Am. Math. Soc. [To appear]
38. Lewis, J.L.: Uniformly fat sets. Trans. Am. Math. Soc. **308**, 177–196 (1988)
39. Lohoué, N.: Une variante de l'inégalité de Hardy. Manuscr. Math. **123**, 73–78 (2007)
40. Maz'ya, V.G.: The negative spectrum of the n-dimentional Schrödinger operator (Russian). Dokl. Akad. Nauk SSSR **144**, 721–722 (1962); English transl.: Sov. Math., Dokl. **3**, 808–810 (1962)
41. Maz'ya, V.G.: Sobolev Spaces. Springer-Verlag, Berlin–Tokyo (1985)
42. Mikkonen, P.: On the Wolff potential and quasilinear elliptic equations involving measures. Ann. Acad. Sci. Fenn., Ser AI, Math. Dissert. **104** (1996)
43. Monti, R., Morbidelli, D.: Regular domains in homogeneous groups. Trans. Am. Math. Soc. **357**, no. 8, 2975–3011 (2005)
44. Nagel, A., Stein, E.M., Wainger, S.: Balls and metrics defined by vector fields I: Basics properties. Acta Math. **155**, 103–147 (1985)
45. Niu, P., Zhang, H., Wang, Y.: Hardy type and Rellich type inequalities on the Heisenberg group. Proc. Am. Math. Soc. **129**, no. 12, 3623–3630 (2001)
46. Rashevsky, P.K.: Any two points of a totally nonholonomic space may be connected by an admissible line (Russian). Uch. Zap. Ped. Inst. Liebknechta, Ser. Phys. Math. **2**, 83–94 (1938)

47. Rothschild, L.P., Stein, E.M.: Hypoelliptic differential operators and nilpotent groups. Acta. Math. **137**, 247–320 (1976)
48. Sanchez-Calle, A.: Fundamental solutions and geometry of sum of squares of vector fields. Invent. Math. **78**, 143–160 (1984)
49. Sobolev, S.L.: On a theorem of functional analysis (Russian). Mat. Sb. **46**, 471–497 (1938); English transl.: Am. Math. Soc., Transl., II. Ser. **34**, 39–68 (1963)
50. Trudinger, N.S., Wang, X.J.: On the weak continuity of elliptic operators and applications to potential theory. Am. J. Math. **124**, 369–410 (2002)
51. Wannebo, A.: Hardy inequalities. Proc. Am. Math. Soc. **109**, 85–95 (1990)
52. Stein, E.M.: Singular Integrals and Differentiability of Functions. Princeton Univ. Press, Princeton, NJ (1970)
53. Stein, E.M.: Harmonic Analysis: Real Variable Methods, Orthogonality and Oscillatory Integrals, Princeton Univ. Press, Princeton, NJ (1993)

Sobolev Embeddings and Hardy Operators

David E. Edmunds and W. Desmond Evans

Abstract Generalized ridged domains (GRD) are defined and examples of domains with irregular (even fractal) boundaries are given. Embedding problems on GRD are reduced to analogous problems on the generalized ridge (a tree in general). The latter problems involve Hardy type operators on trees with weights depending on geometric properties of the original GRD. Approximation and other singular numbers of Hardy type operators, including global bounds and asymptotic limits, are discussed.

1 Introduction

An important quantity in the study of properties of the embedding $E : W_p^1(\Omega) \to L_p(\Omega)$ is the *measure of noncompactness*

$$\alpha(E) := \inf\{\|E - P\| : P \in \mathcal{F}(W_p^1(\Omega), L_p(\Omega))\}, \tag{1.1}$$

where $\mathcal{F}(W_p^1(\Omega), L_p(\Omega))$ denotes the set of bounded linear maps of finite rank from the Sobolev space $W_p^1(\Omega)$ defined on a connected open subset Ω of $\mathbb{R}^n$ into $L_p(\Omega)$. We take the norm on $W_p^1(\Omega), 1 \leqslant p < \infty$, to be

$$\|f\|_{1,p,\Omega} := \Big(\|\nabla f\|_{p,\Omega}^p + \|f\|_{p,\Omega}^p\Big)^{1/p},$$

David E. Edmunds
School of Mathematics Cardiff University Senghennydd Road CARDIFF, Wales, CF24 4AG UK, e-mail: davideedmunds@aol.com

W. Desmond Evans
School of Mathematics Cardiff University Senghennydd Road CARDIFF, Wales, CF24 4AG UK, e-mail: EvansWD@cf.ac.uk

V. Maz'ya (ed.), *Sobolev Spaces in Mathematics I, International Mathematical Series.*
doi: 10.1007/978-0-387-85648-3,

where $\|\cdot\|_{p,\Omega}$ is the standard $L_p(\Omega)$ norm and

$$\|\nabla f\|_{p,\Omega}^p = \int_\Omega \sum_{j=1}^n |(\partial/\partial x_j)f|^p d\mathbf{x}.$$

It follows that $0 \leqslant \alpha(E) \leqslant 1$ and clearly $\alpha(E) = 0$ if and only if E is compact. Furthermore, $0 \leqslant \alpha(E) < 1$ if and only if Ω is of finite volume $|\Omega|$ and the Poincaré inequality holds, namely, there exists a positive constant K (depending on Ω) such that for all $f \in W_p^1(\Omega)$

$$\|f - f_\Omega\|_{p,\Omega} \leqslant K\|\nabla f\|_{p,\Omega}, \tag{1.2}$$

where f_Ω is the integral mean

$$f_\Omega := \frac{1}{|\Omega|}\int_\Omega f(\mathbf{x})d\mathbf{x}. \tag{1.3}$$

The case $p = 2$ of this result was proved in [3], and the general case in [17]. In [17], a class of domains (connected open sets) Ω called *generalized ridged domains* (GRDs) was introduced for which manageable criteria could be given to distinguish between the three cases $\alpha(E) = 0$, $0 \leqslant \alpha(E) < 1$, $\alpha(E) = 1$. The definition is motivated by natural properties of general domains, and the class includes a wide array of domains with irregular (even fractal) boundaries. A characteristic feature of GRDs is the so-called *generalized ridge*, a Lipschitz curve which, roughly speaking, is an axis of symmetry of the domain. The technique developed in [17] for GRDs is to equate the problem for $\alpha(E)$ with an analogous one on the generalized ridge. The generalized ridge is a crude approximation of the *central axis* or *skeleton* of the domain, but whereas, for domains with irregular boundaries, the latter is usually a complex array of curves, the generalized ridge can often be selected to be relatively easy to handle. It is this which makes GRDs so amenable to detailed investigation.

In general, the generalized ridges will be trees, and before formally defining the GRDs, it is necessary to prepare by studying the background analysis and, in particular, Hardy operators on trees.

2 Hardy Operators on Trees

Let Γ be a *tree*, i.e., a connected graph without loops or cycles where the edges are nondegenerate closed line segments whose end-points are vertices. We assume that each vertex is of finite degree, i.e., only a finite number of edges emanate from each vertex. For every $x, y \in \Gamma$ there is a unique polygonal path in Γ which joins x and y. Its length is defined to be the distance between x and y.

For $a \in \Gamma$ we define $t \succeq_a x$ (or equivalently $x \preceq_a t$) to mean that $x \in \Gamma$ lies on the path from a to $t \in \Gamma$; $t \succ_a x$ and $x \prec_a t$ have the obvious meaning. This is a partial ordering on Γ and the ordered graph so formed is referred to as the tree *rooted* at a and denoted by $\Gamma(a)$ when the root needs to be exhibited. If a is not a vertex, we can make it one by replacing the edge on which it lies by two. In this way, every rooted tree Γ is the unique finite union of subtrees which meet only at a. Any connected subset of Γ is a subtree if we adjoin its boundary points to the set of vertices, and hence form new edges from existing ones. *We shall adopt the convention of referring to all connected subtrees as subtrees.*

The path joining two points $x, y \in \Gamma$ may be parameterized by $s(t) = \text{dist}(x,t)$, and for $g \in L_{1,loc}(\Gamma)$ we have, with $\langle x, y\rangle = \{t : x \preceq_a t \preceq_a y\}$,

$$\int_x^y g = \int_{\langle x,y\rangle} g(t)dt = \int_0^{\text{dist}(x,y)} g[(t(s)]ds.$$

The space $L_p(\Gamma)$, $1 \leqslant p \leqslant \infty$, is defined in the natural way.

The Hardy type operator T to be considered is given by

$$T_a f(x) := v(x) \int_a^x u(t)f(t)dt, \quad f \in L_p(\Gamma), \tag{2.1}$$

where Γ is rooted at a and u, v are prescribed real-valued functions defined on Γ.

Definition 2.1. Let K be a connected subset of $\Gamma = \Gamma(a)$ containing the root a. Denote by ∂K the set of its boundary points. A point $t \in \partial K$ is said to be *maximal* if every $x \succ_a t$ lies in $\Gamma \setminus K$. We denote by $\mathcal{I}_a(\Gamma)$ (or simply $\mathcal{I}_a$ when no confusion is likely) the set of all connected subsets K of Γ which contain a and all of whose boundary points are maximal.

The main result proved in [20] (see also [9, Theorem 2.2.1]) on the boundedness of T in the case $p \leqslant q$ is

Theorem 2.1. *Let $1 \leqslant p \leqslant q \leqslant \infty$. Suppose that for all relatively compact $K \in \mathcal{I}_a(\Gamma)$*

$$u \in L_{p'}(K), \quad v \in L_q(\Gamma \setminus K), \tag{2.2}$$

where $p' = p/(p-1)$. Define

$$\alpha_K := \inf\{\|f\|_{p,\Gamma} : \int_a^X |u(t)f(t)|dt = 1 \quad \textit{for all} \quad X \in \partial K\}. \tag{2.3}$$

Then T in (2.1) *is a bounded linear map from $L_p(\Gamma)$ into $L_q(\Gamma)$ if and only if*

$$A := \sup_{K \in \mathcal{I}_a} \left\{ \frac{\|v\chi_{\Gamma\setminus K}\|_{q,\Gamma}}{\alpha_K} \right\} < \infty, \tag{2.4}$$

where χ denotes the characteristic function of the set exhibited, in which case we have

$$A \leqslant \|T\| \leqslant 4A. \tag{2.5}$$

Remark 2.1. If K in Theorem 2.1 is a single edge, say e, then $\alpha_e = \|u\|_{p',e}^{-1}$. For, by Hölder's inequality, we have

$$1 \leqslant \int_e |f|u \leqslant \|f\|_{p,\Gamma}\|u\|_{p',e}$$

which yields $\alpha_e \geqslant \|u\|_{p',e}^{-1}$. If $p > 1$, the reverse inequality is derived on taking $f(x) = u^{p'-1}(x)\|u\|_{p',e}^{-p'}$; for the case $p = 1$ see [9, Theorem 2.2.1]. Consequently, if Γ is an interval $[a,b)$, every $K \in \mathcal{I}_a$ is an interval $[a,c)$ with some $c \in (a,b)$ and (2.4) becomes the well-known criterion

$$A = \sup_{c \in (a,b)} \|u\|_{p',(a,c)}\|v\|_{q,(c,b)} < \infty \tag{2.6}$$

(see [9, Sect. 2.2.8] for some historical remarks). In this case, Opic (see [29, Comment 3.6, p. 27]) has shown that the constant $4A$ on the right-hand side of (2.5) may be replaced by

$$(1 + q/p')^{1/q}(1 + p'/q)^{1/p'}A,$$

which, when $p = q \in (1,\infty)$, gives the optimal constant $p^{1/p}(p')^{1/p'}A$. When $1 < p < q < \infty$, $(a,b) = (0,\infty)$, and $u \notin L_{p'}(0,\infty)$, further improvement was given by Manakov [27] and Read [31], who showed independently that instead of $4A$ one may take

$$\left\{ \frac{\Gamma(q/r)}{\Gamma(1+1/r)\Gamma((q-1)/r)} \right\}^{r/q}, \quad r = \frac{q}{p} - 1,$$

where Γ denotes the gamma function. Moreover, if either $p = 1$ and $q \in [1,\infty]$, or $q = \infty$ and $p \in (1,\infty]$, it is known that (see [28] and [29])

$$\|T\| = A.$$

Remark 2.2. Let $\Gamma = \Gamma_a$ be a tree rooted at a. Define

$$A_1 := \sup_{x \in \Gamma} \|u\|_{p',(a,x)}\|v\|_{q,(a,x)^c}, \tag{2.7}$$

where $(a,x)^c = \{y \in \Gamma : y \succeq_a x\}$, the *shadow* of the point x, and

$$A_2 := \sup_{K \in \mathcal{I}_a} \|u\|_{p',K} \|v\|_{q,\Gamma \setminus K}. \tag{2.8}$$

Then A in (2.4) satisfies

$$A_1 \leqslant A \leqslant A_2. \tag{2.9}$$

For, with $x \in \Gamma$ fixed and $K = \Gamma \setminus (a,x)^c$, x is the only point in ∂K and hence from Remark 2.1 it follows that

$$\alpha_K = \|u\|^{-1}_{p',(a,x)},$$

whence $A_1 \leqslant A$. Next, let $K \in \mathcal{I}_a$ and $\partial K = \{x_j\}_{j=1}^{m(K)}$, where $m(K)$ is a natural number or infinity. Then $K = \Gamma \setminus \bigcup\limits_{j=1}^{m(K)} (a, x_j)^c$. We have

$$\alpha_K \geqslant \inf \left\{ \|f\|_{p,\Gamma} : \int_a^{x_j} |f| u = 1 \right\} = \|u\|^{-1}_{p',(a,x_j)},$$

and this yields $A \leqslant A_2$.

Remark 2.3. It is proved in [20] that neither A_1 nor A_2 is, in general, comparable to A. In [19], they are shown to be comparable if it is assumed that

$$\int_{x \succeq_a t} v^p(x) \left[\int_{y \succeq_a x} v^p(y) dy \right]^{-1/p'} dx \leqslant C \left(\int_{y \succeq_a x} v^p(y) dy \right)^{1/p}. \tag{2.10}$$

The first to deal with the case $p > q$ for an interval $\Gamma = [a,b)$ was Maz'ya [28], who established the following result.

Theorem 2.2. *Let $1 \leqslant q < p \leqslant \infty$, $1/s = 1/q - 1/p$. Suppose that* (2.2) *holds. Then T_a is a bounded linear map from $L_p(\Gamma)$ to $L_q(\Gamma)$ if and only if*

$$B := \left\{ \int_a^b \left(\left(\int_x^b |v(t)|^q \, dt \right)^{1/q} \left(\int_a^x |u(t)|^{p'} \, dt \right)^{1/q'} \right)^s |u(x)|^{p'} \, dx \right\}^{1/s} \tag{2.11}$$

is finite, in which case

$$q^{1/q} (p'q/s)^{1/q'} B \leqslant \|T_a\| \leqslant q^{1/q} (p')^{1/q'} B. \tag{2.12}$$

A characterization in the case $p > q$ which holds for a tree Γ is given in [20, Theorem 5.2]. To explain this, let $B_i, i \in \mathcal{I}$, be nonempty disjoint subsets of Γ which are such that $\Gamma = \bigcup\limits_{i \in \mathcal{I}} B_i$ and $B_i = K_{i+1} \setminus K_i$, where $K_i \in \mathcal{I}_a(\Gamma)$ and

$K_i \subset K_{i+1}$. Denote the set of all such decompositions $\{B_i\}_{i\in\mathcal{I}}$ by $\mathcal{C}(\Gamma)$. In the case where Γ is an interval $[a,b]$, we take $\Gamma = \bigcup_{i\in\mathcal{I}} \overline{\mathcal{B}}_i$ and $B_i = (a_i, a_{i+1})$.

Theorem 2.3. *Let* $1 \leqslant q < p \leqslant \infty$, $1/s = 1/q - 1/p$. *Suppose that* (2.2) *holds. For* $\{B_i\}_{i\in\mathcal{I}} \in \mathcal{C}(\Gamma)$ *define*

$$\alpha_i := \{\|f\|_{p,\Gamma} : \textit{supp}\ f \subset B_{i-1}, \int_a^X |u(t)f(t)|dt = 1 \ \textit{ for all }\ X \in \partial K_i\} \quad (2.13)$$

and

$$\beta_i := \|v\|_{q,B_i}/\alpha_i. \quad (2.14)$$

Then T_a *is a bounded linear map from* $L_p(\Gamma)$ *to* $L_q(\Gamma)$ *if and only if*

$$B := \sup_{\mathcal{C}(\Gamma)} \|\{\beta_i\} \mid l_s(\mathcal{I})\| < \infty, \quad (2.15)$$

where $l_s(\mathcal{I})$ *is the usual sequence space. If* (2.15) *holds, then*

$$B \leqslant \|T_a\| \leqslant 4B. \quad (2.16)$$

The constant A in Theorem 2.1 can also be expressed in the form

$$A = \sup_{\mathcal{C}(\Gamma)} \|\{\beta_i\}|\ell_\infty(\mathcal{I})\|$$

(see [20, Remark 3.4]).

3 The Poincaré Inequality, $\alpha(E)$ and Hardy Type Operators

The Poincaré inequality associated with $W_p^1(\Omega)$, $1 \leqslant p < \infty$, is of the form

$$\|f - f_\Omega\|_{p,\Omega} \leqslant K(\Omega, p, n)\|\nabla f\|_{p,\Omega}, \quad f \in W_p^1(\Omega), \quad (3.1)$$

where the constant $K(\Omega, p, n)$ depends only on Ω, p, and n. It holds, for instance, if Ω is a bounded convex domain in which case $K(\Omega, p, n) = K(p, n)\ \mathrm{diam}(\Omega)$. As noted in Sect. 1, there is an intimate connection between the Poincaré inequality for a domain (an open connected set) Ω and the quantity $\alpha(E)$: $\alpha(E) \in [0, 1)$ if and only if Ω has finite volume $|\Omega|$ and the Poincaré inequality holds.

The value of $\alpha(E)$ depends on the nature of the boundary $\partial\Omega$ of Ω and may be determined by the geometry of Ω in the neighborhood of a single point on $\partial\Omega$, as is the case, for example, in the "Rooms and Passages" domain analyzed in [17, Sect. 6] (see also Example 4.2 below). To deal with such cases, the following refinement of the notion of the singular part of the boundary was used in [17]. Let $\mathcal{A}$ be a *filter base* consisting of relatively closed subsets of Ω which satisfy the following:

(1) for each $A \in \mathcal{A}$, the embedding

$$W_p^1(\Omega) \hookrightarrow L_p(\Omega \setminus A) \tag{3.2}$$

is compact;

(2) $\mathcal{A}$ is finer than the filter base

$$\mathcal{A}_0 := \{A : A \in \Omega \setminus \Omega', \Omega' \subset\subset \Omega\}, \tag{3.3}$$

where $\Omega' \subset\subset \Omega$ means that the closure of Ω' is a compact subset of Ω.

The sets of the form $A \cup \{\infty\}$, $A \in \mathcal{A}_0$, are the closed neighborhoods of the point at infinity in the one-point compactification of Ω. The family of all relatively closed sets A satisfying (1) is a filter base if and only if E is not compact, for then the empty set is not a member. The adherence of this filter base in the Stone–Čech compactification of Ω is the set which can prevent E from being compact. Note that each $A \in \mathcal{A}_0$ satisfies (3.2) since there exists a bounded domain Ω_0 with smooth boundary such that $\Omega \setminus A \subset \Omega_0 \subset \Omega$.

Since it is a filter base, $\mathcal{A}$ is directed by reverse inclusion, i.e., by the order relation $\succ$, where $A_1 \succ A_2$ if $A_1 \subseteq A_2$. We say that a family $\{\psi_A\}$ of real numbers indexed by $\mathcal{A}$ converges to a limit $\psi \in \mathbb{R}$, written $\lim\limits_{\mathcal{A}} \psi_A = \psi$, if for each neighborhood U of ψ in $\mathbb{R}$ there is an $A_0 \in \mathcal{A}$ such that $\psi_A \in U$ for all $A \succ A_0$ in $\mathcal{A}$. It is proved in [17, Corollary 2.5] that

$$\alpha(E) = \lim_{\mathcal{A}} \psi_A, \quad \psi_A = \sup_{f \in W_p^1(\Omega)} \{\|f + Hf\|_{p,A} : \|f\|_{1,p,\Omega} = 1\}, \tag{3.4}$$

where H is any (fixed) compact linear map from $W_p^1(\Omega)$ into $L_p(\Omega)$.

Suppose the Poincaré inequality holds, or equivalently $\alpha(E) < 1$, and define

$$W_{M,p}^1(\Omega) := \{f \in W_p^1(\Omega) : f_\Omega = 0\} \tag{3.5}$$

with norm $\|f\|_{M,p,\Omega} := \|\nabla f\|_{p,\Omega}$. This norm is equivalent to the $W_p^1(\Omega)$ norm on $W_{M,p}^1(\Omega)$, and we have the topological isomorphism

$$W_p^1(\Omega) \simeq \mathcal{C} \oplus W_{M,p}^1(\Omega),$$

where $\mathcal{C}$ denotes the set of constants and $\oplus$ the direct sum. We have the embedding

$$E_M : W_{M,p}^1(\Omega) \to L_{M,p}(\Omega) := \{f \in L_p(\Omega) : f_\Omega = 0\}. \tag{3.6}$$

This is bounded if and only if the Poincaré inequality holds, and so if and only if $\alpha(E) < 1$. The maps E, E_M are compact together, and, setting

$$\alpha(E_M) := \inf\{\|E_M - P\| : P \in \mathcal{F}(W^1_{M,p}(\Omega), L_p(\Omega))\},$$

we have from [17, Theorem 2.10] that if $\alpha(E) < 1$, or equivalently, if E_M is bounded, then

$$\frac{\alpha(E_M)^p}{1 + \|E_M\|^p} \leqslant \alpha(E)^p \leqslant \frac{\alpha(E_M)^p}{1 + \alpha(E_M)^p}.$$

These are the facts which enable the three cases $\alpha(E) = 0$, $\alpha(E) \in (0,1)$, $\alpha(E) = 1$ to be distinguished and motivate the strategy adopted in [17] and [19] of working with the embedding E_M rather than E directly. Of crucial importance is Corollary 2.9 in [17], that

$$\begin{aligned} &\alpha(E_M) = \lim_{\mathcal{A}} \varphi_A, \\ &\varphi_A = \sup\{\|f + Hf\|_{p,A} : f \in W^1_{M,p}(\Omega), \|\nabla f\|_{p,\Omega} = 1\}, \end{aligned} \tag{3.7}$$

where H is any fixed compact linear map from $W^1_{M,p}(\Omega)$ into $L_{M,p}(\Omega)$.

Central to the analysis on GRDs to be discussed in the next section is an inequality of Poincaré type on a tree Γ and a Hardy type operator T_a defined by (2.1) with specific functions u, v determined by the geometry of Ω (see (4.1) below). This operator T_a maps a space $L_p(\Gamma; d\mu)$ into itself, where the measure $d\mu$ is given by $d\mu(x) = v^p(x)dx$ and $L_p(\Gamma; d\mu)$ has norm defined by

$$\|F\|_{p,\Gamma;d\mu} := \left\{ \int_\Gamma |F(x)|^p d\mu(x) \right\}^{1/p}.$$

The functions u, v in (2.1) will be assumed to satisfy the conditions

$$u \in L_{p'}(K),\ (K \in \mathcal{I}_a(\Gamma)) \quad v \in L_p(\Gamma), \tag{3.8}$$

and so Γ has finite μ measure. Also, setting

$$T_a f(x) = v(x)F(x), \quad F(x) = \int_a^x f(t)u(t)dt,$$

we have that $T_a f \in L_p(\Gamma)$ if and only if $F \in L_p(\Gamma; d\mu)$. The Poincaré inequality which has such an important role to play is

$$\|F - F_\Gamma\|_{p,\Gamma;d\mu} \leqslant C\|f\|_{p,\Gamma}, \tag{3.9}$$

where F_Γ is the integral mean of F on Γ, namely

$$F_\Gamma := (1/\mu(\Gamma)) \int\limits_\Gamma F(t)d\mu(t).$$

We require two-sided bounds for

$$B(\Gamma) := \sup\{: \|F - F_\Gamma\|_{p,\Gamma;d\mu} : \|f\|_{p,\Gamma} = 1\}. \tag{3.10}$$

These are given in terms of the norm of T_a under the assumption (3.8) and hence of the related quantity A in (2.4).

First observe that $T_a : L_p(\Gamma) \to L_p(\Gamma)$ is bounded if and only if T_c is, where

$$T_c f(x) = v(x) \int\limits_c^x f(t)u(t)dt$$

for any $c \in \Gamma$, for, by (3.8),

$$\|v \int\limits_c^a fudt\|_{p,\Gamma} \leqslant (\|u\|_{p',(a,c)}\|v\|_{p,\Gamma})\|f\|_{p,\Gamma}.$$

Let Γ' be a subtree of Γ such that $\Gamma'' = \Gamma \setminus \Gamma'$ is also a tree, and let $\overline{\Gamma'} \cap \overline{\Gamma''} = \{c\}$. Define

$$A_c(\Gamma') := \sup\{\|(T_c f)\|_{p,\Gamma'} : \|f\|_{p,\Gamma'} = 1\} \tag{3.11}$$

and $A_c(\Gamma'')$ similarly.

Lemma 3.1. *The bound* $B(\Gamma)$ *in* (3.10) *is finite if and only if* $A_c(\Gamma')$ *and* $A_c(\Gamma'')$ *are finite for every* Γ', Γ'' *of the above form. Also*

$$(1 - 2^{-1/p})A(\Gamma) \leqslant B(\Gamma) \leqslant 2A(\Gamma), \tag{3.12}$$

where

$$A(\Gamma) := \inf_{c\in\Gamma} \max\left(A_c(\Gamma'), A_c(\Gamma'')\right).$$

Proof. Set

$$F_c(x) = \int\limits_c^x f(t)u(t)dt,$$

so that $\|T_c f\|_{p,\Gamma} = \|F_c\|_{p,\Gamma;d\mu}$. Then $F(x) - F_\Gamma = F_c(x) - (F_c)_\Gamma$ and

$$\begin{aligned}\|F - F_\Gamma\|_{p,\Gamma;d\mu} &\leqslant \|F_c\|_{p,\Gamma;d\mu} + |(F_c)_\Gamma|\mu(\Gamma)^{1/p} \leqslant 2\|F_c\|_{p,\Gamma;d\mu}\\ &\leqslant 2\{A_c^p(\Gamma')\|f\chi_{\Gamma'}\|_{p,\Gamma} + A_c^p(\Gamma'')\|f\chi_{\Gamma''}\|_{p,\Gamma}\}^{1/p}\\ &\leqslant 2\max\left(A_c(\Gamma'), A_c(\Gamma'')\right)\|f\|_{p,\Gamma}.\end{aligned}$$

Thus, $B(\Gamma) \leqslant 2\max\left(A_c(\Gamma'), A_c(\Gamma'')\right)$.

Conversely, suppose that $B(\Gamma) < \infty$ and let f have support in Γ'. Then $F_c(x) = 0$ for $x \in \Gamma''$ and

$$B(\Gamma)\|f\chi_{\Gamma'}\|_{p,\Gamma} \geqslant \|(F_c - (F_c)_\Gamma)\chi_{\Gamma'}\|_{p,\Gamma;d\mu}$$
$$\geqslant \|F_c\chi_{\Gamma'}\|_{p,\Gamma;d\mu}\left\{1 - \left[\frac{\mu(\Gamma')}{\mu(\Gamma)}\right]^{1/p}\right\}.$$

Hence

$$A_c(\Gamma') \leqslant B(\Gamma)\left\{1 - \left[\frac{\mu(\Gamma')}{\mu(\Gamma)}\right]^{1/p}\right\}^{-1}$$

and similarly

$$A_c(\Gamma'') \leqslant B(\Gamma)\left\{1 - \left[\frac{\mu(\Gamma'')}{\mu(\Gamma)}\right]^{1/p}\right\}^{-1}.$$

The lemma follows on choosing $c \in \Gamma$ such that $\mu(\Gamma') = \mu(\Gamma'') = (1/2)\mu(\Gamma)$. $\square$

4 Generalized Ridged Domains

The GRDs were introduced in [17] and [19] in the search for a class of domains in $\mathbb{R}^n$ which is amenable to detailed analysis and wide enough to contain domains with highly irregular, even fractal, boundaries which have been the subject of interest and intensive study. Examples include, in particular, "Rooms and Passages," "interlocking combs," infinite horns and spirals, and the Koch snowflake and analogues. An account of how the definition was motivated by various properties of sets in $\mathbb{R}^n$ may be found in the original papers, as well as in [9, Chapt. 5].

In what follows we define the "derivative" of a Lipschitz continuous function everywhere by

$$g'(t) := \limsup_{n\to\infty}\{n[g(t + n^{-1}) - g(t)]\};$$

recall that by Rademacher's theorem, a Lipschitz continuous function is differentiable almost everywhere.

Definition 4.1. A domain Ω in $\mathbb{R}^n$, $n \geqslant 1$, with $|\Omega| < \infty$ is a *generalized ridged domain*, GRD for short, if there exist real-valued functions u, ρ, τ, a tree Γ, and positive constants α, β, γ, δ such that the following conditions are satisfied:

(1) $u : \Gamma \to \Omega, \rho : \Gamma \to (0, \infty)$ are Lipschitz;

(2) $\tau : \Omega \to \Gamma$ is surjective and uniformly locally Lipschitz, i.e., for each $\mathbf{x} \in \Omega$ there exists a neighborhood $V(\mathbf{x})$ such that for all $\mathbf{y} \in V(\mathbf{x})$, $|\tau(\mathbf{y}) - \tau(\mathbf{x})|_\Gamma \leqslant \gamma|\mathbf{x} - \mathbf{y}|$, where $|\cdot|_\Gamma$ denotes the metric on Γ;

(3) $|\mathbf{x} - u \circ \tau(\mathbf{x})| \leqslant \alpha(\rho \circ \tau(\mathbf{x}))$ for all $\mathbf{x} \in \Omega$;

(4) $|u'(t)| + |\rho'(t)| \leqslant \beta$ for all $t \in \Gamma$;

(5) with $B_t := B(u(t), \rho(t))$, the ball with center $u(t)$ and radius $\rho(t)$ in Ω, and $\mathcal{C}(\mathbf{x}) := \{\mathbf{y} : s\mathbf{y} + (1-s)\mathbf{x} \in \Omega \text{ for all } s \in [0,1]\}$, we have that for all $\mathbf{x} \in \Omega, \mathcal{C}(\mathbf{x}) \cap B_{\tau(\mathbf{x})}$ contains a ball $B(\mathbf{x})$ such that $B(\mathbf{x})|/|B_{\tau(\mathbf{x})}| \geqslant \delta > 0$.

The curve $t \mapsto u(t) : \Gamma \to \Omega$ is a *generalized ridge* of Ω.

A positive Borel measure on Γ is defined by the map τ in the definition. Since

$$\int\limits_\Omega F \circ \tau(\mathbf{x}) d\mathbf{x}, \quad F \in C_0(\Gamma),$$

is a positive linear functional on $C_0(\Gamma)$, the set of continuous functions on Γ with compact support, it follows from the Riesz representation theorem for $C_0(\Gamma)$ that there exists a positive finite measure μ on Γ such that

$$\int\limits_\Gamma F(t) d\mu(t) := \int\limits_\Omega F \circ \tau(\mathbf{x}) d\mathbf{x}, \quad F \in C_0(\Gamma).$$

For any open subset Γ_0 of Γ we have

$$\mu(\Gamma_0) = |\tau^{-1}(\Gamma_0)|.$$

The map $F \mapsto F \circ \tau : C_0(\Gamma) \to L_p(\Omega)$ extends by continuity to a map

$$T : L_p(\Gamma; d\mu) \to L_p(\Omega)$$

which satisfies $TF(\mathbf{x}) = F \circ \tau(\mathbf{x})$ for a.e. $\mathbf{x} \in \Omega$. Also, T is an isometry:

$$\|TF\|_{p,\Omega} = \|F\|_{p,\Gamma,d\mu},$$

where

$$\|F\|_{p,\Gamma,d\mu} := \left\{ \int\limits_\Gamma |F(t)|^p d\mu(t) \right\}^{1/p}.$$

In fact, μ is given explicitly by the co-area formula (see [9, Theorem 1.2.4]). For if $\nabla\tau(\mathbf{x}) \neq 0$, a.e.

$$\int\limits_\Omega F \circ \tau(\mathbf{x}) d\mathbf{x} = \int\limits_\Gamma F(t) \int\limits_{\tau^{-1}(t)} |\nabla\tau(\mathbf{x})|^{-1} dH^{n-1}(\mathbf{x}) dt,$$

where H^{n-1} denotes $(n-1)$-dimensional Hausdorff measure. Hence μ is locally absolutely continuous with respect to Lebesgue measure and since $|\nabla\tau(\mathbf{x})| \leqslant \gamma$ by Definition 4.1(2),

$$\frac{d\mu}{dt} = \int_{\tau^{-1}(t)} |\nabla\tau(\mathbf{x})|^{-1} dH^{n-1}(\mathbf{x}) \geqslant \frac{1}{\gamma} H^{n-1}(\tau^{-1}(t)).$$

Therefore, if $|\nabla\tau(\mathbf{x})| \neq 0, a.e., dt$ is locally absolutely continuous with respect to $d\mu$: *we always assume this hereafter.* If $n=2$, and $\tau^{-1}(t)$ is a rectifiable curve in Ω for a.e. $t \in \Gamma$, then its 1-dimensional Hausdorff measure is equal to its length, $l(t)$ say, and hence

$$\frac{d\mu}{dt} \geqslant \gamma^{-1} l(t).$$

Consequently, dt is absolutely continuous with respect to $d\mu$ on any compact subset of Γ on which $l(\cdot)$ is positive.

The Hardy operator associated with the GRD Ω is defined by (2.1) with

$$u(t) := (dt/d\mu)^{1/p}, \quad v(t) := (d\mu/dt)^{1/p} \tag{4.1}$$

and (3.8) is assumed. Recall that we are assuming throughout that dt is locally absolutely continuous with respect to $d\mu$.

From (5) it follows that for each $\varepsilon > 0$ the set

$$\Omega(\varepsilon) := \{\mathbf{x} : \mathbf{x} \in \Omega, \rho \circ \tau(\mathbf{x}) > \varepsilon\}$$

lies in a bounded open subset Ω_ε of Ω which satisfies a cone condition, i.e., there is a cone $\mathcal{C}(\varepsilon)$ such that each $\mathbf{x} \in \overline{\Omega_\varepsilon}$ is the vertex of a cone congruent to $\mathcal{C}(\varepsilon)$ which lies in $\overline{\Omega_\varepsilon}$. In view of Remark 6.3(4) in [2], this is enough to ensure that the embedding $W^1_p(\Omega) \hookrightarrow L_p(\Omega(\varepsilon))$ is compact. Therefore, if the embedding E is not compact it is because of the nature of the set of points where the generalized ridge meets the boundary of Ω.

Let

$$Mf(t) := \frac{1}{|B_t|} \int_{B_t} f(\mathbf{x}) d\mathbf{x}, \quad f \in W^1_p(\Omega), \tag{4.2}$$

where the B_t are the balls in Definition 4.1(5). Then, by [19, Sect. 3], $M : W^1_p(\Omega) \to L^1_p(\Gamma; d\mu)$ is bounded, where $L^1_p(\Gamma; d\mu)$ is the set of functions F which are Lipschitz continuous on Γ and $F, F' \in L_p(\Gamma; d\mu)$. Also, for any measurable subset Ω_1 of Ω,

$$\|f - TMf\|_{p,\Omega_1} \leqslant Kk(\Omega_1)\|\nabla f\|_{p,\Omega},$$

where K is a positive constant and

$$k(\Omega_1) := \sup_{\Omega_1}\{\rho \circ \tau(\mathbf{x})\} < \infty.$$

Hence, as Ω_1 approaches any point on $\partial\Omega$ lying on the generalized ridge, $k(\Omega_1) \to 0$. Since $W_p^1(\Omega) \hookrightarrow L_p(\Omega(\varepsilon))$ is compact, it follows that $E_M - TM$ is compact, so that T and M are *approximate inverses*.

We now select a specific filter base of subsets of Ω which will satisfy (3.2) and (3.3). Let

$$\begin{aligned}\mathcal{A}(\Gamma) := \{\Lambda : \Lambda \subset \Gamma \ \ \text{nonempty and relatively closed,} \ \ \overline{\Gamma \setminus \Lambda} \\ \text{a} \ \ \text{compact subtree of} \ \ \Gamma\}\end{aligned} \tag{4.3}$$

and

$$\mathcal{A}(\Omega) := \{\tau^{-1}(\Lambda) : \Lambda \in \mathcal{A}(\Gamma)\}. \tag{4.4}$$

A subset of Γ is compact if and only if it is closed and meets a finite number of edges. Thus, if Γ has an infinite number of edges, the boundary of $\Gamma \setminus \Lambda$, $\Lambda \in \mathcal{A}(\Gamma)$, is finite and Λ is a finite union of closed disjoint subtrees A_i of Γ which are rooted at the boundary points of $\Gamma \setminus \Lambda : A = \bigcup_{i \in N_A} A_i$ say. If Γ is an interval $[a, b)$, then the sets Λ are given by

$$\Lambda = \Lambda(\varepsilon) = \begin{cases} [b - \varepsilon, b) & \text{if } b < \infty, \\ [\varepsilon^{-1}, \infty) & \text{if } b = \infty \end{cases}$$

for suitable $\varepsilon > 0$. It is shown in [19, Lemma 4.1] that $\mathcal{A}(\Omega)$ is a filter base which satisfies (3.2) and (3.3). Thus, as noted in Sect. 4, $\alpha(E_M) = \lim_{\mathcal{A}(\Omega)} \varphi_A$ where φ_A is given by (3.7) and $A = \tau^{-1}(\Lambda) \in \mathcal{A}(\Omega)$.

The maps T and M and the filter bases $\mathcal{A}(\Gamma), \mathcal{A}(\Omega)$ play a leading role in the analysis of [17] and [19]. The strategy is based on the following steps. We assume throughout that $1 < p < \infty$ and that (3.8) is satisfied.

(A) For $A \in \mathcal{A}(\Omega)$ define

$$h_A f := \sum_{i \in N_A} \chi(A_i) f_{A_i}$$

and set

$$\psi_A : \sup\{\|f - h_A f\|_{p,A} : f \in W^1_{M,p}(\Omega), \|\nabla f\|_{p,\Omega} = 1\}.$$

On choosing $\mathcal{A} = \mathcal{A}(\Omega)$ and $H = h_{A_0}$ in (3.7) with a fixed $A_0 \in \mathcal{A}(\Omega)$, it follows from the fact that $\alpha(E_M) = \lim_{\mathcal{A}} \varphi_A$ that

$$\alpha(E_M) \leqslant \inf_{\mathcal{A}(\Omega)} \psi_A \leqslant \limsup_{\mathcal{A}(\Omega)} \psi_A \leqslant 2\alpha(E_M) \tag{4.5}$$

and this yields, with $A = \tau^{-1}(\Lambda) \in \mathcal{A}(\Omega)$,

$$\gamma^{-1}\theta_\Lambda \leqslant \psi_A \leqslant K\{k(A)+\theta_\Lambda\}, \tag{4.6}$$

where

$$\begin{aligned}\theta_\Lambda &:= \sup\{\|F - H_\Lambda F\|_{p,\Lambda,d\mu} : F\in L_p^1(\Gamma;d\mu), \|F'\|_{p,\Lambda,d\mu}=1\},\\ H_\Lambda F &:= \sum_{i\in N_A}\chi(\Lambda_i)F_{\Lambda_i},\\ F_{\Lambda_i} &:= \frac{1}{\mu(\Lambda_i)}\int\limits_{\Lambda_i} F(t)d\mu(t).\end{aligned} \tag{4.7}$$

(B) Set

$$\theta_+ := \limsup_{\mathcal{A}(\Gamma)} \theta_\Lambda, \quad \theta_- := \liminf_{\mathcal{A}(\Gamma)} \theta_\Lambda.$$

Then

$$\frac{1}{2\gamma}\theta_+ \leqslant \alpha(E_M) \leqslant K\theta_-.$$

This follows from (4.5) and (4.6), on showing that if $\lim\limits_{\mathcal{A}(\Omega)} k(A)\neq 0$, then $\alpha(E_M)=\theta_+=\theta_-=0$.

(C) The following Poincaré inequalities are equivalent:

$$\|f-f_\Omega\|_{p,\Omega} \leqslant C(\Omega)\|\nabla f\|_{p,\Omega} \quad (f\in W_p^1(\Omega))$$

$$\|F-F_\Gamma\|_{p,\Gamma,d\mu} \leqslant c(\Gamma)\|F'\|_{p,\Gamma,d\mu} \quad (F\in L_p^1(\Gamma;d\mu)).$$

Moreover, the optimal constants satisfy

$$\gamma^{-1}c(\Gamma) \leqslant C(\Omega) \leqslant K\{k(\Omega)+c(\Gamma)\}.$$

This equivalence of the two Poincaré inequalities lies at the heart of the technique and is what motivates the choice of the Hardy operator T_a associated with the GRD Ω.

(D) If (2.10) is satisfied, then

$$\alpha(E_M) \asymp \lim_{\mathcal{A}(\Gamma)} J(\Lambda), \tag{4.8}$$

where

$$J(\Lambda) := \sup_{x\in\Lambda}\left\{[\mu\{t : t\in\Lambda, t\succeq_a x\}]^{1/p}\Big[\int\limits_{t\preceq_a x, t\in\Lambda} \psi^{p'/p}(t)dt\Big]^{1/p'}\right\} \tag{4.9}$$

and $\psi(t) := dt/d\mu$.

(E) Under the above conditions, we have the following:

(1) $\alpha(E) < 1$ if and only if $J(\Gamma) < \infty$;

(2) E and E_M are compact if and only if $\lim\limits_{\mathcal{A}(\Gamma)} J(\Lambda) = 0$.

Example 4.1 (horn-shaped domain). Let $\mathbf{x} = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$ be written as $(x_1, \mathbf{x}')$, where $\mathbf{x}' = (x_2, \cdots, x_n) \in \mathbb{R}^{n-1}$. A horn-shaped domain is of the form

$$\Omega := \{\mathbf{x} = (x_1, \mathbf{x}') \in \mathbb{R}^n : 0 < x_1 < \infty, |\mathbf{x}'| < \Phi(x_1)\}, \tag{4.10}$$

where Φ is smooth and bounded. This is a GRD with a generalized ridge $(0, \infty)$. In the notation of Definition 4.1, $u(t) = (t, 0, \cdots)$ and $\tau(x_1, \mathbf{x}') = x_1$. Furthermore, with ω_{n-1} denoting the measure of the unit ball in $\mathbb{R}^{n-1}$,

$$\begin{aligned}\int_0^s F(t)d\mu(t) &= \int_0^s F(x_1)dx_1 \int_{|\mathbf{x}'|<\Phi(x_1)} d\mathbf{x}' \\ &= \omega_{n-1} \int_0^s F(x_1)\Phi(x_1)^{n-1}dx_1,\end{aligned}$$

whence

$$\mu'(t) = \omega_{n-1}\Phi(t)^{n-1} \tag{4.11}$$

and

$$\begin{aligned}&\{\mu(\infty) - \mu(s)\}^{1/p} \left(\int_c^s \psi(t)^{p'/p} dt \right)^{1/p'} \\ &= \left(\int_s^\infty \Phi(r)^{n-1} dr \right)^{1/p} \left(\int_c^s \Phi(t)^{(1-n)/(p-1)} dt \right)^{1/p'} .\end{aligned} \tag{4.12}$$

It follows that

$\Phi(t)^{n-1} = (t+1)^{-\theta}$, $\theta > 1 \Rightarrow \alpha(E) = 1$;

$\Phi(t)^{n-1} = e^{-\lambda t}$, $\lambda > 0 \Rightarrow 0 < \alpha(E) < 1$;

$\Phi(t)^{n-1} = e^{-t^\theta}, \theta > 1 \Rightarrow \alpha(E) = 0$.

Example 4.2 (Rooms and Passages). Let $\{h_k\}_{k\in\mathbb{N}}$ and $\{\delta_{2k}\}_{k\in\mathbb{N}}$ be sequences of positive numbers such that

$$\sum_{k=1}^\infty h_k = b < \infty, \quad 0 < \text{const.} \leqslant h_{k+1}/h_k \leqslant 1, \quad 0 < \delta_{2k} \leqslant h_{2k+1}, \tag{4.13}$$

and let $H_k := \sum_{k=1}^{k} h_j, (k \in \mathbb{N})$. Then Ω is defined to be the union of the rooms R_k and passages P_{k+1} given by

$$\begin{aligned} R_k &:= (H_k - h_k, H_k) \times (-h_k/2, h_k/2) \\ P_{k+1} &= [H_k, H_k + h_{k+1}] \times (-\delta_{k+1}, \delta_{k+1}), \end{aligned} \tag{4.14}$$

for $k = 1, 3, \cdots$. We choose the interval $[0, b)$ to be a generalized ridge, with $u(t) = (t, 0), 0 \leqslant t < b$. In [17, Sect. 6], it is shown that functions τ, ρ can be chosen such that

$$J([c,b)) \asymp \sup_{N \leqslant k < \infty} \left\{ \sum_{j=k}^{\infty} h_{2j}^2 \left[\sum_{i=N}^{k} (h_{2i}/\delta_{2i}^{p'/p} + h_{2i+2}/\delta_{2i+2}^{p'/p}) \right]^{p-1} \right\}^{1/p},$$

when $H_{2N} - (1/2)h_{2N} \leqslant c < H_{2N+1} + (1/2)h_{2N+2}$. This yields

$$\alpha(E_M) \asymp \lim_{N\to\infty} \sup_{N \leqslant k < \infty} \left\{ \sum_{j=k}^{\infty} h_{2j}^2 \left[\sum_{i=N}^{k} (h_{2i}/\delta_{2i}^{p'/p} + h_{2i+2}/\delta_{2i+2}^{p'/p}) \right]^{p-1} \right\}^{1/p}. \tag{4.15}$$

In the special case $\delta_{2i} = ch_{2i}^{\varkappa}$, $\varkappa > 1$, $h_{2i} = C^{-i}$, where c is a positive constant and $C > 1$, we have the following:

$\varkappa \geqslant p + 1 \Rightarrow \alpha(E) = 1;$

$\varkappa = p + 1 \Rightarrow 0 < \alpha(E) < 1;$

$\varkappa < p + 1 \Rightarrow \alpha(E) = 0.$

Example 4.3 (a snowflake type domain). We construct a domain in $\mathbb{R}^2$ from a succession of *generations* Θ_m of closed congruent rectangles Q_m with nonoverlapping interiors, having edge lengths $2\alpha_m \times 2\beta_m$, where $\alpha_m = c^{\varkappa m}$, $\beta_m = c^m$ $(m \in \mathbb{N}_0)$, $\varkappa \geqslant 1$, and to ensure nonoverlapping, suppose that $c^{\varkappa} + 2c^2 < 1$ and $c^{1+\varkappa/p+p/p'^2} < 1/2$ when $\varkappa > p/p'$. The generation Θ_0 consists of a single rectangle, as does Θ_1, a short edge of Q_1 being attached to the middle portion of a long edge of Q_0. For $m \geqslant 1, \Theta_m$ contains 2^{m-1} rectangles and to each long edge of Q_m is attached a short edge of a rectangle Q_{m+1}, these 2^m rectangles Q_{m+1} being the members of Θ_{m+1}. The domain Ω is the interior of the connected set Θ constructed in this way:

$$\Omega = \Theta^o, \Theta = \bigcup_{m \in \mathbb{N}_0} \left(\cup \{Q_m : Q_m \in \Theta_m\} \right). \tag{4.16}$$

The generalized ridge is the tree formed by the major and minor axes of the rectangles. It is of finite degree and $u : \Gamma \to \Omega$ is the identification map. For $\Lambda \in \mathcal{A}(\Gamma)$ we have $\Lambda = \bigcup_{i \in N_\Lambda} \Lambda_i(a_i)$, where N_Λ is a finite set and the $\Lambda_i(a_i)$

are subtrees rooted at a_i. If $a_i \in Q_{k_i}$, then from [19, Sect. 6] it follows that

$$J(\Lambda_i(a_i)) \asymp \sup_{k_i \leqslant s < \infty} J_{k_i,s},$$

where, as $s \to \infty$,

$$J_{k,s} := \left(\sum_{m=s}^{\infty} 2^{m-s} \alpha_m \beta_m \right)^{1/p} \left(\sum_{m=k}^{s} \beta_m \alpha_m^{-p'/p} \right)^{1/p'}$$

$$\asymp \begin{cases} c^s & \text{if } \varkappa > p/p', \\ c^{k+(1+\varkappa)(s-k)/p} & \text{if } \varkappa < p/p', \\ (s-k)^{1/p'} c^{(1+\varkappa)s/p} & \text{if } \varkappa = p/p'. \end{cases}$$

Consequently, $J(\Lambda_i(a_i)) \to 0$ as $k_i \to \infty$ and $\lim\limits_{\mathcal{A}(\Gamma)} J(\Lambda) = 0$. Hence E is compact. Note that, when $\varkappa = 1$, Ω can be shown to be a quasidisc (see [28, Sect. 1.5.1, Example 1]) and hence has the W_p^1-extension property, which in turn implies that E is compact. However, if $\varkappa > 1$, then $\beta_m/\alpha_m \to \infty$ as $m \to \infty$ and so Ω is not a quasi-disc.

In [19, Sect. 6.2], the distribution of the approximation numbers (see the next section for details of these) of E is also investigated, and in the case $p = 2$, the precise growth rate of the error term in the spectral asymptotic formula for the Neumann Laplacian on Ω is given. To state the last result, we first need to define the following terms.

Let

$$(\partial\Omega)^i_\delta := \{\mathbf{x} \in \Omega : \operatorname{dist}(\mathbf{x}, \partial\Omega) < \delta\},$$

$$(\partial\Omega)^o_\delta := \{\mathbf{x} \in \mathbb{R}^n \setminus \Omega : \operatorname{dist}(\mathbf{x}, \partial\Omega) < \delta\},$$

$$\mathcal{M}^i_d(\partial\Omega) := \limsup_{\delta \to 0} \delta^{-(2-d)} |(\partial\Omega)^i_\delta|,$$

$$d_i := \inf\{t : \mathcal{M}^i_t(\partial\Omega) < \infty\},$$

and define $\mathcal{M}^o_d(\partial\Omega)$, d_o similarly. Then $\mathcal{M}^i_d(\partial\Omega), \mathcal{M}^o_d(\partial\Omega)$ are the upper Minkowski contents of $\partial\Omega$ relative to Ω and $\mathbb{R}^n \setminus \Omega$ respectively, and d_i, d_o are respectively the inner and outer Minkowski dimensions of $\partial\Omega$. By [19, Sect. 6.2.2],

$$d_i = \begin{cases} 1 & \text{if } c \leqslant 1/2, \\ 1 + \log(2c)/\varkappa \log(1/c) & \text{if } c > 1/2, \end{cases}$$

and

$$d_o = \begin{cases} 1 & \text{if } c \leqslant 1/2, \\ 1 + \log(2c)/\log(1/c) & \text{if } c > 1/2, \end{cases}$$

The upper Minkowski contents $\mathcal{M}_d^i(\partial\Omega)$ and $\mathcal{M}_d^o(\partial\Omega)$ are both infinite when $c = 1/2$ and finite otherwise.

Denoting the number of eigenvalues of the Neumann Laplacian $-\Delta_{\Omega,N}$ on Ω which are less than λ by $\mathcal{N}(\lambda; -\Delta_{\Omega,N})$, and with a similar notation for the Dirichlet Laplacian $-\Delta_{\Omega,D}$, it is proved in [19, Theorems 6.3 and 6.4] that

$$\mathcal{N}(\lambda; -\Delta_{\Omega,N}) - (1/4\pi)|\Omega|\lambda \begin{cases} = O(\lambda^{1/2}) & \text{if } c < 1/2, \\ = O(\lambda^{1/2}\log\lambda) & \text{if } c = 1/2, \\ \asymp \lambda^{d_o/2} & \text{if } c > 1/2, \end{cases}$$

and

$$\mathcal{N}(\lambda; -\Delta_{\Omega,D}) - (1/4\pi)|\Omega|\lambda = \begin{cases} O(\lambda^{1/2}) & \text{if } c < 1/2, \\ O(\lambda^{1/2}\log\lambda) & \text{if } c = 1/2, \\ O(\lambda^{d_i/2}) & \text{if } c > 1/2. \end{cases}$$

Note that $\partial\Omega$ is fractal when $c > 1/2$ in the sense that the inner and outer Minkowski dimensions lie in $(1, 2)$. It is particularly interesting that the precise growth rate of the Neumann error term is obtained in this case.

5 Approximation and Other s-Numbers of Hardy Type Operators

The Hardy type maps T we consider in this section act between Lebesgue spaces on an interval $I = (a, b)$, where b may be infinite, and are of the form

$$Tf(x) = v(x) \int_a^x u(t) f(t) dt, \tag{5.1}$$

u and v being prescribed functions. Our restriction to intervals I rather than trees is made to simplify the exposition: many of the results to follow have natural analogues for trees. Throughout this section, we assume that $-\infty < a < b \leqslant \infty$ and $p, q \in [1, \infty]$, while u and v are given real-valued functions such that for all $X \in I$,

$$u \in L_{p'}(a, X) \tag{5.2}$$

and

$$v \in L_q(X, b), \tag{5.3}$$

where $1/p' = 1 - 1/p$ (cf. (2.2). Criteria for the boundedness of T as a map from $L_p(I)$ to $L_q(I)$ were given earlier (see Remark 2.1 and Theorem 2.2).

Turning now to characterizations of the compactness of T, we see that there is an even more remarkable difference between the cases $p \leqslant q$ and $p > q$ than for boundedness. For the first of these we have the following result, the proof of which is given in [29, Theorems 7.3 and 7.5] and [9, Theorem 2.3.1].

Theorem 5.1. *Let* $1 \leqslant p \leqslant q < \infty$ *or* $1 < p \leqslant q = \infty$*. Suppose that* (5.2) *and* (5.3) *hold. Put*

$$A(c,d) = \sup_{c<X<d} \left\{ \|u\|_{p',(c,X)} \|v\|_{q,(X,d)} \right\}, \quad a \leqslant c < d \leqslant b. \tag{5.4}$$

Then if T is a bounded linear map from $L_p(I)$ to $L_q(I)$, T is compact if and only if

$$\lim_{c \to a+} A(a,c) = \lim_{d \to b-} A(d,b) = 0. \tag{5.5}$$

Remark 5.1. It will be seen that this result does not cover the case in which $p = 1$ and $q = \infty$. In fact, $T : L_1(I) \to L_\infty(I)$ is never compact. For this, see Remark (b) after Theorem 4 in [12].

When $p > q$, the striking result given next, proved in [29, Theorem 7.5] (see also [9, Theorem 2.3.4]) asserts that boundedness of T is equivalent to compactness.

Theorem 5.2. *Let* $1 \leqslant q < p \leqslant \infty$*. Suppose that* (5.2) *and* (5.3) *hold. Then T is a compact map from $L_p(a,b)$ to $L_q(a,b)$ if and only if it is bounded.*

Note that if $u \in L_{p'}(I)$ and $v \in L_q(I)$, then $T : L_p(I) \to L_q(I)$ is compact for all $p, q \in [1, \infty]$ with $(p,q) \neq (1, \infty)$.

Given a compact map acting between two Banach spaces, it is often desirable to have a quantitative means of assessing "how compact" it is, and one way of doing this is by use of what are called s–numbers. The definition that we give below applies to all bounded linear maps, not merely to compact ones, and is of use in this wider context. We denote by $B(X,Y)$ the family of all bounded linear maps from X to Y, abbreviating this to $B(X)$ if $X = Y$. Let B_X stand for the closed unit ball in X. We write $\|\cdot\|_X$ for the norm in X, omitting the subscript if no ambiguity is likely.

Definition 5.1. A map s which to each bounded linear map S from one Banach space to another such space assigns a sequence $(s_n(S))$ of nonnegative real numbers is called an *s–function* if, for all Banach spaces W, X, Y and Z, it has the following properties:

(i) $\|S\| = s_1(S) \geqslant s_2(S) \geqslant \ldots \geqslant 0$ for all $S \in B(X,Y)$;

(ii) for all $S_1, S_2 \in B(X,Y)$ and all $n \in \mathbb{N}$,

$$s_n(S_1 + S_2) \leqslant s_n(S_1) + \|S_2\|;$$

(iii) for all $S \in B(X,Y)$, $R \in B(Y,Z)$ and $U \in B(Z,W)$ and $n \in \mathbb{N}$,

$$s_n(URS) \leqslant \|U\|\, s_n(R)\, \|S\|\,;$$

(iv) for all $S \in B(X,Y)$ with rank $S < n \in \mathbb{N}$,

$$s_n(S) = 0;$$

(v) $s_n(I_n) = 1$ for all $n \in \mathbb{N}$; here I_n is the identity map of $l_2^n := \{x \in l_2 : x_j = 0 \text{ if } j > n\}$ to itself.

For all $n \in \mathbb{N}$, $s_n(S)$ is called the n^{th} *s–number* of S. An s–function is called *additive* if for all $m, n \in \mathbb{N}$ and all $S_1, S_2 \in B(X,Y)$, where X and Y are arbitrary Banach spaces,

$$s_{m+n-1}(S_1 + S_2) \leqslant s_m(S_1) + s_n(S_2);$$

it is said to be *multiplicative* if for all $m, n \in \mathbb{N}$ and all $S \in B(X,Y)$ and $R \in B(Y,Z)$, where X, Y and Z are arbitrary Banach spaces,

$$s_{m+n-1}(RS) \leqslant s_m(R)s_n(S).$$

All s–numbers coincide for operators acting between Hilbert spaces. Some of the most widely used s–numbers are the following:

(i) the *approximation numbers* $a_n(S)$ given by

$$a_n(S) = \inf\left\{\|S - F\| : F \in B(X,Y),\ \text{rank } F < n\right\};$$

(ii) the *Kolmogorov numbers* $d_n(S)$ defined by

$$d_n(S) = \inf\left\{\left\|Q_M^Y S\right\| : M \text{ is a linear subspace of } Y,\ \dim M < n\right\},$$

where Q_M^Y is the canonical map of Y onto Y/M;

(iii) the *Gelfand numbers* $c_n(S)$, where

$$c_n(S) = \inf\left\{\left\|S J_M^X\right\| : M \text{ is a linear subspace of } X,\ \text{codim } M < n\right\},$$

where J_M^X is the embedding map from M to X;

(iv) the *Bernstein numbers* $b_n(S)$ defined by

$$b_n(S) = \sup\left\{\inf_{x \in X_n \setminus \{0\}} \frac{\|Sx\|_Y}{\|x\|_X} \right.$$

$$\left. :\ X_n \text{ is an } n\text{-dimensional subspace of } X\right\}.$$

The numbers in (i), (ii), and (iii) form additive and multiplicative s–functions; the approximation numbers are the largest s–numbers and satisfy the inequalities $a_n(S^*) \leqslant a_n(S) \leqslant 5a_n(S^*)$ with $a_n(S) = a_n(S^*)$ if S is compact; and the Gelfand and Kolmogorov numbers are related by

$$c_n(S) = d_n(S^*),\ d_n(S) \geqslant c_n(S^*),$$

with equality if S is compact.

In addition to the s–numbers, there are the *entropy* numbers, which play a most useful rôle in connection with the compactness properties of a map. Given any $n \in \mathbb{N}$, the n^{th} *entropy number* of a map $S \in B(X,Y)$ is defined by

$$\begin{aligned} e_n(S) = \inf\{&\varepsilon > 0 \\ &: \ S(B_X) \text{ can be covered by } 2^{n-1} \text{ balls in } Y \text{ of radius } \ \varepsilon\}. \end{aligned}$$

These numbers are monotonic decreasing as n increases, and also have the additive and multiplicative properties mentioned above; but they are not s–numbers as they do not have property (iv) required of such numbers. Indeed, if X is real with dim $X = m < \infty$ and $I : X \to X$ is the identity map, then for all $n \in \mathbb{N}$,

$$1 \leqslant 2^{(n-1)/m} e_n(I) \leqslant 4.$$

Since the $e_n(S)$ are nonnegative and have the monotonicity property, $\beta(S) := \lim\limits_{n\to\infty} e_n(S)$ exists and is the so-called (ball) *measure of noncompactness* of S. The terminology is justified since $\beta(S) = 0$ if and only if S is compact. Of course, every s–number also has the property that $\lim\limits_{n\to\infty} s_n(S)$ exists, but a similar characterization of compactness is not possible: for example, $\alpha(S) := \lim\limits_{n\to\infty} a_n(S)$ may well be positive even though S is compact. The difficulty is that for some spaces X and Y there is a compact map S from X to Y that cannot be approximated arbitrarily closely in the norm sense by finite-dimensional linear maps. However, if the target space Y is an L_p space with $1 \leqslant p < \infty$, this difficulty disappears and $\alpha(S) = 0$ if and only if S is compact. In fact, we then have $\alpha(S) = \beta(S)$ for all $S \in B(X, L_p)$. The linkage with the measure of noncompactness defined in the Introduction is thus complete. Further details and proofs of the assertions made above can be found in [8] and [30].

Given a compact map S with $\alpha(S) = 0$, the speed at which the $a_n(S)$ approach zero as $n \to \infty$ is of obvious interest from the point of view of approximation theory. The same is true for the entropy numbers, with the additional spectral connection brought about by Carl's inequality. To explain this, recall that if S is a compact linear map from X to itself, then its spectrum, apart from the point 0, consists solely of eigenvalues of finite algebraic multiplicity: let $(\lambda_n(S))$ be the sequence of all nonzero eigenvalues of S, repeated according to their algebraic multiplicity and ordered by decreasing modulus. If S has only m ($< \infty$) distinct eigenvalues and M is the sum of their multiplicities, put $\lambda_n(S) = 0$ for all $n \in \mathbb{N}$ with $n > M$. Carl's inequality ([6]; see also [8]) asserts that for all $n \in \mathbb{N}$,

$$|\lambda_n(S)| \leqslant \sqrt{2} e_{n+1}(S).$$

Together with the multiplicative properties of the entropy numbers, this striking result enables information about the behavior of eigenvalues of (possibly degenerate) elliptic operators to be obtained provided that sharp two-sided estimates are available for the entropy numbers of embedding maps between function spaces. A very considerable amount of work has been done on this topic, leading to the availability of these sharp estimates for embeddings between spaces of Sobolev, Besov and Lizorkin–Triebel type, for example. We refer to [16] for further information concerning these estimates and their applications.

We now return to the map $T : L_p(I) \to L_q(I)$ given by (5.1), namely

$$Tf(x) = v(x) \int_a^x u(t) f(t) dt,$$

where $I = (a, b)$ and $p, q \in [1, \infty]$. For simplicity of exposition, we assume henceforth that $b < \infty$ and u, v are positive functions with

$$u \in L_{p'}(I), \quad v \in L_q(I). \tag{5.6}$$

Under these conditions on u and v, it follows from (2.6) and Theorem 2.2 that T is bounded; moreover, by Theorems 5.1 and 5.2, T is compact provided that

$$p, q \in [1, \infty] \text{ with } (p, q) \neq (1, \infty). \tag{5.7}$$

Our first objective is to obtain upper and lower bounds for the approximation numbers of T, and with this in mind we introduce a quantity $\mathbb{A}(J) = \mathbb{A}(J, u, v)$ defined, for any interval $J \subset I$ and any $c \in (a, b]$, by

$$\mathbb{A}(J) = \sup_{f \in L_p(J)\setminus\{0\}} \inf_{\alpha \in \mathbb{C}} \|T_{c,J} f - \alpha v\|_{q,J} / \|f\|_{p,J}, \tag{5.8}$$

where

$$T_{c,J} f(x) := v(x) \chi_J(x) \int_c^x f(t) u(t) \chi_J(t) dt$$

and

$$\mu(J) = \begin{cases} \displaystyle\int_J |v(t)|^q \, dt, & 1 \leqslant q < \infty, \\ \sup_J |v(t)|, & q = \infty. \end{cases}$$

It is clear that the boundedness of T implies that $\mathbb{A}(I) < \infty$. Note also that, as is easily shown, $\mathbb{A}(J)$ is independent of the particular choice of c. We summarize in the following proposition some of the basic properties of $\mathbb{A}(J)$, referring to [9] for their proofs.

Proposition 5.1. (i) *When u and v are constant over an interval J with length $|J|$,*

$$\mathbb{A}(J,u,v) = \frac{1}{2}uv\gamma_{pq}, \tag{5.9}$$

where

$$\gamma_{pq} = r^{1/r-1}(p')^{1/p}q^{1/q'}/\mathbf{B}(1/p',1/q),\ 1/r = 1+1/q-1/p, \tag{5.10}$$

and $\mathbf{B}$ is the beta function, given in terms of the gamma function Γ by $\mathbf{B}(s,t) = \Gamma(s)\Gamma(t)/\Gamma(s+t)$.

(ii) *If $u_1, u_2 \in L_{p'}(J)$ and $v_1, v_2 \in L_q(J)$, then*

$$|\mathbb{A}(J,u_1,v) - \mathbb{A}(J,u_2,v)| \leqslant \|v\|_{q,J}\,\|u_1-u_2\|_{p',J} \tag{5.11}$$

and

$$|\mathbb{A}(J,u,v_1) - \mathbb{A}(J,u,v_2)| \leqslant 2\,\|v_1-v_2\|_{q,J}\,\|u\|_{p',J}\,. \tag{5.12}$$

(iii) *If K_1, K_2 are subintervals of I with $K_1 \subset K_2$, then*

$$|\mathbb{A}(K_1,u,v) - \mathbb{A}(K_2,u,v)| \to 0 \ \text{as}\ |K_2 \backslash K_1| \to 0.$$

The constant γ_{pq} determines the norm of the particular form of the operator T obtained when the functions u and v are identically 1. In fact, denoting by H this particular operator, so that

$$Hf(x) = \int\limits_a^x f(t)dt \quad (f \in L_p(I), x \in I),$$

it turns out that

$$\|H : L_p(I) \to L_q(I)\| = (b-a)^{1/p'+1/q}\gamma_{pq}. \tag{5.13}$$

The special case of this when $p = q$ is denoted by γ_p. Thus,

$$\gamma_p = \pi^{-1}p^{1/p'}(p')^{1/p}\sin(\pi/p). \tag{5.14}$$

Now suppose that $1 < p \leqslant q < \infty$ and for any $\varepsilon > 0$ define

$$N(I,\varepsilon) = N(I,\varepsilon,u,v) = \min\left\{ n : I = \bigcup_{j=1}^{n} I_j, \mathbb{A}(I_j) \leqslant \varepsilon \right\}, \tag{5.15}$$

where the I_j are nonoverlapping subintervals of I. Note that since T is compact, $N(I, \varepsilon) < \infty$. The properties of $\mathbb{A}$ guaranteed by Proposition 5.1 ensure that given any small enough $\varepsilon > 0$, there are intervals $I_j = (c_j, c_{j+1})$ $(j = 1, \dots, N)$ such that $a = c_1 < c_2 < \dots < c_{N+1} = b$ and

$$\mathbb{A}(I_j) = \varepsilon \ (j = 1, \dots, N-1), \ \mathbb{A}(I_N) \leqslant \varepsilon. \tag{5.16}$$

The decomposition of the interval I given by the I_j enables us to construct a finite-dimensional linear approximation to T that leads to the inequality

$$a_{N+1}(T) \leqslant \varepsilon, \text{ where } N = N(I, \varepsilon). \tag{5.17}$$

Another useful quantity is $M(I, \varepsilon) = M(I, \varepsilon, u, v)$, defined to be the maximum $m \in \mathbb{N}$ such that there are nonoverlapping intervals $J_j \subset I \quad (j = 1, 2, \dots, m)$ with $\mathbb{A}(J_j) \geqslant \varepsilon$ for each j. Evidently,

$$M(I, \varepsilon) \geqslant N(I, \varepsilon) - 1. \tag{5.18}$$

In fact, if $J_1, \dots, J_M$ are nonoverlapping subintervals of I such that $\mathbb{A}(J_j) \geqslant \varepsilon$ for each j, then use of the definition of $\mathbb{A}(J_j)$ quickly leads to an estimate from below of the norm of the difference between T and an arbitrary linear map of rank $< M$, and hence to the lower bound

$$a_M(T) \geqslant \varepsilon M^{1/q - 1/p}. \tag{5.19}$$

Combination of (5.17)-(5.19) now gives

Theorem 5.3. *Let $1 < p \leqslant q < \infty$. Then given $\varepsilon > 0$,*

$$a_{N+1}(T) \leqslant \varepsilon \text{ and } a_M(T) \geqslant M^{1/q-1/p}\varepsilon, \tag{5.20}$$

where $N = N(I, \varepsilon)$ and $M \geqslant N - 1$.

Estimates similar to these were first derived in [10]. The corresponding estimates can be obtained when $p = 1$ or $q = \infty$, and also for the case $1 \leqslant q < p \leqslant \infty$; we summarise the position as follows, referring to [21] and [1] for details.

Theorem 5.4. (i) *Let $1 \leqslant p \leqslant q \leqslant \infty$, and let $N = N(I, \varepsilon)$. Then*

$$a_{N+1}(T) \leqslant \varkappa_q \varepsilon \text{ and } a_{[N/2]-1}(T) \geqslant \varepsilon, \tag{5.21}$$

where $\varkappa_2 = 1$ and $\varkappa_q = 2$ if $q \neq 2$.

(ii) *Suppose that $1 \leqslant q < p \leqslant \infty$ and put $N = N(I, \varepsilon)$. Then*

$$a_{N+2}(T) \leqslant (N+1)^{1/q-1/p}\varepsilon \text{ and } a_{N-1}(T) \geqslant \nu_q \varepsilon. \tag{5.22}$$

where $\nu_2 = 1$ and $\nu_q = 1/2$ if $q \neq 2$.

The next task is to sharpen these results and obtain more precisely the dependence of $a_n(T)$ upon n. It turns out that striking results can be obtained when $p = q \in (1, \infty)$, for then asymptotic estimates are available. In fact, we have

Theorem 5.5. *Suppose that* $1 < p < \infty$. *Then the approximation numbers of the map* $T : L_p(I) \to L_p(I)$ *satisfy*

$$\lim_{n\to\infty} n a_n(T) = \frac{1}{2}\gamma_p \int_a^b u(t)v(t)dt. \tag{5.23}$$

The idea of the proof is to show that

$$\lim_{\varepsilon\to 0+} \varepsilon N(I, \varepsilon) = \frac{1}{2}\gamma_p \int_a^b u(t)v(t)dt,$$

and then to use Theorem 5.3. The special case of this when $p = 2$ was first obtained in [11]. For an arbitrary $p \in (1, \infty)$ it was established in [22] as a particular instance of the corresponding result in which the interval I is replaced by a tree. Equation (5.23) also holds with the approximation numbers $a_n(T)$ replaced by the Gelfand numbers $c_n(T)$, the Kolmogorov numbers $d_n(T)$ and the Bernstein numbers $b_n(T)$ (see [24]). Appropriate extra conditions on u and v lead to the next result, which sharpens Theorem 5.5 by obtaining remainder estimates.

Theorem 5.6. *Let* $1 < p < \infty$. *In addition to the standing assumptions that* $u \in L_{p'}(I)$ *and* $v \in L_p(I)$, *suppose that* u, v *have continuous derivatives* u', v' *respectively that satisfy the conditions* $u' \in L_{p'/(p'+1)}(I)$, $v' \in L_{p/(p+1)}(I)$. *Let* $\rho_n(T)$ *stand for* $a_n(T)$, $b_n(T)$, $c_n(T)$ *or* $d_n(T)$. *Then*

$$\limsup_{n\to\infty} n^{1/2} \left| n\rho_n(T) - \frac{1}{2}\gamma_p \int_a^b u(t)v(t)dt \right| \leqslant \frac{3}{2}\gamma_p \int_a^b u(t)v(t)dt + c(p)R, \tag{5.24}$$

where

$$R = \left(\|u'\|_{p'/(p'+1),I} + \|v'\|_{p/(p+1),I}\right)\left(\|u\|_{p',I} + \|v\|_{p,I}\right)$$

and $c(p)$ *is a constant that depends only on* p.

For this we refer to [24], which refines the techniques used in [13] for the case $p = 2$, $\rho_n(T) = a_n(T)$. In fact, in [24] it is also shown that if u and v satisfy the standard conditions $u \in L_{p'}(I)$, $v \in L_p(I)$, together with $v'/v, u'/u \in L_1(I) \cap C(\overline{I})$, then

$$\limsup_{n\to\infty} n \left| n\rho_n(T) - \frac{1}{2}\gamma_p \int_a^b u(t)v(t)dt \right|$$
$$\leqslant \int_a^b u(t)v(t) \left\{ \|u'/u\|_{1,I} + \|v'/v\|_{1,I} + 2\gamma_p + \|u'/u\|_{1,I} \|v'/v\|_{1,I} \right\} dt, \tag{5.25}$$

from which we see that

$$\rho_n(T) = \frac{\gamma_p}{2n} \int_a^b u(t)v(t)dt + O(n^{-2}). \tag{5.26}$$

These results show that when $p = q$, the state of knowledge of various s–numbers of T is reasonably satisfactory. Greater difficulties are presented when $p \neq q$, but substantial progress has been made, as we now indicate. In [26], it is shown that, under appropriate conditions on u and v, the approximation numbers of T satisfy, for all $n \in \mathbb{N}$ and some positive constants c_1 and c_2 (independent of n),

$$c_1 \|uv\|_{r,I} \leqslant \liminf_{n\to\infty} n^\lambda a_n(T) \leqslant \limsup_{n\to\infty} n^\lambda a_n(T) \leqslant c_2 \|uv\|_{r,I}, \tag{5.27}$$

where $1/r = 1 - 1/p + 1/q > 0$, $\lambda = \min\{1, 1/r\}$, and $p > 1$, with either $1 \leqslant q \leqslant p \leqslant \infty$, $1 \leqslant p \leqslant q \leqslant 2$ or $2 \leqslant p \leqslant q \leqslant \infty$. This was established by means of techniques quite different from those sketched above; for similar results derived by procedures closer to those of this paper we refer to [25]. The paper [26] also gives estimates for the entropy numbers of T, namely

$$c_1 \|uv\|_{r,I} \leqslant \liminf_{n\to\infty} ne_n(T) \leqslant \limsup_{n\to\infty} ne_n(T) \leqslant c_2 \|uv\|_{r,I}, \tag{5.28}$$

valid for all $p, q \in [1, \infty]$ with $1/r = 1-1/p+1/q > 0$, under certain conditions on u and v.

The results just mentioned do not give asymptotic formulas for the various numbers; these are not to be expected in the case of the entropy numbers. Some advances have been made by way of sealing this gap and we now describe briefly recent work leading to genuine asymptotic results for the Bernstein numbers, when $p \leqslant q$, and for the approximation and Kolmogorov numbers when $q \leqslant p$.

Theorem 5.7. *Let $1 \leqslant p \leqslant q < \infty$, and let $1/r = 1/q + 1/p'$. Then*

$$\lim_{n\to\infty} nb_n(T) = C \left(\int_a^b (u(t)v(t))^r \, dt \right)^{1/r}, \tag{5.29}$$

where

$$C = \frac{1}{2}(p')^{1/q} q^{1/p'} (p'+q)^{1/p-1/q} / \mathbf{B}(1/q, 1/p').$$

This is proved in [15]. To indicate some of the main features of the proof, we begin with certain generalizations of the trigonometric functions. For $\sigma \in [0, q/2]$ we write

$$F_{p,q}(\sigma) = \frac{q}{2} \int_0^{2\sigma/q} (1-s^q)^{-1/p} ds, \ \pi_{p,q} = 2F_{p,q}(q/2) = \mathbf{B}(1/q, 1/p')$$

and denote by $\sin_{p,q}$ the inverse of the strictly increasing function $F_{p,q}$ on $[0, \pi_{p,q}/2]$. We continue to denote by $\sin_{p,q}$ the extension of this inverse function to $\mathbb{R}$ by evenness about $\pi_{p,q}/2$ to $[0, \pi_{p,q}]$, then by oddness to $[-\pi_{p,q}, \pi_{p,q}]$, and finally to $\mathbb{R}$ by $2\pi_{p,q}$–periodicity. With B representing the closed unit ball in $L_p(I)$, the problem of determining

$$\sup_{g \in T(B)} \|g\|_{q,I}$$

leads to the nonlinear integral problem

$$g(x) = (Tf)(x), \ \varphi_p(f)(x) = \lambda \left(T^* \varphi_q(g)\right)(x), \tag{5.30}$$

where $\varphi_r(h) = |h|^{r-2} h$ $(h \neq 0)$, $\varphi_r(0) = 0$, and T^* is the map defined by

$$(T^* f)(x) = u(x) \int_x^b v(t) f(t) dt;$$

λ is to be thought of as an eigenvalue parameter. When u and v are both identically equal to 1 on I, this integral problem may be transformed into the p, q–Laplacian equation

$$-\left(\varphi_p(w')\right)' = \lambda \varphi_q(w), \tag{5.31}$$

with the boundary condition $w(a) = 0$. What emerges is that given any $\alpha \in \mathbb{R} \backslash \{0\}$, the set of all eigenvalues λ of (5.31) under the conditions

$$w(a) = 0, \quad w'(a) = \alpha, \tag{5.32}$$

is given by (see [7])

$$\lambda_n(\alpha) = \left(\frac{2(n-1/2)\pi_{p,q}}{b-a}\right)^q \cdot \frac{|\alpha|^{p-q}}{p' q^{q-1}} \qquad (n \in \mathbb{N}), \tag{5.33}$$

with the corresponding eigenfunctions

$$w_{n,\alpha}(t) = \frac{\alpha(b-a)}{(n-1/2)\pi_{p,q}} \sin_{p,q}\left(\frac{(n-1/2)\pi_{p,q}}{b-a}t\right) \quad (t \in I). \tag{5.34}$$

In addition, when u and v are both identically equal to 1 on I, a result of ([4]; see also [5]) gives a connection between the Bernstein numbers and these eigenvalues from which the Bernstein numbers $b_n(T)$ can be determined precisely and are given by

$$b_n(T) = \frac{\alpha(b-a)}{(n-1/2)\pi_{p,q}} \left(\frac{p' q^{q-1}}{|\alpha|^{p-q}}\right)^{1/q} \tag{5.35}$$

where α is so chosen that

$$\left\| \frac{\alpha(b-a)}{(n-1/2)\pi_{p,q}} \left(\sin_{p,q}\left(\frac{(n-1/2)\pi_{p,q}}{b-a}t\right)\right)' \right\|_{p,I} = 1.$$

The procedure from this point onwards resembles that outlined for the approximation numbers when $p = q$ in that the interval I is cut up into subintervals to facilitate calculation, but this time, instead of the function $\mathbb{A}$, we now require two functions, $\mathbb{C}_0$ and $\mathbb{C}_+$, defined for all subintervals $J = [c, d]$ of I by

$$\mathbb{C}_0(J) = \sup\{\|Tf\|_{q,J} \,/\, \|f\|_{p,J} : f \in L_p(J)\backslash\{0\}, (Tf)(c) = (Tf)(d) = 0\}$$

and

$$\mathbb{C}_+(J) = \sup\{\|Tf\|_{q,J} \,/\, \|f\|_{p,J} : f \in L_p(J)\backslash\{0\}, (Tf)(c) = 0\}.$$

For the approximation and Kolmogorov numbers we have the following assertion (see [14]).

Theorem 5.8. *Suppose that $1 < q \leqslant p < \infty$. Let $r = 1/q + 1/p'$, and let $\rho_n(T)$ stand for either $a_n(T)$ or $d_n(T)$. Then*

$$\lim_{n\to\infty} n\rho_n(T) = C\left(\int_a^b (u(t)v(t))^{1/r}\, dt\right)^r, \tag{5.36}$$

where C is as in the previous theorem.

So far as the estimates of the approximation numbers (from above and below) and those of the Kolmogorov numbers from above are concerned, the pattern of the proof is a natural adaptation of that used for the Bernstein numbers. However, to estimate the Kolmogorov numbers from below, we use the following result due to Makavoz (see [5, Sect. 3.11]):

Let $U_n \subset \{Tf : \|f\|_{p,I} \leqslant 1\}$ be a continuous and odd image of the unit sphere S^n in $\mathbb{R}^{n+1}$ endowed with the l_1–norm. Then

$$d_n(T) \geqslant \inf\{\|x\|_{q,I} : x \in U_n\}.$$

This enables it to be shown that

$$\liminf_{n\to\infty} nd_n(T) \geqslant \left(\int_a^b |u(t)v(t)|^{1/r}\, dt\right)^r.$$

6 Approximation Numbers of Embeddings on Generalized Ridged Domains

In [9, Chapt. 6], relationships between the approximation numbers of E_M for a GRD Ω and those of the associated Hardy type operator T_a were derived: the approximation numbers of other embedding maps for related spaces on Ω and Γ were also investigated. Of particular interest for the present discussion are the following results. A consequence of Lemma 6.2.3, Lemma 6.2.4 and Theorem 6.4.1 is

$$a_m(E_M) \leqslant K\{k(\Omega) + a_m(I_M)\}, \tag{6.1}$$

where $k(\Omega) = \sup_{x\in\Omega}(\rho o \tau)(x)$ and I_M is the embedding $L^1_{M,p}(\Gamma; d\mu) \to L_{M,p}(\Gamma; d\mu)$, where

$$L^1_{M,p}(\Gamma; d\mu) := \{F \in L^1_p(\Gamma, d\mu), F_\Gamma = 0\}$$

with norm $\|F\|_{M,p;d\mu} = \|F'\|_{p,\Gamma;d\mu}$ and

$$L_{M,p}(\Gamma; d\mu) := \{F \in L_p(\Gamma; d\mu), F_\Gamma = 0\}.$$

Furthermore, from Lemma 6.2.4 and Theorem 6.4.3 in [9] it follows that

$$(1/2)a_{m+1}(T_a) \leqslant a_m(I_M) \leqslant 2a_m(T_a). \tag{6.2}$$

For GRDs, upper and lower bounds for the approximation numbers $a_k(E)$ of the embedding $E : W^1_p(\Omega) \to L_p(\Omega), 1 < p < \infty$, were also obtained in [18, Theorem 5.1]. On applying the results to Examples 4.1 and 4.2, the outcomes are as follows:

Example 6.1 (horn-shaped domain). For this domain Ω, the embedding E is compact in the case

$$\Phi^{n-1}(t) = e^{-t^\theta}, \quad \theta > 1.$$

In [18, Example 6.1], it is shown that

$$a_k(E) \asymp k^{-1/\nu}, \quad \nu = \max\{\frac{\theta}{\theta-1}, n\}. \tag{6.3}$$

Example 6.2 (Rooms and Passages). In the case $\delta_{2j} = ch_{2j}^{\varkappa}$, $h_{2j} = C^{-j}$, $1 \leqslant \varkappa < p+1$, $C > 1$, the embedding E is compact and from [18, Example 6.2] it follows that

$$a_k(E) \asymp k^{-1/2}. \tag{6.4}$$

In [19], a partial analogue of the Dirichlet–Neumann bracketing technique, which is so effective in determining the asymptotic limit of the eigenvalue distribution functions in the case $p = 2$, was developed for estimating the approximation numbers $a_k(E)$, and, in particular, for the function

$$\nu_M(\varepsilon, \Omega) := \max\{k : a_k(E_M) \geqslant \varepsilon\}$$

in $L_{M,p}(\Omega)$ (see also [9, Chapt. 6]). In [23], this technique is applied to a wide class of domains, including GRDs which are not unduly pathological, and the results illustrated by a detailed analysis of the horn-shaped domain and also of the snowflake type domain of Example 4.3 (see also [9, Example 6.4.4]).

References

1. Achache, M.: Some Problems Associated with Linear Differential Operators. M. Sc. thesis, University of Wales (1988)
2. Adams, R.A.: Sobolev Spaces. Academic Press, New York (1975)
3. Amick, C.J.: Some remarks on Rellich's theorem and the Poincaré inequality. J. London Math. Soc. **18**, 81–93 (1978)
4. Buslaev, A.P.: On Bernstein-Nikol'skij inequalities and widths of Sobolev function classes (Russian). Dokl. Akad. Nauk, Ross. Akad. Nauk **323**, no.2, 202–205 (1992); English transl.: Russ. Acad. Sci., Dokl., Math. **45**, no.2, 272–275 (1992)
5. Buslaev, A.P., Tikhomirov, V.M.: Spectra of nonlinear differential equations and widths of Sobolev classes (Russian). Mat. Sb. **181**, 1587–1606 (1990); English transl.: Math. USSR Sb. **71**, 427–446 (1992)
6. Carl, B.: Entropy numbers, s–numbers and eigenvalue problems. J. Funct. Anal. **41**, 290–306 (1981)
7. Drábek, P., Manásevich, R.: On the solution to some p-Laplacian nonhomogeneous eigenvalue problems. Differ. Integral Equ. **12**, 723–740 (1999)
8. Edmunds, D.E., Evans, W.D.: Spectral Theory and Differential Operators. Oxford Univ. Press, Oxford (1987)
9. Edmunds, D.E., Evans, W.D.: Hardy Operators, Function Spaces and Embeddings. Springer, Berlin-Heidelberg-New York (2004)
10. Edmunds, D.E., Evans, W.D., Harris, D.J.: Approximation numbers of certain Volterra integral operators. J. London Math. Soc. **37**, 471–489 (1988)
11. Edmunds, D.E., Evans, W.D., Harris, D.J.: Two-sided estimates of the approximation numbers of certain Volterra integral operators. Stud. Math. **124**, 59–80 (1997)
12. Edmunds, D.E., Gurka, P., Pick, L.: Compactness of Hardy type integral operators in weighted Banach function spaces. Stud. Math. **109**, 73–90 (1994)
13. Edmunds, D.E., Kerman, R., Lang, J.: Remainder estimates for the approximation numbers of weighted Hardy operators acting on L^2. J. d'Analyse Math. **85**, 225–243 (2001)

14. Edmunds, D.E., Lang, J.: Approximation numbers and Kolmogorov widths of Hardy-type operators in a non-homogeneous case. Math. Nachr. **279**, 727–742 (2006)
15. Edmunds, D.E., Lang, J.: Bernstein widths of Hardy-type operators in a non-homogeneous case. J. Math. Anal. Appl. **325**, 1060–1076 (2007)
16. Edmunds, D.E., Triebel, H.: Function Spaces, Entropy Numbers, Differential Operators. Cambridge Univ. Press, Cambridge (1996)
17. Evans, W.D., Harris, D.J.: Sobolev embeddings for generalized ridged domains. Proc. London Math. Soc. (3) **54**, 141–175 (1987)
18. Evans, W.D., Harris, D.J.: On the approximation numbers of Sobolev embeddings for irregular domains. Q. J. Math., Oxf. II. Ser. **40**, 13–42 (1989)
19. Evans, W.D., Harris, D.J.: Fractals, trees and the Neumann Laplacian. Math. Ann. **296**, 493–527 (1993)
20. Evans, W.D., Harris, D.J., Pick, L.: Weighted Hardy and Poincaré inequalities on trees. J. London Math. Soc. (2) **52**, 121–136 (1995)
21. Evans, W.D., Harris, D.J., Lang, J.: Two-sided estimates for the approximation numbers of Hardy-type operators in L^1 and L^∞. Stud. Math. **130**, 171–192 (1998)
22. Evans, W.D., Harris, D.J., Lang, J.: The approximation numbers of Hardy-type operators on trees. Proc. London Math. Soc. **83**, 390–418 (2001)
23. Evans, W.D., Harris, D.J., Saitō, Y.: On the approximation numbers of Sobolev embeddings on singular domains and trees. Q. J. Math., Oxf. II. Ser. **55**, 267–302 (2004)
24. Lang, J.: Improved estimates for the approximation numbers of Hardy-type operators. J. Approx. Theory. **121**, 61–70 (2006)
25. Lang, J., Mendez, O., Nekvinda, A.: Asymptotic behaviour of the approximation numbers of the Hardy-type operator from L^p to L^q. J. Inequal. Pure Appl. Math. **5**, 36–52 (2004)
26. Lifschits, M.A., Linde, W.: Approximation and entropy numbers of Volterra integral operators with applications to Brownian motion. Mem. Am. Math. Soc. no. 745 (2002).
27. Manakov, V.M.: On the best constant in weighted inequalities for Riemann–Liouville integrals. Bull. London Math. Soc. **24**, 442–448 (1992)
28. Maz'ya, V.G.: Sobolev Spaces. Springer-Verlag, Berlin–Tokyo (1985)
29. Opic, B., Kufner, A.: Hardy Type Inequalities. Longman Sci. Tech., Harlow (1990)
30. Pietsch. A.: s–numbers of operators in Banach spaces. Stud. Math. **51**, 201–223 (1974)
31. Read, T.T.: Generalisations of Hardy's Inequality. Preprint

Sobolev Mappings between Manifolds and Metric Spaces

Piotr Hajłasz

Abstract In connection with the theory of p-harmonic mappings, Eells and Lemaire raised a question about density of smooth mappings in the space of Sobolev mappings between manifolds. Recently Hang and Lin provided a complete solution to this problem. The theory of Sobolev mappings between manifolds has been extended to the case of Sobolev mappings with values into metric spaces. Finally analysis on metric spaces, the theory of Carnot–Carathéodory spaces, and the theory of quasiconformal mappings between metric spaces led to the theory of Sobolev mappings between metric spaces. The purpose of this paper is to provide a self-contained introduction to the theory of Sobolev spaces between manifolds and metric spaces. The paper also discusses new results of the author.

1 Introduction

For $\Omega \subset \mathbb{R}^n$ and $1 \leqslant p < \infty$ we denote by $W^{1,p}(\Omega)$ the usual Sobolev space of functions for which $\|u\|_{1,p} = \|u\|_p + \|\nabla u\|_p < \infty$. This definition can easily be extended to the case of Riemannian manifolds $W^{1,p}(M)$. Let now M and N be compact Riemannian manifolds. We can always assume that N is isometrically embedded in the Euclidean space $\mathbb{R}^\nu$ (Nash's theorem). We also assume that the manifold N has no boundary, while M may have boundary. This allows one to define the class of Sobolev mappings between the two manifolds as follows:

$$W^{1,p}(M,N) = \{u \in W^{1,p}(M,\mathbb{R}^\nu) |\ u(x) \in N \text{ a.e.}\}$$

Piotr Hajłasz
University of Pittsburgh, 301 Thackeray Hall, Pittsburgh, PA 15260 USA,
e-mail: `hajlasz@pitt.edu`

V. Maz'ya (ed.), *Sobolev Spaces in Mathematics I, International Mathematical Series.*
doi: 10.1007/978-0-387-85648-3,

$W^{1,p}(M,N)$ is equipped with the metric inherited from the norm $\varrho(u,v) = \|u-v\|_{1,p}$. The space $W^{1,p}(M,N)$ provides a natural setting for geometric variational problems like, for example, *weakly p-harmonic mappings* (called *weakly harmonic mappings* when $p=2$). Weakly p-harmonic mappings are stationary points of the functional

$$I(u) = \int_M |\nabla u|^p \quad \text{for } u \in W^{1,p}(M,N).$$

Because of the constrain in the image (manifold N) one has to clarify how the variation of this functional is defined. Let $\mathcal{U} \subset \mathbb{R}^\nu$ be a tubular neighborhood of N, and let $\pi : \mathcal{U} \to N$ be the smooth nearest point projection. For $\varphi \in C_0^\infty(M, \mathbb{R}^\nu)$, and $u \in W^{1,p}(M,N)$ the mapping $u + t\varphi$ takes on values into $\mathcal{U}$ provided that $|t|$ is sufficiently small. Then we say that u is *weakly p-harmonic* if

$$\frac{d}{dt}\Big|_{t=0} I(\pi(u+t\varphi)) = 0 \quad \text{for all } \varphi \in C_0^\infty(M, \mathbb{R}^\nu).$$

The condition that the mappings take values into the manifold N is a constrain that makes the corresponding Euler-Lagrange system

$$-\text{div}\,(|\nabla u|^{p-2}\nabla u) = |\nabla u|^{p-2} A(u)(\nabla u, \nabla u), \tag{1.1}$$

very nonlinear and difficult to handle. Here, A is a second fundamental form of the embedding of N into the Euclidean space (see, for example, [4, 16, 34, 41, 42, 48, 65, 69, 79, 80, 83, 84]). There is a huge and growing literature on the subject, and it is impossible to list here all relevant papers, but the reader can easily find other papers following the references in the papers cited above.

Our main focus in this paper is the theory of Sobolev mappings between manifolds, and later, the theory of Sobolev mappings between metric spaces, rather than applications of this theory to variational problems, and the above example was just to illustrate one of many areas in which the theory applies.

In connection with the theory of p-harmonic mappings, Eells and Lemaire [18] raised a question about density of smooth mappings $C^\infty(M,N)$ in $W^{1,p}(M,N)$. If $p \geqslant n = \dim M$, then smooth mappings are dense in $W^{1,p}(M,N)$ [73, 74], but if $p < n$, the answer depends on the topology of manifolds M and N. Recently, Hang and Lin [39] found a necessary and sufficient condition for the density in terms of algebraic topology. Their result is a correction of an earlier result of Bethuel [3] and a generalization of a result of Hajłasz [26]. To emphasize the connection of the problem with algebraic topology, let us mention that it is possible to reformulate the Poincaré conjecture (now a theorem) in terms of approximability of Sobolev mappings [25]. The theory of Sobolev mappings between manifolds has been extended to the case of Sobolev mappings with values into metric spaces. The first papers on this subject include the work of Ambrosio [2] on limits of classical variational problems and the work of Gromov and Schoen [24] on Sobolev mappings into the Bruhat–Tits buildings, with applications to rigidity questions for discrete groups. Later, the theory of Sobolev mappings with values into metric spaces was developed in a more elaborated form by Korevaar and Schoen [55] in their approach to the theory of harmonic mappings into Alexandrov spaces of nonpositive curvature. Other papers on Sobolev mappings from a manifold into a metric space include [12, 17, 49, 50, 51, 52, 70, 76]. Finally, analysis on metric spaces, the theory of Carnot–Carathéodory spaces, and the theory of quasiconformal mappings between metric spaces led to the theory of Sobolev mappings between metric spaces [46, 47, 58, 81], among which the theory of Newtonian–Sobolev mappings $N^{1,p}(X,Y)$ is particularly important.

In Sect. 2, we discuss fundamental results concerning the density of smooth mappings in $W^{1,p}(M,N)$. Section 3 is devoted to a construction of the class of Sobolev mappings from a manifold into a metric space. We also show there that several natural questions to the density problem have negative answers when we consider mappings from a manifold into a metric space. In Sect. 4, we explain the construction and basic properties of Sobolev spaces on metric measure spaces and, in final Sect. 5, we discuss recent development of the theory of Sobolev mappings between metric spaces, including results about approximation of mappings.

The notation in the paper is fairly standard. We assume that all manifolds are compact (with or without boundary), smooth, and connected. We always assume that such a manifold is equipped with a Riemannian metric, but since all such metrics are equivalent, it is not important with which metric we work. By a closed manifold we mean a smooth compact manifold without boundary. The integral average of a function u over a set E is denoted by

$$u_E = \fint_E u\,d\mu = \mu(E)^{-1}\int_E u\,d\mu\,.$$

Balls are denoted by B and σB for $\sigma > 0$ denotes the ball concentric with B whose radius is σ times that of B. The symbol C stands for a general constant whose actual value may change within a single string of estimates. We write $A \approx B$ if there is a constant $C \geqslant 1$ such that $C^{-1}A \leqslant B \leqslant CA$.

2 Sobolev Mappings between Manifolds

It is easy to see and is well known that smooth functions are dense in the Sobolev space $W^{1,p}(M)$. Thus, if N is isometrically embedded into $\mathbb{R}^\nu$, it follows that every $W^{1,p}(M,N)$ mapping can be approximated by $C^\infty(M,\mathbb{R}^\nu)$ mappings and the question is whether we can approximate $W^{1,p}(M,N)$ by $C^\infty(M,N)$ mappings. It was answered in the affirmative by Schoen and Uhlenbeck [73, 74] in the case $p \geqslant n = \dim M$.

Theorem 2.1. *If $p \geqslant n = \dim M$, then the class of smooth mappings $C^\infty(M,N)$ is dense in the Sobolev space $W^{1,p}(M,N)$.*

Proof.[1] Assume that N is isometrically embedded in some Euclidean space $\mathbb{R}^\nu$. If $p > n$, then the result is very easy. Indeed, let $u_k \in C^\infty(M,\mathbb{R}^\nu)$ be a sequence of smooth mappings that converge to u in the $W^{1,p}$ norm. Since $p > n$, the Sobolev embedding theorem implies that u_k converges uniformly to u. Hence for $k \geqslant k_0$ values of the mappings u_k belong to a tubular neighborhood $\mathcal{U} \subset \mathbb{R}^\nu$ of N from which there is a smooth nearest point projection $\pi : \mathcal{U} \to N$. Now $\pi \circ u_k \in C^\infty(M,N)$ and $\pi \circ u_k \to \pi \circ u = u$ in the $W^{1,p}$ norm. If $p = n$, then we do not have uniform convergence, but one still can prove that the values of the approximating sequence u_k whose construction is based locally on the convolution approximation belong to the tubular neighborhood of N for all sufficiently large k. This follows from the Poincaré inequality. To see this, it suffices to consider the localized problem where the mappings are defined on an Euclidean ball. Let $u \in W^{1,n}(B^n(0,1),N)$, and let $\overline{u}$ be the extension of u to a neighborhood of the ball (by reflection). We define $u_\varepsilon = \overline{u} * \varphi_\varepsilon$, where φ_ε is a standard mollifying kernel. The Poincaré inequality yields

$$\left(\fint_{B^n(x,\varepsilon)} |\overline{u}(y) - u_\varepsilon(x)|^n \, dy \right)^{1/n} \leqslant Cr \left(\fint_{B^n(x,\varepsilon)} |\nabla \overline{u}|^n \right)^{1/n}$$

$$= C' \left(\int_{B^n(x,\varepsilon)} |\nabla \overline{u}|^n \right)^{1/n} . \tag{2.1}$$

The right-hand side (as a function of x) converges to 0 as $\varepsilon \to 0$ uniformly on $B^n(0,1)$. Since

$$\operatorname{dist}(u_\varepsilon(x), N) \leqslant |\overline{u}(y) - u_\varepsilon(x)|$$

for all y, from (2.1) we conclude that

$$\operatorname{dist}(u_\varepsilon(x), N) \to 0 \quad \text{as } \varepsilon \to 0$$

uniformly on $B^n(0,1)$. Hence for $\varepsilon < \varepsilon_0$ values of the smooth mappings u_ε belong to $\mathcal{U}$ and thus $\pi \circ u_\varepsilon \to \pi \circ u = u$ as $\varepsilon \to 0$. □

Arguments used in the above proof lead to the following result.

---

[1] See also Theorems 3.7 and 5.5.

Proposition 2.2. *If $u \in W^{1,p}(M,N)$ can be approximated by continuous Sobolev mappings $C^0 \cap W^{1,p}(M,N)$, then it can be approximated by smooth $C^\infty(M,N)$ mappings.*

Proof. Indeed, if $v \in C^0 \cap W^{1,p}(M,N)$, then $v_\varepsilon \rightrightarrows v$ uniformly and hence $\pi \circ v_\varepsilon \to \pi \circ v = v$ in $W^{1,p}$. □

A basic tool in the study of Sobolev mappings between manifolds is a variant of the Fubini theorem for Sobolev functions. Let us illustrate it in a simplest setting. Suppose that $u, u_i \in W^{1,p}([0,1]^n)$, $\|u - u_i\|_{1,p} \to 0$ as $i \to \infty$. Denote by (t,x), where $t \in [0,1]$, $x \in [0,1]^{n-1}$, points in the cube. Then

$$\int_{[0,1]^n} |u - u_i|^p + |\nabla u - \nabla u_i|^p$$

$$= \int_0^1 \left(\int_{[0,1]^{n-1}} |u - u_i|^p + |\nabla u - \nabla u_i|^p \, dx \right) dt$$

$$= \int_0^1 F_i(t)\, dt \overset{i\to\infty}{\longrightarrow} 0 \, .$$

Hence $F_i \to 0$ in $L^1(0,1)$ and so there is a subsequence u_{i_j} such that $F_{i_j}(t) \to 0$ for almost all $t \in (0,1)$. That means that for almost all $t \in [0,1]$ we have $u(t,\cdot), u_{i_j}(t,\cdot) \in W^{1,p}([0,1]^{n-1})$ and $u_{i_j}(t,\cdot) \to u(t,\cdot)$ in $W^{1,p}([0,1]^{n-1})$. Clearly, the same argument applies to lower dimensional slices of the cube.

As was already pointed out, if $p < n = \dim M$, then density of smooth mappings does not always hold. The first example of this type was provided by Schoen and Uhlenbeck, and it is actually quite simple. A direct computation shows that the radial projection

$$u_0(x) = x/|x| : B^n(0,1) \setminus \{0\} \to S^{n-1}$$

belongs to the Sobolev space $W^{1,p}(B^n, S^{n-1})$ for all $1 \leqslant p < n$. Shoen and Uhlenbeck [73, 74] proved the following assertion.

Theorem 2.3. *If $n-1 \leqslant p < n$, then the mapping u_0 cannot be approximated by smooth mappings $C^\infty(B^n, S^{n-1})$ in the $W^{1,p}$ norm.*

Proof. Suppose that there is a sequence $u_k \in C^\infty(B^n, S^{n-1})$ such that $\|u_k - u_0\|_{1,p} \to 0$ as $k \to \infty$ for some $n-1 \leqslant p < n$. Then from the Fubini theorem it follows that there is a subsequence (still denoted by u_k) such that for almost every $0 < r < 1$

$$u_k|_{S^{n-1}(0,r)} \to u_0|_{S^{n-1}(0,r)}$$

in the $W^{1,p}(S^{n-1}(0,r))$ norm. If $n-1<p<n$, then the Sobolev embedding theorem into Hölder continuous functions implies that u_k restricted to such spheres converges uniformly to u_0

$$u_k|_{S^{n-1}(0,r)} \rightrightarrows u_0|_{S^{n-1}(0,r)},$$

which is impossible because the Brouwer degree[2] of $u_k|_{S^{n-1}(0,r)}$ is 0 and the degree of $u_0|_{S^{n-1}(0,r)}$ is 1. The case $p=n-1$ needs a different, but related argument. The degree of a mapping $v: M\to N$ between two oriented $(n-1)$-dimensional compact manifolds without boundary can be defined by the integral formula

$$\deg v = \int_M \det Dv/\mathrm{vol}\, N,$$

and from the Hölder inequality it follows that the degree is continuous in the $W^{1,n-1}$ norm. This implies that if $u_k\to u_0$ in $W^{1,n-1}(S^{n-1}(0,r))$, then the degree of $u_k|_{S^{n-1}(0,r)}$ which is 0 converges to the degree of $u_0|_{S^{n-1}(0,r)}$ is 1. Again we obtain a contradiction. □

It turns out, however, that for $1\leqslant p<n-1$ smooth maps are dense in $W^{1,p}(B^n,S^{n-1})$. Indeed, the following result was proved by Bethuel and Zheng [5].

Theorem 2.4. *For $1\leqslant p<k$ smooth mappings $C^\infty(M,S^k)$ are dense in $W^{1,p}(M,S^k)$.*

Proof. Let $u\in W^{1,p}(M,S^k)$. It is easy to see that for every $x\in S^k$ and $\delta>0$ there is a Lipschitz retraction $\pi_{x,\delta}: S^k\to S^k\setminus B(x,\delta)$, i.e., $\pi_{x,\delta}\circ\pi_{x,\delta}=\pi_{x,\delta}$, with the Lipschitz constant bounded by $C\delta^{-1}$. Now we consider the mapping $u_{x,\delta}=\pi_{x,\delta}\circ u$. Since $u_{x,\delta}$ maps M into the set $S^k\setminus B(x,\delta)$ which is diffeomorphic with a closed k dimensional ball, it is easy to see that $u_{x,\delta}$ can be approximated by smooth maps from M to $S^k\setminus B(x,\delta)\subset S^k$. Thus, it remains to prove that for every $\varepsilon>0$ there is $\delta>0$ and $x\in S^k$ such that $\|u-u_{x,\delta}\|_{1,p}<\varepsilon$.

There are $C\delta^{-k}$ disjoint balls of radius δ on S^k. Such a family of balls is denoted by $B(x_i,\delta)$, $i=1,2,\ldots,N_\delta$, where $N_\delta\approx\delta^{-k}$. Note that the mapping $u_{x_i,\delta}$ differs from u on the set $u^{-1}(B(x_i,\delta))$ and this is a family of $N_\delta\approx\delta^{-k}$ disjoint subset of M. Therefore, there is i such that

$$\int_{u^{-1}(B(x_i,\delta))} |u|^p+|\nabla u|^p \leqslant C\delta^k\|u\|_{1,p}^p.$$

[2] The degree is 0 because u_k has continuous (actually smooth) extension to the entire ball.

Using the fact that the Lipschitz constant of $\pi_{x_i,\delta}$ is bounded by $C\delta^{-1}$, it is easy to see that

$$\int\limits_{u^{-1}(B(x_i,\delta))} |\nabla u_{x_i,\delta}|^p \leqslant C\delta^{-p} \int\limits_{u^{-1}(B(x_i,\delta))} |\nabla u|^p \leqslant C\delta^{k-p}\|u\|^p_{1,p}$$

Since $u = u_{x_i,\delta}$ on the complement of the set $u^{-1}(B(x_i,\delta))$, we have

$$\begin{aligned}\|\nabla u - \nabla u_{x_i,\delta}\|_p &= \left(\int\limits_{u^{-1}(B(x_i,\delta))} |\nabla u - \nabla u_{x_i,\delta}|^p\right)^{1/p} \\ &\leqslant \left(\int\limits_{u^{-1}(B(x_i,\delta))} |\nabla u|^p\right)^{1/p} + \left(\int\limits_{u^{-1}(B(x_i,\delta))} |\nabla u_{x_i,\delta}|^p\right)^{1/p} \\ &\leqslant C(\delta^{k/p} + \delta^{(k-p)/p})\|u\|_{1,p}\,.\end{aligned}$$

Since $k - p > 0$, this implies that for given $\varepsilon > 0$ there is $\delta > 0$ and $x \in S^k$ such that $\|\nabla u - \nabla u_{x,\delta}\|_{1,p} < \varepsilon$. It remains to note that the mappings u and $u_{x,\delta}$ are also close in the L^p norm. Indeed, they are both uniformly bounded (as mappings into the unit sphere) and they coincide outside a set of very small measure. □

The above two results show that the answer to the problem of density of smooth mappings in the Sobolev space $W^{1,p}(M,N)$ depends of the topology of the manifold N and perhaps also on the topology of the manifold M. We find now necessary conditions for the density of $C^\infty(M,N)$ in $W^{1,p}(M,N)$.

The density result (Theorem 2.1) implies that if M and N are two smooth oriented compact manifolds without boundary, both of dimension n, then we can define the degree of mappings in the class $W^{1,n}(M,N)$. Indeed, if $u \in C^\infty(M,N)$, then the degree is defined in terms of the integral of the Jacobian and then it can be extended to the entire space $W^{1,n}(M,N)$ by the density of smooth mappings. Thus,

$$\deg : W^{1,n}(M,N) \to \mathbb{Z}$$

is a continuous function and it coincides with the classical degree on the subclass of smooth mappings. It turns out, however, that not only degree, but also homotopy classes can be defined. This follows from the result of White [85].

Theorem 2.5. *Let M and N be closed manifolds, and let $n = \dim M$. Then for every $f \in W^{1,n}(M,N)$ there is $\varepsilon > 0$ such that any two smooth mappings $g_1, g_2 : M \to N$ satisfying $\|f - g_i\|_{1,n} < \varepsilon$ for $i = 1, 2$ are homotopic.*

Note that Theorem 2.5 is also a special case of Theorem 2.8 and Theorem 5.5 below.

We use the above result to find the first necessary condition for the density of smooth mappings in the Sobolev space. The following result is due to Bethuel and Zheng [5] and Bethuel [3]. A simplified proof provided below is taken from [25]. Let $[p]$ denote the largest integer less than or equal to p. In the following theorem, π_k stands for the homotopy group.

Theorem 2.6. *If $\pi_{[p]}(N) \neq 0$ and $1 \leqslant p < n = \dim M$, then the smooth mappings $C^\infty(M,N)$ are not dense in $W^{1,p}(M,N)$.*

Proof. It is easy to construct a smooth mapping $f : B^{[p]+1} \to S^{[p]}$ with two singular points such that f restricted to small spheres centered at the singularities have degree $+1$ and -1 respectively and f maps a neighborhood of the boundary of the ball $B^{[p]+1}$ into a point. We can model the singularities on the radial projection mapping as in Theorem 2.3 so the mapping f belongs to $W^{1,p}$. Let now $g : B^{[p]+1} \times S^{n-[p]-1} \to S^{[p]}$ be defined by $g(b,s) = f(b)$. We can embed the torus $B^{[p]+1} \times S^{n-[p]-1}$ into the manifold M and extend the mapping on the completion of this torus as a mapping into a point. Clearly, $g \in W^{1,p}(M, S^{[p]})$. Let $\varphi : S^{[p]} \to N$ be a smooth representative of a nontrivial homotopy class. We prove that the mapping $\varphi \circ g \in W^{1,p}(M,N)$ cannot be approximated by smooth mappings from $C^\infty(M,N)$. By contrary, suppose that $u_k \in C^\infty(M,N)$ converges to $\varphi \circ f$ in the $W^{1,p}$ norm. In particular, $u_k \to \varphi \circ g$ in $W^{1,p}(B^{[p]+1} \times S^{n-[p]-1}, N)$. From the Fubini theorem it follows that there is a subsequence of u_k (still denoted by u_k) such that for almost every $s \in S^{n-[p]-1}$, u_k restricted to the slice $B^{[p]+1} \times \{s\}$ converges to the corresponding restriction of $\varphi \circ g$ in the Sobolev norm. Take such a slice and denote it simply by $B^{[p]+1}$. Again, by the Fubini theorem, u_k restricted to almost every sphere centered at the $+1$ singularity of f converges to the corresponding restriction of $\varphi \circ g$ in the Sobolev norm. Denote such a sphere by $S^{[p]}$. Hence $u_k|_{S^{[p]}} \to \varphi \circ g|_{S^{[p]}}$ in the space $W^{1,p}(S^{[p]}, N)$. Now the mapping $u_k|_{S^{[p]}} : S^{[p]} \to N$ is contractible (because it has a smooth extension to the ball), while $\varphi \circ g|_{S^{[p]}} : S^{[p]} \to N$ is a smooth representative of a nontrivial homotopy class $\pi_{[p]}(N)$, so $u_k|_{S^{[p]}}$ cannot be homotopic to $\varphi \circ g|_{S^{[p]}}$, which contradicts Theorem 2.5. ☐

It turns out that, in some cases, the condition $\pi_{[p]}(N) = 0$ is also sufficient for the density of smooth mappings. The following statement is due to Bethuel [3].

Theorem 2.7. *If $1 \leqslant p < n$, then smooth mappings $C^\infty(B^n, N)$ are dense in $W^{1,p}(B^n, N)$ if and only if $\pi_{[p]}(N) = 0$.*

Actually, Bethuel [3] claimed a stronger result that $\pi_{[p]}(N) = 0$ is a necessary and sufficient condition for the density of $C^\infty(M, N)$ mappings in $W^{1,p}(M, N)$ for any compact manifold M of dimension $\dim M = n > p$. This, however, turned out to be false: Hang and Lin [38] provided a counterexample to Bethuel's claim by demonstrating that despite the equality $\pi_3(\mathbb{CP}^2) = 0$, $C^\infty(\mathbb{CP}^3, \mathbb{CP}^2)$ is *not* dense in $W^{1,3}(\mathbb{CP}^3, \mathbb{CP}^2)$. Bethuel's claim made people believe that the problem of density of smooth mappings in the Sobolev space has a local nature. However the example of Hang and Lin and Theorem 2.7 shows that there might be global obstacles. Indeed, the mapping constructed by Hang and Lin cannot be approximated by smooth mappings $C^\infty(\mathbb{CP}^3, \mathbb{CP}^2)$, however, since $\pi_3(\mathbb{CP}^2) = 0$, Theorem 2.7 shows that this mapping can be smoothly approximated in a neighborhood of any point in $\mathbb{CP}^3$.

Therefore, searching for a necessary and sufficient condition for the density of smooth mappings, one has to take into account the topology of both manifolds M and N, or rather the interplay between the topology of M and the topology of N. Now we find such a necessary condition for the density of smooth mappings. Before we start, we need to say a few words about the behavior of Sobolev mappings on k-dimensional skeletons of generic smooth triangulations.

Let the manifold M be equipped with a smooth triangulation M^k, $k = 0, 1, 2, \ldots, n = \dim M$. Since the skeletons of the triangulation are piecewise smooth, it is not difficult to define the Sobolev space on skeletons $W^{1,p}(M^k)$. There is no problem with the definition of Sobolev functions in the interiors of the simplexes, but one needs to clarify how the Sobolev functions meet at the boundaries, so that the function belongs to the Sobolev space not only in each of the simplexes, but on the whole skeleton M^k. One possibility is to define the Sobolev norm $\|u\|_{1,p}$ for functions u that are Lipschitz continuous on M^k and then define $W^{1,p}$ by completion. Suppose now that $u \in W^{1,p}(M)$. If $v \in W^{1,p}([0,1]^n)$, then, in general, it is not true that the function v restricted to *each* slice $\{t\} \times [0,1]^{n-1}$ belongs to the Sobolev space $W^{1,p}([0,1]^{n-1})$, but it is true for almost all $t \in [0,1]$. By the same reason, u restricted to M^k does not necessarily belong to the Sobolev space $W^{1,p}(M^k)$. This problem can, however, be handled. Indeed, faces of the k dimensional skeleton M^k can be translated in the remaining directions which form an $n-k$ dimensional space. Hence, roughly speaking, with each skeleton M^k we can associate an $n-k$ dimensional family of skeletons.[3] Now u restricted to almost every skeleton in this family belongs to the Sobolev space $W^{1,p}$ on that skeleton by the Fubini theorem. We briefly summarize this construction by saying that if $u \in W^{1,p}(M)$, then u restricted to a generic k dimensional skeleton M^k belongs to the Sobolev space $W^{1,p}(M^k)$. Moreover, if $u, u_i \in W^{1,p}(M)$, $\|u - u_i\|_{1,p} \to 0$, then there is a subsequence u_{i_j} such that $u_{i_j} \to u$ in

[3] This is not entirely obvious because we translate different faces in different directions and we have to make sure that after all the faces glue together, so that we still have a k-dimensional skeleton. This, however, can be done and there are no unexpected surprises.

$W^{1,p}(M^k)$ on generic k-dimensional skeletons. This follows from the Fubini theorem argument explained above.

We say that two continuous mappings $f, g : M \to N$ are *k-homotopic*, $0 \leqslant k \leqslant n = \dim M$, if the restrictions of both mappings to the k-dimensional skeleton of a triangulation of M are homotopic. Using elementary topology, one can prove that the above definition does not depend on the choice of a triangulation of M (see [39, Lemma 2.1]). Theorem 2.5 is a special case of a more general result of White [85].

Theorem 2.8. *Let M and N be closed manifolds, and let $n = \dim M$. Then for every $f \in W^{1,p}(M, N)$, $1 \leqslant p \leqslant n$, there is $\varepsilon > 0$ such that any two Lipschitz mappings $g_1, g_2 : M \to N$ satisfying $\|f - g_i\|_{1,p} < \varepsilon$, $i = 1, 2$ are $[p]$-homotopic.*

Another result that we frequently use is the homotopy extension theorem. We state it only in a special case.

Theorem 2.9. *Let M be a smooth compact manifold equipped with a smooth trinagulation M^k, $k = 0, 1, 2, \ldots, n = \dim M$. Then for any topological space X every continuous mapping*

$$H : (M \times \{0\}) \cup (M^k \times [0, 1]) \to X$$

has a continuous extension to $\widetilde{H} : M \times [0, 1] \to X$.

In particular, the theorem implies that if $f : M \to N$ is continuous and $g : M^k \to N$ is homotopic to $f|_{M^k}$, then g admits a continuous extension to $\widetilde{g} : M \to N$. We apply this observation below.

In the proof of the necessity of the condition $\pi_{[p]}(N) = 0$, we constructed a map with the $(n - [p] - 1)$-dimensional singularity. The condition we will present now will actually imply $\pi_{[p]}(N) = 0$ and, not surprisingly, our argument will also involve a construction of a map with the $(n - [p] - 1)$-dimensional singularity.

Let $1 \leqslant p < n = \dim M$. Suppose that smooth mappings $C^\infty(M, N)$ are dense in $W^{1,p}(M, N)$. Assume that M is endowed with a smooth triangulation. Let $h : M^{[p]} \to N$ be a Lipschitz mapping. Observe that if $f \in W^{1,p}(S^k, N)$, then the integration in spherical coordinates easily implies that the mapping $\overline{f}(x) = f(x/|x|)$ belongs to $W^{1,p}(B^{k+1}, N)$ provided that $p < k+1$. Clearly, the ball B^{k+1} can be replaced by a $(k+1)$-dimensional simplex and S^k by its boundary. By this reason, the mapping $h : M^{[p]} \to N$ can be extended to a mapping in $W^{1,p}(M^{[p]+1}, N)$. The extension will have singularity consisting of one point in each $([p] + 1)$-dimensional simplex in $M^{[p]+1}$. Next, we can extend the mapping to $W^{1,p}(M^{[p]+2}, N)$. Now, the singularity is one dimensional. We can continue this process by extending the mapping to higher dimensional skeletons. Eventually, we obtain a mapping $\overline{h} \in W^{1,p}(M, N)$ with the $(n - [p] - 1)$-dimensional singularity located on a dual skeleton to $M^{[p]}$.

Let $u_i \in C^\infty(M,N)$ be such that $\|\overline{h} - u_i\|_{1,p} \to 0$ as $i \to \infty$. From the Fubini theorem it follows that there is a subsequence u_{i_j} such that $u_{i_j} \to \overline{h}$ in $W^{1,p}$ on generic $[p]$-dimensional skeletons, so

$$u_{i_j} \to \overline{h} \quad \text{in } W^{1,p}(\widetilde{M}^{[p]}, N)\,,$$

where $\widetilde{M}^{[p]}$ is a “tilt” of $M^{[p]}$. Since $\overline{h}$ and u_{i_j} are Lipschitz, from Theorem 2.8 it follows that u_{i_j} is homotopic to $\overline{h}$ on $\widetilde{M}^{[p]}$ for all $j \geqslant j_0$. Now, from the homotopy extension theorem (see Theorem 2.9) it follows that the mapping $\overline{h}|_{\widetilde{M}^{[p]}}$ admits an extension to a continuous mapping $\overline{\overline{h}} : M \to N$. Hence also $h : M^{[p]} \to N$ can be extended to a continuous mapping $h' : M \to N$.

We proved that every Lipschitz mapping $h : M^{[p]} \to N$ admits a continuous extension $h' : M \to N$. Since every continuous mapping $f : M^{[p]} \to N$ is homotopic to a Lipschitz mapping, another application of the homotopy extension theorem implies that also f has continuous extension. We proved the following assertion.

Proposition 2.10. *If $1 \leqslant p < n = \dim M$ and $C^\infty(M,N)$ is dense in $W^{1,p}(M,N)$, then every continuous mapping $f : M^{[p]} \to N$ can be extended to a continuous mapping $f' : M \to N$.*

The following result provides a characterization of the property described in the above proposition.

We say that M has $(k-1)$-*extension property with respect to* N if for every continuous mapping $f \in C(M^k, N)$, $f|_{M^{k-1}}$ has a continuous extension to $\widetilde{f} \in C(M,N)$.

Proposition 2.11. *If $1 \leqslant k < n = \dim M$, then every continuous mapping $f : M^k \to N$ can be extended to a continuous mapping $f' : M \to N$ if and only if $\pi_k(N) = 0$ and M has the $(k-1)$-extension property with respect to N.*

Proof. Suppose that every continuous mapping $f : M^k \to N$ has a continuous extension to $f' : M \to N$. Then it is obvious that M has the $(k-1)$-extension property with respect to N. We need to prove that $\pi_k(N) = 0$. Suppose that $\pi_k(N) \neq 0$. Let Δ be a $(k+1)$-dimensional simplex in M^{k+1}, and let $\partial\Delta$ be its boundary. It is easy to see that there is a continuous retraction $\pi : M^k \to \partial\Delta$. Let $\varphi : \partial\Delta \to N$ be a representative of a nontrivial element in the homotopy group $\pi_k(N)$. Then φ cannot be extended to Δ. Hence $f = \varphi \circ \pi : M^k \to N$ has no continuous extension to M. We obtain a contradiction.

Now, suppose that $\pi_k(N) = 0$ and M has the $(k-1)$-extension property with respect to N. Let $f : M^k \to N$ be continuous. We need to show that f can be continuously extended to M. Let $\widetilde{f} : M \to N$ be a continuous extension of $f|_{M^{k-1}}$.

The set $M^k \times [0,1]$ is the union of $(k+1)$-dimensional cells $\Delta \times [0,1]$, where Δ is a k-dimensional simplex in M^k. Denote by (x,t) the points in $\Delta \times [0,1]$

and define the mapping on the boundary on each cell as follows:

$$H(x,0) = \widetilde{f}(x) \qquad \text{for } x \in \Delta,$$

$$H(x,1) = f(x) \qquad \text{for } x \in \Delta,$$

$$H(x,t) = \widetilde{f}(x) = f(x) \text{ for } x \in \partial\Delta.$$

Because $\pi_k(N) = 0$, H can be continuously extended to the interior of each cell. Denote by $H : M^k \times [0,1] \to N$ the extension. Now, from the homotopy extension theorem it follows that $f : M^k \to N$ admits a continuous extension $f' : M \to N$. □

Thus, if $f \in W^{1,p}(M,N)$, $1 \leqslant p < n = \dim M$, can be approximated by smooth mappings, then $\pi_{[p]}(N) = 0$ and for every continuous mapping $g : M^{[p]} \to N$, $g|_{M^{[p]-1}}$ has a continuous extension to M.

Actually, this property was used by Hang and Lin [38] to demonstrate that $C^\infty(\mathbb{CP}^3, \mathbb{CP}^2)$ mappings are not dense in $W^{1,3}(\mathbb{CP}^3, \mathbb{CP}^2)$ (despite the fact that $\pi_3(\mathbb{CP}^2) = 0$).

Since the extension property is of topological nature, it is easier to work with the natural CW structure of $\mathbb{CP}^n$ rater than with the triangulation and the extension property can be equivalently formulated for CW structures.

It is well known that $\mathbb{CP}^n$ has a natural CW structure

$$\mathbb{CP}^0 \subset \mathbb{CP}^1 \subset \ldots \subset \mathbb{CP}^n.$$

If $M = \mathbb{CP}^3$, then $M^2 = M^3 = \mathbb{CP}^1$. Now, from the elementary algebraic topology it follows that the identity mapping

$$i : M^3 = \mathbb{CP}^1 \subset \mathbb{CP}^2$$

cannot be continuously extended to $\widetilde{i} : \mathbb{CP}^3 \to \mathbb{CP}^2$ and since $M^2 = M^3$ we also have that $i|_{M^2}$ has no continuous extension. Thus, $C^\infty(\mathbb{CP}^3, \mathbb{CP}^2)$ mappings are not dense in $W^{1,p}(\mathbb{CP}^3, \mathbb{CP}^2)$ (see [38, pp. 327-328] for more details).

It turns out that the above necessary condition for density is also sufficient. Namely, the following result was proved by Hang and Lin [39].

Theorem 2.12. *Assume that M and N are compact smooth Riemannian manifolds without boundary. If $1 \leqslant p < \dim M$, then smooth mappings $C^\infty(M,N)$ are dense in $W^{1,p}(M,N)$ if and only if $\pi_{[p]}(N) = 0$ and M has the $([p]-1)$-extension property with respect to N.*

The following two corollaries easily follow from the theorem (see [39]).

Corollary 2.13. *If $1 \leqslant p < n = \dim M$, k is an integer such that $0 \leqslant k \leqslant [p]-1$, $\pi_i(M) = 0$ for $1 \leqslant i \leqslant k$, and $\pi_i(N) = 0$ for $k+1 \leqslant i \leqslant [p]$, then $C^\infty(M,N)$ is dense in $W^{1,p}(M,N)$.*

Corollary 2.14. *If* $1 \leqslant p < n = \dim M$, $\pi_i(N) = 0$ *for* $[p] \leqslant i \leqslant n-1$*, then* $C^\infty(M, N)$ *is dense in* $W^{1,p}(M, N)$*.*

In particular, Corollary 2.13 with $k = 0$ gives the following result that was previously proved in [26].

Corollary 2.15. *If* $1 \leqslant p < n = \dim M$ *and* $\pi_1(N) = \pi_2(N) = \ldots = \pi_{[p]}(N) = 0$*, then* $C^\infty(M, N)$ *is dense in* $W^{1,p}(M, N)$*.*

The reason why we stated this corollary in addition to Corollary 2.13 is that, in the case of Sobolev mappings from metric spaces supporting Poincaré inequalities into Lipschitz polyhedra, the homotopy condition from Corollary 2.15 turns out to be necessary and sufficient for density (see Theorem 5.6).

Another interesting question regarding density of smooth mappings is the question about the density in the sequential weak topology. We do not discuss this topic here and refer the reader to [26, 36, 37, 39, 40, 66, 67].

3 Sobolev Mappings into Metric Spaces

There were several approaches to the definition of the class of Sobolev mappings from a manifold, or just an open set in $\mathbb{R}^n$ into a metric space (see, for example, [2, 24, 49, 55, 70]). The approach presented here is taken from [35] and it is an elaboration of ideas of Ambrosio [2] and Reshetnyak [70]. One of the benefits of the construction presented here is that the Sobolev space of mappings into a metric space is equipped in a natural way with a metric, so one can ask whether the class of Lipschitz mappings is dense. In the case of mappings into metric spaces, it does not make sense to talk about smooth mappings, so we need to consider Lipschitz mappings instead.

Since every metric space X admits an isometric embedding into a Banach space[4] V, the idea is to define the Sobolev space of functions with values into a Banach space V and then define the Sobolev space of mappings with values into X as

$$W^{1,p}(M, X) = \{f \in W^{1,p}(M, V) |\, f(M) \subset X\}\,.$$

Since $W^{1,p}(M, V)$ is a Banach space, this approach equips $W^{1,p}(M, X)$ with a natural metric inherited from the norm of $W^{1,p}(M, V)$, just like in the case of Sobolev mappings between manifolds. With this metric at hand, we can ask under what conditions the class of Lipschitz mappings $\mathrm{Lip}\,(M, X)$ is dense in $W^{1,p}(M, X)$.

On the other hand, the approach described above depends on the isometric embedding of X into V, so it is useful to find another, equivalent and intrinsic

[4] Every metric space admits an isometric embedding into the Banach space $V = \ell^\infty(X)$ of bounded functions on X. If, in addition, X is separable, then X admits an isometric embedding into ℓ^∞.

approach independent of the embedding. In this section, we describe both such approaches. In our approach, we follow [35], where the reader can find detailed proofs of results stated here.

For the sake of simplicity, we consider Sobolev functions defined on a domain in $\mathbb{R}^n$ rather than on a manifold, but all the statements can easily be generalized to the case of Sobolev functions defined on manifolds.

Before we define the Sobolev space of functions with values into a Banach space, we need briefly recall the notion of the Bochner integral (see [15]).

Let V be a Banach space, $E \subset \mathbb{R}^n$ a measurable set, and $1 \leqslant p \leqslant \infty$. We say that $f \in L^p(E,V)$ if

(1) f is *essentially separable valued*, i.e., $f(E \setminus Z)$ is a separable subset of V for some set Z of Lebesgue measure zero,

(2) f is *weakly measurable*, i.e., for every $v^* \in V^*$, $\langle v^*, f\rangle$ is measurable;

(3) $\|f\| \in L^p(E)$.

If $f = \sum_{i=1}^k a_i \chi_{E_i} : E \to V$ is a simple function, then the Bochner integral is defined by the formula

$$\int_E f(x)\,dx = \sum_{i=1}^k a_i |E_i|$$

and for $f \in L^1(E,V)$ the Bochner integral is defined as the limit of integrals of simple functions that converge to f almost everywhere. The following two properties of the Bochner integral are well known:

$$\left\| \int_E f(x)\,dx \right\| \leqslant \int_E \|f(x)\|\,dx$$

and

$$\left\langle v^*, \int_E f(x)\,dx \right\rangle = \int_E \langle v^*, f(x)\rangle\,dx \quad \text{for all } v^* \in V^*. \tag{3.1}$$

In the theory of the Bochner integral, a measurable set $E \subset \mathbb{R}^n$ can be replaced by a more general measure space. We need such a more general setting later, in Sect. 5.

Let now $\Omega \subset \mathbb{R}^n$ be an open set, and let V be a Banach space. It is natural to define the Sobolev space $W^{1,p}(\Omega,V)$ using the notion of weak derivative, just like in the case of real valued functions. We say that $f \in W^{1,p}(\Omega,V)$ if $f \in L^p(\Omega,V)$ and for $i = 1,2,\ldots,n$ there are functions $f_i \in L^p(\Omega,V)$ such that

$$\int_\Omega \frac{\partial\varphi}{\partial x_i}(x)f(x)\,dx = -\int_\Omega \varphi(x)f_i(x)\,dx \quad \text{for all } \varphi\in C_0^\infty(\Omega).$$

We denote $f_i = \partial f/\partial x_i$ and call these functions *weak partial derivatives.* We also write $\nabla f = (\partial f/\partial x_1, \ldots, \partial f/\partial x_n)$ and

$$|\nabla f| = \left(\sum_{i=1}^n \left\|\frac{\partial f}{\partial x_i}\right\|^2\right)^{1/2}.$$

The space $W^{1,p}(\Omega, V)$ is equipped with the norm

$$\|f\|_{1,p} = \left(\int_\Omega \|f\|^p\right)^{1/p} + \left(\int_\Omega |\nabla f|^p\right)^{1/p}.$$

It is an easy exercise to show that $W^{1,p}(\Omega, V)$ is a Banach space.

The problem with this definition is that it is not clear what conditions are needed to guarantee that Lipschitz functions belong to the Sobolev space. Indeed, a Lipschitz function $f:[0,1]\to V$ need not be differentiable in the Fréchet sense at any point, unless V has the Radon–Nikodym property (see [61, p. 259]). Since we want to work with Sobolev mappings from the geometric point of view, it is a very unpleasant situation.

There is another, more geometric, definition of the Sobolev space of functions with values in Banach spaces which we describe now. The definition below is motivated by the work of Ambrosio [2] and Reshetnyak [70].

Let $\Omega\subset\mathbb{R}^n$ be an open set, V a Banach space, and $1\leqslant p<\infty$. The space $R^{1,p}(\Omega, V)$ is the class of all functions $f\in L^p(\Omega, V)$ such that

(1) for every $v^*\in V^*$, $\|v^*\|\leqslant 1$ we have $\langle v^*, f\rangle \in W^{1,p}(\Omega)$;

(2) there is a nonnegative function $g\in L^p(\Omega)$ such that

$$|\nabla\langle v^*, f\rangle| \leqslant g \quad \text{a.e.} \tag{3.2}$$

for every $v^*\in V^*$ with $\|v^*\|\leqslant 1$.

Using arguments similar to those in the proof of the completeness of L^p, one can easily show that $R^{1,p}(\Omega, V)$ is a Banach space with respect to the norm

$$\|f\|_{R^{1,p}} = \|f\|_p + \inf \|g\|_p$$

where the infimum is over the class of all functions g satisfying the inequality (3.2). Using the definitions and the property (3.1), one can easily prove the following result (see [35]).

Proposition 3.1. *If $\Omega\subset\mathbb{R}^n$ is open and V is a Banach space, then $W^{1,p}(\Omega, V)\subset R^{1,p}(\Omega, V)$ and $\|f\|_{R^{1,p}}\leqslant \|f\|_{1,p}$ for all $f\in W^{1,p}(\Omega, V)$.*

However, we can prove the opposite inclusion only under additional assumptions about the space V (see [35]).

Theorem 3.2. *If $\Omega \subset \mathbb{R}^n$ is open, $V = Y^*$ is dual to a separable Banach space Y, and $1 \leqslant p < \infty$, then $W^{1,p}(\Omega, V) = R^{1,p}(\Omega, V)$ and $\|f\|_{R^{1,p}} \leqslant \|f\|_{1,p} \leqslant \sqrt{n}\|f\|_{R^{1,p}}$.*

Idea of the proof. One only needs to prove the inclusion $R^{1,p} \subset W^{1,p}$ along with the estimate for the norm. Actually, the proof of this inclusion is quite long and it consists of several steps. In the sketch provided below, many delicate steps are omitted.

By the canonical embedding $Y \subset Y^{**} = V^*$, elements of the Banach space Y can be interpreted as functionals on V. Observe that if $u : [0,1] \to V$ is absolutely continuous, then for every $v^* \in Y$ the function $\langle v^*, u\rangle$ is absolutely continuous, so it is differentiable almost everywhere and satisfies the integration by parts formula. Since the space Y is separable, we have almost everywhere the differentiability of $\langle v^*, u\rangle$ and the integration by parts for all v^* from a countable and dense subset of Y. This implies that the function $u : [0,1] \to V$ is differentiable in a certain weak sense known as the w^*-differentiability. Moreover, the w^*-derivative $u' : [0,1] \to V$ satisfies the integration by parts formula

$$\int_0^1 \varphi'(t)u(t)\,dt = -\int_0^1 \varphi(t)u'(t)\,dt.$$

Using this fact and the Fubini theorem, one can prove that a function $f \in L^p(\Omega, V)$ that is absolutely continuous on almost all lines parallel to coordinate axes and such that the w^*-partial derivatives of f satisfy $\|\partial f/\partial x_i\| \leqslant g$ almost everywhere for some $g \in L^p(\Omega)$ belongs to the Sobolev space $W^{1,p}(\Omega, V)$, $\|f\|_{1,p} \leqslant \|f\|_p + \sqrt{n}\|g\|_p$. This fact is similar to the well-known characterization of the Sobolev space $W^{1,p}(\Omega)$ by absolute continuity on lines.

At the last step, one proves that if $f \in R^{1,p}(\Omega, V)$, then f is absolutely continuous on almost all lines parallel to the coordinate axes and the w^*-partial derivatives satisfy $\|\partial f/\partial x_i\| \leqslant g$, where the function $g \in L^p(\Omega)$ satisfies (3.2).

The above facts put together easily imply the result. □

One can prove the following more geometric characterization of the space $R^{1,p}(\Omega, V)$ which is very useful (see [35]).

Theorem 3.3. *Let $\Omega \subset \mathbb{R}^n$ be open, V a Banach space and $1 \leqslant p < \infty$. Then $f \in R^{1,p}(\Omega, V)$ if and only if $f \in L^p(\Omega, V)$ and there is a nonnegative function $g \in L^p(\Omega)$ such that for every Lipschitz continuous function $\varphi : V \to \mathbb{R}$, $\varphi \circ f \in W^{1,p}(\Omega)$ and $|\nabla(\varphi \circ f)| \leqslant \mathrm{Lip}\,(\varphi)g$ almost everywhere.*

Idea of the proof. One implication is obvious. Indeed, if a function f satisfies the condition described in the above theorem, then it belongs to the space $R^{1,p}(\Omega, V)$ because for $v^* \in V^*$, $\|v^*\| \leqslant 1$, $\varphi(v) = \langle v^*, v\rangle$ is 1-Lipschitz continuous and hence $\langle v^*, f\rangle \in W^{1,p}(\Omega)$ with $|\nabla\langle v^*, f\rangle| \leqslant g$ almost everywhere.

In the other implication, we use the fact that $R^{1,p}(\Omega, V)$ functions are absolutely continuous on almost all lines parallel to coordinate axes. This implies that if $\varphi : V \to \mathbb{R}$ is Lipschitz continuous, then also $\varphi \circ f$ is absolutely continuous on almost all lines and hence $\varphi \circ f \in W^{1,p}(\Omega)$ by the characterization of $W^{1,p}(\Omega)$ in terms of absolute continuity on lines. □

Now we are ready to define the Sobolev space of mappings with values into an arbitrary metric space. Let $\Omega \subset \mathbb{R}^n$ be open, and let X be a metric space. We can assume that X is isometrically embedded into a Banach space V. We have now two natural definitions

$$W^{1,p}(\Omega, X) = \{f \in W^{1,p}(\Omega, V) |\, f(\Omega) \subset X\}$$

and

$$R^{1,p}(\Omega, X) = \{f \in R^{1,p}(\Omega, V) |\, f(\Omega) \subset X\}$$

Both spaces $W^{1,p}(\Omega, X)$ and $R^{1,p}(\Omega, X)$ are endowed with the norm metric.

Since every Lipschitz function $\varphi : X \to \mathbb{R}$ can be extended to a Lipschitz function $\widetilde{\varphi} : V \to \mathbb{R}$ with the same Lipschitz constant (McShane extension), we easily see that *if X is compact and Ω is bounded, then $f \in R^{1,p}(\Omega, X)$ if and only if there is a nonnegative function $g \in L^p(\Omega)$ such that for every Lipschitz continuous function $\varphi : X \to \mathbb{R}$ we have $\varphi \circ f \in W^{1,p}(\Omega)$ and $|\nabla(\varphi \circ f)| \leqslant \mathrm{Lip}\,(\varphi) g$ almost everywhere.*

We assume here the compactness of X and boundedness of Ω to avoid problems with the L^p integrability of f.

Observe that the last characterization of the space $R^{1,p}(\Omega, X)$ is independent of the isometric embedding of X into a Banach space.

As a direct application of Theorem 3.2, we have

Theorem 3.4. *If $\Omega \subset \mathbb{R}^n$ is open, $V = Y^*$ is dual to a separable Banach space Y, $1 \leqslant p < \infty$, and $X \subset V$, then $W^{1,p}(\Omega, X) = R^{1,p}(\Omega, X)$.*

The most interesting case is that where the space X is separable. In this case, X admits an isometric embedding to $V = \ell^\infty$ which is dual to a separable Banach space, $\ell^\infty = (\ell^1)^*$ and hence Theorem 3.4 applies.

With a minor effort one can extend the above arguments to the case of Sobolev spaces defined on a manifold, which leads to the spaces $W^{1,p}(M, X)$ and $R^{1,p}(M, X)$.

The following theorem is the main result in [35].

Suppose that any two points $x, y \in X$ can be connected by a curve of finite length. Then $d_\ell(x, y)$ defined as the infimum of lengths of curves connecting x to y is a metric. We call it the *length metric.* Since $d_\ell(x, y) \geqslant d(x, y)$, it easily follows that if X is compact with respect to d_ℓ, then X is compact with respect to d.

Theorem 3.5. *Let X be a metric space, compact with respect to the length metric. If $n \geqslant 2$, then there is a continuous Sobolev mapping $f \in C^0 \cap W^{1,n}([0,1]^n, X)$ such that $f([0,1]^n) = X$.*

3.1 Density

Once the space of Sobolev mappings with values into metric spaces has been defined, we can ask under what conditions Lipschitz mappings $\mathrm{Lip}\,(M,X)$ are dense in $W^{1,p}(M,X)$ or in $R^{1,p}(M,X)$. In this section, we follow [30] and provide several counterexamples to natural questions and very few positive results. For the sake of simplicity, we assume that the metric space X is compact and admits an isometric embedding into the Euclidean space. Thus, $X \subset \mathbb{R}^\nu$ and we simply define

$$W^{1,p}(M,X) = \{f \in W^{1,p}(M,\mathbb{R}^\nu) |\, f(M) \subset X\}.$$

If M and N are smooth compact manifolds, $\dim M = n$, then, as we know (Theorem 2.1), smooth mappings are dense in $W^{1,n}(M,N)$. The key property of N used in the proof was the existence of a smooth nearest point projection from a tubular neighborhood of N. The proof employed the fact that the composition with the smooth nearest point projection is continuous in the Sobolev norm. It turns out that the composition with a Lipschitz mapping need not be continuous in the Sobolev norm [30].

Theorem 3.6. *There is a Lipschitz function $\varphi \in \mathrm{Lip}\,(\mathbb{R}^2)$ with compact support such that the operator $\Phi : W^{1,p}([0,1],\mathbb{R}^2) \to W^{1,p}([0,1])$ defined as composition $\Phi(f) = \varphi \circ f$ is not continuous for any $1 \leqslant p < \infty$.*

The proof of the continuity of composition with a *smooth* function φ is based on the chain rule and continuity of the derivative $\nabla\varphi$. If φ is just Lipschitz continuous, then $\nabla\varphi$ is only measurable, so the proof does not work and the existence of the example as in the theorem above is not surprising after all (see, however, [64]).

Although the composition with a Lipschitz mapping is not continuous in the Sobolev norm, we can still prove that Theorem 2.1 is true if we replace N by a compact Lipschitz neighborhood retract.

We say that a closed set $X \subset \mathbb{R}^\nu$ is a *Lipschitz neighborhood retract* if there is an open neighborhood $\mathcal{U} \subset \mathbb{R}^\nu$ of X, $X \subset \mathcal{U}$, and a Lipschitz retraction $\pi : \mathcal{U} \to X$, $\pi \circ \pi = \pi$.

The following result was proved in [30].

Theorem 3.7. *Let $X \subset \mathbb{R}^\nu$ be a compact Lipschitz neighborhood retract. Then for every smooth compact n-dimensional manifold M Lipschitz mappings* $\mathrm{Lip}\,(M,X)$ *are dense in $W^{1,p}(M,X)$ for $p \geqslant n$.*

Sketch of the proof. If $f \in W^{1,p}(M,X)$ and $f_i \in C^\infty(M,\mathbb{R}^\nu)$ is a smooth approximation based on the mollification, then $\|f_i - f\|_{1,p} \to 0$ and for all sufficiently large i the values of f_i belong to $\mathcal{U}$ (Sobolev embedding for $p > n$ and Poincaré inequality for $p = n$), but there is no reason to claim that $\pi \circ f_i \to \pi \circ f = f$. To overcome this problem, one needs to construct another approximation $f_t \in \mathrm{Lip}\,(M,\mathbb{R}^\nu)$ such that

(1) Lipschitz constant of f_t is bounded by Ct;

(2) $t^p|\{f \neq f_t\}| \to 0$ as $t \to \infty$;

(3) $\sup_{x\in M} \mathrm{dist}\,(f_t(x),X) \to 0$ as $t \to \infty$.

The construction of such an approximation is not easy, but once we have it, a routine calculation shows that $\|f - \pi \circ f_t\|_{1,p} \to 0$ as $t \to \infty$. Indeed,

$$\Big(\int_M |\nabla f - \nabla(\pi \circ f_t)|^p\Big)^{1/p}$$

$$\leqslant \Big(\int_{\{f\neq f_t\}} |\nabla f|^p\Big)^{1/p} + \Big(\int_{\{f\neq f_t\}} |\nabla(\pi \circ f_t)|^p\Big)^{1/p}$$

$$\leqslant \Big(\int_{\{f\neq f_t\}} |\nabla f|^p\Big)^{1/p} + Ct|\{f \neq f_t\}|^{1/p} \to 0 \quad \text{as } t \to \infty.$$

The proof is complete. □

The class of Lipschitz neighborhood retracts contains Lipschitz submanifolds of $\mathbb{R}^\nu$ [63, Theorem 5.13].

In the following example, X is replaced by an n-dimensional submanifold of the Euclidean space such that it is smooth except for a just one point, and we no longer have the density of Lipschitz mappings [30].

Theorem 3.8. *Let $M \subset \mathbb{R}^\nu$ be a closed n-dimensional manifold. Then there is a homeomorphism $\varPhi \in C^\infty(\mathbb{R}^\nu,\mathbb{R}^\nu)$ which is a diffeomorphism in $\mathbb{R}^\nu \setminus \{0\}$ which is identity outside a sufficiently large ball and has the property that Lipschitz mappings* $\mathrm{Lip}\,(M,\widetilde{M})$ *are not dense in $W^{1,n}(M,\widetilde{M})$, where $\widetilde{M} = \varPhi^{-1}(M)$.*

Clearly, $\widetilde{M}$ cannot be a Lipschitz neighborhood retract. The derivative of the mapping Φ is zero at 0 and hence derivative of Φ^{-1} is unbounded in a neighborhood of 0. This causes $\widetilde{M}$ to have highly oscillating smooth "wrinkles" which accumulate at one point. In a neighborhood of that point, $\widetilde{M}$ is the graph of a continuous function which is smooth everywhere except for this point. Actually, the construction is done in such a way that $\widetilde{M}$ is $W^{1,n}$-homeomorphic to M, but, due to high oscillations, there is no Lipschitz mapping from M onto $\widetilde{M}$, and one proves that this $W^{1,n}$-homeomorphism cannot be approximated by Lipschitz mappings. This actually shows that there is a continuous Sobolev mapping from M onto $\widetilde{M}$ which cannot be approximated by Lipschitz mappings, a situation which never occurs in the case of approximation of mappings between smooth manifolds (see Proposition 2.2).

Another interesting question is the stability of density of Lipschitz mappings with respect to bi-Lipschitz modifications of the target.

Assume that X and Y are compact subsets of $\mathbb{R}^\nu$ that are bi-Lipschitz homeomorphic. Assume that M is a closed n-dimensional manifold and Lipschitz mappings $\mathrm{Lip}\,(M, X)$ *are dense in* $W^{1,p}(M, X)$ *for some* $1 \leqslant p < \infty$. *Are the Lipschitz mappings* $\mathrm{Lip}\,(M, Y)$ *dense in* $W^{1,p}(M, Y)$*?*

Since bi-Lipschitz invariance is a fundamental principle in geometric analysis on metric spaces, one expects basic theorems and definitions to remain unchanged when the ambient space is subject to a bi-Lipschitz transformation. Although the composition with a Lipschitz mapping is not continuous in the Sobolev norm, there are several reasons to expect a positive answer to remain in accordance with the principle.

First, if $\Phi : X \to Y$ is a bi-Lipschitz mapping, then $T(f) = \Phi \circ f$ induces bijections

$$T : W^{1,p}(M, X) \to W^{1,p}(M, Y), \quad T : \mathrm{Lip}\,(M, X) \to \mathrm{Lip}\,(M, Y).$$

Second, have the following positive result [31].

Theorem 3.9. *If Lipschitz mappings* $\mathrm{Lip}\,(M, X)$ *are dense in* $W^{1,p}(M, X)$ *in the following strong sense: for every* $\varepsilon > 0$ *there is* $f_\varepsilon \in \mathrm{Lip}\,(M, X)$ *such that* $|\{x|\, f_\varepsilon(x) \neq f(x)\}| < \varepsilon$ *and* $\|f - f_\varepsilon\|_{1,p} < \varepsilon$, *then Lipschitz mappings are dense in* $W^{1,p}(M, Y)$.

The strong approximation property described in the theorem is quite natural because if $f \in W^{1,p}(M, \mathbb{R}^\nu)$, then for every $\varepsilon > 0$ there is a Lipschitz mapping $f_\varepsilon \in \mathrm{Lip}\,(M, \mathbb{R}^\nu)$ such that $|\{x|\, f_\varepsilon(x) \neq f(x)\}| < \varepsilon$ and $\|f - f_\varepsilon\|_{1,p} < \varepsilon$. Such an approximation argument was employed in the proof of Theorem 3.7.

The above facts are convincing reasons to believe that the answer to the stability question should be positive. Surprisingly it is not. The following counterexample was constructed in [30].

Theorem 3.10. *Fix an integer* $n \geqslant 2$. *There is a compact and connected set* $X \subset \mathbb{R}^{n+2}$ *and a global bi-Lipschitz homeomorphism* $\Phi : \mathbb{R}^{n+2} \to \mathbb{R}^{n+2}$ *with* *the property that for any closed n-dimensional manifold M smooth mappings $C^\infty(M,X)$ are dense in $W^{1,n}(M,X)$, but Lipschitz mappings* $\mathrm{Lip}\,(M,Y)$ *are not dense in $W^{1,n}(M,Y)$, where $Y=\Phi(X)$.*

By smooth mappings $C^\infty(M,X)$ we mean smooth mappings from M to $\mathbb{R}^{n+2}$ with the image contained in X.

The space X constructed in the proof is quite irregular: it is the closure of a carefully constructed sequence of smooth submanifolds that converges to a manifold with a point singularity and all the manifolds are connected by a fractal curve. The space X looks like a stack of pancakes. The proof involves also a construction of a mapping $f\in W^{1,p}(M,X)$ which can be approximated by Lipschitz mappings, but the mappings that approximate f do not coincide with f at any point, so the strong approximation property from Theorem 3.9 is not satisfied.

4 Sobolev Spaces on Metric Measure Spaces

In order to define the space of Sobolev mappings between metric spaces, we need first define Sobolev spaces on metric spaces equipped with so-called doubling measures. By the end of the 1970s, it was discovered that a substantial part of harmonic analysis could be generalized such spaces [14]. This included the study of maximal functions, Hardy spaces and BMO, but it was only the zeroth order analysis in the sense that no derivatives were involved. The study of the first order analysis with suitable generalizations of derivatives, fundamental theorem of calculus, and Sobolev spaces, in the setting of metric spaces with a doubling measure was developed since the 1990s. This area is growing and plays an important role in many areas of the contemporary mathematics [43].

We recommend the reader a beautiful expository paper of Heinonen [44], where the significance and broad scope of applications of the first order analysis on metric spaces is carefully explained.

We precede the definition of the Sobolev space with auxiliary definitions and results. The material of Sects. 4.1–4.5 is standard by now. In our presentation, we follow [29], where the reader can find detailed proofs.

4.1 Integration on rectifiable curves

Let (X,d) be a metric space. By a *curve* in X we mean any continuous mapping $\gamma:[a,b]\to X$. The *image of the curve* is denoted by $|\gamma|=\gamma([a,b])$. The *length* of γ is defined by

$$\ell(\gamma) = \sup \left\{ \sum_{i=0}^{n-1} d(\gamma(t_i), \gamma(t_{i+1})) \right\},$$

where the supremum is taken over all partitions $a = t_0 < t_1 < \ldots < t_n = b$. We say that the curve is *rectifiable* if $\ell(\gamma) < \infty$. The *length function* associated with a rectifiable curve $\gamma : [a,b] \to X$ is $s_\gamma : [a,b] \to [0, \ell(\gamma)]$ given by $s_\gamma(t) = \ell(\gamma|_{[a,t]})$. Not surprisingly, the length function is nondecreasing and continuous.

It turns out that every rectifiable curve admits the *arc-length* parametrization.

Theorem 4.1. *If $\gamma : [a,b] \to X$ is a rectifiable curve, then there is a unique curve $\widetilde{\gamma} : [0, \ell(\gamma)] \to X$ such that*

$$\gamma = \widetilde{\gamma} \circ s_\gamma. \tag{4.1}$$

Moreover, $\ell(\widetilde{\gamma}|_{[0,t]}) = t$ for every $t \in [0, \ell(\gamma)]$. In particular, $\widetilde{\gamma} : [0, \ell(\gamma)] \to X$ is a 1-Lipschitz mapping.

We call $\widetilde{\gamma}$ parametrized by the *arc-length* because $\ell(\widetilde{\gamma}|_{[0,t]}) = t$ for $t \in [0, \ell(\gamma)]$.

Now we are ready to define the integrals along the rectifiable curves. Let $\gamma : [a,b] \to X$ be a rectifiable curve, and let $\varrho : |\gamma| \to [0, \infty]$ be a Borel measurable function, where $|\gamma| = \gamma([a,b])$. Then we define

$$\int_\gamma \varrho := \int_0^{\ell(\gamma)} \varrho(\widetilde{\gamma}(t))\, dt,$$

where $\widetilde{\gamma} : [0, \ell(\gamma)] \to X$ is the arc-length parametrization of γ.

It turns out that we can nicely express this integral in any Lipschitz parametrization of γ.

Theorem 4.2. *For every Lipschitz curve $\gamma : [a,b] \to X$ the speed*

$$|\dot{\gamma}|(t) := \lim_{h \to 0} \frac{d(\gamma(t+h), \gamma(t))}{|h|},$$

exists almost everywhere and

$$\ell(\gamma) = \int_a^b |\dot{\gamma}|(t)\, dt. \tag{4.2}$$

Theorem 4.3. *Let $\gamma : [a,b] \to X$ be a Lipschitz curve, and let $\varrho : |\gamma| \to [0, \infty]$ be Borel measurable. Then*

$$\int_\gamma \varrho = \int_a^b \varrho(\gamma(t))|\dot\gamma|(t)\,dt.$$

4.2 Modulus

In the study of geometric properties of Sobolev functions on Euclidean spaces, the absolute continuity on almost all lines plays a crucial role. Thus, there is a need to define a notion of almost all curves also in the setting of metric spaces. This leads to the notion of the modulus of the family of rectifiable curves, which is a kind of a measure in the space of all rectifiable curves.

Let (X, d, μ) be a *metric measure space*, i.e., a metric space with a Borel measure that is positive and finite on every ball.

Let $\mathfrak{M}$ denote the family of all nonconstant rectifiable curves in X. It may happen that $\mathfrak{M} = \varnothing$, but we are interested in metric spaces for which the space $\mathfrak{M}$ is sufficiently large.

For $\varGamma \subset \mathfrak{M}$, let $F(\varGamma)$ be the family of all Borel measurable functions $\varrho : X \to [0,\infty]$ such that

$$\int_\gamma \varrho \geqslant 1 \quad \text{for every } \gamma \in \varGamma.$$

Now for each $1 \leqslant p < \infty$ we define

$$\operatorname{Mod}_p(\varGamma) = \inf_{\varrho \in F(\varGamma)} \int_X \varrho^p \, d\mu.$$

The number $\operatorname{Mod}_p(\varGamma)$ is called *p-modulus* of the family $\varGamma$.

The following result is easy to prove.

Theorem 4.4. Mod_p *is an outer measure on* $\mathfrak{M}$.

If some property holds for all curves $\gamma \in \mathfrak{M} \setminus \varGamma$, where $\operatorname{Mod}_p(\varGamma) = 0$, then we say that the property holds for *p-a.e. curve*.

The notion of p-a.e.curve is consistent with the notion of almost every line parallel to a coordinate axis. Indeed, if $E \subset [0,1]^{n-1}$ is Borel measurable and we consider straight segments passing through E

$$\varGamma_E = \{\gamma_{x'} : [0,1] \to [0,1]^n : \gamma_{x'}(t) = (t, x'),\ x' \in E\}$$

then $\operatorname{Mod}_p(\varGamma_E) = 0$ if and only if the $(n-1)$-dimensional Lebesgue measure of E is zero. This fact easily follows from the definition of the modulus and the Fubini theorem.

4.3 Upper gradient

As before, we assume that (X,d,μ) is a metric measure space. Let $u:X\to\mathbb{R}$ be a Borel function. Following [46], we say that a Borel function $g:X\to[0,\infty]$ is an *upper gradient* of u if

$$|u(\gamma(a))-u(\gamma(b))|\leqslant\int_\gamma g \tag{4.3}$$

for *every* rectifiable curve $\gamma:[a,b]\to X$. We say that g is a *p-weak upper gradient* of u if (4.3) holds on p-a.e. curve $\gamma\in\mathfrak{M}$.

If g is an upper gradient of u and $\widetilde{g}=g$, μ-a.e., is another nonnegative Borel function, then it may be that $\widetilde{g}$ is no longer upper gradient of u. However, we have the following assertion.

Lemma 4.5. *If g is a p-weak upper gradient of u and $\widetilde{g}$ is another nonnegative Borel function such that $\widetilde{g}=g$ μ-a.e., then $\widetilde{g}$ is a p-weak upper gradient of u too.*

It turns out that p-weak upper gradients can be approximated in the L^p norm by upper gradients.

Lemma 4.6. *If g is a p-weak upper gradient of u which is finite almost everywhere, then for every $\varepsilon>0$ there is an upper gradient g_ε of u such that*

$$g_\varepsilon\geqslant g \text{ everywhere} \qquad \text{and} \qquad \|g_\varepsilon-g\|_{L^p}<\varepsilon.$$

We do not require here that $g\in L^p$.

The following result shows that the notion of an upper gradient is a natural generalization of the length of the gradient to the setting of metric spaces (see also Theorem 4.10).

Proposition 4.7. *If $u\in C^\infty(\Omega)$, $\Omega\subset\mathbb{R}^n$, then $|\nabla u|$ is an upper gradient of u. This upper gradient is the least one in the sense that if $g\in L^1_{\rm loc}(\Omega)$ is another upper gradient of u, then $g\geqslant|\nabla u|$ almost everywhere.*

4.4 Sobolev spaces $N^{1,p}$

Let $\widetilde{N}^{1,p}(X,d,\mu)$, $1\leqslant p<\infty$, be the class of all L^p integrable Borel functions on X for which there exists a p-weak upper gradient in L^p. For $u\in\widetilde{N}^{1,p}(X,d,\mu)$ we define

$$\|u\|_{\widetilde{N}^{1,p}}=\|u\|_{L^p}+\inf_g\|g\|_{L^p},$$

where the infimum is taken over all p-weak upper gradients g of u.

Lemma 4.6 shows that in the definition of $\widetilde{N}^{1,p}$ and $\|\cdot\|_{\widetilde{N}^{1,p}}$, p-weak upper gradients can be replaced by upper gradients.

We define the equivalence relation in $\widetilde{N}^{1,p}$ as follows: $u \sim v$ if and only if $\|u - v\|_{\widetilde{N}^{1,p}} = 0$. Then the space $N^{1,p}(X, d, \mu)$ is defined as the quotient $\widetilde{N}^{1,p}(X, d, \mu)/\sim$ and is equipped with the norm

$$\|u\|_{N^{1,p}} := \|u\|_{\widetilde{N}^{1,p}}.$$

The space $N^{1,p}$ was introduced by Shanmugalingam [77].

Theorem 4.8. *$N^{1,p}(X, d, \mu)$, $1 \leqslant p < \infty$, is a Banach space.*

One can prove that functions $u \in N^{1,p}(X, d, \mu)$ are absolutely continuous on almost all curves in the sense that for p-a.e. $\gamma \in \mathfrak{M}$, $u \circ \widetilde{\gamma}$ is absolutely continuous, where $\widetilde{\gamma}$ is the arc-length parametrization of γ. This fact, Proposition 4.7, and the characterization of the classical Sobolev space $W^{1,p}(\Omega)$, by the absolute continuity on lines, lead to the following result.

Theorem 4.9. *If $\Omega \subset \mathbb{R}^n$ is open and $1 \leqslant p < \infty$, then*

$$N^{1,p}(\Omega, |\cdot|, \mathcal{L}^n) = W^{1,p}(\Omega)$$

and the norms are equal.

Here, we consider the space $N^{1,p}$ on Ω regarded as a metric space with respect to the Euclidean metric $|\cdot|$ and the Lebesgue measure $\mathcal{L}^n$. The following result supplements the above theorem.

Theorem 4.10. *Any function $u \in W^{1,p}(\Omega)$, $1 \leqslant p < \infty$, has a representative for which $|\nabla u|$ is a p-weak upper gradient. On the other hand, if $g \in L^1_{\mathrm{loc}}$ is a p-weak upper gradient of u, then $g \geqslant |\nabla u|$ almost everywhere.*

Both above theorems hold also when Ω is replaced by a Riemannian manifold, and also, in this case, $|\nabla u|$ is the least p-weak upper gradient of $u \in W^{1,p}$. Actually, one can prove that there always exists a minimal p-weak upper gradient.

Theorem 4.11. *For any $u \in N^{1,p}(X, d, \mu)$ and $1 \leqslant p < \infty$ there exists the least p-weak upper gradient $g_u \in L^p$ of u. It is smallest in the sense that if $g \in L^p$ is another p-weak upper gradient of u, then $g \geqslant g_u$ μ-a.e.*

4.5 Doubling measures

We say that a measure μ is *doubling* if there is a constant $C_d \geqslant 1$ (called *doubling constant*) such that $0 < \mu(2B) \leqslant C_d\mu(B) < \infty$ for every ball $B \subset X$.

We say that a metric space X is *metric doubling* if there is a constant $M > 0$ such that every ball in X can be covered by at most M balls of half the radius.

If μ is a doubling measure on X, then it easily follows that X is metric doubling. In particular, bounded sets in X are totally bounded. Hence, if X is a complete metric space equipped with a doubling measure, then bounded and closed sets are compact.

The following beautiful characterization of metric spaces supporting doubling measures was proved by Volberg and Konyagin [62, 82].

Theorem 4.12. *Let X be a complete metric space. Then there is a doubling measure on X if and only if X is metric doubling.*

The doubling condition implies a lower bound for the measure of a ball.

Lemma 4.13. *If the measure μ is doubling with the doubling constant C_d and $s = \log_2 C_d$, then*

$$\frac{\mu(B(x,r))}{\mu(B_0)} \geqslant 4^{-s}\left(\frac{r}{r_0}\right)^s \tag{4.4}$$

whenever B_0 is a ball of radius r_0, $x \in B_0$ and $r \leqslant r_0$.

The lemma easily follows from the iteration of the doubling inequality. The exponent s is sharp as the example of the Lebesgue measure shows.

Metric spaces equipped with a doubling measure are called *spaces of homogeneous type* and $s = \log_2 C_d = \log C_d / \log 2$ is called *homogeneous dimension.*

An important class of doubling measures is formed by the so-called *n-regular* measures[5], which are measures for which there are constants $C \geqslant 1$ and $s > 0$ such that $C^{-1}r^s \leqslant \mu(B(x,r)) \leqslant Cr^s$ for all $x \in X$ and $0 < r < \operatorname{diam} X$. The s-regular measures are closely related to the Hausdorff measure $\mathcal{H}^s$ since we have the following assertion.

Theorem 4.14. *If μ is an s-regular measure, then there is a constant $C \geqslant 1$ such that $C^{-1}\mu(E) \leqslant \mathcal{H}^s(E) \leqslant C\mu(E)$ for every Borel set $E \subset X$. In particular, $\mathcal{H}^s$ is s-regular too.*

The proof is based on the so-called $5r$-covering lemma.

For a locally integrable function $g \in L^1_{\mathrm{loc}}(\mu)$ we define the *Hardy–Littlewood maximal function*

$$\mathcal{M}g(x) = \sup_{r>0} \fint_{B(x,r)} |g| \, d\mu \,.$$

[5] Called also *Ahlfors–David* regular measures.

Theorem 4.15. *If μ is doubling, then*

1) $\mu(\{x : \mathcal{M}g(x) > t\}) \leqslant Ct^{-1} \int\limits_X |g| d\mu$ *for every* $t > 0$;

2) $\|\mathcal{M}g\|_{L^p} \leqslant C\|g\|_{L^p}$, *for* $1 < p < \infty$.

4.6 Other spaces of Sobolev type

There are many other definitions of Sobolev type spaces on metric spaces that we describe now (see [19, 27, 29, 32, 33]). Let (X, d, μ) be a metric measure space with a doubling measure.

Following [27], for $0 < p < \infty$ we define $M^{1,p}(X, d, \mu)$ to be the set of all functions $u \in L^p(\mu)$ for which there is $0 \leqslant g \in L^p(\mu)$ such that

$$|u(x) - u(y)| \leqslant d(x, y)(g(x) + g(y)) \quad \mu\text{-a.e.} \tag{4.5}$$

Then we set

$$\|u\|_{M^{1,p}} = \|u\|_p + \inf_g \|g\|_p$$

where the infimum is taken over the class of all g satisfying (4.5). For $p \geqslant 1$, $\|\cdot\|_{1,p}$ is a norm and $M^{1,p}(X, d, \mu)$ is a Banach space.

For a locally integrable function u we define the *Calderón maximal function*

$$u_1^{\#}(x) = \sup_{r>0} r^{-1} \fint_{B(x,r)} |u - u_B| \, d\mu \, .$$

Following [32], we define $C^{1,p}(X, d, \mu)$ to be the class of all $u \in L^p(\mu)$ such that $u_1^{\#} \in L^p(\mu)$. Again, for $p \geqslant 1$, $C^{1,p}(X, d, \mu)$ is a Banach space with respect to the norm

$$\|u\|_{C^{1,p}} = \|u\|_p + \|u_1^{\#}\|_p \, .$$

Following [33], for $0 < p < \infty$ we say that a locally integrable function $u \in L^1_{\text{loc}}$ belongs to the space $P^{1,p}(X, d, \mu)$ if there are $\sigma \geqslant 1$ and $0 \leqslant g \in L^p(\mu)$ such that

$$\fint_B |u - u_B| \, d\mu \leqslant r \left(\fint_{\sigma B} g^p \, d\mu \right)^{1/p} \qquad \text{for every ball } B \text{ of radius } r. \tag{4.6}$$

We do not equip the space $P^{1,p}$ with a norm.

To motivate the above definitions, we observe that $u \in W^{1,p}(\mathbb{R}^n)$ satisfies the pointwise inequality

$$|u(x) - u(y)| \leqslant C|x - y|(\mathcal{M}|\nabla u|(x) + \mathcal{M}|\nabla u|(y)) \quad \text{a.e.},$$

where $\mathcal{M}|\nabla u|$ is the Hardy–Littlewood maximal function, so $g = \mathcal{M}|\nabla u| \in L^p$ for $p > 1$, and actually one can prove [27] that $u \in W^{1,p}(\mathbb{R}^n)$, $p > 1$, if and only if $u \in L^p$ and there is $0 \leqslant g \in L^p$ such that $|u(x) - u(y)| \leqslant |x-y|(g(x)+g(y))$ almost everywhere. Moreover, $\|u\|_{1,p} \approx \|u\|_p + \inf_g \|g\|_p$. Thus, for $p > 1$, $W^{1,p}(\mathbb{R}^n) = M^{1,p}(\mathbb{R}^n)$. In the case $p = 1$, $M^{1,1}(\mathbb{R}^n)$ is not equivalent with $W^{1,1}(\mathbb{R}^n)$ [28] (see, however, [57] and Theorem 4.16 below).

The classical Poincaré inequality

$$\fint_{B(x,r)} |u - u_{B(x,r)}| \leqslant Cr \fint_{B(x,r)} |\nabla u| \tag{4.7}$$

implies that for $u \in W^{1,p}(\mathbb{R}^n)$ the Calderón maximal function is bounded by the maximal function of $|\nabla u|$ and hence it belongs to L^p for $p > 1$. Calderón [10] proved that for $p > 1$, $u \in W^{1,p}(\mathbb{R}^n)$ if and only if $u \in L^p$ and $u_1^\# \in L^p$. Moreover, $\|u\|_{1,p} \approx \|u\|_p + \|u_1^\#\|_p$. Thus, for $p > 1$, $W^{1,p}(\mathbb{R}^n) = C^{1,p}(\mathbb{R}^n)$.

The inequality (4.7) also implies that for $p \geqslant 1$, $\sigma \geqslant 1$, and $u \in W^{1,p}(\mathbb{R}^n)$ we have

$$\fint_B |u - u_B|\, dx \leqslant Cr \left(\fint_{\sigma B} |\nabla u|^p \, dx \right)^{1/p} .$$

Thus, $W^{1,p}(\mathbb{R}^n) \subset P^{1,p} \cap L^p$. On the other hand, it was proved in [56, 19, 28] that $W^{1,p}(\mathbb{R}^n) = P^{1,p} \cap L^p$ for $p \geqslant 1$.

In the case of general metric spaces, we have the following assertion.

Theorem 4.16. *If the measure μ is doubling and $1 \leqslant p < \infty$, then*

$$C^{1,p}(X,d,\mu) = M^{1,p}(X,d,\mu) \subset P^{1,p}(X,d,\mu) \cap L^p(\mu) \subset N^{1,p}(X,d,\mu) .$$

For a proof see [29, Corollary 10.5 and Theorem 9.3], [32, Theorem 3.4 and Lemma 3.6], and [71].

The so-called telescoping argument (infinite iteration of the inequality (4.6) on a decreasing sequence of balls) shows that if $u \in P^{1,p}(X,d,\mu)$, then

$$|u(x) - u(y)| \leqslant Cd(x,y)((\mathcal{M}g^p(x))^{1/p} + (\mathcal{M}g^p(y))^{1/p}) \quad \text{a.e.} \tag{4.8}$$

(see [33]). A version of the same telescoping argument shows also that for $u \in L^1_{\mathrm{loc}}$

$$|u(x) - u(y)| \leqslant Cd(x,y)(u_1^\#(x) + u_1^\#(y)) \quad \text{a.e.}$$

(see [32, Lemma 3.6]). This implies that $C^{1,p} \subset M^{1,p}$ for $p \geqslant 1$. On the other hand, if $u \in M^{1,p}$ and $|u(x) - u(y)| \leqslant d(x,y)(g(x) + g(y))$, then a direct integration with respect to x and y yields

$$\fint_B |u - u_B| \, d\mu \leqslant 4r \fint_B g \, d\mu \leqslant 4r \left(\fint_B g^p \, d\mu \right)^{1/p}.$$

Hence $M^{1,p} \subset P^{1,p} \cap L^p$ and also

$$u_1^{\#} \leqslant 4\mathcal{M}g,$$

which shows that $M^{1,p} \subset C^{1,p}$ for $p > 1$. Thus, $C^{1,p} = M^{1,p}$ for $p > 1$. The case $p = 1$ of this equality is more difficult (see [29, Theorem 9.3] and [71]).

For the proof of the remaining inclusion $P^{1,p} \cap L^p \subset N^{1,p}$ see [29, Corollary 10.5].

If a metric space X has no nonconstant rectifiable curves, then $g = 0$ is an upper gradient of any $u \in L^p$ and hence $N^{1,p}(X, d, \mu) = L^p(\mu)$. On the other hand, the theory of Sobolev spaces $M^{1,p}$, $C^{1,p}$, and $P^{1,p}$ is not trivial in this case. Indeed, a variant of the above telescoping argument leads to the estimate of $|u - u_B|$ by a generalized Riesz potential [33], and hence the fractional integration theorem implies Sobolev embedding theorems. Many results of the classical theory of Sobolev spaces extend to this situation (see, for example, [27, 33, 29]), and we state just one of them.

Theorem 4.17. *Let μ be a doubling measure, and let $s = \log C_d / \log 2$ be the same as in Lemma 4.13. If $u \in L^1_{\text{loc}}(\mu)$, $\sigma \geqslant 1$, and $0 \leqslant g \in L^p(\mu)$, $0 < p < s$ are such that the p-Poincaré inequality*

$$\fint_B |u - u_B| \, d\mu \leqslant r \left(\fint_{\sigma B} g^p \, d\mu \right)^{1/p}$$

holds on every ball B of radius r, then for any $p < q < s$ the Sobolev–Poincaré inequality

$$\left(\fint_B |u - u_B|^{q^*} \, d\mu \right)^{1/q^*} \leqslant Cr \left(\fint_{5\sigma B} g^q \, d\mu \right)^{1/q}$$

holds on every ball B of radius r, where $q^* = sq/(s - q)$ is the Sobolev exponent.

This result implies Sobolev embedding for the spaces $C^{1,p}$, $M^{1,p}$, and $P^{1,p}$, but not for $N^{1,p}$.

Other results for $C^{1,p}$, $M^{1,p}$, and $P^{1,p}$ spaces available in the general case of metric spaces with doubling measure include Sobolev embedding into Hölder continuous functions, Trudinger inequality, compact embedding theorem, embedding on spheres, and extension theorems.

4.7 Spaces supporting the Poincaré inequality

Metric spaces equipped with doubling measures are too general for the theory of $N^{1,p}$ spaces to be interesting. Indeed, if there are no nonconstant rectifiable curves in X, then, as we have already observed, $N^{1,p}(X,d,\mu) = L^p(\mu)$. Thus, we need impose additional conditions on the metric space that will imply, in particular, the existence of many rectifiable curves. Such a condition was discovered by Heinonen and Koskela [46].

We say that (X,d,μ) *supports a p-Poincaré inequality*, $1 \leqslant p < \infty$, if the measure μ is doubling and there exist constants C_P and $\sigma \geqslant 1$ such that for every ball $B \subset X$, every Borel measurable function $u \in L^1(\sigma B)$, and every upper gradient $0 \leqslant g \in L^p(\sigma B)$ of u on σB the following Poincaré type inequality is satisfied:

$$\fint_B |u - u_B|\, d\mu \leqslant C_P r \left(\fint_{\sigma B} g^p\, d\mu \right)^{1/p} . \tag{4.9}$$

Note that this condition immediately implies the existence of rectifiable curves. Indeed, if u is not constant, then $g = 0$ cannot be an upper gradient of u; otherwise, the inequality (4.9) would not be satisfied. More precisely, we have the following assertion (see, for example, [33, Proposition 4.4]).

Theorem 4.18. *If a space X supports a p-Poincaré inequality, then there is a constant $C > 0$ such that any two points $x, y \in X$ can be connected by a curve of length less than or equal to $Cd(x,y)$.*

Clearly, $\mathbb{R}^n$ supports the p-Poincaré inequality for all $1 \leqslant p < \infty$. Another example of spaces supporting Poincaré inequalities is provided by Riemannian manifolds of nonnegative Ricci curvature [8, 72]. There are, however, many examples of spaces supporting Poincaré inequalities which carry some mild geometric structure, but do not resemble Riemannian manifolds [7, 45, 46, 59, 60, 75]. An important class of spaces that support the p-Poincaré inequality is provided by the so-called *Carnot groups* [23, 68, 9] and more general Carnot–Carathéodory spaces [22, 23]. For the sake of simplicity, only the simplest case of the Heisenberg group is described here.

The Heisenberg group $\mathbb{H}_1$ can be identified with $\mathbb{R}^3 \equiv \mathbb{C} \times \mathbb{R}$ equipped with the noncommutative group law $(z_1, t_1) \cdot (z_2, t_2) = (z_1 + z_2, t_1 + t_2 + 2\mathrm{Im}\,(z_1 \overline{z}_2))$. It is equipped with a non-Riemannian metric $d(x,y) = \|a^{-1} \cdot b\|$, where $\|(z,t)\| = (|z|^4 + t^2)^{1/2}$. This metric is bi-Lipschitz equivalent to another so-called Carnot–Carathéodory metric. The metric d is quite exotic because the Hausdorff dimension of $(\mathbb{H}_1, d)$ is 4, while topological dimension is 3. The applications of the Heisenberg group include several complex variables, subelliptic equations and noncommutative harmonic analysis [78]. More recently, it was a subject of an intense study from the perspective of geometric measure theory [21, 9].

If a space (X, d, μ) supports the p-Poincaré inequality, $u \in N^{1,p}(X, d, \mu)$, and $0 \leqslant g \in L^p(\mu)$ is an upper gradient of u, then the p-Poincaré inequality (4.9) is satisfied and hence the Sobolev embedding (Theorem 4.17) holds. One can actually prove that, in this case, we can take $q = p$, i.e.,

$$\left(\fint_B |u - u_B|^{p^*} \, d\mu \right)^{1/p^*} \leqslant Cr \left(\fint_{5\sigma B} g^p \, d\mu \right)^{1/p}$$

on every ball B of radius r, where $1 \leqslant p < s$ and $p^* = sp/(s-p)$ (see [33]).

A direct application of the Hölder inequality shows that if a space supports a p-Poincaré inequality, then it also supports a q-Poincaré inequality for all $q > p$. On the other hand, we have the following important result of Keith and Zhong [54].

Theorem 4.19. *If a complete metric measure space supports a p-Poincaré inequality for some $p > 1$, then it also supports a q-Poincaré inequality for some $1 \leqslant q < p$.*

This important result implies that, in the case of spaces supporting the p-Poincaré inequality, other approaches to Sobolev spaces described in the previous section are equivalent.

Theorem 4.20. *If the space supports the p-Poincaré inequality, $1 < p < \infty$, then $C^{1,p}(X, d, \mu) = M^{1,p}(X, d, \mu) = P^{1,p}(X, d, \mu) \cap L^p(\mu) = N^{1,p}(X, d, \mu)$.*

Indeed, prior to the work of Keith and Zhong it was known that the spaces are equal provided that the space supports the q-Poincaré inequality for some $1 \leqslant q < p$ (see, for example, [29, Theorem 11.3]).

Spaces supporting Poincaré inequalities play a fundamental role in the modern theory of quasiconformal mappings [46, 47], geometric rigidity problems [7], nonlinear subelliptic equations (see, for example, [11, 22, 20, 33, 34]), and nonlinear potential theory [1, 6].

Although the known examples show that spaces supporting a Poincaré inequality can be very exotic, surprisingly, one can prove that such spaces are always equipped with a weak differentiable structure [13, 53].

5 Sobolev Mappings between Metric Spaces

Throughout this section, we assume that (X, d, μ) is a metric measure space equipped with a doubling measure. Let Y be another metric space. The construction of the space of Sobolev mappings between metric spaces $N^{1,p}(X, Y)$ is similar to that in Sect. 3 with the difference that the classical Sobolev space is replaced by the Sobolev space $N^{1,p}$. The space $N^{1,p}(X, Y)$ was introduced in [47].

Let V be a Banach space. Following [47], we say that $F \in \widetilde{N}^{1,p}(X,V)$ if $F \in L^p(X,V)$ (in the Bochner sense) and there is a Borel measurable function $0 \leqslant g \in L^p(\mu)$ such that

$$\|F(\gamma(a)) - F(\gamma(b))\| \leqslant \int\limits_\gamma g$$

for every rectifiable curve $\gamma : [a,b] \to X$. We call g an *upper gradient* of F. We also define

$$\|F\|_{1,p} = \|F\|_p + \inf_g \|g\|_p,$$

where the infimum is taken over all upper gradients of F. Now we define $N^{1,p}(X,V) = \widetilde{N}^{1,p}(X,V)/\sim$, where $F_1 \sim F_2$ when $\|F_1 - F_2\|_{1,p} = 0$.

As in the case of $N^{1,p}(X,d,\mu)$ spaces, the p-upper gradient can be replaced by p-weak upper gradient in the above definition. The following two results were proved in [47] (see also [31] for Theorem 5.2).

Theorem 5.1. *$N^{1,p}(X,V)$ is a Banach space.*

Theorem 5.2. *Suppose that the space (X,d,μ) supports the p-Poincaré inequality for some $1 \leqslant p < \infty$. Then for every Banach space V the pair (X,V) supports the p-Poincaré inequality in the following sense*: *there is a constant $C > 0$ such that for every ball $B \subset X$, for every $F \in L^1(6\sigma B, V)$, and for every $0 \leqslant g \in L^p(6\sigma B)$ being a p-weak upper gradient of F on $6\sigma B$ the following inequality is satisfied*:

$$⨍\limits_B \|F - F_B\|\, d\mu \leqslant C(\operatorname{diam} B) \left(⨍\limits_{6\sigma B} g^p\, d\mu \right)^{1/p}. \tag{5.1}$$

The Poincaré inequality (5.1) and the standard telescoping argument implies the following pointwise inequality: if $F \in N^{1,p}(X,V)$ and $0 \leqslant g \in L^p(\mu)$ is a p-weak upper gradient of F, then

$$\|F(x) - F(y)\| \leqslant C d(x,y)((\mathcal{M}g^p(x))^{1/p} + (\mathcal{M}g^p(y))^{1/p})$$

almost everywhere, where, on the right-hand side, we have the maximal function, just like in the case of the equality (4.8).

In particular, F restricted to the set $E_t = \{x : \mathcal{M}g^p < t^p\}$ is Lipschitz continuous with the Lipschitz constant Ct. Using the Lipschitz extension of $F|_{E_t}$ to the entire space X (McShane extension), one can prove [31] the following assertion.

Theorem 5.3. *Suppose that the space (X,d,μ) supports the p-Poincaré inequality for some $1 \leqslant p < \infty$ and V is a Banach space. If $F \in N^{1,p}(X,V)$,* *then for every* $\varepsilon > 0$ *there is a Lipschitz mapping* $G \in \mathrm{Lip}\,(X, V)$ *such that* $\mu\{x : F(x) \neq G(x)\} < \varepsilon$ *and* $\|F - G\|_{1,p} < \varepsilon$.

As we have seen in the previous section, the Poincaré inequality plays a crucial role in the development of the theory of Sobolev spaces on metric spaces. Since such an inequality is also valid for $N^{1,p}(X, V)$ spaces, Theorem 5.2, many results true for $N^{1,p}(X, d, \mu)$ like, for example, Sobolev embedding theorems can be generalized to $N^{1,p}(X, V)$ spaces (see [47]).

Now, if Y is a metric space isometrically embedded into a Banach space V, $Y \subset V$, we define

$$N^{1,p}(X, Y) = \{F \in N^{1,p}(X, V) : F(X) \subset Y\}.$$

Since $N^{1,p}(X, V)$ is a Banach space, $N^{1,p}(X, Y)$ is equipped with a norm metric.

If X is an open set in $\mathbb{R}^n$ or X is a compact manifold, then the space $N^{1,p}(X, d, \mu)$ is equivalent with the classical Sobolev space (see Theorem 4.9). Hence, in this case, the definition of $N^{1,p}(\Omega, Y)$ (or $N^{1,p}(M, Y)$) is equivalent with that of $R^{1,p}(\Omega, Y)$ (or $R^{1,p}(M, Y)$) described in Sect. 3 (see [47]).

If $F \in N^{1,p}(X, Y)$, then, according to Theorem 5.3, F can be approximated by Lipschitz mappings $\mathrm{Lip}\,(X, V)$ and the question is: Under what conditions F can be approximated by $\mathrm{Lip}\,(X, Y)$ mappings?

This is a question about extension of the theory described in Sect. 2 to the case of Sobolev mappings between metric spaces and it was formulated explicitly by Heinonen, Koskela, Shanmugalingam, and Tyson [47, Remark 6.9].

An answer to this question cannot be easy because, as soon as we leave the setting of manifolds, we have many unpleasant counterexamples like those in Sect. 3. A particularly dangerous situation is created by the lack of stability with respect to bi-Lipschitz deformations of the target (Theorem 3.10). Indeed, in most situations, there is no canonical way to choose a metric on Y and we are free to choose any metric in the class of bi-Lipschitz equivalent metrics.

An example of spaces supporting the p-Poincaré inequality is provided by the Heisenberg group and, more generally, Carnot groups and Carnot–Carathéodory spaces. In this setting, Gromov [23, Sect. 2.5.E] stated as an open problem the extension of the results from Sect. 2 to the case of mappings from Carnot–Carathéodory spaces to Riemannian manifolds. Thus, the question of Heinonen, Koskela, Shanmugalingam, and Tyson can be regarded and a more general form of Gromov's problem.

The following result was proved in [31] (see Theorem 3.9 above).

Theorem 5.4. *Suppose that* (X, d, μ) *is a doubling metric measure space of finite measure* $\mu(X) < \infty$ *and* Y_1, Y_2 *are two bi-Lipschitz homeomorphic metric spaces of finite diameter isometrically embedded into Banach spaces* V_1 *and* V_2 *respectively. Suppose that Lipschitz mappings* $\mathrm{Lip}\,(X, Y_1)$ *are dense in* $N^{1,p}(X, Y_1)$, $1 \leqslant p < \infty$, *in the following strong sense*: *for any* $f \in$

$N^{1,p}(X,Y_1)$ and $\varepsilon > 0$ there is $f_\varepsilon \in \mathrm{Lip}\,(X,Y_1)$ such that $\mu(\{x : f(x) \neq f_\varepsilon(x)\}) < \varepsilon$ and $\|f - f_\varepsilon\|_{1,p} < \varepsilon$. Then the Lipschitz mappings $\mathrm{Lip}\,(X,Y_2)$ are dense in $N^{1,p}(X,Y_2)$.

This result shows that, in the case in which we can prove strong density, there is no problem with the bi-Lipschitz invariance of the density.

It turns out that also White's theorem (Theorem 2.5) and the density result of Schoen and Uhlenbeck (Theorems 2.1 and 3.7) can be generalized to the setting of mappings between metric spaces. Theorem 5.4 plays a crucial role in the proof.

Theorem 5.5. *Let (X,d,μ) be a metric measure space of finite measure $\mu(X) < \infty$ supporting the p-Poincaré inequality. If $p \geqslant s = \log C_d/\log 2$ and Y is a compact metric doubling space which is bi-Lipschitz homeomorphic to a Lipschitz neighborhood retract of a Banach space, then for every isometric embedding of Y into a Banach space Lipschitz mappings $\mathrm{Lip}\,(X,Y)$ are dense in $N^{1,p}(X,Y)$. Moreover, for every $f \in N^{1,p}(X,Y)$ there is $\varepsilon > 0$ such that if $f_1, f_2 \in \mathrm{Lip}\,(X,Y)$ satisfy $\|f - f_i\|_{1,p} < \varepsilon$, $i = 1,2$, then the mappings f_1 and f_2 are homotopic.*

5.1 Lipschitz polyhedra

By a *simplicial complex* we mean a finite collection K of simplexes in some Euclidean space $\mathbb{R}^\nu$ such that

1) if $\sigma \in K$ and τ is a face of σ, then $\tau \in K$;

2) if $\sigma, \tau \in K$, then either $\sigma \cap \tau = \varnothing$ or $\sigma \cap \tau$ is a common face of σ and τ.

The set $|K| = \bigcup_{\sigma \in K} \sigma$ is called a *rectilinear polyhedron.* By a *Lipschitz polyhedron* we mean any metric space which is bi-Lipschitz homeomorphic to a rectilinear polyhedron. The main result of [31] reads as follows.

Theorem 5.6. *Let Y be a Lipschitz polyhedron, and let $1 \leqslant p < \infty$. Then the class of Lipschitz mappings $\mathrm{Lip}\,(X,Y)$ is dense in $N^{1,p}(X,Y)$ for every metric measure space X of finite measure that supports the p-Poincaré inequality if and only if $\pi_1(Y) = \pi_2(Y) = \ldots = \pi_{[p]}(Y) = 0$.*

Observe that the density of Lipschitz mappings does not depend on the particular choice of the metric in Y in the class of bi-Lipschitz equivalent metrics, only on the topology of Y. This is because, in the proof of Theorem 5.6, one shows the strong approximation property described in Theorem 5.4. Theorem 5.6 can be regarded as a partial answer to the problems of Heinonen, Koskela, Shanmugalingam, and Tyson and also to the problem of Gromov.

Acknowledgement. The work was supported by NSF (grant DMS-0500966).

References

1. Aikawa, H., Shanmugalingam, N.: Hölder estimates of p-harmonic extension operators. J. Differ. Equations **220**, 18–45 (2006)
2. Ambrosio, L.: Metric space valued functions of bounded variation. Ann. Sc. Norm. Super. Pisa, Cl. Sci. **17**, 439–478 (1990)
3. Bethuel, F.: The approximation problem for Sobolev maps between two manifolds. Acta Math. **167**, 153–206 (1991)
4. Bethuel, F.: On the singular set of stationary harmonic maps. Manuscr. Math. **78**, 417–443 (1993)
5. Bethuel, F., Zheng, X.M.: Density of smooth functions between two manifolds in Sobolev spaces. J. Funct. Anal. **80**, 60–75 (1988)
6. Björn, A., Björn, J., Shanmugalingam, N.: The Perron method for p-harmonic functions in metric spaces. J. Differ. Equations **195**, 398–429 (2003)
7. Bourdon, M., Pajot, H.: Poincaré inequalities and quasiconformal structure on the boundary of some hyperbolic buildings. Proc. Am. Math. Soc. **127**, 2315–2324 (1999)
8. Buser, P.: A note on the isoperimetric constant. Ann. Sci. École Norm. Sup. (4) **15**, 213–230 (1982)
9. Capogna, L., Danielli, D., Pauls, S.D., Tyson, J.T.: An Introduction to the Heisenberg Group and the Sub-Riemannian Isoperimetric Problem. Birkhäuser, Basel (2007)
10. Calderón, A.P.: Estimates for singular integral operators in terms of maximal functions. Stud. Math. **44** , 563–582 (1972)
11. Capogna, L., Garofalo, N.: Regularity of minimizers of the calculus of variations in Carnot groups via hypoellipticity of systems of Hörmander type. J. Eur. Math. Soc. (JEMS) **5**, 1–40 (2003)
12. Capogna, L., Lin, F.-H.: Legendrian energy minimizers. I. Heisenberg group target. Calc. Var. Partial Differ. Equ. **12** , 145–171 (2001)
13. Cheeger, J.: Differentiability of Lipschitz functions on metric measure spaces. Geom. Funct. Anal. **9**, 428–517 (1999)
14. Coifman, R., Weiss, G.: Analyse Harmonique sur Certains Espaces Homogenes. Lect. Notes Math. **242**, Springer (1971)
15. Diestel, J., Uhl, J.J., Jr.: Vector Measures. Am. Math. Soc., Providence, RI (1977)
16. Duzaar, F., Mingione, G.: The p-harmonic approximation and the regularity of p-harmonic maps. Calc. Var. Partial Differ. Equ. **20**, 235–256 (2004)
17. Eells, J., Fuglede, B.: Harmonic Maps between Riemannian Polyhedra. Cambridge Univ. Press, Cambridge (2001)
18. Eells, J., Lemaire, L.: A report on harmonic maps. Bull. London Math. Soc. **10**, 1–68 (1978)
19. Franchi, B.; Hajłasz, P.; Koskela, P.: Definitions of Sobolev classes on metric spaces. Ann. Inst. Fourier (Grenoble) **49**, 1903–1924 (1999)
20. Franchi, B., Gutiérrez, C.E., Wheeden, R.L.: Weighted Sobolev–Poincaré inequalities for Grushin type operators. Commun. Partial Differ. Equ. **19**, 523–604 (1994)
21. Franchi, B., Serapioni, R., Serra Cassano, F.: Rectifiability and perimeter in the Heisenberg group. Math. Ann. **321**, 479–531 (2001)
22. Garofalo, N., Nhieu, D.-M.: Isoperimetric and Sobolev inequalities for Carnot-Carathéodory spaces and the existence of minimal surfaces. Commun. Pure Appl. Math. **49**, 1081–1144 (1996)

23. Gromov, M.: Carnot-Carathéodory spaces seen from within. In: Sub-Riemannian Geometry, pp. 79–323, Progr. Math. **144**, Birkhäuser, Basel (1996)
24. Gromov, M., Schoen, R.: Harmonic maps into singular spaces and p-adic superrigidity for lattices in groups of rank one. Inst. Hautes Études Sci. Publ. Math. **76**, 165–246 (1992)
25. Hajłasz, P.: Equivalent statement of the Poincaré conjecture, Ann. Mat. Pura Appl. **167**, 25–31 (1994)
26. Hajłasz, P.: Approximation of Sobolev mappings. Nonlinear Anal., Theory Methods Appl. **22**, 1579–1591 (1994)
27. Hajłasz, P.: Sobolev spaces on an arbitrary metric space. Potential Anal. **5**, 403–415 (1996)
28. Hajłasz, P.: A new characterization of the Sobolev space. Stud. Math. **159**, 263–275 (2003)
29. Hajłasz, P.: Sobolev spaces on metric measure spaces. In: Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces (Paris, 2002), pp. 173–218, Contemp. Math. **338**, Am. Math. Soc., Providence, RI (2003)
30. Hajłasz, P.: Sobolev mappings: Lipschitz density is not a bi-Lipschitz invariant of the target. Geom. Funct. Anal. **17**, 435–467 (2007)
31. Hajłasz, P. Density of Lipschitz mappings in the class of Sobolev mappings between metric spaces. Math. Ann. [To appear]
32. Hajłasz, P., Kinnunen, J.: Hölder quasicontinuity of Sobolev functions on metric spaces. Rev. Mat. Iberoam. **14**, 601–622 (1998)
33. Hajłasz, P., Koskela, P.: Sobolev met Poincaré. Mem. Am. Math. Soc. no. 688 (2000)
34. Hajłasz, P., Strzelecki, P.: Subelliptic p-harmonic maps into spheres and the ghost of Hardy spaces, Math. Ann. **312**, 341–362 (1998)
35. Hajłasz, P., Tyson, J.: Sobolev Peano cubes. Michigan Math. J. [To appear]
36. Hang, F.: On the weak limits of smooth maps for the Dirichlet energy between manifolds. Commun. Anal. Geom. **13**, 929–938 (2005)
37. Hang, F.: Density problems for $W^{1,1}(M,N)$. Commun. Pure Appl. Math. **55**, 937–947 (2002)
38. Hang, F., Lin, F.: Topology of Sobolev mappings. Math. Res. Lett. **8**, 321–330 (2001)
39. Hang, F., Lin, F.: Topology of Sobolev mappings II. Acta Math. **191**, 55–107 (2003)
40. Hang, F., Lin, F.: Topology of Sobolev mappings. III. Commun. Pure Appl. Math. **56**, 1383–1415 (2003)
41. Hardt, R.: Singularities of harmonic maps. Bull. Am. Math. Soc. (N.S.) **34**, 15–34 (1997)
42. Hardt, R., Lin, F.-H.: Mappings minimizing the L^p norm of the gradient. Commun. Pure Appl. Math. **40**, no. 5, 555–588 (1987)
43. Heinonen, J.: Lectures on Analysis on Metric Spaces. Springer-Verlag, New York (2001)
44. Heinonen, J.: Nonsmooth calculus. Bull. Am. Math. Soc. (N.S.) **44**, 163–232 (2007)
45. Hanson, B., Heinonen, J. An n-dimensional space that admits a Poincaré inequality but has no manifold points. Proc. Am. Math. Soc. **128**, 3379–3390 (2000)
46. Heinonen, J., Koskela, P.: Quasiconformal maps in metric spaces with controlled geometry. Acta Math. **181**, 1–61 (1998)
47. Heinonen, J., Koskela, P., Shanmugalingam, N., Tyson, J.T.: Sobolev classes of Banach space-valued functions and quasiconformal mappings. J. Anal. Math. **85**, 87–139 (2001)
48. Hélein, F.: Harmonic Maps, Conservation Laws and Moving Frames. Cambridge Univ. Press, Cambridge (2002)
49. Jost, J.: Equilibrium maps between metric spaces. Calc. Var. Partial Differ. Equ. **2**, 173–204 (1994)

50. Jost, J.: Generalized Dirichlet forms and harmonic maps. Calc. Var. Partial Differ. Equ. **5**, 1–19 (1997)
51. Jost, J.: Nonlinear Dirichlet forms. In: New Directions in Dirichlet Forms. pp. 1–47, AMS/IP Stud. Adv. Math. **8**, Am. Math. Soc., Providence, RI (1998)
52. Jost, J., Zuo, K.: Harmonic maps into Bruhat-Tits buildings and factorizations of p-adically unbounded representations of π_1 of algebraic varieties. I. J. Algebraic Geom. **9**, 1–42 (2000)
53. Keith, S.: A differentiable structure for metric measure spaces. Adv. Math. **183**, 271–315 (2004)
54. Keith, S., Zhong, X.: The Poincaré inequality is an open ended condition. Ann. Math. **167**, 575–599 (2008)
55. Korevaar, N.J., Schoen, R.M.: Sobolev spaces and harmonic maps for metric space targets. Commun. Anal. Geom. **1**, 561–659 (1993)
56. Koskela, P., MacManus, P.: Quasiconformal mappings and Sobolev spaces. Stud. Math. **131**, 1–17 (1998)
57. Koskela P., Saksman, E.: Pointwise characterizations of Hardy–Sobolev functions. Math. Res. Lett. [To appear]
58. Kronz, M.: Some function spaces on spaces of homogeneous type. Manuscr. Math. **106**, 219–248 (2001)
59. Laakso, T.J.: Ahlfors Q-regular spaces with arbitrary $Q > 1$ admitting weak Poincaré inequality. Geom. Funct. Anal. **10**, 111–123 (2000)
60. Laakso, T.J.: Erratum to: "Ahlfors Q-regular spaces with arbitrary $Q > 1$ admitting weak Poincaré inequality", GAFA 10 (2000). Geom. Funct. Anal. **12**, 650 (2002)
61. Lindenstrauss, J., Preiss, D.: On Fréchet differentiability of Lipschitz maps between Banach spaces. Ann. Math. (2) **157**, 257–288 (2003)
62. Luukkainen, J., Saksman, E.: Every complete doubling metric space carries a doubling measure. Proc. Am. Math. Soc. **126**, 531–534 (1998)
63. Luukkainen, J., Väisälä, J.: Elements of Lipschitz topology. Ann. Acad. Sci. Fenn. Ser. A I Math. **3**, 85–122 (1977)
64. Marcus, M., Mizel, V.J.: Every superposition operator mapping one Sobolev space into another is continuous. J. Funct. Anal. **33**, 217–229 (1979)
65. Mou, L., Yang, P.: Regularity for n-harmonic maps. J. Geom. Anal. **6**, 91–112 (1996)
66. Pakzad, M.R.: Weak density of smooth maps in $W^{1,1}(M, N)$ for non-Abelian $\pi_1(N)$. Ann. Global Anal. Geom. **23**, 1–12 (2003)
67. Pakzad, M.R., Rivière, T.: Weak density of smooth maps for the Dirichlet energy between manifolds. Geom. Funct. Anal. **13**, 223–257 (2003)
68. Pansu, P.: Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un. Ann. Math. **129**, 1–60 (1989)
69. Rivière, T.: Conservation laws for conformally invariant variational problems. Invent. Math. **168**, 1–22 (2007)
70. Reshetnyak, Yu. G.: Sobolev classes of functions with values in a metric space (Russian). Sib. Mat. Zh. **38**, 657–675 (1997); English transl. Sib. Math. J. **38**, 567–583 (1997)
71. Sadhu, S.: A characterization of Sobolev spaces on metric measure spaces. [In preparation]
72. Saloff-Coste, L.: Aspects of Sobolev Type Inequalities. Cambridge Univ. Press, Cambridge (2002)
73. Schoen, R., Uhlenbeck, K.: The Dirichlet problem for harmonic maps. J. Differ. Geom. **18** 253–268 (1983)
74. Schoen, R., Uhlenbeck, K.: Approximation theorems for Sobolev mappings [Unpublished manuscript]

75. Semmes, S.: Finding curves on general spaces through quantitative topology, with applications to Sobolev and Poincaré inequalities. Sel. Math., New Ser. **2** 155–295 (1996)
76. Serbinowski, T.: Boundary regularity of harmonic maps to nonpositively curved metric spaces. Commun. Anal. Geom. **2**, 139–153 (1994)
77. Shanmugalingam, N.: Newtonian spaces: an extension of Sobolev spaces to metric measure spaces. Rev. Mat. Iberoamericana **16**, 243–279 (2000)
78. Stein, E.M.: Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals. Princeton Univ. Press, Princeton, NJ (1993)
79. Strzelecki, P.: Regularity of generalized sphere valued p-harmonic maps with small mean oscillations. Manuscr. Math. **112**, 473–487 (2003)
80. Toro, T., Wang, C.: Compactness properties of weakly p-harmonic maps into homogeneous spaces. Indiana Univ. Math. J. **44**, 87–113 (1995)
81. Troyanov, M.: Approximately Lipschitz mappings and Sobolev mappings between metric spaces. In: Proceedings on Analysis and Geometry. Novosibirsk, 1999. pp. 585–594. Izdat. Ross. Akad. Nauk Sib. Otd. Inst. Mat., Novosibirsk, 2000.
82. Volberg, A.L., Konyagin, S.V.: On measures with the doubling condition (Russian). Izv. Acad. Nauk SSSR **51**, 666–675 (1987); English transl.: Math. USSR Izv. **30**, 629–638 (1988)
83. Wang, C.: Subelliptic harmonic maps from Carnot groups. Calc. Var. Partial Differ. Equ. **18**, 95–115 (2003)
84. Wang, C.: A compactness theorem of n-harmonic maps. Ann. Inst. H. Poincaré Anal. Non Linéaire **22**, 509–519 (2005)
85. White B.: Infima of energy functionals in homotopy classes of mappings. J. Differ. Geom. **23**, 127–142 (1986)

A Collection of Sharp Dilation Invariant Integral Inequalities for Differentiable Functions

Vladimir Maz'ya and Tatyana Shaposhnikova

In memory of S.L. Sobolev

Abstract We find best constants in several dilation invariant integral inequalities involving derivatives of functions. Some of these inequalities are new and some were known without best constants. In particular, we deal with an estimate for a quadratic form of the gradient, weighted Gårding inequality for the biharmonic operator, dilation invariant Hardy's inequalities with remainder term, a generalized Hardy–Sobolev inequality with sharp constant, and the Hardy inequality with sharp Sobolev remainder term.

1 Introduction

The classical integral inequality

$$\|u\|_{L^q(\mathbb{R}^n)} \leqslant C\,\|\nabla_l u\|_{L^p(\mathbb{R}^n)}, \tag{1.1}$$

where u is an arbitrary function in $C_0^\infty(\mathbb{R}^n)$, obtained by Sobolev [48] for

$$n > lp, \quad p > 1, \quad q = pn(n-lp)^{-1},$$

Vladimir Maz'ya
Ohio State University, Columbus, OH 43210 USA; University of Liverpool, Liverpool L69 7ZL, UK; Linköping University, Linköping SE-58183, Sweden, e-mail: vlmaz@mai.liu.se, vlmaz@math.ohio-state.edu

Tatyana Shaposhnikova
Ohio State University, Columbus, OH 43210 USA; Linköping University, Linköping SE-58183, Sweden, e-mail: tasha@mai.liu.se

V. Maz'ya (ed.), *Sobolev Spaces in Mathematics I, International Mathematical Series.*
doi: 10.1007/978-0-387-85648-3,

possesses the property of invariance under dilations $x \to \lambda x$ with $\lambda = \mathrm{const} \neq 0$. This property obviously ensures that the above value of q is the only possible one. The best constant C was found in [18] and [33] (see also [37, Sect. 1.4.2]) for $l = 1$, $p = 1$ and in [47, 4, 50] for $l = 1$, $p > 1$.

The present article consists of five independent sections dealing with various dilation invariant integral inequalities with optimal constants. We briefly describe the contents, starting with Sect. 1.

Let us recall the Gagliardo–Nirenberg inequality [25, 46]

$$\|v\|_{L^2(\mathbb{R}^2)} \leqslant C\,\|\nabla v\|_{L^1(\mathbb{R}^2)}, \quad v \in C_0^\infty(\mathbb{R}^2). \tag{1.2}$$

Setting $v = |\nabla u|$, we observe that the Dirichlet integral of u admits the estimate

$$\int\limits_{\mathbb{R}^2} |\nabla u|^2 dx \leqslant C \Big(\int\limits_{\mathbb{R}^2} |\nabla_2 u|\, dx\Big)^2, \tag{1.3}$$

where

$$|\nabla_2 u|^2 = |u_{x_1x_1}|^2 + 2|u_{x_1x_2}|^2 + |u_{x_2x_2}|^2.$$

One can see that it is impossible to improve (1.3), replacing $|\nabla_2 u|$ on the right-hand side by $|\Delta u|$. Indeed, it suffices to put a sequence of mollifications of the function $x \to \eta(x)\log|x|$, where $\eta \in C_0^\infty(\mathbb{R}^2)$, $\eta(0) \neq 0$, into the estimate in question in order to check its failure.

However, we show that the estimate of the same nature

$$\Big|\int\limits_{\mathbb{R}^2} \sum_{i,j=1}^{2} a_{i,j}\, u_{x_i}\, \overline{u}_{x_j}\, dx\Big| \leqslant C \Big(\int\limits_{\mathbb{R}^2} |\Delta u|\, dx\Big)^2,$$

where $a_{i,j} = \mathrm{const}$ and u is an arbitrary complex-valued function in $C_0^\infty(\mathbb{R}^2)$, may hold if and only if $a_{11} + a_{22} = 0$. We also find the best constant C in the last inequality. This is a particular case of Theorem 2.1 proved in Sect. 1.

In Sect. 2, we establish a new weighted Gårding type inequality

$$\int\limits_{\mathbb{R}^2} |\nabla_2 u|^2 \log(e^2|x|)^{-1} dx \leqslant \mathrm{Re} \int\limits_{\mathbb{R}^2} \Delta^2 u \cdot \overline{u}\, \log|x|^{-1} dx \tag{1.4}$$

for all $u \in C_0^\infty(\mathbb{R}^2 \backslash \{0\})$. Estimates of such a kind proved to be useful in the study of boundary behavior of solutions to elliptic equations (see [36, 43, 38, 39, 16, 32]).

Before turning to the contents of the next section, we introduce some notation. By $\mathbb{R}^n_+$ we denote the half-space $\{x = (x_1, \dots x_n) \in \mathbb{R}^n, x_n > 0\}$. Also let $\mathbb{R}^{n-1} = \partial\mathbb{R}^n_+$. As usual, $C_0^\infty(\mathbb{R}^n_+)$ and $C_0^\infty(\overline{\mathbb{R}^n_+})$ stand for the spaces of infinitely differentiable functions with compact support in $\mathbb{R}^n_+$ and $\overline{\mathbb{R}^n_+}$ respectively.

In Sect. 3, we are concerned with the inequality

$$\int_{\mathbb{R}^n_+} x_n\, |\nabla u|^2 dx \geqslant \Lambda \int_{\mathbb{R}^n_+} \frac{|u|^2}{(x_{n-1}^2 + x_n^2)^{1/2}} dx, \quad u \in C_0^\infty(\overline{\mathbb{R}^n_+}). \tag{1.5}$$

It was obtained in 1972 by one of the authors and proved to be useful in the study of the generic case of degeneration in the oblique derivative problem for second order elliptic differential operators [34].

Substituting $u(x) = x_n^{-1/2} v(x)$ into (1.5), one deduces with the same Λ that

$$\int_{\mathbb{R}^n_+} |\nabla v|^2 dx \geqslant \frac{1}{4} \int_{\mathbb{R}^n_+} \frac{|v|^2 dx}{x_n^2} + \Lambda \int_{\mathbb{R}^n_+} \frac{|v|^2 dx}{x_n (x_{n-1}^2 + x_n^2)^{1/2}} \tag{1.6}$$

for all $v \in C_0^\infty(\mathbb{R}^n_+)$ (see [37, Sect. 2.1.6]).

Another inequality of a similar nature obtained in [37] is

$$\int_{\mathbb{R}^n_+} |\nabla v|^2 dx \geqslant \frac{1}{4} \int_{\mathbb{R}^n_+} \frac{|v|^2}{x_n^2} dx + C\, \| x_n^\gamma\, v \|^2_{L_q(\mathbb{R}^n_+)}. \tag{1.7}$$

(This is a special case of the inequality (2.1.6/3) in [37].)

Without the second term on the right-hand sides of (1.6) and (1.7), these inequalities reduce to the classical Hardy inequality with the sharp constant 1/4 (see [13]). An interesting feature of (1.6) and (1.7) is their dilation invariance.

Variants, extensions, and refinements of (1.6) and (1.7), usually called Hardy's inequalities with remainder term, became the theme of many subsequent studies [2, 1, 3, 8, 5, 6, 7, 9, 10, 12, 14, 15, 17, 19, 20, 21, 22, 23, 24, 27, 29, 51, 52, 53, 54, 55, 56] *et al*).

In Theorem 4.1 proved in Sect. 3, we find a condition on the function q which is necessary and sufficient for the inequality

$$\int_{\mathbb{R}^n_+} |\nabla v|^2 dx - \frac{1}{4} \int_{\mathbb{R}^n_+} \frac{|v|^2 dx}{x_n^2} \geqslant C \int_{\mathbb{R}^n_+} q\Big(\frac{x_n}{(x_{n-1}^2 + x_n^2)^{1/2}} \Big) \frac{|v|^2 dx}{x_n\, (x_{n-1}^2 + x_n^2)^{1/2}}, \tag{1.8}$$

where v is an arbitrary function in $C_0^\infty(\mathbb{R}^n_+)$. This condition implies, in particular, that the right-hand side of (1.6) can be replaced by

$$C \int_{\mathbb{R}^n_+} \frac{|v|^2 dx}{x_n^2 \Big(1 - \log \dfrac{x_n}{(x_{n-1}^2 + x_n^2)^{1/2}}\Big)^2}.$$

The value $\Lambda = 1/16$ in (1.5) obtained in [34] is not the best possible. Tidblom [54] replaced it by 1/8. As a corollary of Theorem 4.1, we find an expression for the optimal value of Λ.

Let a measure μ_b be defined by

$$\mu_b(K) = \int\limits_K \frac{dx}{|x|^b} \tag{1.9}$$

for any compact set K in $\mathbb{R}^n$. In Sect. 4 we obtain the best constant in the inequality

$$\|u\|_{\mathcal{L}_{\tau,q}(\mu_b)} \leqslant C\Big(\int\limits_{\mathbb{R}^n} |\nabla u(x)|^p \frac{dx}{|x|^a}\Big)^{1/p},$$

where the left-hand side is the quasinorm in the Lorentz space $\mathcal{L}_{\tau,q}(\mu_b)$, i.e.,

$$\|u\|_{\mathcal{L}_{\tau,q}(\mu_b)} = \Big(\int\limits_0^\infty \big(\mu_b\{x : |u(x)| \geqslant t\}\big)^{q/\tau} d(t^q)\Big)^{1/q}.$$

As a particular case of this result, we obtain the best constant in the Hardy–Sobolev inequality

$$\left(\int\limits_{\mathbb{R}^n} |u(x)|^q \frac{dx}{|x|^b}\right)^{1/q} \leqslant \mathcal{C}\Big(\int\limits_{\mathbb{R}^n} |\nabla u(x)|^p \frac{dx}{|x|^a}\Big)^{1/p}, \tag{1.10}$$

first proved by Il'in [30, Theorem 1.4] in 1961 without discussion of the value of $\mathcal{C}$. Our result is a direct consequence of the capacitary integral inequality from [39] combined with an isocapacitary inequality. For particular cases, the best constant $\mathcal{C}$ was found in [11] ($p = 2$), [35, Sect. 2] ($p = 1$, $a = 0$), [26] ($p = 2$, $n = 3$, $a = 0$), [31] ($p = 2$, $n \geqslant 3$, $a = 0$), and [45] ($1 < p < n$, $a = 0$), where different methods were used.

The topic of the concluding section (Sect. 5) is the best constant C in the inequality (1.7), where $u \in C^\infty(\overline{\mathbb{R}^n_+})$ and $u = 0$ on $\mathbb{R}^{n-1}$.

Recently, Tertikas and Tintarev [51] obtained (among other results) the existence of an optimizer in (1.7) in the case $\gamma = 0$, $q = 2n/(n-2)$, $n \geqslant 4$. However, for these values of γ, q, and n the best value of C is unknown. In the case $n = 3$, $\gamma = 0$, $q = 6$ Benguria, Frank, and Loss [7] proved the nonexistence of an optimizer and found the best value of C by an ingenious argument.

We note in Sect. 5 that a similar problem can be easily solved for the special case $q = 2(n+1)/(n-1)$ and $\gamma = -1/(n+1)$.

2 Estimate for a Quadratic Form of the Gradient

Theorem 2.1. *Let $n \geqslant 2$, and let $A = \|a_{ij}\|_{i,j=1}^n$ be an arbitrary matrix with constant complex entries. The inequality*

$$\Big|\int\limits_{\mathbb{R}^n} \langle A\nabla u, \nabla u\rangle_{\mathbb{C}^n}\, dx\Big| \leqslant C\Big(\int\limits_{\mathbb{R}^n} \big|(-\Delta)^{\frac{n+2}{4}} u\big|\, dx\Big)^2, \tag{2.1}$$

where C is a positive constant, holds for all complex-valued $u \in C_0^\infty(\mathbb{R}^n)$ if and only if the trace of A is equal to zero. The best value of C is given by

$$C = \frac{(4\pi)^{-n/2}}{\Gamma\big(\frac{n}{2}+1\big)} \max_{\omega\in S^{n-1}}\Big|\sum_{1\leqslant i,j\leqslant n} a_{ij}\,\omega_i\omega_j\Big|, \tag{2.2}$$

where S^{n-1} is the $(n-1)$-dimensional unit sphere in $\mathbb{R}^n$.

(The notation $(-\Delta)^s$ in (2.1) stands for an integer or noninteger power of $-\Delta$.)

Proof. By $\mathcal{F}$ we denote the unitary Fourier transform in $\mathbb{R}^n$ defined by

$$\mathcal{F}h(\xi) = (2\pi)^{-n/2}\int\limits_{\mathbb{R}^n} h(x)\, e^{-i\,x\cdot\xi}\, dx. \tag{2.3}$$

We set $h = (-\Delta)^{(n+2)/4}u$ and write (2.1) in the form

$$\Big|\int\limits_{\mathbb{R}^n} |\mathcal{F}h(\xi)|^2 \Big\langle A\frac{\xi}{|\xi|}, \frac{\xi}{|\xi|}\Big\rangle_{\mathbb{C}^n} \frac{d\xi}{|\xi|^n}\Big| \leqslant C\Big(\int\limits_{\mathbb{R}^n} |h(x)|\, dx\Big)^2. \tag{2.4}$$

The singular integral on the left-hand side exists in the sense of the Cauchy principal value since

$$\int\limits_{S^{n-1}} \langle A\omega, \omega\rangle_{\mathbb{C}^n}\, ds_\omega = n^{-1}|S^{n-1}|\,\mathrm{Tr}\,A = 0,$$

where $\mathrm{Tr}\,A$ is the trace of A (see, for example, [44, Chapt. 9, Sect. 1] or [49, Theorem 4.7]). Let

$$k(\xi) = |\xi|^{-n}\Big\langle A\frac{\xi}{|\xi|}, \frac{\xi}{|\xi|}\Big\rangle_{\mathbb{C}^n}.$$

The left-hand side of (2.4) is equal to

$$\Big|\int\limits_{\mathbb{R}^n} \mathcal{F}^{-1}\Big(k(\xi)\big(\mathcal{F}h\big)(\xi)\Big)(x)\,\overline{h(x)}\, dx\Big| = (2\pi)^{-n/2}\Big|\int\limits_{\mathbb{R}^n} \big((\mathcal{F}^{-1}k) * h\big)(x)\,\overline{h(x)}\, dx\Big|$$

with $*$ meaning the convolution. Thus, the inequality (2.4) becomes

$$\Big|\int\limits_{\mathbb{R}^n} \big((\mathcal{F}^{-1}k) * h\big)(x)\,\overline{h(x)}\, dx\Big| \leqslant (2\pi)^{n/2}\, C\Big(\int\limits_{\mathbb{R}^n} |h(x)|\, dx\Big)^2. \tag{2.5}$$

We note that for $\xi \in \mathbb{R}^n$

$$k(\xi) = |\xi|^{-n-2}\Big(\sum_{j=1}^{n} a_{jj}\big(\xi_j^2 - n^{-1}|\xi|^2\big) + \sum_{\substack{i,j=1\\ i\neq j}}^{n} a_{ij}\xi_i\xi_j\Big). \tag{2.6}$$

Hence for $n > 2$

$$k(\xi) = \frac{1}{n\,(n-2)} \sum_{i,j=1}^{n} a_{ij} \frac{\partial^2}{\partial\xi_i\partial\xi_j} |\xi|^{2-n}, \tag{2.7}$$

and for $n = 2$

$$k(\xi) = \frac{1}{2} \sum_{i,j=1}^{2} a_{ij} \frac{\partial^2}{\partial\xi_i\partial\xi_j} \log|\xi|^{-1}. \tag{2.8}$$

Applying $\mathcal{F}^{-1}$ to the identity

$$-\Delta_\xi\Big(\frac{\partial^2}{\partial\xi_i\partial\xi_j}\frac{|\xi|^{2-n}}{|S^{n-1}|\,(n-2)}\Big) = \frac{\partial^2}{\partial\xi_i\partial\xi_j}\delta(\xi),$$

where $n > 2$ and δ is the Dirac function, from (2.7) we obtain

$$\big(\mathcal{F}^{-1}k\big)(x) = \frac{-|S^{n-1}|}{n\,(2\pi)^{n/2}} \sum_{i,j=1}^{n} a_{ij}\frac{x_i x_j}{|x|^2}. \tag{2.9}$$

Here, $|S^{n-1}|$ stands for the $(n-1)$-dimensional measure of S^{n-1}:

$$|S^{n-1}| = \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}. \tag{2.10}$$

Hence

$$\big(\mathcal{F}^{-1}k\big)(x) = \frac{-2^{-n/2}}{\Gamma(1+\frac{n}{2})} \sum_{i,j=1}^{n} a_{ij}\frac{x_i x_j}{|x|^2}. \tag{2.11}$$

Similarly, from (2.8) we deduce that (2.11) holds for $n = 2$ as well. Now, (2.1) with C given by (2.2) follows from (2.11) inserted into (2.5).

Next, we show the sharpness of C given by (2.2). Let θ denote a point on S^{n-1} such that

$$\big|\big(\mathcal{F}^{-1}k\big)(\theta)\big| = \max_{\xi\in\mathbb{R}^n\setminus\{0\}} |\mathcal{F}^{-1}k(\xi)|. \tag{2.12}$$

In order to obtain the required lower estimate for C, it suffices to set

$$h(x) = \eta(|x|)\,\delta_\theta\Big(\frac{x}{|x|}\Big),$$

where $\eta \in C_0^\infty[0,\infty)$, $\eta \geqslant 0$, and δ_θ is the Dirac measure on S^{n-1} concentrated at θ, into the inequality (2.5). (The legitimacy of this choice of h can be easily checked by approximation.) Then the estimate (2.5) becomes

$$\Big|\int_0^\infty\!\!\int_0^\infty (\mathcal{F}^{-1}k)\Big(\frac{\rho-r}{|\rho-r|}\theta\Big)\eta(r)\,r^{n-1}\,\eta(\rho)\,\rho^{n-1}drd\rho\Big|$$

$$\leqslant (2\pi)^{n/2}\,C\Big(\int_0^\infty \eta(\rho)\,\rho^{n-1}d\rho\Big)^2. \tag{2.13}$$

In view of (2.9) and (2.11),

$$\big(\mathcal{F}^{-1}k\big)(\pm\theta) = \big(\mathcal{F}^{-1}k\big)(\theta)$$

which, together with (2.12), enables one to write (2.13) in the form

$$\max_{\xi\in\mathbb{R}^n\setminus\{0\}} |\mathcal{F}^{-1}k(\xi)| \leqslant (2\pi)^{n/2}\,C.$$

By (2.9) and (2.11), this can be written as

$$\frac{|S^{n-1}|}{n\,(2\pi)^{n/2}}\max_{\omega\in S^{n-1}}\Big|\sum_{1\leqslant i,j\leqslant n} a_{ij}\,\omega_i\omega_j\Big| \leqslant (2\pi)^{n/2}\,C.$$

The result follows from (2.10). □

Remark 2.1. Let P and Q be functions, positively homogeneous of degrees $2m$ and $m+n/2$ respectively, $m>-n/2$. We assume that the restrictions of P, Q, and $P|Q|^{-2}$ to S^{n-1} belong to $L^1(S^{n-1})$, By the same argument as in Theorem 2.1, one concludes that the condition

$$\int_{S^{n-1}} \frac{P(\omega)}{|Q(\omega)|^2}\,ds_\omega = 0 \tag{2.14}$$

is necessary and sufficient for the inequality

$$\Big|\int_{\mathbb{R}^n} P(D)u\cdot\overline{u}\,dx\Big| \leqslant C\Big(\int_{\mathbb{R}^n}\big|Q(D)\,u\big|\,dx\Big)^2 \tag{2.15}$$

to hold for all $u\in C_0^\infty(\mathbb{R}^n)$. Moreover, using the classical formula for the Fourier transform of a positively homogeneous function of degree $-n$ (see [49, Theorem 4.11][1]), one finds that the best value of C in (2.15) is given by

$$\sup_{\omega\in S^{n-1}}\Big|\int_{S^{n-1}}\Big(\frac{i\pi}{2}\mathrm{sgn}(\theta\cdot\omega)+\log|\theta\cdot\omega|\Big)\frac{P(\theta)}{|Q(\theta)|^2}\,ds_\theta\Big|. \tag{2.16}$$

[1] Note that the definition of the Fourier transform in [49] contains $\exp(-2\pi i\,x\cdot\xi)$ unlike (2.3).

In particular, if $P(\omega)/|Q(\omega)|^2$ is a spherical harmonic, the best value of C in (2.15) is equal to

$$\frac{(4\pi)^{-n/2}\Gamma(m)}{\Gamma\left(\frac{n}{2}+m\right)} \max_{\omega\in S^{n-1}} \frac{|P(\omega)|}{|Q(\omega)|^2}, \tag{2.17}$$

which coincides with (2.2) for $m = 1$, $P(\xi) = A\xi \cdot \xi$, and $Q(\xi) = |\xi|^{1+n/2}$. □

One can notice that the argument used in the proof of (2.15) leads to the stronger estimate

$$\Big|\int_{\mathbb{R}^n} P(D)u \cdot \overline{v}\, dx\Big| \leqslant C\, \|Q(D)u\|_{L^1(\mathbb{R}^n)}\, \|Q(D)v\|_{L^1(\mathbb{R}^n)} \tag{2.18}$$

for all u and v in $C_0^\infty(\mathbb{R}^n)$.

Next, we discuss one more inequality of a similar nature. Let A be a symmetric 2×2-matrix with constant real entries which generates the hyperbolic operator $P(D) = \operatorname{div}(A\nabla)$. Then for all u and v in $C_0^\infty(\mathbb{R}^2)$ the following sharp inequality holds:

$$\Big|\int_{\mathbb{R}^2} P(D)u \cdot \overline{v}\, dx\Big| \leqslant 8^{-1}|\det A|^{-1/2}\, \|P(D)u\|_{L^1(\mathbb{R}^2)}\, \|P(D)v\|_{L^1(\mathbb{R}^2)}. \tag{2.19}$$

This is almost a special case of (2.15). The only difference is that the integral on the left-hand side of (2.14) should be understood as the Cauchy principal value. A more direct way to (2.18) is through the inequality

$$\Big|\int_{\mathbb{R}^2} u_{x_1}\, \overline{v}_{x_2}\, dx\Big| \leqslant 4^{-1}\|u_{x_1x_2}\|_{L^1(\mathbb{R}^2)}\, \|v_{x_1x_2}\|_{L^1(\mathbb{R}^2)}, \tag{2.20}$$

which follows from the obvious sharp estimate

$$\|w\|_{L^\infty(\mathbb{R}^2)} \leqslant 4^{-1}\, \|w_{x_1x_2}\|_{L^1(\mathbb{R}^2)}$$

for $w \in C_0^\infty(\mathbb{R}^2)$.

3 Weighted Gårding Inequality for the Biharmonic Operator

We start with an auxiliary Hardy type inequality.

Lemma 3.1. *Let $u \in C_0^\infty(\mathbb{R}^2)$. Then the following sharp inequality holds*:

$$\Big|\operatorname{Re}\int_{\mathbb{R}^2} (x_1u_{x_1} + x_2u_{x_2})\Delta\overline{u}\, \frac{dx}{|x|^2}\Big| \leqslant \int_{\mathbb{R}^2} |\Delta u|^2 dx. \tag{3.1}$$

Proof. Let (r,φ) denote polar coordinates in $\mathbb{R}^2$, and let

$$u(r,\varphi) = \sum_{k=-\infty}^{\infty} u_k(r) e^{ik\varphi}.$$

Then (3.1) is equivalent to the sequence of inequalities

$$\Big|\mathrm{Re}\int_0^\infty \Big(v'' + \frac{1}{r}v' - \frac{k^2}{r^2}v\Big)\overline{v}'\,dr\Big| \leqslant \int_0^\infty \Big|v'' + \frac{1}{r}v' - \frac{k^2}{r^2}v\Big|^2 r\,dr, \quad k = 0,1,2,\dots, \tag{3.2}$$

where v is an arbitrary function on $C_0^\infty([0,\infty))$. Putting $t = \log r^{-1}$ and $w(t) = v(e^{-t})$, we write (3.2) in the form

$$\Big|\mathrm{Re}\int_{\mathbb{R}^1} \big(w'' - k^2 w\big)\,\overline{w}'\,e^{2t}\,dt\Big| \leqslant \int_{\mathbb{R}^1} \big|w'' - k^2 w\big|^2 e^{2t}\,dt$$

which is equivalent to the inequality

$$\Big|\mathrm{Re}\int_{\mathbb{R}^1} \big(g'' - 2g' + (1-k^2)g\big)\big(\overline{g}' - \overline{g}\big)dt\Big| \leqslant \int_{\mathbb{R}^1} \big|g'' - 2g' + (1-k^2)g\big|^2 dt, \tag{3.3}$$

where $g = e^t w$. Making use of the Fourier transform in t, we see that (3.3) holds if and only if for all $\lambda \in \mathbb{R}^1$ and $k = 0,1,2\dots$

$$\Big|\mathrm{Re}\,\big(-\lambda^2 + 1 - k^2 - 2i\lambda\big)\big(1 - i\lambda\big)\Big| \leqslant \big(\lambda^2 - 1 + k^2\big)^2 + 4\lambda^2,$$

which is the same as

$$\big|3x - 1 + k^2\big| \leqslant x^2 + 2(k^2+1)x + (k^2-1)^2$$

with $x = \lambda^2$. This elementary inequality becomes equality if and only if $k = 0$ and $x = 0$. □

Remark 3.1. In spite of the simplicity of its proof, the inequality (3.1) deserves some interest. Let us denote the integral over $\mathbb{R}^2$ on the left-hand side of (3.1) by $Q(u,u)$ and write (3.1) as

$$|\,\mathrm{Re}\,Q(u,u)| \leqslant \|\Delta u\|_{L^2(\mathbb{R}^2)}.$$

However, the absolute value of the corresponding sesquilinear form $Q(u,v)$ cannot be majorized by $C\|\Delta u\|_{L^2(\mathbb{R}^2)}\|\Delta v\|_{L^2(\mathbb{R}^2)}$. Indeed, the opposite assertion would yield an upper estimate of $\|r^{-1}\partial u/\partial r\|_{L^2(\mathbb{R}^2)}$ by the norm of Δu in $L^2(\mathbb{R}^2)$, which is wrong for a function linear near the origin. □

Remark 3.2. Note that, under the additional orthogonality assumption

$$\int_0^{2\pi} u(r,\varphi)d\varphi = 0 \quad \text{for } r > 0, \tag{3.4}$$

the above proof of Lemma 3.1 provides the inequality (3.1) with the sharp constant factor 3/4 on the right-hand side. Moreover, (3.4) implies

$$\mathrm{Re}\int_{\mathbb{R}^2} (x_1 u_{x_1} + x_2 u_{x_2})\,\Delta\overline{u}\,\frac{dx}{|x|^2} \leqslant 0. \qquad \square$$

Using (3.1), we establish a new weighted Gårding type inequality.

Theorem 3.1. *Let* $u \in C_0^\infty(\mathbb{R}^2\backslash\{0\})$*. Then the inequality* (1.4) *holds.*

Proof. Clearly, the right-hand side of (1.4) is equal to

$$\mathrm{Re}\int_{\mathbb{R}^2} \Delta\overline{u}\,\Delta\,(u\log|x|^{-1})\,dx$$

$$= \int_{\mathbb{R}^2} |\Delta u|^2 \log|x|^{-1} dx + 2\mathrm{Re}\int_{\mathbb{R}^2} \Delta\overline{u}\cdot\nabla u\cdot\nabla\log|x|^{-1} dx.$$

Combining this identity with (3.1), we arrive at the inequality

$$\int_{\mathbb{R}^2} |\Delta u|^2 \log(e^2\,|x|)^{-1} dx \leqslant \mathrm{Re}\int_{\mathbb{R}^2} \Delta\overline{u}\,\Delta(u\,\log|x|^{-1})\,dx. \tag{3.5}$$

Note that

$$\int_{\mathbb{R}^2} \Delta u\cdot\Delta\overline{u}\cdot\log|x|^{-1} = -\int_{\mathbb{R}^2} \nabla u\cdot\nabla(\Delta\overline{u}\cdot\log|x|^{-1})\,dx$$

$$= \mathrm{Re}\int_{\mathbb{R}^2}\sum_{j=1}^{2}\Big(\nabla\frac{\partial u}{\partial x_j}\cdot\frac{\partial}{\partial x_j}\nabla\overline{u}\cdot\log|x|^{-1} + \nabla u\cdot\frac{\partial}{\partial x_j}\nabla\overline{u}\cdot\frac{\partial}{\partial x_j}\big(\log|x|^{-1}\big)\Big)dx,$$

which is equal to

$$\int_{\mathbb{R}^2}\Bigg(\sum_{j,k=1}^{2}\Big|\frac{\partial^2 u}{\partial x_j\partial x_k}\Big|^2\log|x|^{-1} + \frac{1}{2}\sum_{j=1}^{2}\frac{\partial}{\partial x_j}|\nabla u|^2\cdot\frac{\partial}{\partial x_j}\Big(\log|x|^{-1}\Big)\Bigg)dx.$$

Integrating by parts in the second term, we see that it vanishes. Thus, we conclude that

$$\int\limits_{\mathbb{R}^2} |\Delta u|^2 \log |x|^{-1} dx = \int\limits_{\mathbb{R}^2} |\nabla_2 u|^2 \log |x|^{-1} dx$$

which, together with (3.1) and the obvious identity

$$\int\limits_{\mathbb{R}^2} |\Delta u|^2 dx = \int\limits_{\mathbb{R}^2} |\nabla_2 u|^2 dx,$$

completes the proof of (1.4).

In order to see that no constant less than 1 is admissible in front of the integral on the right-hand side of (1.4), it suffices to put

$$u(x) = e^{i\langle x,\xi\rangle}\, \eta(x)$$

with $\eta \in C_0^\infty(\mathbb{R}^2)\backslash\{0\}$ into (1.4) and take the limit as $|\xi| \to \infty$. □

Remark 3.3. If the condition $u = 0$ near the origin in Theorem 3.1 is removed, the above proof gives the additional term

$$\pi\Big(|\nabla u(0)|^2 - 2\,\mathrm{Re}\big(u(0)\,\Delta\overline{u}(0)\big)\Big)$$

on the right-hand side of (1.4). □

4 Dilation Invariant Hardy's Inequalities with Remainder Term

Theorem 4.1. (i) *Let q denote a locally integrable nonnegative function on* $(0,1)$*. The best constant in the inequality*

$$\int\limits_{\mathbb{R}^n_+} x_n\, |\nabla u|^2 dx \geqslant C \int\limits_{\mathbb{R}^n_+} q\Big(\frac{x_n}{(x_{n-1}^2 + x_n^2)^{1/2}}\Big)\frac{|u|^2}{(x_{n-1}^2 + x_n^2)^{1/2}} dx, \tag{4.1}$$

for all $u \in C_0^\infty(\overline{\mathbb{R}^n_+})$, which is equivalent to (1.8)*, is given by*

$$\lambda := \inf \frac{\displaystyle\int_0^{\pi/2} \Big(\big|y'(\varphi)\big|^2 + \frac{1}{4}\big|y(\varphi)\big|^2\Big)\sin\varphi\, d\varphi}{\displaystyle\int_0^{\pi/2} \big|y(\varphi)\big|^2 q(\sin\varphi)\, d\varphi}, \tag{4.2}$$

where the infimum is taken over all smooth functions on $[0,\pi/2]$.

(ii) *The inequalities* (4.1) *and* (1.8) *with positive* C *hold if and only if*

$$\sup_{t\in(0,1)} (1-\log t)\int_0^t q(\tau)\,d\tau < \infty. \tag{4.3}$$

Moreover,

$$\lambda \sim \Big(\sup_{t\in(0,1)} (1-\log t)\int_0^t q(\tau)\,d\tau\Big)^{-1}, \tag{4.4}$$

where $a \sim b$ *means that* $c_1 a \leqslant b \leqslant c_2 a$ *with absolute positive constants* c_1 *and* c_2.

Proof. (i) Let $U \in C_0^\infty(\overline{\mathbb{R}^2_+})$, $\zeta \in C_0^\infty(\mathbb{R}^{n-2})$, $x' = (x_1,\dots x_{n-2})$, and let $N = \text{const} > 0$. Putting

$$u(x) = N^{(2-n)/2}\zeta(N^{-1}x')U(x_{n-1},x_n)$$

into (4.1) and passing to the limit as $N\to\infty$, we see that (4.1) is equivalent to the inequality

$$\int_{\mathbb{R}^2_+} x_2\big(|U_{x_1}|^2+|U_{x_2}|^2\big)dx_1dx_2 \geqslant C\int_{\mathbb{R}^2_+} q\Big(\frac{x_2}{(x_1^2+x_2^2)^{1/2}}\Big)\frac{|U|^2dx_1dx_2}{(x_1^2+x_2^2)^{1/2}}, \tag{4.5}$$

where $U \in C_0^\infty(\overline{\mathbb{R}^2_+})$. Let (ρ,φ) be the polar coordinates of $(x_1,x_2)\in\mathbb{R}^2_+$. Then (4.5) can be written as

$$\int_0^\infty\int_0^\pi \Big(|U_\rho|^2+\rho^{-2}|U_\varphi|^2\Big)\sin\varphi\,d\varphi\,\rho^2d\rho \geqslant C\int_0^\infty\int_0^\pi |U|^2q(\sin\varphi)\,d\varphi d\rho.$$

By the substitution $U(\rho,\varphi)=\rho^{-1/2}v(\rho,\varphi)$, the left-hand side becomes

$$\int_0^\infty\int_0^\pi\Big(|\rho v_\rho|^2+|v_\varphi|^2+\frac14|v|^2\Big)\sin\varphi\,d\varphi\frac{d\rho}{\rho} - \operatorname{Re}\int_0^\pi\int_0^\infty \overline{v}\,v_\rho\,d\rho\,\sin\varphi\,d\varphi. \tag{4.6}$$

Since $v(0)=0$, the second term in (4.6) vanishes. Therefore, (4.5) can be written in the form

$$\int_0^\infty\int_0^\pi\Big(|\rho v_\rho|^2+|v_\varphi|^2+\frac14|v|^2\Big)\sin\varphi\,d\varphi\frac{d\rho}{\rho} \geqslant C\int_0^\infty\int_0^\pi |v|^2q(\sin\varphi)\,d\varphi\frac{d\rho}{\rho}. \tag{4.7}$$

Now, the definition (4.2) of λ shows that (4.7) holds with $C=\lambda$.

In order to show the optimality of this value of C, put $t = \log\rho$ and $v(\rho,\varphi) = w(t,\varphi)$. Then (4.7) is equivalent to

$$\int_{\mathbb{R}^1}\int_0^\pi \Big(|w_t|^2 + |w_\varphi|^2 + \frac{1}{4}|w|^2\Big)\sin\varphi\, d\varphi\, dt \geqslant C \int_{\mathbb{R}^1}\int_0^\pi |w|^2 q(\sin\varphi) d\varphi\, dt. \quad (4.8)$$

Applying the Fourier transform $w(t,\varphi) \to \widehat{w}(s,\varphi)$, we obtain

$$\int_{\mathbb{R}^1}\int_0^\pi \Big(|\widehat{w}_\varphi|^2 + \Big(|s|^2 + \frac{1}{4}\Big)|\widehat{w}|^2\Big)\sin\varphi\, d\varphi\, ds \geqslant C \int_{\mathbb{R}^1}\int_0^\pi |\widehat{w}|^2 q(\sin\varphi) d\varphi\, ds. \quad (4.9)$$

Putting here

$$\widehat{w}(s,\varphi) = \varepsilon^{-1/2}\eta(s/\varepsilon)y(\varphi),$$

where $\eta \in C_0^\infty(\mathbb{R}^1)$, $\|\eta\|_{L^2(\mathbb{R}^1)} = 1$, and y is a function on $C^\infty([0,\pi])$, and passing to the limit as $\varepsilon \to 0$, we arrive at the estimate

$$\int_0^\pi \Big(|y'(\varphi)|^2 + \frac{1}{4}|y(\varphi)|^2\Big)\sin\varphi\, d\varphi \geqslant C \int_0^\pi |y(\varphi)|^2 q(\sin\varphi) d\varphi, \quad (4.10)$$

where π can be changed for $\pi/2$ by symmetry. This, together with (4.2), implies $\varLambda \leqslant \lambda$. The proof of (i) is complete.

(ii) Introducing the new variable $\xi = \log\cot\frac{\varphi}{2}$, we write (4.2) as

$$\lambda = \inf_z \frac{\displaystyle\int_0^\infty \Big(|z'(\xi)|^2 + \frac{|z(\xi)|^2}{4\,(\cosh\xi)^2}\Big) d\xi}{\displaystyle\int_0^\infty |z(\xi)|^2\, q\Big(\frac{1}{\cosh\xi}\Big)\frac{d\xi}{\cosh\xi}}. \quad (4.11)$$

Since

$$|z(0)|^2 \leqslant 2\int_0^1 \big(|z'(\xi)|^2 + |z(\xi)|^2\big) d\xi$$

and

$$\int_0^\infty |z(\xi)|^2 \frac{e^{2\xi}}{(1+e^{2\xi})^2} d\xi \leqslant 2\int_0^\infty |z(\xi) - z(0)|^2 \frac{d\xi}{\xi^2} + 2\,|z(0)|^2 \int_0^\infty \frac{e^{2\xi}}{(1+e^{2\xi})^2} d\xi$$

$$\leqslant 8\int_0^\infty |z'(\xi)|^2 d\xi + |z(0)|^2,$$

from (4.11) it follows that

$$\lambda \sim \inf_z \frac{\int\limits_0^\infty |z'(\xi)|^2 d\xi + |z(0)|^2}{\int\limits_0^\infty |z(\xi)|^2\, q\Big(\frac{1}{\cosh\xi}\Big)\frac{d\xi}{\cosh\xi}} \,. \tag{4.12}$$

Setting $z(\xi) = 1$ and $z(\xi) = \min\{\eta^{-1}\xi, 1\}$ for all positive ξ and fixed $\eta > 0$ into the ratio of quadratic forms in (4.12), we deduce that

$$\lambda \leqslant \min\Big\{ \Big(\int\limits_0^\infty q\Big(\frac{1}{\cosh\xi}\Big)\frac{d\xi}{\cosh\xi}\Big)^{-1},\ \Big(\sup_{\eta>0} \eta \int\limits_\eta^\infty q\Big(\frac{1}{\cosh\xi}\Big)\frac{d\xi}{\cosh\xi}\Big)^{-1}\Big\}.$$

Hence

$$\lambda \leqslant c\Big(\sup_{t\in(0,1)} (1-\log t) \int\limits_0^t q(\tau)\, d\tau\Big)^{-1}.$$

In order to obtain the converse estimate, note that

$$\int\limits_0^\infty |z(\xi)|^2\, q\Big(\frac{1}{\cosh\xi}\Big)\frac{d\xi}{\cosh\xi}$$

$$\leqslant 2\,|z(0)|^2 \int\limits_0^\infty q\Big(\frac{1}{\cosh\xi}\Big)\frac{d\xi}{\cosh\xi} + 2\int\limits_0^\infty |z(\xi)-z(0)|^2\, q\Big(\frac{1}{\cosh\xi}\Big)\frac{d\xi}{\cosh\xi}\,.$$

The second term on the right-hand side is dominated by

$$8 \sup_{\eta>0}\Big(\eta \int\limits_\eta^\infty q\Big(\frac{1}{\cosh\xi}\Big)\frac{d\xi}{\cosh\xi}\Big) \int\limits_0^\infty |z'(\xi)|^2 d\xi$$

(see, for example, [37, Sect. 1.3.1]). Therefore,

$$\int\limits_0^\infty |z(\xi)|^2\, q\Big(\frac{1}{\cosh\xi}\Big)\frac{d\xi}{\cosh\xi} \leqslant 8 \max\Big\{ \int\limits_0^\infty q\Big(\frac{1}{\cosh\xi}\Big)\frac{d\xi}{\cosh\xi}\,,$$

$$\sup_{\eta>0} \eta \int\limits_\eta^\infty q\Big(\frac{1}{\cosh\sigma}\Big)\frac{d\sigma}{\cosh\sigma}\Big\}\Big(\int\limits_0^\infty |z'(\xi)|^2 d\xi + |z(0)|^2\Big)$$

which, together with (4.12), leads to the lower estimate

$$\lambda \geqslant \min\Big\{ \Big(\int_0^\infty q\Big(\frac{1}{\cosh\xi}\Big)\frac{d\xi}{\cosh\xi}\Big)^{-1},\ \Big(\sup_{\eta>0}\eta\int_\eta^\infty q\Big(\frac{1}{\cosh\xi}\Big)\frac{d\xi}{\cosh\xi}\Big)^{-1}\Big\}.$$

Hence

$$\lambda \geqslant c\Big(\sup_{t\in(0,1)}(1-\log t)\int_0^t q(\tau)\,d\tau\Big)^{-1}.$$

The proof of (ii) is complete. □

Since (4.3) holds for $q(t) = t^{-1}(1-\log t)^{-2}$, Theorem 4.1 (ii) leads to the following assertion.

Corollary 4.1. *There exists an absolute constant $C > 0$ such that the inequality*

$$\int_{\mathbb{R}^n_+} |\nabla v|^2 dx - \frac{1}{4}\int_{\mathbb{R}^n_+}\frac{|v|^2 dx}{x_n^2} \geqslant C\int_{\mathbb{R}^n_+}\frac{|v|^2 dx}{x_n^2\Big(1-\log\dfrac{x_n}{(x_{n-1}^2+x_n^2)^{1/2}}\Big)^2} \tag{4.13}$$

holds for all $v \in C_0^\infty(\mathbb{R}^n_+)$. The best value of C is equal to

$$\lambda := \inf \frac{\displaystyle\int_0^\pi \Big[\big|y'(\varphi)\big|^2 + \frac{1}{4}\big|y(\varphi)\big|^2\Big]\sin\varphi\,d\varphi}{\displaystyle\int_0^\pi \big|y(\varphi)\big|^2(\sin\varphi)^{-1}\big(1-\log\sin\varphi\big)^{-2}d\varphi}, \tag{4.14}$$

where the infimum is taken over all smooth functions on $[0,\pi/2]$. By numerical approximation, $\lambda = 0.16\ldots$

A particular case of Theorem 4.1 corresponding to $q = 1$ is the following assertion.

Corollary 4.2. *The sharp value of Λ in* (1.5) *and* (1.6) *is equal to*

$$\lambda := \inf \frac{\displaystyle\int_0^\pi \Big[\big|y'(\varphi)\big|^2 + \frac{1}{4}\big|y(\varphi)\big|^2\Big]\sin\varphi\,d\varphi}{\displaystyle\int_0^\pi \big|y(\varphi)\big|^2 d\varphi}, \tag{4.15}$$

where the infimum is taken over all smooth functions on $[0,\pi]$. By numerical approximation, $\lambda = 0.1564\ldots$

Remark 4.1. Let us consider the Friedrichs extension $\widetilde{\mathcal{L}}$ of the operator

$$\mathcal{L} : z \to -\big((\sin\varphi)z'\big)' + \frac{\sin\varphi}{4} z \tag{4.16}$$

defined on smooth functions on $[0,\pi]$. It is a simple exercise to show that the energy space of $\widetilde{\mathcal{L}}$ is compactly imbedded into $L^2(0,\pi)$. Hence the spectrum of $\widetilde{\mathcal{L}}$ is discrete and λ defined by (4.15) is the smallest eigenvalue of $\widetilde{\mathcal{L}}$. □

Remark 4.2. The argument used in the proof of Theorem 4.1 (i) with obvious changes enables one to obtain the following more general fact. Let P and Q be measurable nonnegative functions in $\mathbb{R}^n$, positive homogeneous of degrees 2μ and $2\mu - 2$ respectively. The sharp value of C in

$$\int\limits_{\mathbb{R}^n} P(x)|\nabla u|^2 dx \geqslant C \int\limits_{\mathbb{R}^n} Q(x)|u|^2 dx, \quad u \in C_0^\infty(\mathbb{R}^n), \tag{4.17}$$

is equal to

$$\lambda := \inf \frac{\int\limits_{S^{n-1}} P(\omega)\Big(|\nabla_\omega Y|^2 + \big(\mu - 1 + \frac{n}{2}\big)^2 |Y|^2\Big) ds_\omega}{\int\limits_{S^{n-1}} Q(\omega)|Y|^2 ds_\omega},$$

where the infimum is taken over all smooth functions on the unit sphere S^{n-1}. □

A direct consequence of this assertion is the following particular case of (4.17).

Remark 4.3. Let p and q stand for locally integrable nonnegative functions on $(0,1]$, and let $\mu \in \mathbb{R}^1$. If $n > 2$, the best value of C in

$$\int\limits_{\mathbb{R}^n} |x|^{2\mu} p\Big(\frac{x_n}{|x|}\Big)|\nabla u|^2 dx \geqslant C \int\limits_{\mathbb{R}^n} |x|^{2\mu-2} q\Big(\frac{x_n}{|x|}\Big)|u|^2 dx, \tag{4.18}$$

where $u \in C_0^\infty(\mathbb{R}^n)$, is equal to

$$\inf \frac{\int\limits_0^\pi \Big(|y'(\theta)|^2 + \big(\mu - 1 + \frac{n}{2}\big)^2 |y(\theta)|^2\Big) p(\cos\theta)(\sin\theta)^{n-2} d\theta}{\int\limits_0^\pi |y(\theta)|^2 q(\cos\theta)(\sin\theta)^{n-2} d\theta}, \tag{4.19}$$

with the infimum taken over all smooth functions on the interval $[0,\pi]$.

Formula (4.19) enables one to obtain a necessary and sufficient condition for the existence of a positive C in (4.18). Let us assume that the function

$$\theta \to \frac{(\sin\theta)^{2-n}}{p(\cos\theta)}$$

is locally integrable on $(0,\pi)$. We make the change of variable $\xi = \xi(\theta)$, where

$$\xi(\theta) = \int_{\pi/2}^{\theta} \frac{(\sin\tau)^{2-n}}{p(\cos\tau)} d\tau,$$

and suppose that $\xi(0) = -\infty$ and $\xi(\pi) = \infty$. Then (4.19) can be written in the form

$$\lambda = \inf_z \frac{\int_{\mathbb{R}^1} |z'(\xi)|^2 d\xi + \big(\mu - 1 + \frac{n}{2}\big)^2 \int_{\mathbb{R}^1} |z(\xi)|^2 \big(p(\cos\theta(\xi))\,(\sin\theta(\xi))^{n-2}\big)^2 d\xi}{\int_{\mathbb{R}^1} |z(\xi)|^2\, p(\cos\theta(\xi))\, q(\cos\theta(\xi)) (\sin\theta(\xi))^{2(n-2)}\, d\xi},$$

where $\theta(\xi)$ is the inverse function of $\xi(\theta)$. By [42, Theorem 1],

$$\lambda \sim \inf_{\substack{\xi\in\mathbb{R}^1\\ d>0,\delta>0}} \frac{\frac{1}{\delta} + \int_{\theta(\xi-d-\delta)}^{\theta(\xi+d+\delta)} p(\cos\theta)\,(\sin\theta)^{n-2}\, d\theta}{\int_{\theta(\xi-d)}^{\theta(\xi+d)} q(\cos\theta)\,(\sin\theta)^{n-2}\, d\theta}. \tag{4.20}$$

Here, the equivalence $a \sim b$ means that $c_1 b \leqslant a \leqslant c_2 b$, where c_1 and c_2 are positive constants depending only on μ and n. Hence (4.18) holds with a positive Λ if and only if the infimum (4.20) is positive. □

Remark 4.4. In the case $n = 2$, the best constant in (4.18) is equal to

$$\lambda := \inf_y \frac{\int_0^{2\pi} \Big(|y'(\varphi)|^2 + \mu^2 |y(\varphi)|^2\Big) p(\sin\varphi) d\varphi}{\int_0^{2\pi} |y(\varphi)|^2 q(\sin\varphi) d\varphi}, \tag{4.21}$$

where the infimum is taken over all smooth functions on the interval $[0, 2\pi]$. Note that (4.2) is a particular case of (4.21) with $\mu = 1/2$ and $p(t) = t$.

As another application of (4.21), we obtain the following special case of the inequality (4.18) with $n = 2$.

For all $u \in C_0^\infty(\mathbb{R}^2)$ *the inequality*

$$\int_{\mathbb{R}^2} \frac{x_2^2\,|u(x)|^2\,dx}{(x_1^2+x_2^2)\Big(\dfrac{\pi^2}{4} - \Big(\arcsin\dfrac{x_1}{(x_1^2+x_2^2)^{1/2}}\Big)^2\Big)} \leqslant \frac{1}{2}\int_{\mathbb{R}^2} x_2^2\,|\nabla u(x)|^2 dx \tag{4.22}$$

holds, where $1/2$ *is the best constant.*

In order to prove (4.22), we choose $p(t) = t^2$ and $\mu = 1$ in (4.21), which becomes

$$\lambda = \inf_y \frac{\int\limits_0^{2\pi}\Big(|y'(\varphi)|^2 + |y(\varphi)|^2\Big)(\sin\varphi)^2 d\varphi}{\int\limits_0^{2\pi} |y(\varphi)|^2 q(\sin\varphi)d\varphi}, \tag{4.23}$$

where y is an arbitrary smooth 2π-periodic function. Putting $\eta(\varphi) = y(\varphi)\sin\varphi$, we write (4.23) in the form

$$\lambda = \inf_\eta \frac{\int\limits_0^{2\pi} |\eta'(\varphi)|^2 d\varphi}{\int\limits_0^{2\pi} |\eta(\varphi)|^2 q(\sin\varphi)(\sin\varphi)^{-2} d\varphi} \tag{4.24}$$

with the infimum taken over all 2π-periodic functions satisfying $\eta(0) = \eta(\pi) = 0$. Let

$$q(\sin\varphi) = \frac{(\sin\varphi)^2}{\varphi(\pi-\varphi)}.$$

In view of the well-known sharp inequality

$$\int\limits_0^1 \frac{|z(t)|^2}{t(1-t)}dt \leqslant \frac{1}{2}\int\limits_0^1 |z'(t)|^2 dt$$

(see [28, Theorem 262]), we have $\lambda = 2$ in (4.24). Therefore, (4.18) becomes (4.22). $\square$

5 Generalized Hardy–Sobolev Inequality with Sharp Constant

Let Ω denote an open set in $\mathbb{R}^n$, and let $p \in [1, \infty)$. By the (p, a)-capacity of a compact set $K \subset \Omega$ we mean the set function

$$\mathrm{cap}_{p,a}(K, \Omega) = \inf\Big\{ \int_\Omega |\nabla u|^p |x|^{-a}\, dx : \ u \in C_0^\infty(\Omega),\ u \geqslant 1 \text{ on } K \Big\}.$$

In the case $a = 0$, $\Omega = \mathbb{R}^n$, we write simply $\mathrm{cap}_p(K)$.

The following inequality is a particular case of a more general one obtained in [41], where Ω is an open subset of an arbitrary Riemannian manifold and $|\Phi(x, \nabla u(x)|$ plays the role of $|\nabla u(x)|\, |x|^{-a/p}$.

Theorem 5.1. (see [35] for $q = p$ and [41] for $q \geqslant p$) (i) *Let $q \geqslant p \geqslant 1$, and let Ω be an open set in $\mathbb{R}^n$. Then for an arbitrary $u \in C_0^\infty(\Omega)$,*

$$\left(\int_0^\infty \big(\mathrm{cap}_{p,a}(M_t, \Omega)\big)^{q/p} d(t^q) \right)^{1/q} \leqslant \mathcal{A}_{p,q} \left(\int_\Omega |\nabla u(x)|^p |x|^{-a} dx \right)^{1/p}, \quad (5.1)$$

where $M_t = \{x \in \Omega : \ |u(x)| \geqslant t\}$ and

$$\mathcal{A}_{p,q} = \left(\frac{\Gamma\big(\frac{pq}{q-p}\big)}{\Gamma\big(\frac{q}{q-p}\big)\Gamma\big(p\frac{q-1}{q-p}\big)} \right)^{1/p-1/q} \quad (5.2)$$

for $q > p$, and

$$\mathcal{A}_{p,p} = p(p-1)^{(1-p)/p}. \quad (5.3)$$

(ii) *The sharpness of this constant is checked by a sequence of radial functions in $C_0^\infty(\Omega)$. Moreover, there exists a radial optimizer vanishing at infinity if $\Omega = \mathbb{R}^n$.*

Being combined with the isocapacitary inequality

$$\mu(K)^\gamma \leqslant \Lambda_{p,\gamma}\, \mathrm{cap}_{p,a}(K, \Omega) \quad (5.4)$$

where μ is a Radon measure in Ω, (5.1) implies the estimate

$$\left(\int_0^\infty \big(\mu(M_t)\big)^{\gamma q/p} d(t^q) \right)^{1/q} \leqslant \mathcal{A}_{p,q} \Lambda_{p,\gamma}^{1/p} \left(\int_\Omega |\nabla u(x)|^p |x|^{-a} dx \right)^{1/p} \quad (5.5)$$

for all $u \in C_0^\infty(\Omega)$.

This estimate of u in the Lorentz space $\mathcal{L}_{p/\gamma,q}(\mu)$ becomes the estimate in $L_q(\mu)$ for $\gamma = p/q$:

$$\|u\|_{L_q(\mu)} \leqslant \mathcal{A}_{p,q} \Lambda_{p,\gamma}^{1/p} \left(\int_{\Omega} |\nabla u(x)|^p |x|^{-a} dx \right)^{1/p}.$$

In the next assertion, we find the best value of $\Lambda_{p,\gamma}$ in (5.4) for the measure $\mu = \mu_b$ defined by (1.9).

Lemma 5.1. *Let*

$$1 \leqslant p < n, \quad 0 \leqslant a < n-p, \quad \text{and} \quad a+p \geqslant b \geqslant \frac{an}{n-p}. \tag{5.6}$$

Then

$$\left(\int_{\mathbb{R}^n} \frac{dx}{|x|^b} \right)^{\frac{n-p-a}{n-b}} \leqslant \Big(\frac{p-1}{n-p-a} \Big)^{p-1} \frac{|S^{n-1}|^{\frac{b-p-a}{n-b}}}{(n-b)^{\frac{n-p-a}{n-b}}} \operatorname{cap}_{p,a}(K). \tag{5.7}$$

The value of the constant factor in front of the capacity is sharp and the equality in (5.7) *is attained at any ball centered at the origin.*

Proof. Introducing spherical coordinates (r, ω) with $r > 0$ and $\omega \in S^{n-1}$, we have

$$\operatorname{cap}_{p,a}(K) = \inf_{u|_K \geqslant 1} \int_{S^{n-1}} \int_0^\infty \Big(\Big| \frac{\partial u}{\partial r} \Big|^2 + \frac{1}{r^2} |\nabla_\omega u|^2 \Big)^{\frac{p}{2}} r^{n-1-a} dr ds_\omega. \tag{5.8}$$

Let us put here $r = \rho^{1/\varkappa}$, where

$$\varkappa = \frac{n-p-a}{n-p}$$

and $y = (\rho, \omega)$. The mapping $(r, \omega) \to (\rho, \omega)$ will be denoted by σ. Then (5.8) takes the form

$$\operatorname{cap}_{p,a}(K) = \varkappa^{p-1} \inf_v \int_{\mathbb{R}^n} \Big(\Big| \frac{\partial u}{\partial \rho} \Big|^2 + (\varkappa \rho)^{-2} |\nabla_\omega u|^2 \Big)^{\frac{p}{2}} dy, \tag{5.9}$$

where the infimum is taken over all $v = u \circ \sigma^{-1}$. Since $0 \leqslant \varkappa \leqslant 1$ owing to the conditions $p < n$, $0 < a < n-p$, and $a \geqslant 0$, the inequality (5.9) implies

$$\operatorname{cap}_{p,a}(K) \geqslant \varkappa^{p-1} \inf_v \int_{\mathbb{R}^n} |\nabla u|^p dy \geqslant \varkappa^{p-1} \operatorname{cap}_p(\sigma(K)) \tag{5.10}$$

which, together with the isocapacitary property of cap_p (see [37, Corollary 2.2.3/2]), leads to the estimate

$$\operatorname{cap}_p(\sigma(K)) \geqslant \Big(\frac{n-p}{p-1}\Big)^{p-1} |S^{n-1}|^{\frac{p}{n}} n^{\frac{n-p}{n}} \big(\operatorname{mes}_n(\sigma(K)\big)^{\frac{n-p}{n}}. \tag{5.11}$$

Clearly,

$$\mu_b(K) = \frac{1}{\varkappa} \int\limits_{\sigma(K)} \frac{dy}{|y|^\alpha}$$

with

$$\alpha = n - \frac{n-b}{\varkappa} = \frac{b(n-p)-an}{n-p-a} \geqslant 0. \tag{5.12}$$

Furthermore, one can easily check that

$$\mu_b(K) \leqslant \frac{n^{1-\frac{\alpha}{n}}}{n-b} |S^{n-1}|^{\frac{\alpha}{n}} \big(\operatorname{mes}_n(\sigma(K)\big)^{1-\frac{\alpha}{n}} \tag{5.13}$$

(see, for example, [40, Example 2.2]). Combining (5.13) with (5.11), we find

$$\big(\mu_b(K)\big)^{\frac{n-p-a}{n-b}} \leqslant \Big(\frac{p-1}{n-p}\Big)^{p-1} \frac{|S^{n-1}|^{\frac{b-p-a}{n-b}}}{(n-b)^{\frac{n-p-a}{n-b}}} \operatorname{cap}_p(\sigma(K)) \tag{5.14}$$

which, together with (5.10), completes the proof of (5.7). □

The main result of this section is as follows.

Theorem 5.2. *Let the conditions* (5.6) *hold, and let* $q \geqslant p$. *Then for all* $u \in C_0^\infty(\mathbb{R}^n)$

$$\Big(\int\limits_0^\infty \big(\mu_b(M_t)\big)^{\frac{(n-p-a)q}{(n-b)p}} d(t^q)\Big)^{\frac{1}{q}} \leqslant \mathcal{C}_{p,q,a,b} \Big(\int\limits_{\mathbb{R}^n} |\nabla u(x)|^p \frac{dx}{|x|^a}\Big)^{\frac{1}{p}}, \tag{5.15}$$

where

$$\mathcal{C}_{p,q,a,b} = \Bigg(\frac{\Gamma\big(\frac{pq}{q-p}\big)}{\Gamma\big(\frac{q}{q-p}\big)\Gamma\big(p\frac{q-1}{q-p}\big)}\Bigg)^{\frac{1}{p}-\frac{1}{q}} \Big(\frac{p-1}{n-p-a}\Big)^{1-\frac{1}{p}} \Bigg(\frac{\Gamma(\frac{n}{2})}{2\pi^{\frac{n}{2}}(n-b)^{\frac{n-p-a}{p+a-b}}}\Bigg)^{\frac{p+a-b}{(n-b)p}}. \tag{5.16}$$

The constant (5.16) *is best possible which can be shown by constructing a radial optimizing sequence in* $C_0^\infty(\mathbb{R}^n)$.

Proof. The inequality (5.15) is obtained by substitution of (5.2) and (5.7) into (5.5). The sharpness of (5.16) follows from part (ii) of Theorem 5.1 and the fact that the isocapacitary inequality (5.7) becomes equality for balls. □

The last theorem contains the best constant in the Il'in inequality (1.10) as a particular case $q = (n-b)p/(n-p-a)$. We formulate this as the following assertion.

Corollary 5.1. *Let the conditions* (5.6) *hold. Then for all* $u \in C_0^\infty(\mathbb{R}^n)$

$$\Big(\int\limits_{\mathbb{R}^n} |u(x)|^{\frac{(n-b)p}{(n-p-a)}} \frac{dx}{|x|^b}\Big)^{\frac{n-p-a}{n-b}} \leqslant \mathcal{C}_{p,a,b}\Big(\int\limits_{\mathbb{R}^n} |\nabla u(x)|^p \frac{dx}{|x|^a}\Big)^{\frac{1}{p}}, \tag{5.17}$$

where

$$\mathcal{C}_{p,a,b}=\Big(\frac{p-1}{n-p-a}\Big)^{1-\frac{1}{p}}\Big(\frac{\Gamma(\frac{n}{2})}{2\pi^{\frac{n}{2}}(n-b)^{\frac{n-p-a}{p+a-b}}}\Big)^{\frac{p+a-b}{(n-b)p}} \times \Big(\frac{\Gamma\big(\frac{p(n-b)}{p-b}\big)}{\Gamma\big(\frac{n-b}{p-b}\big)\Gamma\big(1+\frac{(n-b)(p-1)}{p-b}\big)}\Big)^{\frac{p-b}{p(n-b)}}.$$

This constant is best possible, which can be shown by constructing a radial optimizing sequence in $C_0^\infty(\mathbb{R}^n)$.

6 Hardy's Inequality with Sharp Sobolev Remainder Term

Theorem 6.1. *For all* $u \in C^\infty(\overline{\mathbb{R}^n_+})$, $u = 0$ *on* $\mathbb{R}^{n-1}$, *the following sharp inequality holds*:

$$\int\limits_{\mathbb{R}^n_+} |\nabla u|^2 dx \geqslant \frac{1}{4}\int\limits_{\mathbb{R}^n_+} \frac{|u|^2}{x_n^2}dx + \frac{\pi^{n/(n+1)}(n^2-1)}{4\big(\Gamma\big(\frac{n}{2}+1\big)\big)^{2/(n+1)}}\|x_n^{-1/(n+1)}u\|^2_{L_{\frac{2(n+1)}{n-1}}(\mathbb{R}^n_+)}. \tag{6.1}$$

Proof. We start with the Sobolev inequality

$$\int\limits_{\mathbb{R}^{n+1}} |\nabla w|^2 dz \geqslant \mathcal{S}_{n+1}\|w\|^2_{L_{\frac{2(n+1)}{n-1}}(\mathbb{R}^{n+1})} \tag{6.2}$$

with the best constant (see [47, 4, 50])

$$\mathcal{S}_{n+1} = \frac{\pi^{(n+2)/(n+1)}(n^2-1)}{4^{n/(n+1)}\big(\Gamma\big(\frac{n}{2}+1\big)\big)^{2/(n+1)}}. \tag{6.3}$$

Let us introduce the cylindrical coordinates (r, φ, x'), where $r \geqslant 0$, $\varphi \in [0, 2\pi)$, and $x' \in \mathbb{R}^{n-1}$. Assuming that w does not depend on φ, we write (6.2) in the form

$$2\pi \int\limits_{\mathbb{R}^{n-1}} \int\limits_0^\infty \Big(\Big| \frac{\partial w}{\partial r} \Big|^2 + |\nabla_{x'} w|^2 \Big) r\, dr dx'$$

$$\geqslant (2\pi)^{(n-1)/(n+1)} \mathcal{S}_{n+1} \Big(\int\limits_{\mathbb{R}^{n-1}} \int\limits_0^\infty |w|^{2(n+1)/(n-1)} r\, dr dx' \Big)^{(n-1)/(n+1)}.$$

Replacing r by x_n, we obtain

$$\int\limits_{\mathbb{R}^n_+} |\nabla w|^2 x_n\, dx \geqslant (2\pi)^{-2/(n+1)} \mathcal{S}_{n+1} \Big(\int\limits_{\mathbb{R}^n_+} |w|^{2(n+1)/(n-1)} x_n\, dx \Big)^{(n-1)/(n+1)}.$$

It remains to set $w = x_n^{1/2} v$ and use (6.3). □

Acknowledgement. The first author was partially supported by the USA National Science Foundation (grant DMS 0500029) and the UK Engineering and Physical Sciences Research Council (grant no. EP/F005563/1).

References

1. Adimurthi: Hardy–Sobolev inequalities in $H^1(\Omega)$ and its applications. Commun. Contemp. Math. **4**, no. 3, 409–434 (2002)
2. Adimurthi, Chaudhuri, N., Ramaswamy, M.: An improved Hardy–Sobolev inequality and its applications. Proc. Am. Math. Soc. **130** no. 489–505 (2002)
3. Adimurthi, Grossi, M., Santra, S.: Optimal Hardy–Rellich inequalities, maximum principle and related eigenvalue problems. J. Funct. Anal. **240**, no.1, 36–83 (2006)
4. Aubin, T.: Problèmes isopérimètriques et espaces de Sobolev. C.R. Acad. Sci., Paris **280**, no. 5, 279–281 (1975)
5. Barbatis, G., Filippas, S., Tertikas, A.: Series expansion for L^p Hardy inequalities. Indiana Univ. Math. J. **52**, no. 1, 171–190. (2003)
6. Barbatis, G., Filippas, S., Tertikas, A.: A unified approach to improved L^p Hardy inequalities with best constants. Trans. Am. Math. Soc. **356**, no. 6, 2169–2196. (2004)
7. Benguria, R.D., Frank, R.L., Loss, M.: The sharp constant in the Hardy–Sobolev–Maz'ya inequality in the three dimensional upper half-space. arXiv: math/0705 3833.
8. Brandolini, B., Chiacchio, F., Trombetti, C.: Hardy inequality and Gaussian measure. Commun. Pure Appl. Math. **6**, no. 2, 411–428, (2007)
9. Brezis, H., Marcus, M.: Hardy's inequalities revisited. Dedicated to Ennio De Giorgi. Ann. Sc. Norm. Super. Pisa, Cl. Sci. (4) **25**. no. 1-2, 217–237 (1997)
10. Brezis, H., Vázquez, J.L.: Blow-up solutions of some noinlinear elliptic problems. Rev. Mat. Univ. Comp. Madrid **10**, 443–469 (1997)
11. Catrina, F., Wang, Z.-Q.: On the Caffarelli–Kohn–Nirenberg inequalities: sharp constants, existence (and nonexistence), and symmetry of extremal functions. Commun. Pure Appl. Math. **54**, 229–258, (2001)

12. Costin, O., Maz'ya, V.: Sharp Hardy–Leray inequality for axisymmetric divergence-free fields. Calc. Var. Partial Differ. Equ. [To appear]; arXiv:math/0703116
13. Davies, E.B.: A review of Hardy inequalities. In: The Maz'ya Anniversary Collection. Vol. 2 (Rostock, 1998), pp. 55–67, Oper. Theory Adv. Appl. **110**, Birkhäuser, Basel (1999)
14. Dávila, J., Dupaigne, L.: Hardy type inequalities. J. Eur. Math. Soc. **6**, no. 3, 335–365 (2004)
15. Dou, J., Niu, P., Yuan, Z.: A Hardy inequality with remainder terms in the Heisenberg group and weighted eigenvalue problem. J. Inequal. Appl. **2007**, ID 32585 (2007)
16. Eilertsen, S.: On weighted positivity and the Wiener regularity of a boundary point for the fractional Laplacian. Ark. Mat. **38**, 53–75 (2000)
17. Evans, W.E., Lewis, R.T.: Hardy and Rellich inequalities with remainders. J. Math. Ineq. **1**, no. 4, 473–490 (2007)
18. Federer, H., Fleming, W.H.: Normal and integral currents. Ann. Math. **72**, 458–520 (1960)
19. Filippas, S., Maz'ya, V., Tertikas, A.: Sharp Hardy–Sobolev inequalities. C.R. Acad. Sci., Paris **339**, no. 7, 483–486 (2004)
20. Filippas, S., Maz'ya, V., Tertikas, A.: On a question of Brezis and Marcus. Calc. Var. Partial Differ. Equ. **25**, no. 4, 491–501 (2006)
21. Filippas, S., Maz'ya, V., Tertikas, A.: Critical Hardy–Sobolev inequalities J. Math. Pure Appl. **87**, 37–56 (2007)
22. Filippas, S., Tertikas, A.: Optimizing improved Hardy inequalities. J. Funct. Anal. **192**, 186–233 (2002)
23. Filippas, S., Tertikas, A., Tidblom, J.: On the structure of Hardy–Sobolev–Maz'ya inequalities. arXiv: math/0802.0986.
24. Frank, R.L., Seiringer, R.: Nonlinear ground state representations and sharp Hardy inequalities. arXiv: math/0803.0503.
25. Gagliardo, E.; Ulteriori proprietà di alcune classi di funzioni in più variabili. Ric. Mat. **8**, no. 1, 24–51 (1959)
26. Glazer, V., Martin, A., Grosse, H., Thirring, W.: A family of optimal conditions for the absence of bound states in a potential. In: Studies in Mathematical Physics, pp. 169–194. Princeton Univ. Press, Princeton, NJ (1976)
27. Gazzola, F., Grunau, H.-Ch., Mitidieri, E.: Hardy inequalities with optimal constants and remainder terms. Trans. Am. Math. Soc. **356**, 2149–2168 (2004)
28. Hardy, G.H., Littlewood, J.E., Polya, G.: Inequalities. Cambridge Univ. Press, Cambridge (1952)
29. Hoffmann-Ostenhof, M., Hoffmann-Ostenhof, T., Laptev, A.: A geometrical version of Hardy's inequality. J. Funct. Anal. **189**, 539–548 (2002)
30. Il'in, V.P.: Some integral inequalities and their applications in the theory of differentiable functions of several variables (Russian). Mat. Sb, **54**, no. 3, 331–380 (1961)
31. Lieb, E.H., Sharp constants in the Hardy–Littlewood-Sobolev and related inequalities. Ann. Math. **118**, 349–374 (1983)
32. Mayboroda, S., Maz'ya, V.: Boundedness of the Hessian of a biharmonic function in a convex domain. Commun. Partial Differ. Equ. [To appear]; arXiv:math/0611058
33. Maz'ya, V.: Classes of domains and imbedding theorems for function spaces (Russian). Dokl. Akad. Nauk SSSR **133**, 527–530 (1960); English transl.: Sov. Math., Dokl. **1**, 882–885 (1960)
34. Maz'ya, V.: On a degenerating problem with directional derivative (Russian). Mat. Sb. **87**, 417–454, (1972); English transl.: Math. USSR Sb. **16**, no. 3, 429–469 (1972)
35. Maz'ya, V.: On certain integral inequalities for functions of many variables (Russian). Probl. Mat. Anal. **3**, 33–68 (1972); English transl.: J. Math. Sci. **1**, no. 2, 205–234 (1973)

36. Maz'ya, V.: Behavior of solutions to the Dirichlet problem for the biharmonic operator at a boundary point. Equadiff IV (Proc. Czechoslovak Conf. Differential Equations and their Applications, Prague, 1977) pp. 250–262. Lect. Notes Math. **703**. Springer, Berlin (1979)
37. Maz'ya, V.G.: Sobolev Spaces. Springer-Verlag, Berlin–Tokyo (1985)
38. Maz'ya, V.: On the Wiener-type regularity of a boundary point for the polyharmonic operator. Appl. Anal. **71**, 149–165 (1999)
39. Maz'ya, V.: The Wiener test for higher order elliptic equations. Duke Math. J. **115**. no. 3, 479–512 (2002)
40. Maz'ya, V.: Lectures on isoperimetric and isocapacitary inequalities in the theory of Sobolev spaces. Contemp. Math. **338**, 307–340 (2003)
41. Maz'ya, V.: Conductor and capacitary inequalities for functions on topological spaces and their applications to Sobolev type imbeddings. J. Funct. Anal.. **224**, no. 2, 408–430 (2005)
42. Maz'ya, V.: Conductor inequalities and criteria for Sobolev type two-weight imbeddings. J. Comp. Appl. Math.. **194**, no. 1. 94–114 (2006).
43. Maz'ya, V., Tashchiyan, G.: On the behavior of the gradient of the solution of the Dirichlet problem for the biharmonic equation near a boundary point of a three-dimensional domain (Russian). Sib. Mat. Zh. **31**, no. 6, 113–126 (1991); English transl.: Sib. Math. J. **31**, no. 6, 970–983 (1991)
44. Mikhlin, S., Prössdorf, S.: Singular Integral Operators. Springer (1986)
45. Nazarov, A.I.: Hardy–Sobolev inequalities in a cone. J. Math. Sci. **132**, no. 4, 419–427 (2006)
Hardy–Sobolev inequalities in a cone (Russian). Probl. Mat. Anal. **31**, 39–46 (2005); English transl.: J. Math. Sci., New York **132**, no. 4, 419–427 (2006)
46. Nirenberg, L.: On elliptic partial differential equations, Lecture II. Ann. Sc. Norm. Super. Pisa, Ser. 3 **13**, 115–162 (1959)
47. Rosen, G.: Minimum value for C in the Sobolev inequality $\|\varphi^3\| \leqslant C\|\text{grad}\varphi\|^3$, SIAM J. Appl. Math. **21**, no. 1, 30–33 (1971)
48. Sobolev, S.L.: On a theorem of functional analysis (Russian). Mat. Sb. **46**, 471–497 (1938); English transl.: Am. Math. Soc., Transl., II. Ser. **34**, 39–68 (1963)
49. Stein, E.M., Weiss, G.: Introduction to Fourier Analysis on Euclidean Spaces. Princeton Univ. Press, Princeton, NJ (1971)
50. Talenti, G.: Best constant in Sobolev inequality. Ann. Mat. Pura Appl. **136**, 353–172 (1976)
51. Tertikas, A., Tintarev, K.: On existence of minimizers for the Hardy–Sobolev–Maz'ya inequality. Ann. Mat. Pura Appl. **186**, no. 4, 645–662 (2007)
52. Tertikas, A., Zographopoulos, N.B.: Best constants in the Hardy-Rellich inequalities and related improvements. Adv. Math. **209**, no. 2, 407–459 (2007)
53. Tidblom, J.: A geometrical version of Hardy's inequality for $W_0^{1,p}$. Proc. Am. Math. Soc. **138**, no. 8, 2265–2271 (2004)
54. Tidblom, J.: A Hardy inequality in the half-space. J. Funct. Anal. **221**, 482–495 (2005)
55. Vázquez, J.L., Zuazua, E.: The Hardy inequality and the asymptotic behavior of the heat equation with an inverse-square potential. J. Funct. Anal. **173**, 103–153 (2000)
56. Yaotian, S., Zhihui, C.: General Hardy inequalities with optimal constants and remainder terms. J. Inequal. Appl. no. 3, 207–219 (2005)

Optimality of Function Spaces in Sobolev Embeddings

Luboš Pick

Abstract We study the optimality of function spaces that appear in Sobolev embeddings. We focus on rearrangement-invariant Banach function spaces. We apply methods of interpolation theory.

It is a great honor for me to contribute to this volume dedicated to the centenary of S.L. Sobolev, one of the greatest analysts of the XXth century. The paper concerns a topic belonging to an area bearing the name, called traditionally Sobolev inequalities or Sobolev embeddings. The focus will be on the sharpness or optimality of function spaces appearing in these embeddings. The results presented in this paper were established in recent years. Most of them were obtained in collaboration with Ron Kerman and Andrea Cianchi.

1 Prologue

Sobolev embeddings, or Sobolev inequalities, constitute a very important part of the modern functional analysis.

Suppose that Ω is a bounded domain in $\mathbb{R}^n$, $n \geqslant 2$, with Lipschitz boundary. In the classical form, the *Sobolev inequality* asserts that, given $1 < p < n$ and setting $p^* = \frac{np}{n-p}$, there exists $C > 0$ such that

Luboš Pick
Charles University, Sokolovská 83, 186 75 Praha 8, Czech Republic,
e-mail: pick@karlin.mff.cuni.cz

V. Maz'ya (ed.), *Sobolev Spaces in Mathematics I, International Mathematical Series.*
doi: 10.1007/978-0-387-85648-3,

$$\left(\int_\Omega |u(x)|^{p^*}\,dx\right)^{\frac{1}{p^*}} \leqslant C\left(\int_\Omega |(\nabla u)(x)|^p + |u(x)|^p\,dx\right)^{\frac{1}{p}} \quad \text{for all } W^{1,p}(\Omega).$$

(Throughout the paper, C denotes a constant independent of important quantities, not necessarily the same at each occurrence.) We can restate this result in the form of a *Sobolev embedding*, namely,

$$W^{1,p}(\Omega) \hookrightarrow L_{p^*}(\Omega), \qquad 1 < p < n, \tag{1.1}$$

where $W^{1,p}(\Omega)$ is the classical *Sobolev space*, i.e., a collection of weakly differentiable functions on Ω such that $u \in L_p(\Omega)$ and $|\nabla u| \in L_p(\Omega)$, endowed with the norm

$$\|u\|_{W^{1,p}(\Omega)} = \|u\|_{L_p(\Omega)} + \|\nabla u\|_{L_p(\Omega)},$$

and $\|\cdot\|_{L_p(\Omega)}$ is the usual Lebesgue norm.

We say that the space on the left-hand side of (1.1) is a *Sobolev space built upon* $L_p(\Omega)$. In this sense, we recognize $L_p(\Omega)$ as the *domain space* of the embedding and the space $L_{p^*}(\Omega)$ on the right-hand side as its *range space*.

We focus on the following question: How *sharp* are the domain space and the range space in the Sobolev embedding?

First thing we note is that this question is dependent on an environment within which it is investigated. For example, the embedding (1.1) cannot be improved within the environment of Lebesgue spaces. This should be understood as follows: if we replace the domain space $L_p(\Omega)$ in (1.1) by a larger Lebesgue space, say, $L_q(\Omega)$ with $q < p$, then the resulting embedding

$$W^{1,q}(\Omega) \hookrightarrow L_{p^*}(\Omega)$$

can no longer be true. Likewise, if we replace the range space $L_{p^*}(\Omega)$ by a smaller Lebesgue space, say, $L_r(\Omega)$, $r > p^*$, then again the resulting embedding

$$W^{1,p}(\Omega) \hookrightarrow L_r(\Omega)$$

does not hold any more. In this sense, the embedding (1.1) is, at least within the environment of Lebesgue spaces, *sharp* (or *optimal*), and it cannot be effectively improved. In other words, if we want to improve the embedding and to get thereby a finer result, we need to use classes of function spaces finer than the Lebesgue scale.

The fact that the Lebesgue scale is simply not delicate enough in order to describe all the interesting details about embeddings, is perhaps best illustrated by the so-called *limiting* or *critical* case of the embedding (1.1) corresponding to the case $p = n$. When we let p tend to n from the left, then, of course, p^* tends to ∞. However, the limiting embedding

$$W^{1,n}(\Omega) \hookrightarrow L_\infty(\Omega)$$

is unfortunately not true. It is well known that one can have unbounded functions (typically, with logarithmic singularities) in $W^{1,n}(\Omega)$. Therefore, the only information which we can formulate in the Lebesgue spaces environment for the limiting embedding is

$$W^{1,n}(\Omega) \hookrightarrow L_q(\Omega) \qquad \text{for every } q < \infty. \tag{1.2}$$

Again, this information is optimal within the environment of Lebesgue spaces, where no improvement is available. However, it is quite clear that this result is very unsatisfactory as it does not provide any definite range function space. Such a space can be obtained, but not among Lebesgue spaces. We need a refinement of the Lebesgue scale. One of the most well-known and most widely used such refinements of Lebesgue spaces are Orlicz spaces. We first shortly recall their definition.

Given any *Young function* $A : [0,\infty) \to [0,\infty)$, namely a convex increasing function vanishing at 0, the *Orlicz space* $L_A(\Omega)$ is the rearrangement-invariant space of all measurable functions u in Ω such that the *Luxemburg norm*

$$\|u\|_{L_A(\Omega)} = \inf\left\{ \lambda > 0; \int_\Omega A\left(\frac{|u(x)|}{\lambda}\right) dx \leqslant 1 \right\}$$

is finite. Of course, if $A(t) = t^p$, then we recover Lebesgue spaces. Other important examples of Orlicz spaces are the *logarithmic Zygmund classes* $L_p(\log L)^\alpha(\Omega)$, generated by the Young function

$$A(t) = t^p(\log(e+t))^\alpha, \quad t \in (1,\infty),$$

with $p \in [1,\infty)$ and $\alpha \in \mathbb{R}$, and the *exponential Zygmund classes* $\exp L^\alpha(\Omega)$, generated by the Young function

$$A(t) = \exp(t^\alpha), \quad t \in (1,\infty), \quad \alpha > 0.$$

Equipped with Orlicz spaces, we can formulate the following limiting case of the Sobolev embedding:

$$W^{1,n}(\Omega) \hookrightarrow \exp L^{n'}(\Omega), \tag{1.3}$$

where

$$n' = \frac{n}{n-1}.$$

This result is traditionally attributed to Trudinger [44], however, in a certain modified form, it appeared earlier in works of Yudovich [45], Peetre [37], and Pokhozhaev [40].

We can now, again, ask how sharp this result is. It turns out, that, remarkably, the range space $\exp L^{n'}(\Omega)$ is sharp within the environment of Orlicz spaces. In other words, it is the smallest possible Orlicz space that still renders this embedding true. This optimality result is due to Hempel, Morris, and Trudinger [24].

In this context, it might be of interest to ask whether also the classical Sobolev embedding (1.1) has the optimal *Orlicz* range space. Of course, as we already know, it is the optimal *Lebesgue* range space, but now we are asking about optimality in a much broader sense, so the question is sensible. The answer is positive, as follows from the result of Cianchi [11].

However, it turns out that nontrivial improvements of both (1.1) and (1.3) are still available. To this end, we have to introduce function spaces whose norms involve the so-called nonincreasing rearrangement.

We denote by $\mathcal{M}(\Omega)$ the class of real-valued measurable functions on Ω and by $\mathcal{M}_+(\Omega)$ the class of nonnegative functions in $\mathcal{M}(\Omega)$. Given $f \in \mathcal{M}(\Omega)$, its *nonincreasing rearrangement* is defined by

$$f^*(t) = \inf\{\lambda > 0;\ |\{x \in \Omega;\ |f(x)| > \lambda\}| \leqslant t\}, \quad t \in [0, \infty).$$

We also define the *maximal nonincreasing rearrangement* of f by

$$f^{**}(t) = t^{-1} \int_0^t f^*(s)\, ds, \quad t \in [0, \infty).$$

We note that below we use the rearrangements defined only on $(0, |\Omega|)$, but it can be as well defined on $[0, \infty)$, extended by zero for $t > |\Omega|$.

We work with several classes of function spaces defined with the help of the operation $f \mapsto f^*$. The first such an example will be the scale of the two-parameter Lorentz spaces.

Assume that $0 < p, q \leqslant \infty$. The *Lorentz space* $L_{p,q}(\Omega)$ is the collection of all $f \in \mathcal{M}(\Omega)$ such that $\|f\|_{L_{p,q}(\Omega)} < \infty$, where

$$\|f\|_{L_{p,q}(\Omega)} = \|t^{\frac{1}{p}-\frac{1}{q}} f^*(t)\|_{L_q(0,1)}.$$

The Lorentz spaces are nested in the following sense. For every $p \in (0, \infty]$ and $0 < q < r \leqslant \infty$ we have

$$L_{p,q(\Omega)} \hookrightarrow L_{p,r(\Omega)}, \tag{1.4}$$

and this embedding is strict.

With the help of Lorentz spaces, we have the following refinement of (1.1):

$$W^{1,p}(\Omega) \hookrightarrow L_{p^*,p}(\Omega), \qquad 1 < p < n. \tag{1.5}$$

Note that, thanks to (1.4) and the obvious inequality $p < p^*$, this is a nontrivial improvement of the range space in (1.1). The embedding (1.5) is due to Peetre [37], and it can be also traced in works of O'Neil [35] and Hunt [26].

A natural question arises, whether a similar Lorentz-type refinement is possible also for the limiting embedding (1.3). The answer is positive again, but we need to introduce a yet more general function scale.

Let $0 < p, q \leqslant \infty$ and $\alpha \in \mathbb{R}$. The *Lorentz–Zygmund space* $L_{p,q;\alpha}(\Omega)$ is the collection of all $f \in \mathcal{M}(\Omega)$ such that $\|f\|_{L_{p,q;\alpha}(\Omega)} < \infty$, where

$$\|f\|_{L_{p,q;\alpha}(\Omega)} := \|t^{\frac{1}{p}-\frac{1}{q}} \log^{\alpha}(e/t) f^*(t)\|_{L_q(0,1)}.$$

Occasionally, we have to work with a modification of Lorentz–Zygmund spaces in which f^* is replaced by f^{**}. We denote such a space by $L_{(p,q;\alpha)}(\Omega)$. Hence

$$\|f\|_{L_{(p,q;\alpha)}(\Omega)} := \|t^{\frac{1}{p}-\frac{1}{q}} \log^{\alpha}(e/t) f^{**}(t)\|_{L_q(0,1)}.$$

These spaces were introduced and studied by Bennett and Rudnick [4].

Equipped with Lorentz–Zygmund spaces, we have the following refinement of the Trudinger embedding (1.3):

$$W^{1,n}(\Omega) \hookrightarrow L_{\infty,n;-1}(\Omega). \tag{1.6}$$

The first one to note this fact was Maz'ya who formulated it in a somewhat implicit form involving capacitary estimates (see [34, pp. 105 and 109]). Explicit formulations were given by Hansson [25] and Brézis–Wainger [6], the result can be also traced in the work of Brudnyi [7]. A more general assertion was proved by Cwikel and Pustylnik [18].

The range space in (1.6) is a very interesting function space. It is not a Zygmund class of neither logarithmic nor exponential type. Moreover, as the relations between Lorentz–Zygmund spaces from [4] show, it satisfies

$$L_{\infty,n;-1}(\Omega) \hookrightarrow \exp L^{n'}(\Omega),$$

and this inclusion is strict. We thus get a nontrivial improvement of (1.3).

The embedding (1.6) can be viewed in some sense as the limiting case of (1.5) as $p \to n+$. Indeed, both these results allow us a unified approach, as shown in [33] (again, restricted to functions vanishing on the boundary), where it was noticed that, for $1 < p < n$, we have

$$\int\limits_0^1 t^{\frac{p}{p^*}-1} u^*(t)^p \, dt \leqslant C \int\limits_{\Omega} |\nabla u(x)|^p \, dx$$

for all $u \in W_0^{1,p}(\Omega)$, while, in the limiting case, we have

$$\int\limits_0^1 \left(\frac{u^*(t)}{\log\left(\frac{e}{t}\right)} \right)^n \frac{dt}{t} \leqslant C \int\limits_{\Omega} |\nabla u|^n(x) \, dx$$

for all $u \in W_0^{1,n}(\Omega)$. Both these results were proved in an elementary way by first establishing a weak version of the Sobolev–Gagliardo–Nirenberg embedding, namely

$$\lambda\left(|\{|u| \geqslant \lambda\}|\right)^{\frac{1}{n'}} \leqslant C \int_{\Omega} |\nabla u|\, dx, \qquad u \in W_0^{1,1}(\Omega),\ \lambda > 0,$$

and then using a truncation argument due to Maz'ya.

In the course of the proof it turned out that yet the further improvement of (1.6) is possible, Namely, it was shown that

$$W_0^{1,n}(\Omega) \hookrightarrow W_n(\Omega),$$

where, for $0 < p \leqslant \infty$, the space $W_p(\Omega)$ is defined as the family of all measurable functions on Ω for which

$$\|u\|_{W_p(\Omega)} = \begin{cases} \left(\int_0^1 \left(u^*(\frac{t}{2}) - u^*(t)\right)^p \frac{dt}{t} \right)^{\frac{1}{p}} < \infty & \text{when } p < \infty; \\ \sup_{0<t<1} \left(u^*(\frac{t}{2}) - u^*(t)\right) & \text{when } p = \infty. \end{cases}$$

It was shown that the space $W_p(\Omega)$ has some interesting properties, for example:

(i) $\|\chi_E\|_{W_p(\Omega)} = (\log 2)^{\frac{1}{p}}$ for every measurable $E \subset \Omega$ and $p \in (0, \infty)$;

(ii) $L_\infty = W_1(\Omega)$;

(iii) for $p \in [1, \infty)$ each integer-valued $u \in W_p(\Omega)$ is bounded;

(iv) for $p \in (1, \infty)$, $W_p(\Omega)$ is not a linear set;

(v) for $p \in (1, \infty)$, $W_p(\Omega) \subsetneqq L_{\infty,p;-1}(\Omega)$;

(vi) $W_p(\Omega) \subsetneqq W_q(\Omega)$ for every $0 < p < q \leqslant \infty$.

The norm of the space $W_p(\Omega)$ involves the functional $f^*(\frac{t}{2}) - f^*(t)$. Bastero, Milman, and Ruiz [2] showed that it can be equivalently replaced with $f^{**}(t) - f^*(t)$. The quantity $f^{**}(t) - f^*(t)$, which measures, in some sense, the oscillation of f, was used in the theory of function spaces before. Function spaces involving this functional have been particularly popular since 1981 when Bennett, DeVore and Sharpley [3] introduced the "weak L_∞," the rearrangement-invariant space of functions for which $f^{**}(t) - f^*(t)$ is bounded.

The problem of optimality of function spaces in Sobolev embeddings can be also viewed from a reversed angle. So far we have focused solely on the question of optimality of the range space in various contexts. However, one can also ask whether the *domain* space is optimal. For example, it is clear that (1.1) and (1.5) have the best possible Lebesgue domain spaces. We can however ask whether these domain spaces are also optimal as *Orlicz* spaces. The answer is interesting and perhaps even surprising. While, in the nonlimiting embedding (1.1), the space $L_p(\Omega)$ is indeed the optimal Orlicz range for $L_{p^*}(\Omega)$ ([39, Corollary 4.9]), the situation in the limiting case is quite different. Not only that $L_n(\Omega)$ *is not* the largest Orlicz space for which the Trudinger inequality (1.3) holds, but, oddly enough, there is no such an optimal Orlicz space at all. This should be understood as follows: for every Orlicz space $L_A(\Omega)$ such that

$$W^1 L_A(\Omega) \hookrightarrow \exp L^{n'}(\Omega),$$

there exists another, strictly larger Orlicz space $L_B(\Omega)$ such that

$$W^1 L_B(\Omega) \hookrightarrow \exp L^{n'}(\Omega).$$

A construction of the Young function B which generates such an Orlicz space $L_B(\Omega)$ from a given A can be found in [39, Theorem 4.5]. In a way, this result resembles the unsatisfactory situation with Lebesgue range partners in the limiting embedding (1.2), where one has an "open set of range spaces," and illustrates thereby that not even the (apparently rather fine) class of Orlicz spaces is delicate enough to provide satisfactory answers. We can use this as a motivation to look for optimal function spaces in a broader general context.

The last example shows that the investigation of the optimality of domain spaces in well-known embeddings can bring unexpected surprises. Another such a situation, although quite different by nature, occurs when we ask about the optimality of the domain $L_n(\Omega)$ in the Trudinger embedding (1.3). Indeed, it was shown in [21] that, interestingly, from the scaling property of Lorentz–Zygmund spaces one can deduce the following embedding:

$$W^1 \left(L_{n,1;-\frac{1}{n'}} + L_{n,\infty;\frac{1}{n}} \right) (\Omega) \hookrightarrow \exp L^{n'}(\Omega).$$

Complemented with

$$L_n(\Omega) \hookrightarrow \left(L_{n,1;-\frac{1}{n'}} + L_{n,\infty;\frac{1}{n}} \right) (\Omega),$$

the inclusion being strict, this gives a rather unexpected nontrivial improvement of the domain space in the Trudinger embedding, quite different from the above-mentioned one, built on Orlicz spaces.

All these examples call for considering some reasonable common environment that would provide a roof for all or, at least, most of the function spaces mentioned so far and for considering global optimality within this context. For us, such an environment is that of the so-called rearrangement-invariant (r.i.) spaces.

Furthermore, we should be interested also in higher order Sobolev embeddings (note that all the illustrative examples mentioned so far were first order embeddings). Higher order embeddings are important in applications and, as it turns out, considerably more difficult to handle than the first order ones. This is caused by the fact that for the first order embedding one has the *Pólya–Szegö inequality* for which there is, regrettably, no equally powerful analogue for higher order embeddings.

2 Preliminaries

The context of function spaces in which we study the optimality of Sobolev embeddings is that of the so-called rearrangement-invariant spaces. Before stating exact definitions, let us just mention that most of the function spaces mentioned above, namely, those of Lebesgue, Orlicz, Zygmund, Lorentz and Lorentz-Zygmund, are, at least for some reasonable parameters, r.i. spaces, with a notable exception of the space $W_p(\Omega)$, which is not even linear. Therefore, r.i. spaces constitute a common roof for many important classes of functions, it is a rich collection of general function spaces, yet they are pleasantly modeled upon the example of Lebesgue spaces, inheriting many of their wonderful properties.

Throughout the paper, we assume, unless stated otherwise, that Ω is a bounded domain having Lipschitz boundary and satisfying $|\Omega| = 1$. (If the measure is finite and different from 1, everything can be easily modified in an obvious way by the change of variables $t \mapsto |\Omega|t$.)

A Banach space $X(\Omega)$ of functions defined on Ω, equipped with the norm $\|\cdot\|_{X(\Omega)}$, is said to be *rearrangement-invariant* if the following axioms hold:

$$0 \leqslant g \leqslant f \text{ a.e. implies } \|g\|_{X(\Omega)} \leqslant \|f\|_{X(\Omega)}; \tag{P1}$$

$$0 \leqslant f_n \nearrow f \text{ a.e. implies } \|f_n\|_{X(\Omega)} \nearrow \|f\|_{X(\Omega)}; \tag{P2}$$

$$\|\chi_\Omega\|_{X(\Omega)} < \infty, \text{ where } \chi_E \text{ denotes the characteristic function of } E; \tag{P3}$$

$$\text{for every } E \subset \Omega, \text{with } |E| < \infty, \text{ there exists a constant } C_E \tag{P4}$$

$$\text{such that } \int_E f(x)\,dx \leqslant C_E \|f\|_{X(\Omega)} \quad \text{for every } f \in X(\Omega);$$

$$\|f\|_{X(\Omega)} = \|g\|_{X(\Omega)} \text{ whenever } f^* = g^*. \tag{P5}$$

A basic tool for working with rearrangement-invariant spaces is the *Hardy–Littlewood–Pólya* (HLP) *principle* treated in [5, Chapt. 2, Theorem 4.6]). It asserts that $f^{**}(t) \leqslant g^{**}(t)$ for every $t \in (0,1)$ implies $\|f\|_{X(\Omega)} \leqslant \|g\|_{X(\Omega)}$ for every r.i. space $X(\Omega)$.

The *Hardy–Littlewood inequality* states that

$$\int_\Omega |f(x)g(x)|\,dx \leqslant \int_0^1 f^*(t)g^*(t)\,dt, \qquad f,g\in\mathcal{M}(\Omega). \tag{2.1}$$

Given an r.i. space $X(\Omega)$, the set

$$X'(\Omega)=\Big\{f\in\mathcal{M}(\Omega);\ \int_\Omega |f(x)g(x)|\,dx<\infty \text{ for every } g\in X(\Omega)\Big\},$$

equipped with the norm

$$\|f\|_{X'(\Omega)}=\sup_{\|g\|_{X(\Omega)}\leqslant 1}\int_\Omega |fg|,$$

is called the *associate space* of $X(\Omega)$. Then always $X''(\Omega)=X(\Omega)$ and the *Hölder inequality*

$$\int_\Omega |f(x)g(x)|\,dx\leqslant \|f\|_{X(\Omega)}\|g\|_{X'(\Omega)}$$

holds.

For every r.i. space $X(\Omega)$ there exists a unique r.i. space $\overline{X}(0,1)$ on $(0,1)$ satisfying $\|f\|_{X(\Omega)}=\|f^*\|_{\overline{X}(0,1)}$. Such a space, endowed with the norm

$$\|f\|_{\overline{X}(0,1)}=\sup_{\|g\|_{X(\Omega)}\leqslant 1}\int_0^1 f^*(t)g^*(t)\,dt,$$

is called the *representation space* of $X(\Omega)$.

Let $X(\Omega)$ be an r.i. space. Then the function $\varphi_X:[0,1]\to[0,\infty)$ given by

$$\varphi_X(t)=\begin{cases}\|\chi_{(0,t)}\|_{\overline{X}(0,1)}, & \text{for } t\in(0,1],\\ 0 & \text{for } t=0\end{cases}$$

is called the *fundamental function* of $X(\Omega)$. For every r.i. space $X(\Omega)$ its fundamental function φ_X is *quasiconcave* on $[0,1]$, i.e., it is nondecreasing on $[0,1]$, $\varphi_X(0)=0$, and $\frac{\varphi_X(t)}{t}$ is nonincreasing on $(0,1]$. Moreover,

$$\varphi_X(t)\varphi_{X'}(t)=t \qquad \text{for } t\in[0,1].$$

Given an r.i. space $X(\Omega)$, we can define the *Marcinkiewicz space* $M_X(\Omega)$ corresponding to $X(\Omega)$ as the set of all $f\in\mathcal{M}(\Omega)$ such that

$$\|f\|_{M_X(\Omega)} := \sup_{t\in[0,1]} \varphi_X(t) f^{**}(t) < \infty.$$

Then again, $M_X(\Omega)$ is an r.i. space whose fundamental function is φ_X, and it is the largest such an r.i. space. In particular, when $Z(\Omega)$ is any other r.i. space whose fundamental function is also φ_X, then necessarily

$$Z(\Omega) \hookrightarrow M_X(\Omega).$$

For a comprehensive treatment of r.i. spaces we refer the reader to [5].

3 Reduction Theorems

Recall that Ω is a bounded domain in $\mathbb{R}^n$ having Lipschitz boundary and satisfying $|\Omega| = 1$ and m is an integer satisfying $1 \leqslant m \leqslant n-1$. The basic idea is, again, to compare the size of u with t hat of its mth *gradient* $|D^m u|$ in norms of two function spaces, where $D^m u = \left(\frac{\partial^\alpha u}{\partial x^\alpha}\right)_{0\leqslant|\alpha|\leqslant m}$ and $|D^m u|$ is its Euclidean length. More precisely, we are interested in determining those r.i. spaces $X(\Omega)$ and $Y(\Omega)$ for which

$$\|u\|_{Y(\Omega)} \leqslant C \, \| |D^m u|^*(t) \|_{\overline{X}(0,1)}, \qquad u \in W^m X(\Omega)$$

or, written as a Sobolev embedding,

$$W^m X(\Omega) \hookrightarrow Y(\Omega). \tag{3.1}$$

More specifically, we would like to know that $X(\Omega)$ and $Y(\Omega)$ are *optimal* in the sense that $X(\Omega)$ cannot be replaced by an essentially larger r.i. space and $Y(\Omega)$ cannot be replaced by an essentially smaller one.

The principal idea of our approach to embeddings can be formulated as follows. Our goal is to reduce everything to a one-dimensional inequality involving certain integral operator and then use the available knowledge about weighted inequalities for one-dimensional Hardy type operators on various function spaces. For the first order embedding this was done in [20].

Although the results in [20] are formulated only for Sobolev spaces of functions vanishing on the boundary of Ω, by the combination of the Stein extension theorem [1, Theorem 5.24] with an interpolation argument based on the DeVore–Scherer theorem [19] or [5, Chapt. 5, Theorem 5.12, p. 360], they can be relatively easily extended to bounded domains with Lipschitz boundary. In this approach, the Sobolev space $W^m X(\Omega)$ is extended to $W^m X(\mathbb{R}^n)$ and then restricted again to $W^m X(\Omega_1)$ with $\Omega_1 \supset \Omega$. The details can be found in [28, proof of Theorem 4.1].

The key result in [20] is the following *reduction theorem.*

Theorem 3.1. *Let $X(\Omega)$ and $Y(\Omega)$ be r.i. spaces. Then, in order that the Sobolev embedding*

$$W^1X(\Omega)\hookrightarrow Y(\Omega)$$

holds, it is necessary and sufficient that there exist $C>0$ for which

$$\left\| \int_t^1 f(s)s^{\frac{1}{n}-1}\,ds \right\|_{\overline{Y}(0,1)} \leqslant C\,\|f\|_{\overline{X}(0,1)}\,, \qquad f\in\mathcal{M}_+(0,1).$$

This theorem concerns only the first order embeddings. A natural important question now is, how to obtain a higher order version of the reduction theorem. While the "only if" part is rather straightforward and easily adaptable, the proof of the "if" part of Theorem 3.1 involves a version of the Pólya–Szegö inequality due to Talenti [43], whose higher order version is unavailable without certain restrictions. In 2004, Cianchi [12] obtained the reduction theorem for the case $m=2$ by overcoming certain considerable technical difficulties and using some special estimates for second order derivatives. Finally, in [28], the following general version of the reduction theorem was obtained by a new method using interpolation techniques and properties of special Hardy type operators involving suprema (see the operator $T_{\frac{n}{m}}$ treated below).

Theorem 3.2. *Let $X(\Omega)$ and $Y(\Omega)$ be r.i. spaces. Then the Sobolev embedding* (3.1) *holds if and only if*

$$\left\| \int_t^1 f(s)s^{\frac{m}{n}-1}\,ds \right\|_{\overline{Y}(0,1)} \leqslant C\,\|f\|_{\overline{X}(0,1)}\,, \qquad f\in\mathcal{M}_+(0,1).$$

The proof of Theorem 3.2 is quite involved. We first define the *weighted Hardy operator* $H_{\frac{n}{m}}$, given as

$$(H_{\frac{n}{m}}f)(t) := \int_t^1 f(s)s^{\frac{m}{n}-1}\,ds$$

and its *dual operator* with respect to the L_1 pairing, defined by

$$(H_{\frac{n}{m}}{}'f)(t) := t^{\frac{m}{n}-1}\int_0^t f(s)\,ds, \qquad t\in(0,1),\ f\in\mathcal{M}_+(0,1).$$

Note that when applied to a nonincreasing function f^*, we get

$$(H_{\frac{n}{m}}{}'f^*)(t) = t^{\frac{m}{n}}f^{**}(t) \qquad t\in(0,1),\ f\in\mathcal{M}(\Omega).$$

We observe that the functional

$$\left\|t^{\frac{m}{n}} g^{**}(t)\right\|_{\overline{X}'(0,1)}, \quad g \in \mathcal{M}(\Omega),$$

is an r.i. norm on (Ω). This is easy to verify as the only nontrivial part is the triangle inequality, which follows from the well-known subadditivity of the operation $g \to g^{**}$. Therefore, given an r.i. space $X(\Omega)$, we can define the space $X_\omega(\Omega)$ determined by the functional

$$\|f\|_{X_\omega(\Omega)} := \left\|H_{\frac{n}{m}}{}' f^*\right\|_{\overline{X}(0,1)} = \left\|t^{\frac{m}{n}} f^{**}(t)\right\|_{\overline{X}(0,1)}, \qquad f \in \mathcal{M}(\Omega).$$

Using the same ideas as in [20, Theorem 4.5], it can be shown that $X_\omega(\Omega)$ is also an r.i. space, being, in fact, essentially the largest r.i. space $Y(\Omega)$ satisfying

$$\left\|H_{\frac{n}{m}}{}' f\right\|_{\overline{X}(0,1)} \leqslant C\|f\|_{\overline{Y}(0,1)}, \qquad f \in \mathcal{M}_+(0,1).$$

By duality, $(X')'_\omega(\Omega)$, the associate space of $(X')_\omega(\Omega)$, is essentially the smallest r.i. space $Z(\Omega)$ satisfying

$$\left\|H_{\frac{n}{m}} f\right\|_{\overline{Z}(0,1)} \leqslant C\|f\|_{\overline{X}(0,1)}, \qquad f \in \mathcal{M}_+(0,1).$$

Next, we introduce a special *supremum operator* $T_{\frac{n}{m}}$ by

$$\left(T_{\frac{n}{m}} f\right)(t) := t^{-\frac{m}{n}} \sup_{t \leqslant s < 1} s^{\frac{m}{n}} f^*(s), \qquad f \in \mathcal{M}(0,1),\ t \in (0,1).$$

One readily shows that $T_{\frac{n}{m}}$ is bounded on $L_1(0,1)$ and also on the Lorentz space $L_{\frac{n}{m},\infty}(0,1)$. The key result concerning this operator is that it is bounded on $\overline{X_\omega}(0,1)$ for *absolutely arbitrary* r.i. space $X(\Omega)$. As a consequence, we conclude that for any r.i. space $X(\Omega)$

$$H_{\frac{n}{m}} := \overline{X}(0,1) \to \overline{(X'_\omega)'}(0,1)$$

and $(X'_\omega)'(\Omega)$ is the optimal (smallest) such an r.i. space. The proof of Theorem 3.2 is then completed by combining the obtained estimates with the inequality

$$\int_0^t s^{-\frac{m}{n}} u^*(s)\, ds \leqslant C \int_0^t s^{-\frac{m}{n}} \int_{\frac{s}{2}}^1 |D^m u|^*(y) y^{\frac{m}{n}-1}\, dy\, ds,$$
$$t \in (0,1),\ u \in W^m X(\Omega),$$

which follows from the endpoint Sobolev embeddings, the Holmstedt formulae and the DeVore-Scherer expression for the K-functional between Sobolev spaces.

4 Optimal Range and Optimal Domain of Rearrangement-Invariant Spaces

Now, we show how Theorem 3.2 can be used to characterize the largest r.i. domain space and the smallest r.i. range space in the Sobolev embedding (3.1). Note that Theorem 3.2 implies the following chain of equivalent statements:

$$\begin{aligned} W^m X(\Omega) \hookrightarrow Y(\Omega) \quad &\Leftrightarrow \quad H_{\frac{n}{m}} : \overline{X}(0,1) \to \overline{Y}(0,1) \\ &\Leftrightarrow \quad H_{\frac{n}{m}}{}' : \overline{Y}'(0,1) \to \overline{X}'(0,1) \\ &\Leftrightarrow \quad \left\| t^{\frac{m}{n}} g^{**}(t) \right\|_{\overline{X}'(0,1)} \leqslant C \, \|g\|_{Y'(\Omega)}, \quad g \in \mathcal{M}(\Omega). \end{aligned}$$

The first equivalence is Theorem 3.2 and the second one is duality. The last equivalence is not entirely obvious; the implication "⇒" is restriction to monotone functions, while the converse one follows from the estimate

$$\int_0^t g(s)\, ds \leqslant \int_0^t g^*(s)\, ds,$$

which is just a special case of (2.1). It is of interest to note that when we replace the operator $H_{\frac{n}{m}}{}'$ by $H_{\frac{n}{m}}$, then the corresponding equivalence is no longer true. More precisely, the inequality

$$\left\| H_{\frac{n}{m}} g \right\|_{\overline{Y}(0,1)} \leqslant C \|g\|_{\overline{X}(0,1)}, \quad g \in \mathcal{M}(0,1)$$

implies

$$\left\| H_{\frac{n}{m}} g^* \right\|_{\overline{Y}(0,1)} \leqslant C \|g\|_{\overline{X}(0,1)}, \quad g \in \mathcal{M}(0,1),$$

but *not* vice versa. This illustrates that a Sobolev embedding is a rather delicate process that does not permit a direct duality.

All these ideas are summarized in the following theorem.

Theorem 4.1. *Let $X(\Omega)$ be an r.i. space. Let $Y(\Omega)$ be the r.i. space whose associate space $Y'(\Omega)$ has the norm*

$$\|f\|_{Y'(\Omega)} := \|t^{\frac{m}{n}} f^{**}(t)\|_{\overline{X}'(0,1)}, \qquad f \in \mathcal{M}(\Omega).$$

Then the Sobolev embedding (3.1) *holds, and $Y(\Omega)$ is the optimal* (*i.e., the smallest possible*) *such an r.i. space.*

Theorem 4.1 constitutes an important and rather nice theoretical breakthrough in our search for optimal Sobolev embeddings. On the other hand, it does not easy apply to special examples. Generally speaking, in order to determine $Y(\Omega)$, we have to be able to characterize the associate space of the space whose norm is given by a rather complicated functional

$g \mapsto \left\|t^{\frac{m}{n}} g^{**}(t)\right\|_{\overline{X}'(0,1)}$. That does not have to be easy. Even in the simplest possible instance when $X(\Omega) = L_p(\Omega)$, we can get an explicit formula for $Y(\Omega)$ only by using the duality argument of Sawyer [42], which is highly nontrivial. When $X(\Omega)$ is, for instance, an Orlicz space, the task becomes nearly impossible (however, see Cianchi [13]). In [20], the class of the so-called *Lorentz–Karamata spaces* was introduced and the explicit formulas for the optimal range space were given in the case where the domain space is one of these. The Lorentz–Karamata spaces are a generalization of Lorentz–Zygmund spaces which instead of logarithmic functions involve more general *slowly-varying functions*.

Now, we apply Theorem 4.1 to a particular example, a higher order version of the Maz'ya–Hansson–Brézis–Wainger embedding (1.6).

Example 4.2. Let $X(\Omega) = L_{\frac{n}{m}}(\Omega)$. Then, by Theorem 4.1, its optimal range partner $Y(\Omega)$ is the associate space of $Y'(\Omega)$ determined by the norm

$$\|g\|_{Y'(\Omega)} = \|f^{**}(t)t^{\frac{m}{n}}\|_{L_{\frac{n}{n-m}}(0,1)} = \|f\|_{L_{(1,\frac{n}{n-m})}(\Omega)}.$$

Now, by the duality principle of Sawyer [42], we obtain

$$Y(\Omega) = L_{\infty,\frac{n}{m};-1}(\Omega).$$

For $m = 1$ we recover (1.6). We add a new information that this range space is the best possible among r.i. spaces. As mentioned above already, $W_n(\Omega)$ is still a slightly better range, but it is not an r.i. space for not being linear. The optimality of the range space in a yet broader context was proved by Cwikel and Pustylnik [18].

Another achievement of the reduction theorem is the following characterization of the optimal domain space in a Sobolev embedding.

Theorem 4.3. *Let $Y(\Omega)$ be an r.i. space such that $Y(\Omega) \hookrightarrow L_{\frac{n}{n-m},1}(\Omega)$. Then the function space $X(\Omega)$ generated by the norm*

$$\|f\|_{X(\Omega)} = \sup_{h^*=f^*} \left\|H_{\frac{n}{m}} h\right\|_{\overline{Y}(0,1)}, \qquad f \in \mathcal{M}(\Omega),\ h \in \mathcal{M}(0,1), \tag{4.1}$$

is an r.i. space such that

$$H_{\frac{n}{m}} : \overline{X}(0,1) \to \overline{Y}(0,1)$$

(*hence $W^m X(\Omega) \hookrightarrow Y(\Omega)$*). *Moreover, it is an optimal* (*largest*) *such a space.*

The requirement of the embedding of $Y(\Omega)$ into $L_{\frac{n}{n-m},1}(\Omega)$ is not restrictive as the space $L_{\frac{n}{n-m},1}(\Omega)$ is the range partner for the space $L_1(\Omega)$, the largest of all r.i. spaces. Therefore, larger spaces than $L_{\frac{n}{n-m},1}(\Omega)$ are not interesting range candidates.

Likewise Theorem 4.1, Theorem 4.3 can hardly be directly applied to a particular example since to evaluate $X(\Omega)$ from the quite implicit formula (4.1) involving the supremum over equimeasurable functions is practically impossible. In the search of a simplification, several methods have been applied. Among functions in $\mathcal{M}(0,1)$ that are equimeasurable to a given function f in $\mathcal{M}(\Omega)$, there is one with an exceptional significance, namely f^* itself. So, a natural question arises: Under what conditions one can replace in (4.1) $\sup_{h^*=f^*} \left\|H_{\frac{n}{m}} h\right\|_{\overline{Y}(0,1)}$ by $\left\|H_{\frac{n}{m}} f^*\right\|_{\overline{Y}(0,1)}$? If we could do that without loosing, it would mean a great simplification of the formula (4.1). Of course, only the inequality

$$\sup_{h^*=f^*} \left\|H_{\frac{n}{m}} h\right\|_{\overline{Y}(0,1)} \leqslant C \left\|H_{\frac{n}{m}} f^*\right\|_{\overline{Y}(0,1)}$$

is in question, the converse one is trivial. However, this idea contains one hidden danger: the quantity on the right is not necessarily a norm (recall that the operation $f \mapsto f^*$ is not subadditive, so the triangle inequality is not guaranteed), and, indeed, there are r.i. spaces $Y(\Omega)$ for which it is not. Probably, the simplest example of such $Y(\Omega)$ is $L_1(\Omega)$; it is easy to verify that $\left\|H_{\frac{n}{m}} f^*\right\|_{L_1(0,1)}$ is not a norm. In [20], a sufficient condition was established, namely

$$\left\|H_{\frac{n}{m}} f^{**}\right\|_{\overline{Y}(0,1)} \leqslant C \left\|H_{\frac{n}{m}} f^*\right\|_{\overline{Y}(0,1)}. \tag{4.2}$$

Replacing f^* with f^{**} immediately solves the triangle inequality problem since the operation $f \mapsto f^{**}$ *is* subadditive, but the condition is unsatisfactory (too strong) because it rules out important limiting cases. (It is easy to see that, for example for $Y(\Omega) = L_{\frac{n}{n-m}}(\Omega)$, (4.2) is not true.) In [38], another approach using special operators was elaborated. Finally, in [29], it was shown that a reasonable sufficient condition is the boundedness of the supremum operator $T_{\frac{n}{m}}$ on an associate space of $\overline{Y}(0,1)$.

Theorem 4.4. *Let $Y(\Omega)$ be an r.i. space satisfying*

$$T_{\frac{n}{m}} : \overline{Y}'(0,1) \to \overline{Y}'(0,1). \tag{4.3}$$

In this case, the optimal domain r.i. space $X(\Omega)$ corresponding to $Y(\Omega)$ in (3.1)*, satisfies*

$$\|f\|_{X(\Omega)} \approx \left\| \int_t^1 f^*(s) s^{\frac{m}{n}-1}\, ds \right\|_{\overline{Y}'(0,1)}, \quad f \in \mathcal{M}(\Omega).$$

Here and below, we denote by $\approx$ the comparability of norms.

The condition (4.3) is reasonable and it does not rule out important limiting examples. Moreover, as we shall see, it is quite natural.

Example 4.5. Let us return to the Maz'ya–Hansson–Brézis–Wainger embedding (1.6) or, more precisely, to its higher order modification. Starting with $X(\Omega) = L_{\frac{n}{m}}(\Omega)$, then the corresponding optimal range r.i. space $Y(\Omega)$ is $L_{\infty,\frac{n}{m};-1}(\Omega)$, as it was shown in Example 4.2. In order to be able to apply Theorem 4.4, we must show that $T_{\frac{n}{m}}$ is bounded on the associate space of $\overline{Y}(0,1)$, which, as already observed in Example 4.2, happens to be $L_{(1,\frac{n}{n-m})}(0,1)$. In order to prove the boundedness of $T_{\frac{n}{m}}$ on $L_{(1,\frac{n}{n-m})}(0,1)$, we first note that $T_{\frac{n}{m}}$ is bounded on $L_{1,\frac{n}{n-m}}(0,1)$, which is easier and which can be done either by a standard interpolation argument or by using conditions for the weighted norm inequalities involving the supremum operators from [14] or [22]. Next we show that $(T_{\frac{n}{m}}g)^{**}$ is comparable to $T_{\frac{n}{m}}(g^{**})$. Combining these two facts, we get the desired boundedness of $T_{\frac{n}{m}}$ on $L_{(1,\frac{n}{n-m})}(0,1)$, which is $\overline{Y}'(0,1)$. Hence, according to Theorem 4.4, the optimal r.i. domain partner space $\widetilde{X}(\Omega)$ has the norm

$$\|g\|_{\widetilde{X}(\Omega)} = \|H_{\frac{n}{m}}g^*\|_{L_{\infty,\frac{n}{m};-1}(0,1)}.$$

Now, several interesting facts can be observed about this space. First, it indeed is strictly larger than $X(\Omega) = L_{\frac{n}{m}}(\Omega)$. In fact, it even has an essentially different fundamental function. Moreover, it is a qualitatively new type of function space. In [39], several interesting properties of this space were established, for example its incomparability to several related known function spaces of Orlicz and Lorentz–Zygmund type.

5 Formulas for Optimal Spaces Using the Functional $f^{**} - f^*$

In practice, one often wants to solve the following problem: given m and an r.i. space $X(\Omega)$, find its optimal range r.i. partner, let us call it $Y_X(\Omega)$, so that the Sobolev embedding

$$W^m X(\Omega) \hookrightarrow Y_X(\Omega) \tag{5.1}$$

holds and $Y_X(\Omega)$ is the smallest possible such an r.i. space. A less frequent task, but also of interest, is the converse one; given m and an r.i. space $Y(\Omega)$, find its optimal domain r.i. partner, let us call it $X_Y(\Omega)$, for $Y(\Omega)$ so that

$$W^m X_Y(\Omega) \hookrightarrow Y(\Omega)$$

holds and $X_Y(\Omega)$ is the largest possible such an r.i. space.

At this stage, we have formulas for both $Y_X(\Omega)$ and $X_Y(\Omega)$ given by Theorems 4.1 and 4.3 respectively. As we have already noticed, these formulas are too implicit to allow for some practical use. Theorem 4.3 is particularly bad. In this section, we show that significant simplifications of these formulas, such as the one given by Theorem 4.4, are possible if we *a priori* know that the given space has been chosen in such a way that it is an optimal partner for some other r.i. space.

We first need to introduce one more supremum operator. Let

$$\left(S_{\frac{n}{m}} f\right)(t) := t^{\frac{m}{n}-1} \sup_{0<s\leqslant t} s^{1-\frac{m}{n}} f^*(s), \quad f \in \mathcal{M}(0,1),\ t \in (0,1).$$

Then $S_{\frac{n}{m}}$ has the following endpoint mapping properties:

$$S_{\frac{n}{m}} : L_{\frac{n}{n-m},\infty}(0,1) \to L_{\frac{n}{n-m},\infty}(0,1) \qquad \text{and} \qquad S_{\frac{n}{m}} : L_\infty(0,1) \to L_\infty(0,1).$$

Our point of departure will be the following result from [29].

Theorem 5.1. *Let $X(\Omega)$ be an r.i. space, whose associate space satisfies $X'(\Omega)\hookrightarrow L_{\frac{n}{n-m},\infty}(\Omega)$. Then*

$$\|f\|_{Y_X(\Omega)} \approx \sup_{\|S_{\frac{n}{m}} g\|_{\overline{X}'(0,1)}\leqslant 1} \int_0^1 t^{-\frac{m}{n}} \left[f^{**}(t) - f^*(t)\right] g^*(t)\,dt + \|f\|_{L_1(\Omega)}, \quad (5.2)$$

where $f \in \mathcal{M}(\Omega)$, $g \in \mathcal{M}_+(0,1)$.

The most innovative part of Theorem 5.1 is the new formula (5.2). The L_1-norm has just a cosmetic meaning, its role is to guarantee that the resulting functional is a norm. The main term is formulated as some kind of duality involving the operator $S_{\frac{n}{m}}$. In the case where $S_{\frac{n}{m}}$ can be peeled off, the whole expression is considerably simpler.

Theorem 5.2. *An r.i. space $X(\Omega)$ is the optimal domain partner in* (3.1) *for some other r.i. space $Y(\Omega)$ if and only if*

$$S_{\frac{n}{m}} : \overline{X}'(0,1) \to \overline{X}'(0,1).$$

In this case,

$$\|f\|_{Y_X(\Omega)} \approx \left\|t^{-\frac{m}{n}} \left[f^{**}(t) - f^*(t)\right]\right\|_{\overline{X}(0,1)} + \|f\|_{L_1(\Omega)}, \qquad f \in \mathcal{M}(\Omega).$$

Again, an r.i. space $Y(\Omega)$ is the optimal range partner in (3.1) *for some other r.i. space $X(\Omega)$ if and only if $T_{\frac{n}{m}} : \overline{Y}'(0,1) \to \overline{Y}'(0,1)$.*

In this case,

$$\|f\|_{X_Y(\Omega)} \approx \left\| \int_t^1 f^*(s) s^{\frac{m}{n}-1}\,ds \right\|_{\overline{Y}(0,1)}, \qquad f \in \mathcal{M}(\Omega).$$

This result enables us to apply a new approach. We start with a given r.i. space $X(\Omega)$. We find the corresponding optimal range r.i. partner $Y_X(\Omega)$. Now, the embedding (5.1) has an optimal range, but it does not necessarily have an optimal domain, as Example 4.5 clearly shows. We thus take one more step in order to get the optimal domain r.i. partner for $Y_X(\Omega)$, let us call it $\widetilde{X}(\Omega)$. At this stage however, instead of the rather unpleasant Theorem 4.3, we can use the far more friendly Theorem 4.4 because $Y_X(\Omega)$ is now already known to be the optimal range partner for $X(\Omega)$, and Theorem 5.2 tells us that this is equivalent to the required boundedness of $T_{\frac{n}{m}}$ on $\overline{Y_X}'(0,1)$. Altogether, we have

$$W^m X(\Omega) \subset W^m \widetilde{X}(\Omega) \hookrightarrow Y(\Omega),$$

and $\widetilde{X}(\Omega)$ now can be either equivalent to $X(\Omega)$ or strictly larger. In any case, after these two steps, the couple $(\widetilde{X}(\Omega), Y(\Omega))$ forms an optimal pair in the Sobolev embedding and no further iterations of the process can bring anything new.

The functional $f^{**}(t) - f^*(t)$ appearing in (5.2) should cause some natural concern. It is known [9] that function spaces whose norms involve this functional often do not enjoy nice properties such as linearity, lattice property, or normability. For example, for $X(\Omega) = L_{\frac{n}{m}}(\Omega)$ (see [9, Remark 3.2]) all these properties for $Y_X(\Omega)$ are lost. It is instructive to compare this fact with Theorem 4.4, where this case is ruled out by the assumption $S_{\frac{n}{m}} : \overline{X}'(0,1) \to \overline{X}'(0,1)$. This makes the significance of the supremum operator more transparent; $S_{\frac{n}{m}}$ is bounded on $L_{\frac{n}{n-m},\infty}(0,1)$ but *not* on $L_{\frac{n}{n-m}}(0,1)$. This example is typical, and it illustrates the general principle: the boundedness of $S_{\frac{n}{m}}$ on $\overline{X}'(0,1)$ guarantees that $Y_X(\Omega)$ is a norm.

Incidentally, certain care must be exercised always when the norm of a given function space depends on f^* (for illustration of this fact see [17]). Let us just add that a detailed study of weighted function spaces based on the functional $f^{**} - f^*$ can be found in [9, 10].

Theorem 5.2 can be used to obtain a new description of the space $\widetilde{X}(\Omega)$.

Theorem 5.3. *Let $X(\Omega)$ be an r.i. space, and let $\widetilde{X}(\Omega)$ be defined as above. Define the space $Z(\Omega)$ by*

$$\|g\|_{Z(\Omega)} := \|S_{\frac{n}{m}} g^{**}\|_{\overline{X}'(0,1)}, \qquad g \in \mathcal{M}(\Omega).$$

Then

$$\widetilde{X}(\Omega) = Z'(\Omega).$$

The proofs of Theorems 5.2 and 5.3 reveal a very interesting link between the optimality of r.i. spaces in Sobolev embeddings and their interpolation properties. It is obtained through the following theorem.

Theorem 5.4. *Let $X(\Omega)$ be an r.i. space. Then the operator $T_{\frac{n}{m}}$ is bounded on $\overline{X}'(0,1)$ if and only if $X(\Omega)$ is an interpolation space with respect to the pair $(L_{\frac{n}{n-m},1}(\Omega), L_\infty(\Omega))$, a fact which is written as*

$$X(\Omega) \in \operatorname{Int}(L_{\frac{n}{n-m},1}(\Omega), L_\infty(\Omega)).$$

Similarly, the operator $S_{\frac{n}{m}}$ is bounded on $\overline{X}(0,1)$ if and only if

$$X(\Omega) \in \operatorname{Int}(L_{\frac{n}{n-m},\infty}(\Omega), L_\infty(\Omega)).$$

In other words, r.i. spaces in a Sobolev embedding can be optimal (domains or range) partners for some other r.i. spaces if and only if they satisfy certain interpolation properties. Of course, for example, a very large space, which does not satisfy the interpolation property, can also be a range in a Sobolev embedding, but not the optimal one.

The formulas for optimal spaces given by Theorems 5.2 and 5.3 are still not as explicit as one would desire, but, at least, they show the problem in a new light. They also enable us to obtain explicit formulas for some examples such as Orlicz spaces, previously unavailable. We complete this section by an example that can be computed by using Theorem 5.2.

Theorem 5.5. *Let A be a Young function for which there exists $r > 1$ with*

$$\widetilde{A}(rt) \geqslant 2r^{\frac{n}{n-m}}\widetilde{A}(t), \qquad t \geqslant 1.$$

Then the r.i. spaces $X(\Omega) = L_A(\Omega)$ and $Y(\Omega)$ whose norm is given by

$$\|f\|_{Y(\Omega)} := \|t^{-\frac{m}{n}}[f^{**}(t) - f^*(t)]\|_{L_A(0,1)} + \|f\|_{L_1(\Omega)}, \qquad f \in \mathcal{M}(\Omega)$$

are optimal in (3.1).

6 Explicit Formulas for Optimal Spaces in Sobolev Embeddings

Our goal in this section is to establish explicit formulas for the spaces $Y_X(\Omega)$ and $\widetilde{X}(\Omega)$, given an r.i. space $X(\Omega)$. We recall that the formulas for these spaces which we have so far, are expressed in terms of their associate spaces, namely,

$$\|f\|_{Y_X'(\Omega)} := \|f^{**}(t)t^{\frac{m}{n}}\|_{\overline{X}'(0,1)}, \qquad f \in \mathcal{M}(\Omega), \tag{6.1}$$

and

$$\|g\|_{\widetilde{X}'(\Omega)} := \|S_{\frac{n}{m}} g^{**}\|_{\overline{X}'(0,1)}, \qquad g \in \mathcal{M}(\Omega). \tag{6.2}$$

Our focus is now on the problem how to get these constructions explicit. We first note that the expression for $Y_X'(\Omega)$ turns out to be unsatisfactory in that the function

$$t \to t^{\frac{m}{n}-1} \int_0^t g^*(s)\, ds, \qquad t \in [0,\infty),\ g \in \mathcal{M}(\Omega),$$

need not be nonincreasing. This complicates the construction of explicit formulas for $Y_X(\Omega)$. (However, see [20, Sect. 4] and [29, Sect. 4].) Our next theorem from [31] overcomes this difficulty.

Theorem 6.1. *Suppose that $X(\Omega)$ is an r.i. space satisfying*

$$X(\Omega) \supset L_{\frac{n}{m},1}(\Omega).$$

Define the space $Z_X(\Omega)$ by

$$\|g\|_{Z_X(\Omega)} := \left\| t^{\frac{m}{n}-1} \int_0^t g^*(s) s^{-\frac{m}{n}}\, ds \right\|_{\overline{X}'(0,1)}, \qquad g \in \mathcal{M}(\Omega).$$

Then

$$\|f\|_{Y_X(\Omega)} \approx \left\| t^{-\frac{m}{n}} f^*(t) \right\|_{\overline{Z}'_X(0,1)}, \qquad f \in \mathcal{M}(\Omega).$$

We note that this eliminates the above-mentioned problem since the function

$$t \mapsto t^{\frac{m}{n}-1} \int_0^t g^*(s) s^{-\frac{m}{n}}\, ds, \qquad t \in [0,\infty),$$

is nonincreasing, being a weighted average of a nonincreasing function.

Theorem 6.1 is, again, rather involved. The proof uses delicate estimates and previously obtained optimality results for various integral and supremum operators.

Hence the remaining task is to compute associate spaces of $\widetilde{X}'(\Omega)$ and $Y'_X(\Omega)$. To this end, we use the Brudnyi–Kruglyak duality theory [8] and the interpolation methods using the k-functional, elaborated recently in [27].

The main result reads as follows.

Theorem 6.2. *Suppose that $X(\Omega)$ is an r.i. space. Define the space $V_X(\Omega)$ by*

$$\|g\|_{V_X(\Omega)} := \|g^{**}(t^{1-\frac{m}{n}})\|_{\overline{X}'(0,1)}, \qquad g \in \mathcal{M}(\Omega).$$

Then

$$\|g\|_{\widetilde{X}(\Omega)} \approx \big(k(t, g^*; L_1(0,1), L_{\frac{n}{m},1}(0,1))\big\|_{\overline{V}'_X(0,1)}, \qquad g \in \mathcal{M}(\Omega).$$

Moreover,

$$\|f\|_{Y_X(\Omega)} \approx \left\| k(t, s^{-\frac{m}{n}} f^*(s); L_1(0,1), L_{\frac{n}{m},\infty}(0,1)) \right\|_{\overline{V}'_X(0,1)}, \quad f \in \mathcal{M}(\Omega).$$

Theorem 6.2 can be applied to construct the spaces $Y_X(\Omega)$ and $\widetilde{X}(\Omega)$ explicitly.

Let us now briefly indicate how the interpolation K-method comes in.

Let X_1 and X_2 be Banach spaces, compatible in the sense that they are embedded in a common Hausdorff topological vector space H. Suppose that $x \in X_1 + X_2$ and $t \in [0, \infty)$. The *Peetre K-functional* is defined by

$$K(t, x; X_1, X_2) := \inf_{x=x_1+x_2} \left(\|x_1\|_{X_1} + t\|x_2\|_{X_2}\right), \qquad t > 0.$$

It is an increasing concave function of t on $[0, \infty)$, so that

$$k\left(t, x; X_1, X_2\right) := \frac{d}{dt} K\left(t, x; X_1, X_2\right)$$

is nonincreasing on $[0, \infty)$.

Given an r.i. space Z on $\mathcal{M}_+([0, \infty))$, for which $\left\|\frac{1}{1+t}\right\|_Z < \infty$, the space X, with $\|x\|_X$ defined at $x \in X_1 + X_2$ by

$$\|x\|_X := \left\|t^{-1} K\left(t, x; X_1, X_2\right)\right\|_Z$$

satisfies

$$X_1 \cap X_2 \subset X \subset X_1 + X_2;$$

moreover, for any linear operator T defined on $X_1 + X_2$

$$T : X_i \to X_i, \quad i = 1, 2, \qquad \text{implies} \qquad T : X \to X.$$

We say that the space X is generated by the K-method of interpolation.

The asserted connection of the duality theory for the K-method with our task is through certain reformulations of (6.1) and (6.2), namely

$$\|f\|_{Y_X'(\Omega)} \approx \|t^{\frac{m}{n}-1} K(t^{1-\frac{m}{n}}, f; L_{\frac{n}{n-m},\infty}(0,1), L_\infty(0,1))\|_{\overline{X}'(0,1)}, \quad f \in \mathcal{M}(\Omega),$$

and

$$\|g\|_{\widetilde{X}'(\Omega)} \approx \|t^{\frac{m}{n}-1} K(t^{1-\frac{m}{n}}, g; L_{\frac{n}{n-m},1}(0,1), L_\infty(0,1))\|_{\overline{X}'(0,1)}, \quad g \in \mathcal{M}(\Omega).$$

We complete with an example involving Orlicz spaces.

Theorem 6.3. *Let A be a Young function. Assume that $A(t) = t^q$ near* 0 *and*

$$t^{\frac{m}{n}-1} \notin L_{\widetilde{A}}([0, \infty)).$$

Define B through the equation

$$B(\gamma(t)) := \left(\frac{m}{n} - 1\right) \widetilde{A}\left(t^{\frac{m}{n}-1}\right) \frac{\gamma t)}{t\gamma'(t))}$$

where

$$\gamma(t) := t^{-\frac{m}{n}} \int_t^\infty \widetilde{A}(s^{\frac{m}{n}-1})\, ds, \qquad t \in [0,\infty).$$

Define the space $Z(\Omega)$ *by*

$$\|g\|_{Z(\Omega)} := \|t^{\frac{m}{n}-1} \int_0^{t^{1-\frac{m}{n}}} g^*(s)\, ds\|_{L_{\widetilde{A}}(0,1)}, \qquad g \in \mathcal{M}(\Omega).$$

Then B *is a Young function and*

$$\|f\|_{Z'(\Omega)} \approx \left\| t^{-\frac{m}{n}} f^* \left(t^{1-\frac{m}{n}}\right)\right\|_{\overline{X}(0,1)}, \qquad f \in \mathcal{M}(\Omega).$$

It is of interest to compare this result with that of Cianchi [13] who obtained a description of $Y_X(\Omega)$ different from ours by the use of techniques specific to the Orlicz context.

We note that the results of this section can be applied also to other examples of function spaces such as classical Lorentz spaces of type Gamma and Lambda (details can be found in [31]). However, the formulas are rather complicated and therefore omitted here.

7 Compactness of Sobolev Embeddings

The most important characteristics of Sobolev spaces is not only whether they embed into other function spaces, but also whether they embed compactly.

Let $X(\Omega)$ and $Y(\Omega)$ be two r.i. spaces. We say that $W^mX(\Omega)$ is *compactly embedded* into $Y(\Omega)$ and write

$$W^mX(\Omega) \hookrightarrow\hookrightarrow Y(\Omega)$$

if for every sequence $\{f_k\}$ bounded in $W^mX(\Omega)$ there exists a subsequence $\{f_{k_j}\}$ which is convergent in $Y(\Omega)$.

In the case where $X(\Omega)$ and $Y(\Omega)$ are Lebesgue spaces, we have a theorem, which originated in a lemma of Rellich [41] and was proved specifically for Sobolev spaces by Kondrachov [32], and which asserts that

$$W^{m,p}(\Omega) \hookrightarrow\hookrightarrow L_q(\Omega) \tag{7.1}$$

if $q < p^* = \frac{np}{n-mp}$. Standard examples (see [1]) show that it is not compact when $q = \frac{np}{n-mp}$.

As for embeddings into Orlicz spaces, Hempel, Morris, and Trudinger [24] showed that the embedding (1.3) is not compact. By a standard argument using a uniform absolute continuity of a norm, it can be proved that

$$W^{1,n}(\Omega) \hookrightarrow\hookrightarrow L_B(\Omega)$$

whenever B is a Young function satisfying, with $A(t) := \exp(t^{n'})$ for large values of t,

$$\lim_{t\to\infty} \frac{A(\lambda t)}{B(t)} = \infty$$

for every $\lambda > 0$.

Considering Lorentz spaces, it is of interest to notice that even the Sobolev embedding

$$W^{m,p}(\Omega) \hookrightarrow\hookrightarrow L_{p^*,\infty}(\Omega)$$

is still not compact. (This is not difficult to verify; in fact, standard examples that demonstrate the noncompactness of (7.1) with $q = p^*$ (see, for example, [1]) are sufficient.) The space $L_{p^*,\infty}(\Omega)$ is of course considerably larger than $L_{p^*}(\Omega)$, but it simply is not "larger enough." This observation is a good point of departure since it raises interesting questions.

For example, we may ask whether the space $L_{p^*,\infty}(\Omega)$ is the "*gateway to compactness*" in the sense that every strictly larger space is already a compact range for $W^{m,p}(\Omega)$. It even makes a good sense to formulate this problem in a broader context of r.i. spaces. (Recall that when the Lebesgue space $L_{p^*}(\Omega)$ is replaced by an arbitrary r.i. space $Y(\Omega)$, the role of $L_{p^*,\infty}(\Omega)$ is taken over by the endpoint Marcinkiewicz space $M_Y(\Omega)$.)

We can formulate the following general question (which we have answered for the particular example above).

Let $X(\Omega)$ be an r.i. space, and let $Y_X(\Omega)$ be the corresponding optimal range r.i. space. Let $M_{Y_X}(\Omega)$ be the Marcinkiewicz space corresponding to $Y_X(\Omega)$. Then, of course, $W^m X(\Omega) \hookrightarrow M_{Y_X}(\Omega)$. Can this embedding ever be compact? If not, is the Marcinkiewicz space the gateway to compactness in the above-mentioned sense?

It is clear that in order to obtain satisfactory answers to these and other questions we need a reasonable characterization of pairs of spaces $X(\Omega), Y(\Omega)$ for which we have the compact Sobolev embedding

$$W^m X(\Omega) \hookrightarrow\hookrightarrow Y(\Omega).$$

From various analogous results in less general situations it can be guessed that one such a characterization might be the compactness of $H_{\frac{n}{m}}$ from $\overline{X}(0,1)$ to $\overline{Y}(0,1)$, and another one might be the uniform absolute continuity of the norms of the $H_{\frac{n}{m}}$-image of the unit ball of $\overline{X}(0,1)$ in $\overline{Y}(0,1)$. This guess turns out to be reasonable, but the proof is deep and difficult and contains many unexpected pitfalls. Moreover, the case where the range space is $L_\infty(\Omega)$ must be treated separately.

Theorem 7.1. *Let $X(\Omega)$ and $Y(\Omega)$ be r.i. spaces. Assume that $Y(\Omega) \neq L_\infty(\Omega)$. Then the following three statements are equivalent*:

$$W^m X(\Omega) \hookrightarrow\hookrightarrow Y(\Omega); \tag{7.2}$$

$$H_{\frac{n}{m}} : \overline{X}(0,1) \to\to \overline{Y}(0,1); \tag{7.3}$$

$$\lim_{a\to 0+} \sup_{\|f\|_{\overline{X}(0,1)}\leqslant 1} \left\| \chi_{(0,a)}(t) \int_t^1 f^*(s) s^{\frac{m}{n}-1}\, ds \right\|_{\overline{Y}(0,1)} = 0. \tag{7.4}$$

The case $Y(\Omega) = L_\infty(\Omega)$ is different and, as such, is treated in

Theorem 7.2. *Let $X(\Omega)$ be an r.i. space. Then the following three statements are equivalent*:

$$W^m X(\Omega) \hookrightarrow\hookrightarrow L_\infty(\Omega);$$

$$H_{\frac{n}{m}} : \overline{X}(0,1) \to\to L_\infty(0,1);$$

$$\lim_{a\to 0+} \sup_{\|f\|_{\overline{X}(0,1)}\leqslant 1} \int_0^a f^*(t) t^{\frac{m}{n}-1}\, dt = 0.$$

The most important and involved part is the sufficiency of (7.4) for (7.2). When trying to prove this implication, we discovered an unpleasant technical difficulty. All the methods which we tried to apply, and which would naturally solve the problem, seemed to require

$$\overline{Y}(0,1) \in \operatorname{Int}\,(L_{\frac{n}{n-m},1}(0,1), L_\infty(0,1)),$$

a restriction that does not offer any obvious circumvention. Such a requirement, however, is simply too much to ask. A candidate for a compact range can be as large as it pleases (consider $L_1(\Omega)$) and, in particular, it may by all means lay far outside from the required interpolation sandwich. This obstacle proved to be surprisingly difficult. At the end, it was overcome by the discovery of a useful fact that, given an r.i. space $Y(\Omega)$, we can always construct another one, $Z(\Omega)$, possibly smaller than $Y(\Omega)$, such that the condition (7.4) is still valid, but which already has the required interpolation properties. We formulate this result as a separate theorem because it is of independent interest.

Theorem 7.3. *Let $X(\Omega)$ and $Y(\Omega)$ be r.i. spaces satisfying* (7.4). *Then there exists another r.i. space $Z(\Omega)$ with*

$$\overline{Z}(0,1) \in \operatorname{Int}\,(L_{\frac{n}{n-m},1}(0,1), L_\infty(0,1))$$

such that $Z(\Omega) \hookrightarrow Y(\Omega)$ and

$$\lim_{a\to 0+} \sup_{\|f\|_{X(\Omega)}\leqslant 1} \left\| \chi_{(0,a)}(t) \int_t^1 f^*(s) s^{\frac{m}{n}-1}\, ds \right\|_{Z(\Omega)} = 0.$$

The rest of the proof of the main results uses sharp estimates for supremum operators, various optimality results from the preceding sections, and the Arzela–Ascoli theorem. The proof of Theorem 7.3 is very involved and delicate and requires extensive preparations. The details can be found in [30].

At one stage of the proof, the necessity of the vanishing *Muckenhoupt condition* is shown.

Theorem 7.4. *Let $X(\Omega)$ and $Y(\Omega)$ be r.i. spaces. Assume that $Y(\Omega) \neq L_\infty(\Omega)$. Then each of* (7.2), (7.3), *and* (7.4) *implies*

$$\lim_{a\to 0+} \big\|\chi_{(0,a)}\big\|_{\overline{Y}(0,1)} \big\|t^{\frac{m}{n}-1}\chi_{(a,1)}(t)\big\|_{\overline{X}'(0,1)} = 0.$$

This result shows that a candidate $Y(\Omega)$ for a compact range must have an essentially smaller fundamental function than the optimal embedding space $Y_X(\Omega)$, hence also than the Marcinkiewicz space $M_{Y_X}(\Omega)$. In other words, we must have

$$\lim_{t\to 0+} \frac{\varphi_Y(t)}{\varphi_{Y_X}(t)} = 0.$$

This solves the above question: the embedding

$$W^m X(\Omega) \hookrightarrow M_{Y_X}(\Omega)$$

is always true, but *never* (for any choice of $X(\Omega)$) compact.

Likewise, the “gateway” problem has the negative answer: a counterexample is easily constructed by taking appropriate fundamental functions and using corresponding Marcinkiewicz spaces. It turns out that not even a space which contains $M_{Y_X}(\Omega)$ properly and whose fundamental function is strictly smaller than that of $Y_X(\Omega)$ guarantees compactness.

The connection between a candidate $Y(\Omega)$ for a compact range for a given Sobolev space $W^m X(\Omega)$ and the optimal range $Y_X(\Omega)$ that does imply compactness can be found, but it has to be formulated in terms of a uniform absolute continuity.

Theorem 7.5. *Suppose that $X(\Omega)$ and $Y(\Omega)$ are two r.i. spaces. Assume that $Y(\Omega) \neq L_\infty(\Omega)$. Let $Y_X(\Omega)$ be the optimal r.i. embedding space for $W^m X(\Omega)$. Then* (7.2) *holds if and only if the functions in the unit ball of $Y_X(\Omega)$ have uniformly absolutely continuous norms in $Y(\Omega)$ or, what is the same,*

$$\lim_{a\to 0+} \sup_{\|f\|_{Y_X(\Omega)}\leqslant 1} \big\|\chi_{(0,a)} f^*\big\|_{\overline{Y}(0,1)} = 0, \qquad f \in \mathcal{M}(\Omega). \tag{7.5}$$

Theorem 7.5 gives a necessary and sufficient condition for the compactness of a Sobolev embedding. However, an application of the criterion would involve examination of a uniform absolute continuity of many functions, which may be difficult to verify. It is thus worth looking for a more manageable condition, sufficient for the compactness of the embedding and not too far from being also necessary, which could be used in practical examples. Such a condition is provided by our next theorem. In some sense, it substitutes the negative outcome of the gateway problem.

Theorem 7.6. *Let $X(\Omega)$ and $Y(\Omega)$ be r.i. spaces. Set*

$$\varphi_R(t) := \frac{dc}{dt},$$

where $c(t)$ is the least concave majorant of

$$t\|s^{\frac{m}{n}-1}\chi_{(t,1)}(s)\|_{\overline{X}'(0,1)}.$$

Then the condition

$$\lim_{a\to 0+} \left\|\chi_{(0,a)}\varphi_R\right\|_{\overline{Y}(0,1)} = 0 \tag{7.6}$$

suffices for

$$W^m X(\Omega) \hookrightarrow\hookrightarrow Y(\Omega).$$

Observe that the condition (7.6) can be simply verified in particular examples since it requires to consider just one function rather than the whole unit ball as in (7.5).

Among many examples that can be extracted from these results, we present just one, concerning Orlicz spaces.

Theorem 7.7. *Suppose that A and $\widetilde{A}$ are complementary Young functions and*

$$\int_1^\infty \frac{\widetilde{A}(s)}{s^{1+\frac{n}{n-m}}}\, ds = \infty.$$

Define the Young function $A_R(t)$ for t large by

$$A_R^{-1}(t) := \frac{t^{1-\frac{m}{n}}}{E^{-1}(t)}$$

with

$$E(t) := t^{\frac{n}{n-m}} \int_1^t \frac{\widetilde{A}(s)}{s^{1+\frac{n}{n-m}}}\, ds, \qquad t \geqslant 1.$$

Then

$$W^m L_A(\Omega) \hookrightarrow\hookrightarrow L_B(\Omega)$$

for a given Young function B if and only if

$$\lim_{t\to\infty} \frac{A_R(\lambda t)}{B(t)} = \infty \tag{7.7}$$

for every $\lambda > 0$.

We finally note that, in terms of the explicitly known functions B and E, (7.7) can be expressed by

$$\lim_{t\to\infty} \frac{B\left((\lambda t)^{-1}E(t)^{1-\frac{m}{n}}\right)}{E(t)} = 0 \qquad \text{for every } \lambda > 0.$$

8 Boundary Traces

One of the main applications of Sobolev space techniques is in the field of traces of functions defined on domains. The theory of boundary traces in Sobolev spaces has a number of applications, especially to boundary-value problems for partial differential equations, in particular when the Neumann problem is studied. The *trace operator* defined by

$$\mathrm{Tr}\ u = u_{|\partial\Omega}$$

for a continuous function u on $\overline{\Omega}$ can be extended to a bounded linear operator

$$\mathrm{Tr} : W^{1,1}(\Omega) \to L_1(\partial\Omega),$$

where $L_1(\partial\Omega)$ denotes the Lebesgue space of summable functions on $\partial\Omega$ with respect to the $(n-1)$–dimensional Hausdorff measure $\mathcal{H}^{n-1}$. There exist many powerful methods for proving trace embedding theorems for the trace operator Tr, usually however quite dependent on a particular norms involved. For specific limiting situations other (for example, potential) methods were used, but there does not seem to exist a unified flexible approach that would cover the whole range of situations of interest in applications.

In [15], we developed a new method for obtaining sharp trace inequalities in a general context based on the ideas elaborated in the preceding sections.

Again, the key result is a reduction theorem.

Theorem 8.1. *Let $X(\Omega)$ and $Y(\partial\Omega)$ be r.i. spaces. Then*

$$\left\| \int_{t^{n'}}^{1} f(s) s^{\frac{m}{n}-1}\, ds \right\|_{\overline{Y}(0,1)} \leqslant C\|f\|_{\overline{X}(0,1)}, \qquad f \in \mathcal{M}_+(0,1),$$

if and only if

$$\|\,\mathrm{Tr}\, u\|_{Y(\partial\Omega)} \leqslant C\|u\|_{W^m X(\Omega)} \tag{8.1}$$

for every $u \in W^m X(\Omega)$.

Thus, when dealing with boundary traces, the role of the operator $H_{\frac{n}{m}}$ is taken over by the operator $\int\limits_{t^{n'}}^{1} f(s)s^{\frac{m}{n}-1}\,ds$. Using appropriate interpolation methods, we can characterize the optimal trace range on $\partial\Omega$.

Theorem 8.2. *Let $X(\Omega)$ be an r.i. space. Then the r.i. space $Y(\partial\Omega)$ whose associate norm is given by*

$$\|g\|_{Y'(\partial\Omega)} = \left\| t^{\frac{m-1}{n}} g^{**}(t^{\frac{1}{n'}}) \right\|_{\overline{X}'(0,1)}$$

for every $\mathcal{H}^{n-1}$–measurable function g on $\partial\Omega$ is optimal in (8.1).

Our trace results recover many known examples, prove their optimality that had not been known before, and bring new ones (see [15] for details).

9 Gaussian Sobolev Embeddings

In connection with some specific problems in physics such as quantum fields and hypercontractivity semigroups, it turns out that it would be of interest to extend classical Sobolev embeddings in $\mathbb{R}^n$ to an infinite-dimensional space. The motivation for such things stems from the fact that, in certain circumstances, the study of quantum fields can be reduced to operator or semigroup estimates which are in turn equivalent to inequalities of Sobolev type in infinitely many variables (see [36] and the references therein). However, when we let $n \to \infty$, we have then $\frac{np}{n-p} \to p+$ and so the gain in integrability will apparently be lost. Even more serious, the Lebesgue measure on an infinite-dimensional space is meaningless.

These problems were overcome in the fundamental paper of Gross [23] who replaced the Lebesgue measure by the Gauss one. Note that the Gauss measure γ is defined on $\mathbb{R}^n$ by

$$d\gamma(x) = (2\pi)^{-\frac{n}{2}}\, e^{\frac{-|x|^2}{2}}\, dx.$$

Now, $\gamma(\mathbb{R}^n) = 1$ for every $n \in \mathbb{N}$, hence the extension as $n \to \infty$ is meaningful. The idea was then to seek a version of the Sobolev inequality that would hold on the probability space $(\mathbb{R}^n, \gamma)$ with a constant independent of n. Gross proved [23] an inequality of this kind, which, in particular, entails that

$$\|u - u_\gamma\|_{L_2 \mathrm{Log} L(\mathbb{R}^n,\gamma)} \leqslant C \|\nabla u\|_{L_2(\mathbb{R}^n,\gamma)} \tag{9.1}$$

for every weakly differentiable function u making the right-hand side finite, where

$$u_\gamma = \int_{\mathbb{R}^n} u(x) d\gamma(x),$$

the mean value of u, and $L_2\mathrm{Log}L(\mathbb{R}^n, \gamma)$ is the Orlicz space of those functions u such that $|u|^2|\log|u||$ is integrable in $\mathbb{R}^n$ with respect to γ. Interestingly, (9.1) still provides some slight gain in integrability from $|\nabla u|$ to u, even though it is no longer a power-gain.

In [16], we studied problems concerning the optimality of function spaces in first order Sobolev embeddings on the Gaussian space, namely

$$\|u - u_\gamma\|_{Y(\mathbb{R}^n,\gamma)} \leqslant C\|\nabla u\|_{X(\mathbb{R}^n,\gamma)}. \tag{9.2}$$

As usual, we start with a reduction theorem. This time, the role of the operator $H_{\frac{n}{m}}$ is taken by the operator

$$\int_t^1 \frac{f(s)}{s\sqrt{1+\log(1/s)}}\, ds.$$

The reduction theorem then reads as follows.

Theorem 9.1. *Let $X(\mathbb{R}^n, \gamma)$ and $Y(\mathbb{R}^n, \gamma)$ be r.i. spaces. Then*

$$\|u - u_\gamma\|_{Y(\mathbb{R}^n,\gamma)} \leqslant C \,\|\nabla u\|_{X(\mathbb{R}^n,\gamma)}$$

for every $u \in W^1X(\mathbb{R}^n, \gamma)$ if and only if

$$\left\| \int_t^1 \frac{f(s)}{s\sqrt{1+\log(1/s)}}\, ds \right\|_{\overline{Y}(0,1)} \leqslant C \,\|f\|_{\overline{X}(0,1)}$$

for every $f \in \overline{X}(0, 1)$.

Then the characterization of the optimal range r.i. space for the Gaussian Sobolev embedding when the domain space is obtained via the usual scheme.

Theorem 9.2. *Let $X(\mathbb{R}^n, \gamma)$ be an r.i. space, and let $Z(\mathbb{R}^n, \gamma)$ be the r.i. space equipped with the norm*

$$\|g\|_{Z(\mathbb{R}^n,\gamma)} := \left\| \frac{g^{**}(s)}{\sqrt{1+\log\frac{1}{s}}} \right\|_{\overline{X}'(0,1)}$$

for any measurable function u on $\mathbb{R}^n$. Let $Y(\mathbb{R}^n, \gamma) = Z'(\mathbb{R}^n, \gamma)$. Then $Y(\mathbb{R}^n, \gamma)$ is the optimal range space in the Gaussian Sobolev embedding (9.2)*.*

The role of the operator $T_{\frac{n}{m}}$ is in the Gaussian setting taken over by the operator

$$(Tf)(t) = \sqrt{1+\log\frac{1}{t}}\sup_{t\leqslant s\leqslant 1}\frac{f^*(s)}{\sqrt{1+\log\frac{1}{s}}} \qquad \text{for } t\in(0,1).$$

With the help of the operator T, we can characterize the optimal domain space.

Theorem 9.3. *Let $Y(\mathbb{R}^n,\gamma)$ be an r.i. space such that*

$$\exp L^2(\mathbb{R}^n,\gamma)\hookrightarrow Y(\mathbb{R}^n,\gamma)\hookrightarrow L\,(\log L)^{\frac{1}{2}}\,(\mathbb{R}^n,\gamma)$$

and

c̄enterlineT is bounded on $\overline{Y}'(0,1)$.

Let $X(\mathbb{R}^n,\gamma)$ be the r.i. space equipped with the norm

$$\|u\|_{X(\mathbb{R}^n,\gamma)} = \left\| \int_t^1 \frac{u^*(s)}{s\sqrt{1+\log\frac{1}{s}}}\,ds \right\|_{\overline{Y}(0,1)}.$$

Then $X(\mathbb{R}^n,\gamma)$ is the optimal domain space for $Y(\mathbb{R}^n,\gamma)$ in the Gaussian Sobolev embedding (9.2).

We now collect the basic examples.

Example 9.4. (i) Let $1\leqslant p<\infty$. Then the spaces $X(\mathbb{R}^n,\gamma) = L_p(\mathbb{R}^n,\gamma)$ and $Y(\mathbb{R}^n,\gamma) = L_p(\log L)^{\frac{p}{2}}(\mathbb{R}^n,\gamma)$ form an optimal pair in the Gaussian Sobolev embedding (9.2).

(ii) The spaces $X(\mathbb{R}^n,\gamma) = L_\infty(\mathbb{R}^n,\gamma), Y(\mathbb{R}^n,\gamma) = \exp L^2(\mathbb{R}^n,\gamma)$ form an optimal pair in the Gaussian Sobolev embedding (9.2).

(iii) Let $\beta>0$. Then the spaces $(\exp L^\beta(\mathbb{R}^n,\gamma), \exp L^{\frac{2\beta}{2+\beta}}(\mathbb{R}^n,\gamma))$ form an optimal pair in the Gaussian Sobolev embedding (9.2).

These examples demonstrate a surprising phenomenon: while there is a *gain* in integrability when the domain space is a Lebesgue space, there is actually a *loss* near L_∞. This fact is caused by the nature of the Gaussian measure which rapidly decreases at infinity.

Acknowledgement. This research was supported by the Czech Ministry of Education (project MSM 0021620839) and the Grant Agency of the Czech Republic (grant no. 201/08/0383).

References

1. Adams, R.A., Fournier, J.J.F.: Sobolev Spaces. Academic Press, Amsterdam (2003)
2. Bastero, J., Milman, M., Ruiz, F.: A note in $L(\infty,q)$ spaces and Sobolev embeddings. Indiana Univ. Math. J. **52**, 1215–1230 (2003)

3. Bennett, C., DeVore, R., Sharpley, R.: Weak L^∞ and BMO. Ann. Math. **113**, 601–611 (1981)
4. Bennett, C., Rudnick, K.: On Lorentz–Zygmund spaces. Dissert. Math. **175**, 1–72 (1980)
5. Bennett, C., Sharpley, R.: Interpolation of Operators. Academic Press, Boston (1988)
6. Brézis, H., Wainger, S.: A note on limiting cases of Sobolev embeddings and convolution inequalities. Commun. Partial Differ. Equ. **5**, 773–789 (1980)
7. Brudnyi, Yu.A.: Rational approximation and imbedding theorems (Russian). Dokl. Akad. Nauk SSSR **247**, 269–272 (1979); English transl.: Sov. Math., Dokl. **20**, 681–684 (1979)
8. Brudnyi, Yu.A., Kruglyak, N.Ya.: Interpolation Functors and Interpolation Spaces, Vol. 1. North-Holland (1991)
9. Carro, M., Gogatishvili, A., Martín, J., Pick, L.: Functional properties of rearrangement invariant spaces defined in terms of oscillations. J. Funct. Anal. **229**, no. 2, 375–404 (2005)
10. Carro, M., Gogatishvili, A., Martín, J., Pick, L.: Weighted inequalities involving two Hardy operators with applications to embeddings of function spaces. J. Operator Theory **59**, no. 2, 101–124 (2008)
11. Cianchi, A.: A sharp embedding theorem for Orlicz–Sobolev spaces. Indiana Univ. Math. J. **45**, 39–65 (1996)
12. Cianchi, A.: Symmetrization and second order Sobolev inequalities. Ann. Mat. Pura Appl. **183**, 45–77 (2004)
13. Cianchi, A.: Optimal Orlicz–Sobolev embeddings. Rev. Mat. Iberoam. **20**, 427–474 (2004)
14. Cianchi, A., Kerman, R., Opic, B., Pick, L.: A sharp rearrangement inequality for fractional maximal operator. Stud. Math. **138**, 277–284 (2000)
15. Cianchi, A., Kerman, R., Pick, L.: Boundary trace inequalities and rearrangements. J. d'Analyse Math. [To appear]
16. Cianchi, A., Pick, L.: Optimal Gaussian Sobolev embeddings. Manuscript (2008)
17. Cwikel, M., Kamińska, A., Maligranda, L., Pick, L.: Are generalized Lorentz "spaces" really spaces? Proc. Am. Math. Soc. **132**, 3615–3625 (2004)
18. Cwikel, M., Pustylnik, E.: Weak type interpolation near "endpoint" spaces. J. Funct. Anal. **171**, 235–277 (1999)
19. DeVore, R.A., Scherer, K.: Interpolation of linear operators on Sobolev spaces. Ann. Math. **109**, 583–599 (1979)
20. Edmunds, D.E., Kerman, R., Pick, L.: Optimal Sobolev imbeddings involving rearrangement-invariant quasinorms. J. Funct. Anal. **170**, 307–355 (2000)
21. Evans, W.D., Opic, B., Pick, L.: Interpolation of operators on scales of generalized Lorentz–Zygmund spaces. Math. Nachr. **182**, 127–181 (1996)
22. Gogatishvili, A., Opic, B., Pick, L.: Weighted inequalities for Hardy type operators involving suprema. Collect. Math. **57**, no. 3, 227–255 (2006)
23. Gross, L.: Logarithmic Sobolev inequalities. Am. J. Math. **97**, 1061–1083 (1975)
24. Hempel, J.A., Morris, G.R., Trudinger, N.S.: On the sharpness of a limiting case of the Sobolev imbedding theorem. Bull. Australian Math. Soc. **3**, 369–373 (1970)
25. Holmstedt, T.: Interpolation of quasi-normed spaces. Math. Scand. **26**, 177–199 (1970)
26. Hunt, R.: On $L(p, q)$ spaces. Enseign. Math. **12**, 249–276 (1966)
27. Kerman, R., Milman, M., Sinnamon, G.: On the Brudnyi–Krugljak duality theory of spaces formed by the K-method of interpolation. Rev. Mat. Compl. **20**, 367–389 (2007)
28. Kerman, R., Pick, L.: Optimal Sobolev imbeddings. Forum Math **18**, no. 4, 535–570 (2006)

29. Kerman, R., Pick, L.: Optimal Sobolev imbedding spaces. Preprint No. MATH–KMA–2005/161, Department of Mathematical Analysis, Faculty of Mathematics and Physics, Charles University, Prague, 2005, 1–19, http://www.karlin.mff.cuni.cz/˜rokyta/preprint/index.php. [Submitted for publication]
30. Kerman, R., Pick, L.: Compactness of Sobolev imbeddings involving rearrangement-invariant norms. Stud. Math. [To appear]
31. Kerman, R., Pick, L.: Explicit formulas for optimal rearrangement-invariant norms in Sobolev imbedding inequalities. (2008) [In preparation]
32. Kondrachov, V.I.: Certain properties of functions in the space L^p (Russian, French). Dokl. Akad. Nauk SSSR **48**, 535–538 (1945)
33. Malý, J., Pick, L.: An elementary proof of sharp Sobolev embeddings. Proc. Am. Math. Soc. **130**, 555–563 (2002)
34. Maz'ya, V.G.: Sobolev Spaces. Springer-Verlag, Berlin–Tokyo (1985)
35. O'Neil, R.: Convolution operators and $L(p, q)$ spaces. Duke Math. J. **30**, 129–142 (1963)
36. Nelson, E.: The free Markoff field. J. Funct. Anal. **12**, 221–227 (1973)
37. Peetre, J.: Espaces d'interpolation et théorème de Soboleff. Ann. Inst. Fourier **16**, 279–317 (1966)
38. Pick, L.: Supremum operators and optimal Sobolev inequalities. In: Function Spaces, Differential Operators and Nonlinear Analysis, Jpp. 207–219. Math. Inst. AS CR, Prague (2000)
39. Pick, L.: Optimal Sobolev Embeddings. Rudolph–Lipschitz–Vorlesungsreihe no. 43, Rheinische Friedrich–Wilhelms–Universität Bonn (2002)
40. Pokhozhaev, S.I.: Eigenfunctions of the equation $\Delta u + \lambda f(u) = 0$ (Russian). Dokl. Akad. Nauk SSSR **165**, 36–39 (1965); English transl.: Sov. Math., Dokl. **6**, 1408–1411 (1965)
41. Rellich, F.: Ein Satz über mittlere Konvergenz. Gött. Nachr. 30–35 (1930)
42. Sawyer, E.: Boundedness of classical operators on classical Lorentz spaces. Stud. Math. **96**, 145–158 (1990)
43. Talenti, G.: Best constant in Sobolev inequality. Ann. Mat. Pura Appl. **110**, 353–372 (1976)
44. Trudinger, N.S.: On imbeddings into Orlicz spaces and some applications. J. Math. Mech. **17**, 473–483 (1967)
45. Yudovich, V.I.: Some estimates connected with integral operators and with solutions of elliptic equations (Russian). Dokl. Akad. Nauk SSSR **138**, 805–808 (1961); English transl.: Sov. Math., Dokl. **2**, 746–749 (1961)

On the Hardy–Sobolev–Maz'ya Inequality and Its Generalizations

Yehuda Pinchover and Kyril Tintarev

Abstract The paper deals with natural generalizations of the Hardy–Sobolev–Maz'ya inequality and some related questions, such as the optimality and stability of such inequalities, the existence of minimizers of the associated variational problem, and the natural energy space associated with the given functional.

1 Introduction

The term "inequalities of Hardy–Sobolev type" refers, somewhat vaguely, to families of inequalities that in some way interpolate the Hardy inequality

$$\int\limits_{\Omega} |\nabla u(x)|^p \, \mathrm{d}x \geqslant C(N,p,K,\Omega) \int\limits_{\Omega} \frac{|u(x)|^p}{\operatorname{dist}(x,K)^p} \, \mathrm{d}x, \qquad u \in C_0^\infty(\Omega \setminus K), \tag{1.1}$$

where $\Omega \subset \mathbb{R}^N$ is an open domain and $K \subset \overline{\Omega}$ is a nonempty closed set, and the Sobolev inequality

$$\int\limits_{\Omega} |\nabla u(x)|^p \, \mathrm{d}x \geqslant C \left(\int\limits_{\Omega} |u(x)|^{p^*} \, \mathrm{d}x \right)^{p/p^*}, \qquad u \in C_0^\infty(\Omega), \tag{1.2}$$

Yehuda Pinchover
Technion – Israel Institute of Technology, Haifa 32000, Israel,
e-mail: pincho@techunix.technion.ac.il

Kyril Tintarev
Uppsala University, P.O. Box 480, SE-751 06 Uppsala, Sweden,
e-mail: kyril.tintarev@math.uu.se

V. Maz'ya (ed.), *Sobolev Spaces in Mathematics I, International Mathematical Series.*
doi: 10.1007/978-0-387-85648-3,

where $1 < p < N$ and $p^* \stackrel{\mathrm{def}}{=} pN/(N-p)$ is the corresponding Sobolev exponent. Throughout the paper, we repeatedly consider the following particular case.

Example 1.1. Let $\Omega = \mathbb{R}^N = \mathbb{R}^n \times \mathbb{R}^m$, where $1 < m \leqslant N$, and let $K = \mathbb{R}^n \times \{0\}$. We denote by y and z the variables of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively and set $\mathbb{R}^N_0 \stackrel{\mathrm{def}}{=} \mathbb{R}^n \times (\mathbb{R}^m \setminus \{0\})$. It is well known that the Hardy inequality (1.1) holds with the best constant

$$C(N, p, \mathbb{R}^n \times \{0\}, \mathbb{R}^N) = \left|\frac{m-p}{p}\right|^p. \tag{1.3}$$

An elementary family of Hardy–Sobolev inequalities can be obtained by Hölder interpolation between the Hardy inequality and the Sobolev inequality. More significant inequalities of Hardy–Sobolev type with the best constant in the Hardy term can be derived as consequences of Caffarelli–Kohn–Nirenberg inequality [5, 13] that provides estimates in terms of the weighted gradient norm $\displaystyle\int |\xi|^\alpha |\nabla u|^p \mathrm{d}\xi$. The substitution $u = |y|^\beta v$ into the Caffarelli–Kohn–Nirenberg inequality can be used to produce inequalities that combine terms with the critical exponent and with the Hardy potential. Such inequalities are known as *Hardy–Sobolev–Maz'ya inequalities* (HSM inequalities for brevity). In particular, Maz'ya [16, Sect. 2.1.6, Corollary 3] proved the HSM inequality

$$\int\limits_{\mathbb{R}^N_0} |\nabla u|^2 \,\mathrm{d}y\,\mathrm{d}z - \left(\frac{m-2}{2}\right)^2 \int\limits_{\mathbb{R}^N_0} \frac{|u|^2}{|y|^2}\,\mathrm{d}y\,\mathrm{d}z \geqslant C\left(\int\limits_{\mathbb{R}^N_0} |u|^{2^*}\,\mathrm{d}y\,\mathrm{d}z\right)^{2/2^*}, \qquad u \in C_0^\infty(\mathbb{R}^N_0), \tag{1.4}$$

where $2 < N$ and $1 \leqslant m < N$. This HSM inequality is false for $m = N$ and reduces to the Sobolev inequality for $m = 2$. Since the left-hand side of (1.4) induces a Hilbert norm, the inequality holds on $\mathcal{D}^{1,2}(\mathbb{R}^N_0)$, the completion of $C_0^\infty(\mathbb{R}^N_0)$ in the gradient norm, which coincides with $\mathcal{D}^{1,2}(\mathbb{R}^N)$ for all $m > 1$, in particular, $C_0^\infty(\mathbb{R}^N_0)$ may be replaced by $C_0^\infty(\mathbb{R}^N)$ unless $m = 1$.

The joint paper [8] by Filippas, Maz'ya, and Tertikas gives the following generalization of the HSM inequality (1.4).

Example 1.2. Let $2 \leqslant p < N$, $p \neq m < N$, and let $\Omega \subset \mathbb{R}^N$ be a bounded domain. Let K be a compact C^2-manifold without boundary embedded in $\mathbb{R}^N$ of codimension m such that $K \Subset \Omega$ for $1 < m < N$ (i.e., K is compact in Ω) or $K = \partial\Omega$ for $m = 1$. Assume further that

$$-\Delta_p \operatorname{dist}(\cdot, K)^{(p-m)/(p-1)} \geqslant 0 \qquad \text{in} \;\; \Omega \setminus K, \tag{1.5}$$

where $\Delta_p(u) \stackrel{\text{def}}{=} \nabla \cdot (|\nabla u|^{p-2} \nabla u)$ is the p-Laplacian.

Then for all $u \in C_0^\infty(\Omega \setminus K)$ we have

$$\int\limits_\Omega |\nabla u(x)|^p \,\mathrm{d}x - \left|\frac{m-p}{p}\right|^p \int\limits_\Omega \frac{|u(x)|^p}{\operatorname{dist}(x, K)^p} \,\mathrm{d}x \geqslant C \left(\int\limits_\Omega |u(x)|^{p^*} \,\mathrm{d}x \right)^{p/p^*}. \tag{1.6}$$

For $N = 3$ Benguria, Frank, and Loss [3] have shown recently that the best constant C in (1.4) is the Sobolev constant S_3. Mancini and Sandeep [14] have studied the analog of HSM on the hyperbolic space and its close connection to the original HSM inequality.

In the present paper, we consider a nonnegative functional Q of the form

$$Q(u) \stackrel{\text{def}}{=} \int\limits_\Omega (|\nabla u|^p + V|u|^p) \,\mathrm{d}x, \qquad u \in C_0^\infty(\Omega), \tag{1.7}$$

where $\Omega \subseteq \mathbb{R}^N$ is a domain, $V \in L^\infty_{\text{loc}}(\Omega)$, and $1 < p < \infty$. We study several questions related to extensions of the inequalities (1.4) and (1.6). In Sect. 2, we deal with generalizations of these HSM inequalities for the functional Q. It turns out that, in the subcritical case, a *weighted* HSM inequality holds, where the weight appears in the Sobolev term. In the critical case, one needs to add a Poincaré type term (a one-dimensional p-homogeneous functional), and we call it the *Hardy–Sobolev–Maz'ya–Poincaré* inequality. We show that under "small" perturbations such Hardy–Sobolev–Maz'ya type inequalities are preserved (with the original Sobolev weight). We also address the question concerning the optimal weight in the generalized Hardy–Sobolev–Maz'ya inequality.

In Sect. 3, we study a natural energy space $\mathcal{D}_V^{1,2}(\Omega)$ for nonnegative singular Schrödinger operators and discuss the existence of minimizers for the Hardy–Sobolev–Maz'ya inequality in this space, i.e., minimizers of the equivalent Caffarelli–Kohn–Nirenberg inequality. Finally, in Sect. 4, we prove that a related functional $\widehat{Q}$ which satisfies $C^{-1}Q \leqslant \widehat{Q} \leqslant CQ$ for some $C > 0$ induces a norm on the cone of nonnegative $C_0^\infty(\Omega)$-functions. For $p = 2$ this norm coincides (on the above cone) with the $\mathcal{D}_V^{1,2}(\Omega)$-norm defined in [18]. It is our hope that this approach paves the way to circumvent the general lack of convexity of the nonnegative functional Q for $p \neq 2$.

2 Generalization of the Hardy–Sobolev–Maz'ya Inequality

We need the following definition.

Definition 2.1. Let $\Omega \subseteq \mathbb{R}^N$ be a domain, $V \in L^\infty_{\text{loc}}(\Omega)$, and $1 < p < \infty$. Assume that the functional

$$Q(u) = \int\limits_\Omega \left(|\nabla u|^p + V|u|^p\right) \mathrm{d}x \tag{2.1}$$

is nonnegative on $C_0^\infty(\Omega)$. A function $\varphi \in C^1(\Omega)$ is a *ground state* for the functional Q if φ is an L^p_{loc}-limit of a nonnegative sequence $\{\varphi_k\} \subset C_0^\infty(\Omega)$ satisfying

$$Q(\varphi_k) \to 0 \qquad \text{and} \qquad \int\limits_B |\varphi_k|^p \,\mathrm{d}x = 1$$

for some fixed $B \Subset \Omega$ (such a sequence $\{\varphi_k\}$ is called a *null sequence*). The functional (1.7) is called *critical* if Q admits a ground state and *subcritical* or *weakly coercive* if it does not.

The following statement (see [20]) is a generalization of the Hardy–Sobolev–Maz'ya inequality. The inequality (2.4) might be called a Hardy–Sobolev–Maz'ya–Poincaré type inequality.

Theorem 2.2. *Let Q be a nonnegative functional on $C_0^\infty(\Omega)$ of the form* (1.7), *and let $1 < p < N$.*

(i) *The functional Q does not admit a ground state if and only if there exists a positive continuous function W such that*

$$Q(u) \geqslant \left(\int\limits_\Omega W|u|^{p^*} \mathrm{d}x \right)^{p/p^*}, \qquad u \in C_0^\infty(\Omega). \tag{2.2}$$

(ii) *If Q admits a ground state φ, then φ is a unique global positive* (*super*)*-solution of the Euler–Lagrange equation*

$$Q'(u) \overset{\text{def}}{=} -\Delta_p(u) + V|u|^{p-2}u = 0 \quad \textit{in} \quad \Omega. \tag{2.3}$$

Moreover, there exists a positive continuous function W such that for every function $\psi \in C_0^\infty(\Omega)$ with

$$\int\limits_\Omega \psi\varphi \,\mathrm{d}x \neq 0$$

the following inequality holds:

$$Q(u) + C\left|\int_\Omega \psi u \,\mathrm{d}x\right|^p \geqslant \left(\int_\Omega W|u|^{p^*}\mathrm{d}x\right)^{p/p^*}, \qquad u \in C_0^\infty(\Omega) \tag{2.4}$$

for some suitable constant $C > 0$.

Remark 2.3. For relationships between the criticality of Q in Ω and the p-capacity (with respect to the functional Q) of closed balls see [20, Theorem 4.5] and [24, 25].

Theorem 2.2 applies to the case of $\Omega = \mathbb{R}_0^N$ and the Hardy potential (see Example 1.1 and, in particular, (1.4)), but it does not specify that the weight W in the Sobolev term is the constant function. We note that Example 1.2 provides another Hardy type functional satisfying the Hardy–Sobolev–Maz'ya inequality with the weight $W =$ constant.

On the other hand, let $\Omega = \mathbb{R}_0^N$ with $m = N$. Then the corresponding Hardy functional admits a ground state $\varphi(x) = |x|^{(p-N)/p}$, and therefore the Hardy–Sobolev–Maz'ya inequality does not hold with any weight. Moreover, the Hardy–Sobolev–Maz'ya–Poincaré inequality (2.4) which, by Theorem 2.2, holds with some weight W is false with the weight $W =$ constant (see [9] and Example 2.5 below).

Let us present few other examples which illustrate further the question of admissible weights in the Hardy–Sobolev–Maz'ya inequality and the Hardy–Sobolev–Maz'ya–Poincaré inequality. The first two examples are elementary, but general. In the first one, the Hardy–Sobolev–Maz'ya inequality (2.2) holds with the constant weight function, while in the second example (Example 2.5 below) such an inequality is false.

Example 2.4. Consider a nonnegative functional Q of the form (1.7), where $V \in L^\infty_{\mathrm{loc}}(\Omega)$ is a nonzero function and $1 < p < N$. For $\lambda \in \mathbb{R}$ we denote

$$Q_\lambda(u) \overset{\text{def}}{=} \int_\Omega (|\nabla u|^p + \lambda V|u|^p)\,\mathrm{d}x.$$

Then for every $\lambda \in (0, 1)$ there exists $C > 0$ such that

$$Q_\lambda(u) \geqslant C\|u\|_{p^*}^p, \qquad u \in C_0^\infty(\Omega), \tag{2.5}$$

where $C = C(N, p, \lambda) > 0$. This HSM inequality follows from

$$Q_\lambda(u) = (1-\lambda)\int_\Omega |\nabla u|^p\,\mathrm{d}x + \lambda Q(u) \geqslant (1-\lambda)\int_\Omega |\nabla u|^p\,\mathrm{d}x$$

and the Sobolev inequality.

Example 2.5. Let $Q \geqslant 0$ be of the form (1.7), where $1 < p < N$. Suppose that Q admits ground state $\varphi \notin L^{p^*}(\Omega)$. Let $\{\varphi_k\}$ be a null sequence (see Definition 2.1) such that $\varphi_k \to \varphi$ locally uniformly in Ω (for the existence of a locally uniform convergence null sequence see [20, Theorem 4.2]). Let $V_1 \in L^\infty(\Omega)$ be a nonzero nonnegative function with compact support. Then

$$Q(\varphi_k) + \int_\Omega V_1 |\varphi_k|^p \,\mathrm{d}x \to \int_\Omega V_1 |\varphi|^p \,\mathrm{d}x < \infty,$$

while the Fatou lemma implies that $\|\varphi_k\|_{p^*} \to \infty$. Therefore, the subcritical functional

$$Q_{V_1}(u) \stackrel{\mathrm{def}}{=} Q(u) + \int_\Omega V_1 |u|^p \,\mathrm{d}x$$

does not satisfy the HSM inequality (2.2) with the constant weight. A similar argument shows that the critical functional Q does not satisfy the Hardy–Sobolev–Maz'ya–Poincaré inequality with the constant weight.

Remark 2.6. Example 2.5 can be slightly generalized by replacing the assumption $\varphi \notin L^{p^*}(\Omega)$ with $\varphi \notin L^{p^*}(\Omega, W\,\mathrm{d}x)$, where W is a continuous positive weight function. Under this assumption it follows that the functionals Q_{V_1} and Q do not satisfy the Hardy–Sobolev–Maz'ya inequality and respectively the Hardy–Sobolev–Maz'ya–Poincaré inequality with the weight W.

Example 2.7. Filippas, Tertikas, and Tidblom [10, Theorem C] proved that a nonnegative functional Q of the form (1.7) with $p = 2$ and $N > 1$ satisfies the HSM inequality in a smooth domain Ω with $W =$ constant if the equation $Q'(u) = 0$ has a positive C^2-solution φ such that the following L^1-Hardy type inequality holds:

$$\int_\Omega \varphi^{2(N-1)/(N-2)} |\nabla u| \,\mathrm{d}x \geqslant C \int_\Omega \varphi^{N/(N-2)} |\nabla \varphi|\, |u| \,\mathrm{d}x, \qquad u \in C_0^\infty(\Omega).$$

Example 2.8. Consider the function

$$X(r) \stackrel{\mathrm{def}}{=} (|\log r|)^{-1}, \qquad r > 0.$$

Let $\Omega \subset \mathbb{R}^N$, $N > 2$, be a bounded domain, and let $D \stackrel{\mathrm{def}}{=} \sup_{x \in \Omega} |x|$. The following inequality is due to Filippas and Tertikas [9, Theorem A]:

$$\int_\Omega |\nabla u|^2\,\mathrm{d}x - \left(\frac{N-2}{2}\right)^2 \int_\Omega \frac{|u|^2}{|x|^2}\,\mathrm{d}x$$

$$\geqslant C\left(\int_\Omega |u|^{2^*} X(|x|/D)^{1+N/(N-2)}\,\mathrm{d}x\right)^{2/2^*}, \qquad u\in C_0^\infty(\Omega). \quad (2.6)$$

Moreover, the exponent $1+N/(N-2)$ in (2.6) cannot be decreased. In particular, in this case, the HSM inequality does not hold with $W=$ constant (cf. Example 2.5 and Remark 2.6).

We now consider the question whether the weight W in the HSM inequality (2.2) is preserved (up to a constant multiple) under small perturbations.

Theorem 2.9. *Let Ω be a domain in $\mathbb{R}^N$, $N>2$, and let $V\in L^\infty_{\mathrm{loc}}(\Omega)$. Assume that the following functional Q satisfies the* HSM *inequality*

$$Q(u) \stackrel{\mathrm{def}}{=} \int_\Omega \left(|\nabla u|^2 + V|u|^2\right)\,\mathrm{d}x \geqslant \left(\int_\Omega W|u|^{2^*}\,\mathrm{d}x\right)^{2/2^*}, \qquad u\in C_0^\infty(\Omega) \quad (2.7)$$

with some positive continuous function W. Let $\widetilde{V}\in L^\infty_{\mathrm{loc}}(\Omega)$ be a nonzero potential satisfying

$$|\widetilde{V}|^{N/2} W^{(2-N)/2} \in L^1(\Omega). \quad (2.8)$$

Consider the one-parameter family of functionals $\widetilde{Q}_\lambda$ defined by

$$\widetilde{Q}_\lambda(u) \stackrel{\mathrm{def}}{=} Q(u) + \lambda\int_\Omega \widetilde{V}|u|^2\,\mathrm{d}x,$$

where $\lambda\in\mathbb{R}$.

(i) *If $\widetilde{Q}_\lambda$ is nonnegative on $C_0^\infty(\Omega)$ and does not admit a ground state, then*

$$\widetilde{Q}_\lambda(u) \geqslant C\left(\int_\Omega W|u|^{2^*}\,\mathrm{d}x\right)^{2/2^*}, \qquad u\in C_0^\infty(\Omega), \quad (2.9)$$

where C is a positive constant.

(ii) *If $\widetilde{Q}_\lambda$ is nonnegative on $C_0^\infty(\Omega)$ and admits a ground state v, then for every $\psi\in C_0^\infty(\Omega)$ such that $\int_\Omega \psi v\,\mathrm{d}x \neq 0$ we have*

$$\widetilde{Q}_\lambda(u) + C_1 \left(\int_\Omega \psi u \,\mathrm{d}x \right)^2 \geqslant C \left(\int_\Omega W|u|^{2^*} \,\mathrm{d}x \right)^{2/2^*}, \qquad u \in C_0^\infty(\Omega) \tag{2.10}$$

with suitable positive constants $C, C_1 > 0$.

(iii) *The set*

$$S \stackrel{\mathrm{def}}{=} \{\lambda \in \mathbb{R} \mid \widetilde{Q}_\lambda \geqslant 0 \text{ on } C_0^\infty(\Omega)\}$$

is a closed interval with nonempty interior which is bounded if and only if $\widetilde{V}$ changes its sign on a set of positive measure in Ω. Moreover, $\lambda \in \partial S$ if and only if $\widetilde{Q}_\lambda$ is critical in Ω.

Proof. (i)–(ii) Let $\mathcal{D}^{1,2}_{\lambda\widetilde{V}}(\Omega)$ denote the completion of $C_0^\infty(\Omega)$ with respect to the norm defined by the square root of the left-hand side of (2.9) if $\widetilde{Q_\lambda}$ does not admit a ground state, and by the square root of the left-hand side of (2.10) if $\widetilde{Q_\lambda}$ admits a ground state (see [18]). Similarly, we denote by $\mathcal{D}^{1,2}_V(\Omega)$ the completion of $C_0^\infty(\Omega)$ with respect to the norm defined by the square root of the left-hand side of (2.7). We denote by $\|\cdot\|_{\mathcal{D}^{1,2}_{\lambda\widetilde{V}}}$ and $\|\cdot\|_{\mathcal{D}^{1,2}_V}$ the norms on $\mathcal{D}^{1,2}_{\lambda\widetilde{V}}(\Omega)$ and $\mathcal{D}^{1,2}_V(\Omega)$ respectively.

Assume that (2.9) (respectively, (2.10)) does not hold. Then there exists a sequence $\{u_k\} \subset C_0^\infty(\Omega)$ such that

$$\|u_k\|_{\mathcal{D}^{1,2}_{\lambda\widetilde{V}}} \to 0 \qquad \text{and} \qquad \int_\Omega W|u_k|^{2^*} \,\mathrm{d}x = 1. \tag{2.11}$$

By [18, Proposition 3.1], the space $\mathcal{D}^{1,2}_{\lambda\widetilde{V}}(\Omega)$ is continuously imbedded into $W^{1,2}_{\mathrm{loc}}(\Omega)$. Therefore, $u_k \to 0$ in $W^{1,2}_{\mathrm{loc}}(\Omega)$. Consequently, for any $K \Subset \Omega$ we have

$$\lim_{k\to\infty} \int_K |\widetilde{V}||u_k|^2 \,\mathrm{d}x = 0. \tag{2.12}$$

On the other hand, (2.8) and the Hölder inequality imply that for any $\varepsilon > 0$ there exists $K_\varepsilon \Subset \Omega$ such that

$$\left| \int_{\Omega\setminus K_\varepsilon} \widetilde{V}|u_k|^2 \,\mathrm{d}x \right| \leqslant \left(\int_{\Omega\setminus K_\varepsilon} |\widetilde{V}|^{N/2} W^{(2-N)/2} \,\mathrm{d}x \right)^{2/N} \left(\int_\Omega W|u_k|^{2^*} \,\mathrm{d}x \right)^{2/2^*} < \varepsilon. \tag{2.13}$$

Since

$$\|u_k\|_{\mathcal{D}^{1,2}_V} \leqslant \|u_k\|_{\mathcal{D}^{1,2}_{\lambda\widetilde{V}}} + \left|\int_\Omega \lambda\widetilde{V}|u_k|^2\,\mathrm{d}x\right|^{1/2},$$

from (2.11)–(2.13) it follows that $u_k \to 0$ in $\mathcal{D}^{1,2}_V(\Omega)$. Therefore, (2.7) implies that

$$\int_\Omega W|u_k|^{2^*}\,\mathrm{d}x \to 0,$$

which contradicts the assumption

$$\int_\Omega W|u_k|^{2^*}\,\mathrm{d}x = 1.$$

Consequently, (2.9) (respectively, (2.10)) holds.

(iii) From [19, Proposition 4.3] it follows that S is an interval and that $\lambda \in \operatorname{int} S$ implies that Q_λ is subcritical in Ω. The claim concerning the boundedness of S is trivial and is left to the reader.

On the other hand, suppose that for some $\lambda \in \mathbb{R}$ the functional $\widetilde{Q}_\lambda$ is subcritical. By part (i), $\widetilde{Q}_\lambda$ satisfies the HSM inequality with the weight W. Therefore, (2.13) (with $K_\varepsilon = \varnothing$) implies that

$$\widetilde{Q}_\lambda(u) \geqslant C\left(\int_\Omega W|u|^{2^*}\,\mathrm{d}x\right)^{2/2^*} \geqslant C_1\left|\int_\Omega \widetilde{V}|u|^2\,\mathrm{d}x\right|, \qquad u \in C_0^\infty(\Omega). \tag{2.14}$$

Therefore, $\lambda \in \operatorname{int} S$. Consequently, $\lambda \in \partial S$ implies that $\widetilde{Q}_\lambda$ is critical in Ω. In particular, $0 \in \operatorname{int} S$. □

Example 2.10. Let $\Omega = \mathbb{R}^N$, where $N \geqslant 3$, and let $V \in L^{N/2}(\mathbb{R}^N)$ such that $V \not\geqslant 0$ (so, V is a short range potential). Fix $\mu < (N-2)^2/4$. Then the classical Hardy inequality, together with Example 2.4 and Theorem 2.9, implies that there exists $\lambda^* > 0$ such that for $\lambda < \lambda^*$ we have the following HSM inequality:

$$\int_{\mathbb{R}^N} |\nabla u|^2\,\mathrm{d}x - \mu\int_{\mathbb{R}^N} \frac{|u|^2}{|x|^2}\,\mathrm{d}x + \lambda\int_{\mathbb{R}^N} V(x)|u|^2\,\mathrm{d}x$$

$$\geqslant C_\lambda\left(\int_{\mathbb{R}^N} |u|^{2^*}\,\mathrm{d}x\right)^{2/2^*}, \qquad u \in C_0^\infty(\mathbb{R}^N). \tag{2.15}$$

On the other hand, if $\lambda = \lambda^*$, then the associated functional is critical and satisfies the corresponding Hardy–Sobolev–Maz'ya–Poincaré inequality with the weight function W = constant. Recall that the Hardy–Sobolev–Maz'ya inequality and Hardy–Sobolev–Maz'ya–Poincaré inequality for $\mu = (N-2)^2/4$ are false with the weight W = constant (see Example 2.5 and [9]).

Example 2.11. Consider again Example 1.2 with $p = 2 < N$ and $2 \neq m < N$. By [8, Theorem 1.1], there exists $M \leqslant 0$ such that the following HSM inequality holds:

$$Q(u) \stackrel{\text{def}}{=} \int_\Omega |\nabla u|^2 \,\mathrm{d}x - \left(\frac{m-2}{2}\right)^2 \int_\Omega \frac{|u|^2}{\operatorname{dist}(x,K)^2} \,\mathrm{d}x - M \int_\Omega |u|^2 \,\mathrm{d}x$$

$$\geqslant C \left(\int_\Omega |u|^{2^*} \,\mathrm{d}x \right)^{2/2^*}, \quad u \in C_0^\infty(\Omega \setminus K). \tag{2.16}$$

We note that if (1.5) is satisfied, then (2.16) holds with $M = 0$.

Let $V \in L^\infty_{\mathrm{loc}}(\Omega) \cap L^{N/2}(\Omega)$ be a nonzero function. Consider the one-parameter family of functionals Q_λ defined by

$$Q_\lambda(u) \stackrel{\text{def}}{=} Q(u) + \lambda \int_\Omega V|u|^2 \,\mathrm{d}x,$$

where $\lambda \in \mathbb{R}$. By Theorem 2.9, the set S of all λ such that Q_λ is nonnegative on $C_0^\infty(\Omega)$ is a nonempty closed interval with nonempty interior. Moreover, for $\lambda \in \operatorname{int} S$ there exists a positive constant c_λ such that

$$Q_\lambda(u) \geqslant c_\lambda \left(\int_\Omega |u|^{2^*} \,\mathrm{d}x \right)^{2/2^*}, \qquad u \in C_0^\infty(\Omega \setminus K). \tag{2.17}$$

On the other hand, if $\lambda \in \partial S$, then Q_λ admits a ground state v. Therefore, Theorem 2.9 implies that for every $\psi \in C_0^\infty(\Omega \setminus K)$ satisfying $\int_\Omega \psi v \,\mathrm{d}x \neq 0$ there exist constants $C, C_1 > 0$ such that

$$Q_\lambda(u) + C \left(\int_\Omega u\psi \,\mathrm{d}x \right)^2 \geqslant c_1 \left(\int_\Omega |u|^{2^*} \,\mathrm{d}x \right)^{2/2^*}, \quad u \in C_0^\infty(\Omega \setminus K).$$

We note that if $K = \partial\Omega$ is smooth (i.e., $m = 1$) and $V = \mathbf{1}$, one actually deals with the case considered by Brezis and Marcus in [4, Theorem 1.1]. In particular, let λ^* be the supremum of all $\lambda \in \mathbb{R}$ such that the inequality

$$\int_{\Omega} |\nabla u|^2 \, \mathrm{d}x - \frac{1}{4} \int_{\Omega} \frac{|u|^2}{\mathrm{dist}\,(x, \partial\Omega)^2} \, \mathrm{d}x - \lambda \int_{\Omega} |u|^2 \, \mathrm{d}x \geqslant 0, \qquad u \in C_0^\infty(\Omega), \quad (2.18)$$

holds ($\lambda^* > -\infty$ and is attained by [4, Theorem 1.1]). Then Theorem 2.9 implies that for each $\lambda < \lambda^*$ there exists $C_\lambda > 0$ such that

$$\int_{\Omega} |\nabla u|^2 \, \mathrm{d}x - \frac{1}{4} \int_{\Omega} \frac{|u|^2}{\mathrm{dist}\,(x, \partial\Omega)^2} \, \mathrm{d}x - \lambda \int_{\Omega} |u|^2 \, \mathrm{d}x \geqslant C_\lambda \Bigg(\int_{\Omega} |u|^{2^*} \, \mathrm{d}x \Bigg)^{2/2^*}, \quad u \in C_0^\infty(\Omega). \quad (2.19)$$

Moreover, Theorem 2.9 implies that for $\lambda = \lambda^*$, the functional defined by the left-hand side of (2.19) is critical and satisfies the Hardy–Sobolev–Maz'ya–Poincaré inequality with the weight $W =$ constant. In particular, the corresponding Euler–Lagrange equation $Q'_{\lambda^*}(u) = 0$ in Ω admits a unique positive (super)-solution.

Theorem 1.1 in [4] was extended by Marcus and Shafrir in [15, Theorem 1.2] to the case $1 < p < \infty$ and a perturbation $0 < V(x) = O(\mathrm{dist}\,(x, \partial\Omega)^\gamma)$, where $\gamma > -p$ (cf. our assumption (2.8), where $p = 2$). Following [15], let λ^* be the supremum of all $\lambda \in \mathbb{R}$ such that the following inequality holds:

$$\int_{\Omega} |\nabla u|^2 \, \mathrm{d}x - \frac{1}{4} \int_{\Omega} \frac{|u|^2}{\mathrm{dist}\,(x, \partial\Omega)^2} \, \mathrm{d}x - \lambda \int_{\Omega} V(x)|u|^2 \, \mathrm{d}x \geqslant 0, \quad u \in C_0^\infty(\Omega). \quad (2.20)$$

It follows that Theorem 2.9 with the constant weight applies also to this functional if, in addition, $V \in L^\infty_{\mathrm{loc}}(\Omega) \cap L^{N/2}(\Omega)$.

Remark 2.12. We note that even under the less restricted assumptions of [15, Theorem 1.2] with $p = 2$ and $\lambda = \lambda^*$, one can show that the positive solution u_* of Equation (1.14) in [15] is actually a ground state. Therefore, u_* is a unique (up to a multiplicative constant) global positive supersolution of that equation and the corresponding functional is critical.

Indeed, Lemma 5.1 of [15] implies that any positive supersolution of [15, Equation (1.14)] satisfies

$$Cu(x) \geqslant \mathrm{dist}\,(x, \partial\Omega)^{1/2}, \qquad x \in \Omega. \quad (2.21)$$

On the other hand, Theorem 1.2 in [15] implies that the positive solution u_* satisfies

$$u_*(x) \asymp \mathrm{dist}\,(x, \partial\Omega)^{1/2}, \qquad x \in \Omega, \quad (2.22)$$

where $f \asymp g$ means that there exists a positive constant C such that $C^{-1} \leqslant f/g \leqslant C$ in Ω. Now, we take a positive supersolution u. Let ε be the maximal positive number such that $u - \varepsilon u_* \geqslant 0$ in Ω. Note that, by (2.21) and (2.22), ε is well defined. By the strong maximum principle, either $u = \varepsilon u_*$ or $u - \varepsilon u_* > 0$. Consequently, (2.21) and (2.22) imply that there exists a positive constant C_1 such that

$$u - \varepsilon u_* \geqslant C \text{dist}\,(x, \partial\Omega)^{1/2} \geqslant C_1 u_* \qquad \text{in} \quad \Omega,$$

which contradicts the definition of ε.

3 The Space $\mathcal{D}_V^{1,2}(\Omega)$ and Minimizers for the Hardy–Sobolev–Maz'ya Inequality

Consider again the Hardy–Sobolev–Maz'ya inequality (1.4). This inequality clearly extends to $\mathcal{D}^{1,2}(\mathbb{R}^N)$ for $m > 2$ and to $\mathcal{D}^{1,2}(\mathbb{R}^N_0)$ for $m = 1$, but since the quadratic form $Q(u)$ on the left-hand side of (1.4) induces a scalar product on $C_0^\infty(\mathbb{R}^N_0)$, the natural domain of Q is the completion of $C_0^\infty(\mathbb{R}^N_0)$ with respect to the norm $Q(\cdot)^{1/2}$. Recall [18] that for a given general subcritical functional Q of the form (1.7) (with $p = 2$) such a completion is denoted by $\mathcal{D}_V^{1,2}(\Omega)$. Similarly to the standard definition of $\mathcal{D}^{1,2}(\mathbb{R}^N)$ for $N = 1, 2$, when Q admits a ground state, one appends to $Q(u)$ a correction term of the form

$$\left(\int\limits_\Omega \psi u \, \mathrm{d}x \right)^2 .$$

Hence, by (2.2) and (2.4), the space $\mathcal{D}_V^{1,2}(\Omega)$ is continuously imbedded into a weighted L^{2^*}-space.

In the particular case (1.4), V is the Hardy potential $[(m-2)/2]^2|y|^{-2}$. By (1.4), the space $\mathcal{D}_V^{1,2}(\mathbb{R}^N_0)$ is continuously imbedded into $L^{2^*}(\mathbb{R}^N_0)$. Thus, its elements can be identified as measurable functions. The substitution $u = |y|^{(2-m)/2}v$ transforms the Hardy–Sobolev–Maz'ya inequality (1.4) into the following inequality of Caffarelli–Kohn–Nirenberg type:

$$\int\limits_{\mathbb{R}^N} |y|^{2-m} |\nabla v|^2 \, \mathrm{d}y \, \mathrm{d}z$$
$$\geqslant C \left(\int\limits_{\mathbb{R}^N} |y|^{(2-m)2^*/2} |v|^{2^*} \, \mathrm{d}y \, \mathrm{d}z \right)^{2/2^*}, \quad v \in \mathcal{D}^{1,2}(\mathbb{R}^N_0, |y|^{2-m} \, \mathrm{d}y \, \mathrm{d}z). \tag{3.1}$$

The left-hand side of (3.1) defines a Hilbert space isometric to $\mathcal{D}_V^{1,2}(\mathbb{R}_0^N)$. However, the Lagrange density

$$|\nabla u|^2 - \left(\frac{m-2}{2}\right)^2 \frac{|u|^2}{|y|^2} \tag{3.2}$$

is no longer integrable for an arbitrary $u \in \mathcal{D}_V^{1,2}(\mathbb{R}_0^N)$. The integrable Lagrange density of (3.1), $|y|^{2-m}|\nabla(u|y|^{(m-2)/2})|^2$ can be equated to (3.2) by partial integration when $u \in C_0^\infty(\mathbb{R}_0^N)$, but this connection does not extend to the whole of $\mathcal{D}_V^{1,2}(\mathbb{R}_0^N)$ as the terms that mutually cancel in the partial integration on $C_0^\infty(\mathbb{R}_0^N)$ might become infinite. In particular, it should not be expected a priori that the minimizer for the Hardy–Sobolev–Maz'ya inequality in $\mathcal{D}_V^{1,2}(\mathbb{R}_0^N)$ would have a finite gradient in $L^2(\mathbb{R}_0^N, \mathrm{d}x)$.

The existence of minimizers for the variational problem associated with (3.1) is proved in [23] for all codimensions $0 < m < N$, where $N > 3$. The existence proof is based on the concentration compactness argument which utilizes invariance properties of the problem. Similarly to other problems, where lack of compactness stems from a noncompact equivariant group of transformations, some general domains and potentials admit minimizers and some do not, and an analogy with similar elliptic problems in $\mathcal{D}^{1,2}(\mathbb{R}^N)$ provides useful insights (see, for example, [21]).

4 Convexity Properties of the Functional Q for $p > 2$

The definition of $\mathcal{D}_V^{1,2}(\Omega)$ cannot be applied to other values of p since for $p \neq 2$ the positivity of the functional Q on $C_0^\infty(\Omega)$ does not necessarily imply its convexity, and thus it does not give rise to a norm. For the lack of convexity when $p > 2$ see an elementary one-dimensional counterexample at the end of [6] and also the proof of Theorem 7 in [12]. For $p < 2$ see [11, Example 2].

On the other hand, by [19, Theorem 2.3], the functional Q is nonnegative on $C_0^\infty(\Omega)$ if and only if the equation $Q'(u) = 0$ in Ω admits a positive global solution v. With the help of such a solution v, one has the identity [7, 1, 2]

$$Q(u) = \int\limits_{\Omega} L_v(w)\,\mathrm{d}x, \qquad u \in C_{0+}^\infty(\Omega),$$

where $w \stackrel{\text{def}}{=} u/v$, the Lagrangian $L_v(w)$ is defined by

$$L_v(w) \stackrel{\text{def}}{=} |v\nabla w + w\nabla v|^p - w^p|\nabla v|^p - pw^{p-1}v|\nabla v|^{p-2}\nabla v \cdot \nabla w \geqslant 0, \quad w \in C_{0+}^\infty(\Omega), \tag{4.1}$$

and $C_{0+}^\infty(\Omega)$ denotes the cone of all nonnegative functions in $C_0^\infty(\Omega)$.

The following proposition claims that the nonnegative Lagrangian $L_v(w)$, which contains indefinite terms, is bounded from above and from below by multiples of a simpler Lagrangian.

Proposition 4.1 ([17, Lemma 2.2]). *Let v be a positive solution of the equation $Q'(u) = 0$ in Ω. Then*

$$L_v(w) \asymp v^2|\nabla w|^2 \left(w|\nabla v| + v|\nabla w|\right)^{p-2} \qquad \forall w \in C_{0+}^{\infty}(\Omega). \tag{4.2}$$

In particular, for $p \geqslant 2$

$$L_v(w) \asymp \widehat{L}_v(w) \stackrel{\text{def}}{=} v^p|\nabla w|^p + v^2|\nabla v|^{p-2}w^{p-2}|\nabla w|^2 \qquad \forall w \in C_{0+}^{\infty}(\Omega). \tag{4.3}$$

Define the simplified energy $\widehat{Q}$ by

$$\widehat{Q}(u) \stackrel{\text{def}}{=} \int_{\Omega} \widehat{L}_v(w)\,\mathrm{d}x, \qquad w = u/v \in C_{0+}^{\infty}(\Omega). \tag{4.4}$$

It is shown in [17] that for $p > 2$ neither of the terms in the simplified energy $\widehat{Q}$ is dominated by the other.

From Proposition 4.1 it follows that

$$Q(u) = Q(|u|) \asymp \widehat{Q}(|u|), \qquad u \in C_0^{\infty}(\Omega).$$

In [22], the solvability of the equation $Q'(u) = f$ is proved in the class of functions u satisfying $Q^{**}(u) < \infty$, where $Q^{**} \leqslant Q$ is the second convex conjugate (in the sense of Legendre transformation) of Q. If the inequality $Q \leqslant CQ^{**}$ is true, then ${Q^{**}}^{1/p}(u)$ would define a norm and Q would extend to a Banach space, which should be regarded as the natural energy space for the functional Q.

On the other hand, if $p > 2$, it is not clear whether the functional $\widehat{Q}$ is convex due to the second term in (4.3). It has, however, the following convexity property.

Proposition 4.2. *Assume that $p \geqslant 2$. Let $v \in C_{\mathrm{loc}}^1(\Omega)$ be a fixed positive function. Consider the functional*

$$\mathcal{Q}(\psi) \stackrel{\text{def}}{=} \widehat{Q}(v\psi^{2/p}), \qquad \psi \in C_{0+}^{\infty}(\Omega),$$

where $\widehat{Q}$ is defined by (4.3) and (4.4). Then the functional $\mathcal{Q}$ is convex on $C_{0+}^{\infty}(\Omega)$.

Proof. We first split each of the functionals $\widehat{Q}$ and $\mathcal{Q}$ into the sum of two functionals

$$\widehat{Q}_1(u) \stackrel{\text{def}}{=} \int_{\Omega} v^p|\nabla w|^p\mathrm{d}x, \ \ \widehat{Q}_2(u) \stackrel{\text{def}}{=} \int_{\Omega} v^2|\nabla v|^{p-2}w^{p-2}|\nabla w|^2\mathrm{d}x, \ \ w = u/v \in C_{0+}^{\infty}(\Omega),$$

$$\mathcal{Q}_1(\psi) \stackrel{\text{def}}{=} \widehat{Q}_1(v\psi^{2/p}) = \int\limits_{\Omega} v^p |\nabla(\psi^{2/p})|^p \,\mathrm{d}x, \qquad \psi \in C_{0+}^\infty(\Omega),$$

$$\mathcal{Q}_2(\psi) \stackrel{\text{def}}{=} \widehat{Q}_2(v\psi^{2/p}) = \int\limits_{\Omega} v^2 |\nabla v|^{p-2} \psi^{2(p-2)/p} |\nabla(\psi^{2/p})|^2 \,\mathrm{d}x, \quad \psi \in C_{0+}^\infty(\Omega).$$

Thus, $\widehat{Q} = \widehat{Q}_1 + \widehat{Q}_2$ and $\mathcal{Q} = \mathcal{Q}_1 + \mathcal{Q}_2$.

For $t \in [0,1]$ and $w_0, w_1 \in C_{0+}^\infty(\Omega)$ let

$$w_t \stackrel{\text{def}}{=} \Big[(1-t)w_0^{p/2} + t w_1^{p/2}\Big]^{2/p}.$$

Then

$$\nabla w_t = \frac{(1-t)w_0^{p/2-1}\nabla w_0 + t w_1^{p/2-1}\nabla w_1}{\Big[(1-t)w_0^{p/2} + t w_1^{p/2}\Big]^{1-2/p}}.$$

Therefore,

$$|\nabla w_t| \leqslant \frac{[(1-t)^{2/p} w_0]^{p/2-1}(1-t)^{2/p}|\nabla w_0| + (t^{2/p} w_1)^{p/2-1} t^{2/p} |\nabla w_1|}{\Big[(1-t)w_0^{p/2} + t w_1^{p/2}\Big]^{1-2/p}}. \tag{4.5}$$

Applying the Hölder inequality to the sum in the numerator of (4.5) (with the terms $(1-t)^{2/p}|\nabla w_0|$ and $t^{2/p}|\nabla w_1|$ raised to the power $p/2$) and taking into account that the conjugate of $p/2$ is reciprocal to $1-2/p$, we have

$$|\nabla w_t|^{p/2} \leqslant (1-t)|\nabla w_0|^{p/2} + t|\nabla w_1|^{p/2}. \tag{4.6}$$

From (4.6) it easily follows that

$$|\nabla w_t|^p \leqslant (1-t)|\nabla w_0|^p + t|\nabla w_1|^p.$$

Setting $\psi_t \stackrel{\text{def}}{=} w_t^{p/2}$, $t \in [0,1]$, we immediately conclude that $\mathcal{Q}_1$ is convex as a function of ψ. The same conclusion extends to $\mathcal{Q}_2$ once we note that

$$w^{p-2}|\nabla w|^2 = (2/p)^2 |\nabla w^{p/2}|^2$$

and use (4.6) for $p = 4$. □

Let

$$N(\psi) \stackrel{\text{def}}{=} [\mathcal{Q}(\psi)]^{1/2} = \Big[\widehat{Q}(v\psi^{2/p})\Big]^{1/2}, \qquad \psi \in C_{0+}^\infty(\Omega). \tag{4.7}$$

It is immediate that $N(\psi) > 0$ for $\psi \in C_{0+}^\infty(\Omega)$, unless $\psi = 0$, and that $N(\lambda\psi) = \lambda N(\psi)$ for $\lambda \geqslant 0$. By Proposition 4.2, the functional $N(\cdot)$ satisfies the triangle inequality

$$N(\psi_1 + \psi_2) \leqslant N(\psi_1) + N(\psi_2), \qquad \psi_1, \psi_2 \in C_{0+}^{\infty}(\Omega).$$

Thus, we equipped the cone $C_{0+}^{\infty}(\Omega)$ with a norm. For $p = 2$ the functional $Q = \widehat{Q}$ is a positive quadratic form, and thus convex. Consequently, in the subcritical case, $Q^{1/2}$ extends the functional N to a norm on the whole $C_0^{\infty}(\Omega)$ and then, by completion, to the Hilbert space $\mathcal{D}_V^{1,2}(\Omega)$. It would be interesting to introduce $\mathcal{D}_V^{1,p}(\Omega)$ for $p > 2$, once one finds an extension of N to $C_0^{\infty}(\Omega)$.

Acknowledgement. Part of this research was done while K. Tintarev was visiting the Technion. K. Tintarev would like to thank the Technion for the kind hospitality. Y. Pinchover acknowledges the support of the Israel Science Foundation (grant no. 587/07) founded by the Israeli Academy of Sciences and Humanities, and the B. and G. Greenberg Research Fund (Ottawa).

References

1. Allegretto, M., Huang, Y.X.: A Picone's identity for the p-Laplacian and applications. Nonlinear Anal., Theory Methods Appl. **32**, 819–830 (1998)
2. Allegretto, M., Huang, Y.X.: Principal eigenvalues and Sturm comparison via Picone's identity. J. Differ. Equations **156**, 427–438 (1999)
3. Benguria, R., Frank, R., Loss, M.: The Sharp Constant in the Hardy–Sobolev–Maz'ya Inequality in the Three Dimensional Upper Half-Space. Preprint, arXiv:0705.3833.
4. Brezis, H., Marcus, M.: Hardy's inequalities revisited. Dedicated to Ennio De Giorgi. Ann. Sc. Norm. Super. Pisa, Cl. Sci. (4) **25**, 217–237 (1997)
5. Caffarelli, L., Kohn, R., Nirenberg, L.: First order interpolation inequalities with weight. Compos. Math. **53**, 259–275 (1984)
6. del Pino, M., Elgueta, M., Manasevich, R.: A homotopic deformation along p of a Leray–Schauder degree result and existence for $(\|u'\|^{p-2}u')' + f(t,u) = 0,\ u(0) = u(T) = 0,\ p > 1$. J. Differ. Equations **80**, 1–13 (1989)
7. Diaz, J.I., Saá, J.E.: Existence et unicité de solutions positives pour certaines équations elliptiques quasilinéaires. C. R. Acad. Sci., Paris Ser. I Math. **305**, 521–524 (1987)
8. Filippas, S., Maz'ya, V., Tertikas, A.: Critical Hardy–Sobolev inequalities. J. Math. Pures Appl. (9) **87**, 37–56 (2007)
9. Filippas, S., Tertikas, A.: Optimizing improved Hardy inequalities. J. Funct. Anal. **192**, 186–233 (2002)
10. Filippas, S., Tertikas, A., Tidblom, J.: On the Structure of Hardy–Sobolev–Maz'ya Inequalities. Preprint arXiv:0802.0986.
11. Fleckinger-Pellé, J., Hernández, J., Takáč, P., de Thélin, F.: Uniqueness and positivity for solutions of equations with the p-Laplacian. In: Proc. Conf. Reaction-Diffusion Equations (Trieste, 1995), pp.141–155. Lect. Notes Pure Appl. Math. **194**, Marcel Dekker, New York (1998)
12. García-Melián, J., Sabina de Lis, J.: Maximum and comparison principles for operators involving the p-Laplacian. J. Math. Anal. Appl. **218**, 49–65 (1998)

13. Il'in, V.P.: Some integral inequalities and their applications in the theory of differentiable functions of several variables (Russian). Mat. Sb. **54**, no. 3, 331–380 (1961)
14. Mancini, G., Sandeep, K.: On a Semilinear Elliptic Equation in $\mathbb{H}^n$. Preprint (2007)
15. Marcus, M., Shafrir, I.: An eigenvalue problem related to Hardy's L^p inequality. Ann. Sc. Norm. Super. Pisa, Cl. Sci. (4) **29**, 581–604 (2000)
16. Maz'ya, V.: Sobolev Spaces. Springer-Verlag, Berlin (1985)
17. Pinchover, Y., Tertikas, A., Tintarev, K.: A Liouville-type theorem for the p-Laplacian with potential term. Ann. Inst. H. Poincaré. Anal. Non Linéaire **25**, 357–368 (2008)
18. Pinchover, Y., Tintarev, K.: Ground state alternative for singular Schrödinger operators. J. Funct. Anal. **230**, 65–77 (2006)
19. Pinchover, Y., Tintarev, K.: Ground state alternative for p-Laplacian with potential term. Calc. Var. Partial Differ. Equ. **28** (2007), 179–201.
20. Pinchover, Y., Tintarev, K.: On positive solutions of minimal growth for singular p-Laplacian with potential term. Adv. Nonlin. Studies **8**, 213–234 (2008)
21. Smets, D.: Nonlinear Schrödinger equations with Hardy potential and critical nonlinearities. Trans. Am. Math. Soc. **357**, 2909–2938 (2005)
22. Takáč, P., Tintarev, K.: Generalized minimizer solutions for equations with the p-Laplacian and a potential term. Proc. R. Soc. Edinb., Sect. A, Math. **138**, 201–221 (2008)
23. Tertikas, A., Tintarev, K.: On existence of minimizers for the Hardy–Sobolev–Maz'ya inequality. Ann. Mat. Pura Appl. (4) **183**, 165–172 (2004)
24. Troyanov, M.: Parabolicity of manifolds. Sib. Adv. Math. **9**, 125–150 (1999)
25. Troyanov, M.: Solving the p-Laplacian on manifolds. Proc. Am. Math. Soc. **128**, 541–545 (2000)

Sobolev Inequalities in Familiar and Unfamiliar Settings

Laurent Saloff-Coste

Abstract The classical Sobolev inequalities play a key role in analysis in Euclidean spaces and in the study of solutions of partial differential equations. In fact, they are extremely flexible tools and are useful in many different settings. This paper gives a glimpse of assortments of such applications in a variety of contexts.

1 Introduction

There are few articles that have turned out to be as influential and truly important as S.L. Sobolev 1938 article [93] (the American translation appeared in 1963), where he introduces his famed inequalities. It is the idea of a functional inequality itself that Sobolev brings to life in his paper, as well as the now so familiar notion of an a priori inequality, i.e., a functional inequality established under some strong hypothesis and that might be extended later, perhaps almost automatically, to its natural domain of definition. (These ideas are also related to the theory of distributions which did not exist at the time and whose magnificent development by L. Schwartz was, in part, anticipated in the work of S.L. Sobolev.)

The most basic and important applications of Sobolev inequalities are to the study of partial differential equations. Simply put, Sobolev inequalities provide some of the very basic tools in the study of the existence, regularity, and uniqueness of the solutions of all sorts of partial differential equations, linear and nonlinear, elliptic, parabolic, and hyperbolic. I leave to others, much better qualified than me, to discuss these beautiful developments. Instead, my aim in this paper is to survey briefly an assortments of perhaps less familiar applications of Sobolev inequalities (and related inequalities) to problems and

Laurent Saloff-Coste
Cornell University, Mallot Hall, Ithaca, NY 14853, USA, e-mail: lsc@math.cornell.edu

V. Maz'ya (ed.), *Sobolev Spaces in Mathematics I*,
International Mathematical Series.

doi: 10.1007/978-0-387-85648-3,

in settings that are not always directly related to PDEs, at least not in the most classical sense. The inequalities introduced by S.L. Sobolev have turned out to be extremely useful flexible tools in surprisingly diverse settings. My hope is to be able to give to the reader a glimpse of this diversity. The reader must be warned that the collection of applications of Sobolev inequalities described below is very much influenced by my own interest, knowledge, and limitations. I have not tried at all to present a complete picture of the many different ways Sobolev inequalities have been used in the literature. That would be a very difficult task.

2 Moser's Iteration

2.1 The basic technique

This section is included mostly for those readers that are not familiar with the use of Sobolev inequalities. It illustrates some aspects of one of the basic techniques associated with their use. To the untrained eyes, the fundamental nature of Sobolev inequalities is often lost in the technicalities surrounding their use. Indeed, outside analysis, L^p spaces other than L^1, L^2, and L^∞ still appear quite exotic to many. As the following typical example illustrates, they play a key role in extracting the information contained in Sobolev inequalities.

Recall that Hölder's inequality states that

$$\int |fg|dx \leqslant \|f\|_p\|g\|_q$$

as long as $1 \leqslant p, q \leqslant \infty$ and $1/p + 1/q = 1$ (these are called conjugate exponents). A somewhat clever use of this inequality yields

$$\|f\|_r \leqslant \|f\|_s^\theta \|f\|_t^{1-\theta}$$

as long as $1 \leqslant r, s, t \leqslant \infty$ and $1/r = \theta/s + (1-\theta)/t$. These basic inequalities are used extensively in conjunction with Sobolev inequalities.

Let $\Delta = \sum(\partial/\partial x_i)^2$ be the Laplacian in $\mathbb{R}^n$. Consider a bounded domain $\Omega \subset \mathbb{R}^n$, $\lambda \geqslant 0$, and the (Dirichlet) eigenfunction/eigenvalue problem:

$$\Delta u = -\lambda u \ \text{ in } \ \Omega, \quad u\big|_{\partial\Omega} = 0. \tag{2.1}$$

Our goal is to show how the remarkable inequality

$$\sup_\Omega\{|u|^2\} \leqslant A_n \lambda^{n/2} \int_\Omega |u|^2 dx \tag{2.2}$$

(for solutions of (2.1)) follows from the most classical Sobolev inequality, namely, the inequality (2.5) below. For a normalized eigenfunction u with $\|u\|_2 = 1$ the inequality (2.2) bounds the size of u in terms of the associated eigenvalue. The technique illustrated below is extremely flexible and can be adapted to many situations.

In fact, we only assume that $u \in H_0^1(\Omega)$, i.e., u is the limit of smooth compactly supported functions in Ω in the norm

$$\|u\| = \left(\int_\Omega [|u|^2 + \sum_1^n |\partial u/\partial x_i|^2] dx \right)^{1/2}$$

and that

$$\int_\Omega \sum_1^n \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_i} dx = \lambda \int_\Omega uv dx \tag{2.3}$$

for any $v \in H_0^1(\Omega)$. We set $\nabla u = (\partial u/\partial x_i)_1^n$, $|\nabla u|^2 = \sum\limits_1^n |\partial u/\partial x_i|^2$. To avoid additional technical arguments, we assume a priori that u is bounded on Ω. For $1 \leqslant p < \infty$ we take $v = |u|^{2p-1}(u/|u|)$ in (2.3). This yields

$$\lambda \int_\Omega |u|^{2p} = (2p-1) \int_\Omega |u|^{2p-2} |\nabla u|^2 dx = \frac{2p-1}{p^2} \int_\Omega |\nabla |u|^p|^2 dx. \tag{2.4}$$

As our starting point, we take the most basic Sobolev inequality

$$\forall\, f \in H^1(\mathbb{R}^n), \quad \left(\int |f|^{2q_n} dx \right)^{1/q_n} \leqslant C_n^2 \int |\nabla f|^2 dx, \quad q_n = n/(2-n). \tag{2.5}$$

If $k_n = (1 + 2/n)$, then $1/k_n = \theta_n/q_n + (1 - \theta_n)$ with $\theta_n = n/(n+2)$, and Hölder's inequality yields

$$\int |f|^{2k_n} dx \leqslant \left(\int |f|^{2q_n} dx \right)^{1/q_n} \left(\int |f|^2 dx \right)^{2/n}.$$

Together with the previous Sobolev inequality, we obtain

$$\int |f|^{2(1+2/n)} dx \leqslant C_n^2 \int |\nabla f|^2 dx \left(\int |f|^2 dx \right)^{2/n}. \tag{2.6}$$

For a discussion of this type of "multiplicative" inequality see, for example, [75, Sect. 2.3].

Now, for a solution u of (2.1) the inequalities (2.6) and (2.4) yield

$$\int |u|^{2p(1+2/n)} dx \leqslant C_n^2 p \lambda \left(\int |u|^{2p} dx \right)^{1+2/n}.$$

This inequality can obviously be iterated by taking $p_i = (1+2/n)^i$, and we get

$$\left(\int |u|^{2p_i}dx\right)^{1/p_i} \leqslant (1+2/n)^{\sum_1^i (j-1)p_j^{-1}} \left(3C_n^2\lambda\right)^{\sum_1^i p_j^{-1}} \int |u|^2 dx.$$

Note that $\sum_1^\infty p_j^{-1} = n/2$ and $\lim_{p\to\infty} \||u|^2\|_p = \|u^2\|_\infty$. The desired conclusion (2.2) follows.

2.2 Harnack inequalities

The technique illustrated above is the simplest instance of what is widely known as Moser's iteration technique. In a series of papers [77]–[80], Moser developed this technique as the basis for the study of divergence form uniformly elliptic operators in $\mathbb{R}^n$, i.e., operators of the form (we use $\partial_i = \partial/\partial x_i$)

$$L_a = \sum_{i,j} \partial_i(a_{i,j}(x)\partial_j)$$

with real matrix-valued function a satisfying the ellipticity condition

$$\forall\, x \in \Omega, \quad \begin{cases} \sum_{i,j} a_{i,j}(x)\xi_i\xi_j \geqslant \varepsilon|\xi|^2, \\ \sum_{i,j} a_{i,j}(x)\xi_i\xi'_j \leqslant \varepsilon^{-1}|\xi||\xi'|, \end{cases}$$

where $\varepsilon > 0$ and the coefficients $a_{i,j}$ are simply bounded real measurable functions. Because of the low regularity of the coefficients, the most basic question in this context is that of the boundedness and continuity of solutions of the equation $L_a u = 0$ in the interior of an open set Ω. This was solved earlier by De Giorgi [34] (and by Nash [81] in the parabolic case), but Moser proposed an alternative method, squarely based on the use of Sobolev inequality (2.5). To understand why one might hope this is possible, observe that the argument given in the previous section works without essential changes if, in (2.1), one replaces the Laplacian Δ by L_a.

Let u be a solution of $L_a u = 0$ in a domain Ω, in the sense that for any open relatively compact set Ω_0 in Ω, $u \in H^1(\Omega_0)$ and for all $v \in H_0^1(\Omega_0)$

$$\int_\Omega \sum_{i,j} a_{i,j} \frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_j} dx = 0.$$

In [77], Moser observed that the interior boundedness and continuity of such a solution follow from the Harnack inequality that provides a constant $C(n,\varepsilon)$

such that if u as above is nonnegative in Ω and the ball B satisfies $2B \subset \Omega$, then

$$\sup_B\{u\} \leqslant C(n,\varepsilon)\inf_B\{u\} \tag{2.7}$$

(a priori, the supremum and infimum should be understood here as essential supremum and essential infimum. The ball $2B$ is concentric with B with twice the radius of B). He then proceeded to prove this Harnack inequality by variations on the argument outlined in the previous section. In his later papers [78]–[80], Moser obtained a parabolic version of the above Harnack inequality. Namely, he proved that there exists a constant $C(n,\varepsilon)$ such that any nonnegative solution u of the heat equation $(\partial_t - L_a)u = 0$ in a time-space cylinder $Q = (s-4r^2, s)\times 2B$ satisfies

$$\sup_{Q_-}\{u\} \leqslant C(n,\varepsilon)\inf_{Q_+}\{u\}, \tag{2.8}$$

where $Q_- = (s-3r^2, s-2r^2)\times B$ and $Q_+ = (s-r^2/2, s)\times B$.

Moser's iteration technique has been adapted and used in hundreds of papers studying various PDE problems. Some early examples are [2, 3, 90]. The books [42, 69, 76] contain many applications of this circle of ideas, as well as further references. The survey paper [83] deals specifically with the heat equation and is most relevant for the purpose of the present paper.

The basic question we want to explore in the next two subsections is: what exactly are the crucial ingredients of Moser's iteration? This question is motivated by our desire to use this approach in other settings such as Riemannian manifolds or more exotic spaces. Early uses of Moser's iteration technique on manifolds as in the influential papers [22, 23] were actually limited by a misunderstanding of what is really needed to run this technique successfully. Interesting early works that explored the flexibility of Moser's iteration beyond the classical setting are related to degenerated elliptic operators as in [56]–[58] (see also [39] and the references therein).

2.3 Poincaré, Sobolev, and the doubling property

Moser's technique in $\mathbb{R}^n$ uses only three crucial ingredients:

(1) The Sobolev inequality in the form (2.6), i.e.,

$$\forall\, f\in \mathcal{C}_0^\infty(\mathbb{R}^n),\quad \int |f|^{2(1+2/n)}dx \leqslant C_n^2 \int |\nabla f|^2 dx \left(\int |f|^2 dx\right)^{2/n}.$$

(2) The Poincaré inequality in the unit ball B, i.e.,

$$f \in \mathcal{C}^\infty(B), \quad \int_B |f - f_B|^2 dx \leqslant P_n \int_B |\nabla f|^2 dx,$$

where f_B stands for the average of f over B.

(3) Translations and dilations.

Some might be surprised that the interesting John–Nirenberg inequality that appears to be a crucial tool in [77, 78] is not mentioned above. However, as Moser himself pointed out in [80], it can be avoided altogether by using a clever, but very elementary observation of Bombieri and Giusti. Somewhat unfortunately, this important simplification has been ignored by a large part of the later literature!

Obviously, in order to use the method in a larger context, one wants to replace the use of translations and dilations by hypotheses that are valid at all scales and locations. For instance, the needed Poincaré inequality takes the form

$$\forall\, z,\ \forall\, r > 0,\ f \in \mathcal{C}^\infty(B(z,r)), \int_{B(z,r)} |f - f_{B(z,r)}|^2 dx \leqslant P_n r^2 \int_{B(z,r)} |\nabla f|^2 dx,$$

where f_B stands for the average of f over B. A correct generalization is less obvious in the case of the Sobolev inequality. As stated, the inequality (2.6) turns out to be too restrictive and not strong enough, both at the same time!

For instance, consider a complete Riemannian manifold (M, g) of dimension n. We set $|\nabla f|^2 = g(\nabla f, \nabla f)$, where the gradient ∇f is the vector field defined by $g_x(\nabla f, X) = df(X)$ for any tangent vector X at x. Let μ be the Riemannian measure, $B(x,r)$ the geodesic ball with center x and radius r, and

$$V(x,r) = \mu(B(x,r)).$$

If we assume that the inequality analogous to (2.6) holds on M, i.e.,

$$\forall\, f \in \mathcal{C}_0^\infty(M), \quad \int |f|^{2(1+2/n)} d\mu \leqslant C_M^2 \int |\nabla f|^2 d\mu \left(\int |f|^2 d\mu \right)^{2/n}, \qquad (2.9)$$

then it turns out that this implies the existence of a constant $c_M > 0$ such that

$$\forall\, x \in M, \ \ \forall\, r > 0, \quad \mu(B(x,r)) = V(x,r) \geqslant c_M r^n$$

(see [17, Proposition 2.4] and [87, Theorem 3.15]). This rules out simple manifolds such as $\mathbb{R}^{n+k}/\mathbb{Z}^k$ or $\mathbb{R}^{n-k} \times \mathbb{S}^k$ (on which, for other reasons, one knows that the above-mentioned analogs of the Harnack inequalities (2.7), (2.8) hold). Let us observe that when $n \geqslant 3$, (2.9) is, in fact, equivalent to the more standard Sobolev inequality

$$\forall\, f \in \mathcal{C}_0^\infty(M), \quad \left(\int |f|^{2q_n} d\mu\right)^{1/q_n} \leqslant C_M^2 \int |\nabla f|^2 d\mu, \quad q_n = n/(n-2), \tag{2.10}$$

where the constant C_M may be different in (2.9) and in (2.10).

In the other direction, (2.10) and thus (2.9) holds in the case of hyperbolic spaces (with dimension $n > 2$ for (2.10)), but the desired Harnack inequalities fail to hold uniformly at large scale in such spaces.

Definition 2.1. We say that a complete Riemannian manifold M satisfies a *scale invariant family* of Sobolev inequalities if there is a constant C_M and a real number $q = \nu/(\nu - 2) > 1$ such that for any $x \in M$, $r > 0$, and $B = B(x, r)$ we have

$$\forall\, f \in \mathcal{C}_0^\infty(B), \quad \left(\int_B |f|^{2q} d\mu\right)^{1/q} \leqslant \frac{C_M r^2}{\mu(B)^{2/\nu}} \int_B \left[|\nabla f|^2 + r^{-2}|f|^2\right] d\mu. \tag{2.11}$$

Remark 2.1. The inequality (2.11) can be written in the form: for all $f \in \mathcal{C}_0^\infty(B)$

$$\left(\frac{1}{\mu(B)} \int_B |f|^{2q} d\mu\right)^{1/q} \leqslant C_M r^2 \left(\frac{1}{\mu(B)} \int_B \left[|\nabla f|^2 + r^{-2}|f|^2\right] d\mu\right).$$

Remark 2.2. In this definition, the exact value of q is not very important and ν appears here as a technical parameter. If (2.11) holds for some $q = \nu/(\nu - 2) > 1$, then the Jensen inequality shows that it also holds for all $1 < q' = \nu'/(\nu' - 2) \leqslant q$, i.e., for all finite $\nu' \geqslant \nu$.

Remark 2.3. In general, (2.10) does not imply (2.11). However, (2.10) does imply (2.11) with $\nu = n$ when the manifold M has an Euclidean type volume growth, i.e., there exists $0 < v_M \leqslant V_M < \infty$ such that $v_M r^n \leqslant V(x, r) \leqslant V_M r^n$ for all $x \in M$ and $r > 0$. This is obviously a very restrictive and undesirable hypothesis. This is exactly the point that restricted the use of Moser's iteration technique to very local results in some early applications of the technique to analysis on Riemannian manifolds as in [22, 23].

Remark 2.4. There are many equivalent forms of (2.11). We mention three. The first one is analogous to (2.6) and reads

$$\int_B |f|^{2(1+2/\nu)} d\mu \leqslant \frac{C_M r^2}{\mu(B)^{2/\nu}} \int_B \left[|\nabla f|^2 + r^{-2}|f|^2\right] d\mu \left(\int_B |f|^2 d\mu\right)^{2/\nu}.$$

The second is in the form of the so-called *Nash inequality* and reads

$$\int_B |f|^{2(1+2/\nu)} d\mu \leqslant \frac{C_M r^2}{\mu(B)^{2/\nu}} \int_B \left[|\nabla f|^2 + r^{-2}|f|^2\right] d\mu \left(\int_B |f| d\mu\right)^{4/\nu}$$

(see [81] and [75, Sect. 2.3]). The third is often referred to as a *Faber–Krahn inequality* (see [44]) and reads

$$\lambda_D(\Omega) \geqslant \frac{c_M}{r^2} \left(\frac{\mu(\Omega)}{\mu(B)}\right)^{2/\nu},$$

where $\lambda_D(\Omega)$ is the lowest Dirichlet eigenvalue in Ω, an arbitrary subset of the ball B of radius r. In each case, r is the radius of B and the inequality must hold uniformly for all geodesic balls B. The exact value of the constants varies from one type of inequality to another. Many results in the spirit of these equivalences can be found in [75] in the context of Euclidean domains. A discussion in a very general setting is in [4] (see also [87, Chapt. 3]).

The following theorem describes some of the noteworthy consequences of (2.11). Let Δ_M be the Laplace operator on M, and let $h(t,x,y)$ be the (minimal) fundamental solution of the heat equation $(\partial_t - \Delta_M)u = 0$ on M, i.e., the kernel of the heat semigroup $e^{t\Delta_M}$. For complete discussions, surveys, and variants, see [43, 44, 45, 46, 48, 85, 86, 87].

Theorem 2.1. *Assume that (M,g) is a complete Riemannian manifolds which satisfies the scale invariant family of Sobolev inequalities* (2.11) (*with some parameter $\nu > 2$*). *Then the following properties hold.*

- *There exists a constant V_M such that for any two concentric balls $B \subset B'$ with radii $0 < r < r' < \infty$*

$$\mu(B') \leqslant V_M (r'/r)^\nu \mu(B). \tag{2.12}$$

- *There exists a constant C_M such that for all $x \in M$ and $r > 0$ any positive subsolution u of the heat equation in a time-space cylinder $Q = (s - 4r^2, s) \times B(x, 2r)$ satisfies*

$$\sup_{Q'}\{u^2\} \leqslant C_M \frac{1}{r^2 \mu(B)} \int_Q |u|^2 d\mu ds, \tag{2.13}$$

  where $Q' = (s - r^2, s) \times B(x, r)$.

- *For any integer $k \geqslant 0$ there is a constant $A(M,k)$ such that for all points $x, y \in M$ and $t > 0$*

$$|\partial_t^k h(t,x,y)| \leqslant \frac{A(M,k)}{t^k V(x,\sqrt{t})} \left(1 + d(x,y)^2/t\right)^{\nu+k} \exp\left(-\frac{d(x,y)^2}{4t}\right). \tag{2.14}$$

Remark 2.5. The inequality (2.13) can be obtained by a straightforward application of Moser's iteration technique. One of many possible applications of (2.13) is (2.14).

Remark 2.6. The volume inequality (2.12), together with the heat kernel bound

$$\forall\, x \in M,\, t > 0, \quad h(t,x,x) < \frac{A_M}{V(x,\sqrt{t})},$$

implies the Sobolev inequality (2.11).

Definition 2.2. A complete Riemannian manifold has the *doubling volume property* if there exists a constant V_M such that

$$\forall\, x \in M,\ r > 0,\ \ V(x,2r) \leqslant V_M V(x,r).$$

Remark 2.7. It is easy to see that the doubling property implies (2.12) with $\nu = \log_2 V_M$.

Definition 2.3. A complete Riemannian manifold admits a *scale invariant Poincaré inequality* (in L^2) if there exists a constant P_M such that

$$\forall\, x \in M,\ r > 0, \quad \int\limits_B |f - f_B|^2 d\mu \leqslant P_M r^2 \int\limits_B |\nabla f|^2 d\mu, \tag{2.15}$$

where $B = B(x,r)$ and f_B is the average of f over B.

Remark 2.8. This Poincaré inequality can be stated in terms of the spectrum of minus the Neumann Laplacian in geodesic balls. For minus the Neumann Laplacian (understood in an appropriate sense) in a ball B, the lowest eigenvalue is 0 (associated with constant functions). The L^2 Poincaré inequality above is equivalent to say that the second eigenvalue $\lambda_N(B(x,r))$ is bounded from below by cr^{-2}, where r is the radius of B.

Remark 2.9. Keeping Moser's iteration in mind, it is a very important and remarkable fact that if M satisfies both the doubling property and a scale invariant Poincaré inequality, then it satisfies (2.11) (see [85]–[87]). In this case, one can take ν to be an arbitrary number greater than 2 and such that (2.12) holds.

Definition 2.4. A complete Riemannian manifold admits a *scale invariant parabolic Harnack inequality* if there exists a constant C_M such that for any $x \in M$, $r > 0$, and $s \in \mathbb{R}$ and for any nonnegative solution u of the heat equation in the time-space cylinder $Q = (s - 4r^2, s) \times B(x,2r)$

$$\sup_{Q_-}\{u\} \leqslant C_M \inf_{Q_+}\{u\}$$

with $Q_- = (s - 3r^2, s - 2r^2) \times B(x,r)$ and $Q_+ = (s - r^2, s) \times B(x,r)$.

In this setting, a version of Moser's iteration methods gives one half of the following result (see [43, 85] and a detailed discussion in [87, Sect. 5.5]).

Theorem 2.2. *Let (M, g) be a complete Riemannian manifold. The following properties are equivalent.*

- *The doubling property and a scale invariant L^2 Poincaré inequality.*
- *The scale invariant parabolic Harnack inequality.*
- *The two-sided heat kernel bound*

$$\frac{c}{V(x,\sqrt{t})}\exp\left(-A\frac{d(x,y)^2}{t}\right) \leqslant h(t,x,y) \leqslant \frac{C}{V(x,\sqrt{t})}\exp\left(-a\frac{d(x,y)^2}{t}\right)$$

  for constants $0 < a, A, c, C < \infty$.

One may asked how the above properties are related to the elliptic version of Harnack inequality. This is not entirely understood, but the following result involving the Sobolev inequality (2.11) sheds some light on this question (see [61]).

Theorem 2.3. *Let M be a complete manifold satisfying the Sobolev inequality* (2.11) *for some $q > 1$. Then the following properties are equivalent.*

- *The scale invariant L^2 Poincaré inequality.*
- *The scale invariant elliptic Harnack inequality.*
- *The scale invariant parabolic Harnack inequality.*

We conclude with results concerning global harmonic functions.

Theorem 2.4. *Let M be a manifold satisfying the doubling volume property and a scale invariant L^2 Poincaré inequality.*

- *Any harmonic functions on M that is bounded from below must be constant.*
- *There exists $a_0 > 0$ such that for any fixed point $x \in M$ any harmonic function satisfying $\sup_y\{u(y)/(1+d(x,y))^{a_0}\} < \infty$ must be constant.*
- *For any $a > 0$ and a fixed point $x \in M$ the space of harmonic functions on M satisfying $\sup_y\{u(y)/(1+d(x,y))^{a}\} < \infty$ is finite dimensional.*

Remark 2.10. The first two statements are standard consequences of the (scale invariant) elliptic Harnack inequality which follows from the assumptions of the theorem. The last statement is a recent result due to Colding and Minicozzi [24, 25, 72, 71]. The proof of the last statement makes explicit the use of the Poincaré inequality and the doubling volume property. A number of interesting variations on this result are discussed in [24, 25, 72, 71]. A different viewpoint concerning Liouville theorems, restricted to some special circumstances, but very interesting nonetheless is developed in [68].

Example 2.1. Euclidean spaces are the model examples for manifolds that satisfy both the doubling condition and the Poincaré inequality. Larger classes of examples will be described in the next section. Interesting examples where the Poincaré inequality fails are obtained by considering manifolds M that are the connected sum of two (or more) Euclidean spaces. Here, we write $M = \mathbb{R}^n \# \mathbb{R}^n$ to mean a complete Riemannian manifold that can be decomposed in the disjoint union $E_1 \cup K \cup E_2$, where E_1, E_2 are each isometric to the outside of some compact domain with smooth boundary in $\mathbb{R}^n$ and K is a smooth compact manifold with boundary. In words, $\mathbb{R}^n \# \mathbb{R}^n$ is made of two copies of $\mathbb{R}^n$ smoothly attached together through a compact "collar." The following facts (that are not too difficult to check) make these examples interesting.

- $M = \mathbb{R}^n \# \mathbb{R}^n$ has the doubling property. In fact, obviously, $V(x,r) \simeq r^n$.
- $M = \mathbb{R}^n \# \mathbb{R}^n$ satisfies (2.11) with ν being any positive real that is both at least n and greater than 2. In fact, (2.9) holds on $\mathbb{R}^n \# \mathbb{R}^n$ for any n, and (2.10) holds if $n > 2$. This means that Theorem 2.1 applies.
- Except for the trivial case $n = 1$, the scale invariant Poincaré inequality (2.15) does not hold on $\mathbb{R}^n \# \mathbb{R}^n$. More precisely, if o is a fixed point in the collar of $\mathbb{R}^n \# \mathbb{R}^n$ and $B_r = B(o, r)$, then for large $r \gg 1$, we have

$$\lambda_N(B_r) \simeq \begin{cases} (r^2 \log r)^{-1} & \text{if } n = 2, \\ r^{-n} & \text{if } n > 2, \end{cases}$$

  where $\lambda_N(B_r)$ is the second lowest eigenvalue of the Neumann Laplacian in B_r. This means that the best Poincaré inequality in B_r has a constant that is in $r^2 \log r$ if $n = 2$ and r^n if $n \geqslant 3$ (instead of the desired r^2).
- For $n > 1$, $M = \mathbb{R}^n \# \mathbb{R}^n$ does not satisfy the elliptic Harnack inequality (again, it fails for nonnegative harmonic functions in the balls B_r as above, when r tends to infinity).
- For $n \leqslant 2$ there are no nonconstant positive harmonic functions, but for $n \geqslant 3$ there are nonconstant bounded harmonic function on $M = \mathbb{R}^n \# \mathbb{R}^n$.
- For $n \geqslant 2$ let o be a point in the collar of $M = \mathbb{R}^n \# \mathbb{R}^n$, and let x and y be, respectively, in the first and second copies of $\mathbb{R}^n$ constituting M, at distance about $r = \sqrt{t}$ from o. Then the heat kernel $h(t, x, y)$ satisfies $h(t, x, y) \simeq t^{-n+1}$. This should be compare with the Euclidean heat kernel at time t for points x, y about $\sqrt{t}$ apart which is of size about $t^{-n/2}$. For more on this we refer the reader to [49]–[51].

2.4 Examples

We briefly discuss various examples that illustrate the above-described results.

Example 2.2 (manifolds with nonnegative Ricci curvature). The Ricci curvature Ric is a symmetric $(0, 2)$-tensor (obtained by contraction of the full curvature tensor) that contains a lot of useful information. Two well-known early examples of that are:

(1) Meyers' theorem (more on this later) stating that a complete Riemannian manifold with $\mathrm{Ric} \geqslant Kg$ with $K > 0$ must be compact and

(2) Bishop's volume inequality asserting that if $\mathrm{Ric} \geqslant Kg$ for some $k \in \mathbb{R}$, then the volume function on M, $V(x, r)$, is bounded from above by the volume function $V_{K/(n-1)}(r)$ of the simply connected space of the same dimension and constant sectional curvature $K/(n-1)$ (see, for example, [20, p.73], [21, Theorem 3.9] and [41, 3.85; 3.101]).

Theorem 2.5 ([14, 22, 74]). *A complete Riemannian manifold (M, g) with nonnegative Ricci curvature satisfies the equivalent properties of Theorem* 2.2.

It is interesting to note that the equivalent properties of Theorem 2.2 where proved independently for manifolds with nonnegative Ricci curvature. The doubling property follows from the more precise Bishop–Gromov volume inequality of [22]. Namely, if $\mathrm{Ric} \geqslant k(n-1)g$, then

$$\forall x \in M,\ s > r > 0, \quad \frac{V(x,s)}{V(x,r)} \leqslant \frac{V_k(s)}{V_k(r)}. \tag{2.16}$$

If $k = 0$, this gives $V(x, s) \leqslant (s/r)^n V(x, r)$ for all $x \in M$, $s > r > 0$. The Poincaré inequality follows from the result in [14] (see also [21, Theorems 3.10 and 6.8] and [87, Theorem 5.6.5]). The Harnack inequality and two-sided heat kernel estimate follow from the gradient estimate of Li and Yau [74]. Of course, these results imply that the various conclusions of Theorem 2.4 hold for Riemannian manifolds with nonnegative Ricci curvature. In this setting, the last statement in Theorem 2.4 (due to Colding and Minicozzi) solves a conjecture of Yau (see [24, 25, 72, 71]).

Example 2.3. Let G be a connected real Lie group equipped with a left-invariant Riemannian metric g. Note that the Riemannian measure is also a left-invariant Haar measure. We say that G has polynomial volume growth if there exist $C, a \in (0, \infty)$ such that $V(e, r) \leqslant Cr^a$ for all $r \geqslant 1$. A group G with polynomial volume growth must be unimodular (left-invariant Haar measures are also right-invariant) and, by a theorem of Guivarc'h [55], there exists an integer N such that $c_0 \leqslant r^{-N} V(e, r) \leqslant C_0$ for all $r \geqslant 1$. It follows that (G, g) satisfies the volume doubling property. By a simple direct argument (see, for example, [87, Theorem 5.6.1]), the scale invariant Poincaré inequality also holds. Hence one can apply Theorem 2.2. In fact, in this setting, one has the following result.

Theorem 2.6. *Let G be a connected real unimodular Lie group equipped with a Riemannian metric g. The following properties are equivalent.*

- *The group G has polynomial volume growth.*
- *Any positive harmonic function on G is constant.*
- *The scale invariant elliptic Harnack inequality holds.*
- *The scale invariant Sobolev inequality* (2.11) *holds for some $q > 1$.*
- *The scale invariant parabolic Harnack inequality holds.*

Proof. Connected Lie groups have either strict polynomial growth $V(e,r) \simeq r^N$ for all $r \geqslant 1$ for some integer N or exponential volume growth (see [55]). Thus, if the volume growth is polynomial, it must be strictly polynomial and the doubling volume property follows. As already mentioned, it is also very easy to prove the scale invariant Poincaré inequality on a connected Lie group of polynomial growth (see, for example, [87, Theorem 5.6.2]). By Theorem 2.2, this shows that polynomial volume growth implies the parabolic Harnack inequality in this context. The parabolic Harnack inequality implies all the other mentioned properties (see Theorem 2.2 and the various remarks in the previous section). The Sobolev inequality (2.11) implies the doubling volume property, hence polynomial volume growth in this context. The elliptic Harnack property implies the triviality of positive harmonic functions. This, in turns, implies polynomial volume growth by [13, Theorem 1.4 or 1.6]. The stated theorem follows. □

For more general results in this setting see [103]. Harmonic functions of polynomial growth on Lie groups of polynomial growth are studied in [1].

Example 2.4 (coverings of compact manifolds). Let (M,g) be a complete Riemannian manifold such that there exists a discrete group of isometries Γ acting freely and properly on (M,g) with compact quotient N. The discrete group Γ must be finitely generated. Its volume growth is defined by using the word metric and counting measure.

Theorem 2.7. *Let (M,g) be a complete Riemannian manifold such that there exists a discrete group of isometries Γ acting freely and properly on (M,g) with compact quotient M/Γ. The following properties are equivalent.*

- *The group Γ has polynomial volume growth.*
- *The scale invariant elliptic Harnack inequality holds.*
- *The scale invariant Sobolev inequality* (2.11) *holds for some $q > 1$.*
- *The scale invariant parabolic Harnack inequality holds.*

For a complete discussion see [88, Theorem 5.15]. Note the similarity and differences between this result and Theorem 2.6. The main difference is that, for coverings of a compact manifold, there is no known criterion based on the triviality of positive harmonic functions. This is due to the fact that the group Γ may not be linear (or close to a linear group) (see [13, 88]).

3 Analysis and Geometry on Dirichlet Spaces

3.1 First order calculus

One of the recent developments in the theory of Sobolev spaces concerns the definitions and properties of such spaces under minimal hypotheses. The most general setting is that of metric measure spaces. There are very good reasons to try to understand what can be done in that setting including important applications to problems coming from different areas of mathematics and even to questions concerning classical Sobolev spaces. In what follows, I only discuss a very special class of metric measure spaces, but it is useful to keep in mind the more general setting. Indeed, the theory of Sobolev spaces on metric measure spaces is also of interest because of the many similar, but different setting it unifies. We refer the reader to the entertaining books [59, 62, 89] and the review paper [63] for glimpses of the general viewpoint on "first order calculus."

There are many interesting natural metric spaces (of finite dimension type) on which one wants to do some analysis and that are not Riemannian manifolds. Some appear as limit of Riemannian manifolds, for example, manifolds equipped with sub-Riemannian structures and more exotic objects appearing through various geometric precompactness results. Others are very familiar (polytopal complexes seem to appear in real life as often, if not more often, than true manifolds), but have not been studied in much detail as far as analysis is concerned. One natural structure that captures a good number of such examples and provides many natural analytic objects to study (beyond first order calculus) is the structure of Dirichlet spaces. The earliest detailed reference on Dirichlet spaces is [36]. We refer the reader to [40] for a detailed introduction to Dirichlet spaces.

3.2 Dirichlet spaces

This subsection describes a restricted class of Dirichlet spaces that provides nice metric measure spaces. There are several interesting possible variations on this theme, and we only discuss here the strongest possible version.

We start with a locally compact separable metric space M equipped with a Radon measure μ such that any open relatively compact nonempty set has positive measure. The original metric will not play any important role.

In addition, we are given a symmetric bilinear form $\mathcal{E}$ defined on a dense subset $\mathcal{D}(\mathcal{E})$ of $L^2(M, d\mu)$ such that $(u, u) \geqslant 0$ for any $u \in \mathcal{D}(\mathcal{E})$. We assume that $\mathcal{E}$ is closed, i.e., $\mathcal{D}(\mathcal{E})$ equipped with the norm

$$\mathcal{E}_1(u,u)^{1.2} = \sqrt{\|u\|_2^2 + \mathcal{E}(u,u)}$$

is complete (i.e., is a Hilbert space). In addition, we assume that the unit contraction

$$u \mapsto v_u = \inf\{1, \sup\{0, u\}\}$$

operates on $(\mathcal{E}, \mathcal{D}(\mathcal{E}))$ in the sense that

$$u \in \mathcal{D}(\mathcal{E}) \Longrightarrow v_u \in \mathcal{D}(\mathcal{E}) \text{ and } \mathcal{E}(v_u, v_u) \leqslant \mathcal{E}(u,u).$$

Such a form is called a *Dirichlet form* and is associated with a self-adjoint strongly continuous semigroup of contractions H_t, $t > 0$, on $L^2(M, d\mu)$ with the additional property that $0 \leqslant u \leqslant 1$ implies $0 \leqslant H_t u \leqslant 1$. Namely, if A is the infinitesimal generator so that $H_t = e^{tA}$ (in the sense of spectral theory, say), then $\mathcal{D}(\mathcal{E}) = \mathrm{Dom}((-A)^{1/2})$ and $\mathcal{E}(u, v) = \langle (-A)^{1/2}u, (-A)^{1/2}v\rangle$.

We assume that the form $\mathcal{E}$ is strongly local, i.e., $\mathcal{E}(u, v) = 0$ if $u, v \in \mathcal{D}(\mathcal{E})$ have compact support and v is constant on a neighborhood of the support of u. Finally, we assume that $(M, \mathcal{E}, \mathcal{D}(\mathcal{E}))$ is regular. This means that the space $\mathcal{C}_c(M)$ of continuous compactly supported functions on M has the property that $\mathcal{D}(\mathcal{E}) \cap \mathcal{C}_c(M)$ is dense in $\mathcal{C}_c(M)$ in the sup norm $\|u\|_\infty = \sup\limits_M\{|u|\}$ and is dense in $\mathcal{D}(\mathcal{E})$ in the norm $\mathcal{E}_1^{1/2}$. Note that this is a hypothesis that concerns the interaction between $\mathcal{E}$ and the topology of M. We call $(M, \mu, \mathcal{E}, \mathcal{D}(\mathcal{E}))$ a *strictly local regular Dirichlet space.*

Under these hypotheses, there exists a bilinear form $\varGamma$ defined on $\mathcal{D}(\mathcal{E}) \times \mathcal{D}(\mathcal{E})$ with the values in signed Radon measures on M such that

$$\mathcal{E}(u,v) = \int\limits_M d\varGamma(u,v).$$

For $u \in \mathcal{D}(\mathcal{E}) \cap L^\infty(M)$, $\varGamma(u, u)$ is defined by

$$\forall\, \varphi \in \mathcal{D}(\mathcal{E}) \cap \mathcal{C}_c(M), \quad \int\limits_M \varphi d\varGamma(u,u) = \mathcal{E}(u, \varphi u) - (1/2)\mathcal{E}(u^2, \varphi).$$

Although the measure $\varGamma(u, v)$ might be singular with respect to μ, it behaves much like $g(\nabla u, \nabla v)dx$ on a Riemannian manifold. For instance, versions of the chain rule and Leibniz rule apply. In what follows, we work under additional assumptions that imply that the set of those u in $\mathcal{D}(\mathcal{E})$ such that $d\Gamma/d\mu$ exists is rich enough (see [10, 97] for further details).

We now introduce a key ingredient to our discussion: the intrinsic distance.

Definition 3.1. Let $(M, \mu, \mathcal{E}, \mathcal{D}(\mathcal{E}))$ be a strictly local regular Dirichlet space. For x, y in M we set

$$\rho(x,y) = \sup\{u(x) - u(y) : u \in \mathcal{D}(\mathcal{E}) \cap \mathcal{C}_c(M),\ d\Gamma(u,u) \leqslant d\mu\}.$$

Here, the condition $d\Gamma(u,u) \leqslant d\mu$ means that the measure $\Gamma(u,u)$ is absolutely continuous with respect to μ with Radon–Nykodim derivative bounded by 1 almost everywhere. It is obvious that ρ is symmetric in x, y and satisfies the triangle inequality. It might well be either 0 or ∞ for some x, y. If ρ is finite and $\rho(x,y) = 0$ only if $x = y$, then ρ is a distance function.

Qualitative hypotheses.

Throughout the paper, we assume that

(A1) The function $\rho : M \times M \to [0, \infty]$ is finite, continuous, satisfies

$$\rho(x,y) = 0 \Rightarrow x = y,$$

and defines the topology of M.

(A2) The metric space (M, ρ) is a complete metric space.

With these hypotheses, one can show that the metric space (M, ρ) is a length space (i.e., $\rho(x,y)$ can be computed as the minimal length of continuous curves joining x to y, where the length of a curve is defined using ρ in a natural manner). Denote by $B(x,r)$ the open balls in (M, ρ). Each $B(x,r)$ is precompact with compact closure given by the associated closed ball. Set $V(x,r) = \mu(B(x,r))$. For each fixed $x \in M$, $r > 0$ the function $\delta(y) = \max\{0, r - \rho(x,y)\}$ is in $\mathcal{D}(\mathcal{E}) \cap \mathcal{C}_c(M)$ and satisfies $d\Gamma(\delta, \delta) \leqslant d\mu$ (see [10, 11, 12, 94, 95, 96, 97] for details).

3.3 Local weak solutions of the Laplace and heat equations

Recall that A is the infinitesimal generator of the semigroup of operators associated to our Dirichlet form. Identify $L^2(M, \mu)$ with its dual using the scalar product.

Let V be a nonempty open subset of M. Consider the subspace $\mathcal{F}_c(V) \subset \mathcal{D}(\mathcal{E})$ of those functions with compact support in V. Note that $\mathcal{F}_c(V) \subset$

$\mathcal{D}(\mathcal{E}) \subset L^2(M,\mu)$ and consider their duals $L^2(X,\mu) \subset \mathcal{D}(\mathcal{E})' \subset \mathcal{F}_c(V)'$. We use the brackets $\langle \cdot, \cdot \rangle$ to denote duality pairing between these spaces. Let $\mathcal{F}_{\text{loc}}(V)$ be the space of functions $u \in L^2_{\text{loc}}(V)$ such that for any compact set $K \subset V$ there exists a function $u_K \in \mathcal{D}(\mathcal{E})$ that coincides with u almost everywhere on K.

Definition 3.2. Let V be a nonempty open subset of X. Let $f \in \mathcal{F}_c(V)'$. A function $u : V \mapsto \mathbb{R}$ is a *weak* (local) solution of $Au = f$ in V if

1. $u \in \mathcal{F}_{\text{loc}}(V)$;
2. for any function $\varphi \in \mathcal{F}_c(V)$ we have $\mathcal{E}(\varphi, u) = \langle \varphi, f \rangle$.

Remark 3.1. If f can be represented by a locally integrable function in V and u is such that there exists a function $u^* \in \text{Dom}(A)$ (the domain of the infinitesimal generator A) satisfying $u = u^*|_V$, then u is a weak local solution of $Au = f$ if and only if $Au^*|_V = f$ a.e in V.

Remark 3.2. The notion of weak local solution defined above may contain implicitly a Neumann type boundary condition if M has a natural boundary. Consider, for example, the case where M is the closed upper-half plane $P_+ = \overline{\mathbb{R}^2_+}$ equipped with its natural Dirichlet form

$$\mathcal{E}(f,f) = \iint_{\mathbb{R}^2_+} \left(\left| \frac{\partial f}{\partial x} \right|^2 + \left| \frac{\partial f}{\partial y} \right|^2 \right) dxdy, \quad f \in W^1(\mathbb{R}^2_+).$$

Let $V = \{z = (x,y) : x^2 + y^2 < 1; y \geqslant 0\} \subset P_+$. Note that V is open in P_+. Let u be a local weak solution of $\Delta u = 0$ in V. Then it is easy to see that u is smooth in V and must have vanishing normal derivative along the segment $(-1,1)$ of the real axis.

Next, we discuss local weak solutions of the heat equation $\partial_t u = Au$ in a time-space cylinder $I \times V$, where I is a time interval and V is a nonempty open subset of X. Given a Hilbert space H, let $L^2(I \to H)$ be the Hilbert space of functions $v : I \mapsto H$ such that

$$\|v\|_{L^2(I \to H)} = \left(\int_I \|v(t)\|_H^2 dt \right)^{1/2} < \infty.$$

Let $W^1(I \to H) \subset L^2(I \to H)$ be the Hilbert space of those functions $v : I \mapsto H$ in $L^2(I \to H)$ whose distributional time derivative v' can be represented by functions in $L^2(I \to H)$, equipped with the norm

$$\|v\|_{W^1(I \to H)} = \left(\int_I (\|v(t)\|_H^2 + \|v'(t)\|_H^2) dt \right)^{1/2} < \infty.$$

Given an open time interval I, we set

$$\mathcal{F}(I \times X) = L^2(I \to \mathcal{D}(\mathcal{E})) \cap W^1(I \to \mathcal{D}(\mathcal{E})').$$

Given an open time interval I and an open set $V \subset X$ (both nonempty), let

$$\mathcal{F}_{\mathrm{loc}}(I \times V)$$

be the set of all functions $v : I \times V \to \mathbb{R}$ such that for any open interval $I' \subset I$ relatively compact in I and open subset V' relatively compact in V there exists a function $u^{\#} \in \mathcal{F}(I \times X)$ satisfying $u = u^{\#}$ a.e. in $I' \times V'$. Finally, let

$$\mathcal{F}_c(I \times V) = \{v \in \mathcal{F}(I \times X) : v(t, \cdot) \text{ has compact support in } V \text{ for a.a. } t \in I\}.$$

Definition 3.3. Let I be an open time interval. Let V be an open subset in X, and let $Q = I \times V$. A function $u : Q \mapsto \mathbb{R}$ is a *weak* (local) solution of the heat equation $(\partial_t - A)u = 0$ in Q if

1. $u \in \mathcal{F}_{\mathrm{loc}}(Q)$;
2. for any open interval J relatively compact in I and $\varphi \in \mathcal{F}_c(Q)$

$$\int_J \int_V \varphi \partial_t u d\mu dt + \int_J \mathcal{E}(\varphi(t, \cdot), u(t, \cdot))dt = 0.$$

As noted in the elliptic case, this definition may contain implicitly some Neumann type boundary condition along a natural boundary of X (see [94, 96] for a detailed discussion).

3.4 Harnack type Dirichlet spaces

The following is the main definition of this section.

Definition 3.4. We say that a regular strictly local Dirichlet form $(\mathcal{E}, \mathcal{D}(\mathcal{E}))$ on $L^2(M, \mu)$ is of *Harnack type* if the distance ρ satisfies the qualitative conditions (A1), (A2), and the following scale invariant parabolic Harnack inequality holds. There exists a constant C such that for any $z \in M$, $r > 0$ and weak nonnegative solution u of the heat equation $(\partial_t - A)u = 0$ in $Q = (s - 4r^2, s) \times B(z, 2r)$ we have

$$\sup_{(t,x) \in Q_-} u(t, x) \leqslant C \inf_{(t,x) \in Q_+} u(t, x), \tag{3.1}$$

where $Q_- = (s - 3r^2, s - 2r^2) \times B(z, r)$, $Q_+ = (s - r^2, s) \times B(z, r)$ and both sup and inf are essential, i.e. are computed up to sets of measure zero.

Any Harnack type Dirichlet form $(\mathcal{E},\mathcal{D}(\mathcal{E}))$ obviously satisfies the following elliptic Harnack inequality (with the same constant C as in (3.1)). For any $z \in X$ and $r > 0$ and weak nonnegative solution u of the equation $Lu = 0$ in $B(z, 2r)$ we have

$$\sup_{B(z,r)} u \leqslant C \inf_{B(z,r)} u. \tag{3.2}$$

This elliptic Harnack inequality is weaker than its parabolic counterpart.

One of the simple, but important consequences of the Harnack inequality (3.1) is the following quantitative Hölder continuity estimate (see, for example, [87, Theorem 5.4.7] and [94]).

Theorem 3.1. *Assume that $(\mathcal{E},\mathcal{D}(\mathcal{E}))$ is a Harnack type Dirichlet form on $L^2(M,\mu)$. Then there exists $\alpha \in (0,1)$ and $A > 0$ such that any local (weak) solution of the heat equation $(\partial_t - A)u = 0$ in $Q = (s-4r^2, s) \times B(x,2r)$, $x \in X$, $r > 0$ has a continuous representative and satisfies*

$$\sup_{(t,y),(t',y')\in Q'} \left\{ \frac{|u(y,t)-u(y',t')|}{[|t-t'|^{1/2} + \rho_{\mathcal{E}}(y,y')]^{\alpha}} \right\} \leqslant \frac{A}{r^{\alpha}} \sup_{Q} |u|,$$

where $Q' = (s-3r^2, s-r^2) \times B(x,r)$.

A crucial consequence of this is that, on a Harnack type Dirichlet space, local weak solutions of the Laplace equation or the heat equation are continuous functions (in the sense that they admit a continuous representative).

Definition 3.5. Let $(M,\mu,\mathcal{E},\mathcal{D}(\mathcal{E}))$ be a regular strictly local Dirichlet space satisfying the qualitative conditions (A1), (A2).

- We say that the *doubling volume property* holds if there is a constant D_0 such that $V(x,2r) \leqslant D_0 V(x,r)$ for all $x \in M$ and $r > 0$.
- We say that the *scale invariant L^2 Poincaré inequality* holds if there is a constant P_0 such that for any ball $B = B(x,r)$ in (M,ρ)

$$\forall\, u \in \mathcal{F}_{\mathrm{loc}}(B(x,r)),\quad \int_B |u-u_B|^2 d\mu \leqslant P_o r^2 \int_B d\Gamma(u,u),$$

  where u_B denotes the average of u over B.
- We say that these properties hold *uniformly at small scales* if they hold under the restriction that $r \in (0,1)$.

We can now state the main result of this section which is a direct generalization of Theorem 2.2 in the setting of strictly local regular Dirichlet spaces (see [94]).

Theorem 3.2. *Let $(M,\mu,\mathcal{E},\mathcal{D}(\mathcal{E}))$ be a regular strictly local Dirichlet space satisfying the qualitative conditions* (A1), (A2). *The following properties are equivalent.*

- *The space* $(M,\mu,\mathcal{E},\mathcal{D}(\mathcal{E}))$ *is a Harnack type Dirichlet space.*
- *The doubling volume property and the scale invariant Poincaré inequality are satisfied on* $(M,\mu,\mathcal{E},\mathcal{D}(\mathcal{E}))$.
- *The heat semigroup* e^{tA} *admits a transition kernel* $h(t,x,y)$ *satisfying the two-sided bound*

$$\frac{c}{V(x,\sqrt{t})}\exp\left(-A\frac{\rho(x,y)^2}{t}\right)\leqslant h(t,x,y)\leqslant \frac{C}{V(x,\sqrt{t})}\exp\left(-a\frac{\rho(x,y)^2}{t}\right)$$

for constants $0<a,A,c,C<\infty$.

As in the classical case, if one uses Moser's iteration techniques, one of the first steps of the proof that the doubling property and Poincaré inequality imply the parabolic Harnack inequality is that they imply the family of Sobolev inequalities

$$\forall\, f\in\mathcal{F}_c(B),\quad \left(\int_B |f|^{2q}d\mu\right)^{1/q}\leqslant \frac{C_M r^2}{\mu(B)^{2/\nu}}\left(\int_B d\Gamma(f,f)+\int_B r^{-2}|f|^2 d\mu\right) \tag{3.3}$$

for some $q>1$ and $\nu>2$ related to q by $q=\nu/(\nu-2)$. This inequality implies the volume estimate

$$\forall\, x\in M,\ r>s.0,\quad V(x,r)\leqslant C(r/s)^{\nu}V(x,s).$$

Furthermore, a precise analog of Theorem 2.1 holds in this setting, as well as the following version of Theorem 2.3 (see [61]).

Theorem 3.3. *Let* $(M,\mu,\mathcal{E},\mathcal{D}(\mathcal{E}))$ *be a regular strictly local Dirichlet space satisfying the qualitative conditions* (A1), (A2) *and* (3.3). *The following properties are equivalent.*

- *The scale invariant* L^2 *Poincaré inequality.*
- *The scale invariant elliptic Harnack inequality.*
- *The scale invariant parabolic Harnack inequality.*

3.5 Imaginary powers of $-A$ and the wave equation

This section is merely a pointer to some interesting related results and literature regarding the wave equation. In the classical setting of $\mathbb{R}^n$, the wave equation is the PDE $(\partial_t^2-\Delta)u=0$. One of its main properties is the finite propagation speed property which asserts that if a solution u has support in the ball $B(x_0,r_0)$ at time t_0, then, at time t, its has support in

$B(x_0, r_0 + (t - t_0))$. Although this property can be proved in a number of elegant ways in $\mathbb{R}^n$, its generalization to other settings is not quite straightforward. Basic solutions of the wave equation can be obtain as follows. Using Fourier transform, consider the operator $\cos(t\sqrt{-\Delta})$ acting on $L^2(\mathbb{R}^n)$. Then for any smooth φ with compact support

$$u(t, \cdot) = \cos(t\sqrt{-\Delta})\varphi$$

is a solution of the wave equation with $u(0, \cdot) = \varphi$. This construction generalizes using spectral theory to any (nonpositive) self-adjoint operator, in particular, to the infinitesimal generator A of a Markov semigroup associated with a strictly local regular Dirichlet space $(M, \mu, \mathcal{E}, \mathcal{D}(\mathcal{E}))$. In this general setting, it is not entirely clear how to discuss the finite speed propagation property of the wave equation

$$(\partial_t^2 - A)u = 0.$$

Given a distance function d on $M \times M$ (assumed, at the very least, to be a measurable function on $M \times M$), one says that the wave equation (associated to A) has unit propagation speed with respect to d if for any functions $u_1, u_2 \in L^2(M, \mu)$ compactly supported in S_1, S_2, respectively, with

$$d(S_1, S_2) = \min\{d(s_1, s_2) : s_1 \in S_1, s_2 \in S_2\} > t$$

we have

$$\langle \cos(t\sqrt{-A})u_1, u_2 \rangle_\mu = 0.$$

The following theorem follows from the techniques and results in [91, 92].

Theorem 3.4. *Let* $(M, \mu, \mathcal{E}, \mathcal{D}(\mathcal{E}))$ *be a regular strictly local Dirichlet space satisfying the qualitative conditions* (A1), (A2). *Then the associated wave equation has unit propagation speed with respect to the distance* ρ *introduced in Definition* 3.1.

This result plays an important role in the study of continuity properties on L^p spaces of various operators defined via spectral theory by the functional calculus formula

$$m(-A) = \int_0^\infty m(\lambda) dE_\lambda,$$

where E_λ stands for a spectral resolution of the self-adjoint operator $-A$. This formula defines a bounded operator on $L^2(M, \mu)$ for any bounded function m. The question then is to examine what further properties of m imply additional continuity properties of $m(-A)$. The finite speed propagation property is very helpful in the study of these questions. We refer the reader to [37, 38, 92], where earlier references and detailed discussions of the literature can be found. As an illustrative example, we state the following result. For a function m defined on $[0,\infty)$ we set $m_t(u) = m(tu)$ and $\|m\|_{(s)} = \|(I - (d/du)^2)^{s/2} m\|_\infty$.

Theorem 3.5. *Let $(M, \mu, \mathcal{E}, \mathcal{D}(\mathcal{E}))$ be a regular strictly local Dirichlet space satisfying the qualitative conditions* (A1), (A2)*. Assume that the Sobolev inequality* (3.3) *holds for some $q > 1$ and ν given by $q = \nu/(\nu - 2)$.*

Fix a function $\eta \in \mathcal{C}_c^\infty((0,\infty))$, not identically 0. If m is a bounded function such that

$$\sup_{t>0} \|\eta m_t\|_{(s)} < \infty$$

for some $s > \nu/2$, then the operator $m(-A)$ is bounded on $L^p(M,\mu)$ for each $p \in (1,\infty)$. The operators $(-A)^{i\alpha}$, $\alpha \in \mathbb{R}$, are all bounded on $L^p(M,\mu)$, $1 < p < \infty$, and there exists a constant C such that the norm of $(-A)^{i\alpha}$ on $L^p(M,\mu)$ is at most $C(1 + |\alpha|)^{\nu/2}$, for all $\alpha \in \mathbb{R}$ and $1 < p < \infty$.

3.6 Rough isometries

One of the strengths of the techniques and results discussed in this paper is their robustness. In the present context, the idea of rough isometry was introduced by Kanai [64, 66, 65] and developed further in [32]. It has also been made very popular by the work of M. Gromov. Note that rough isometries as defined below do not preserve the small scale structure of the space.

Definition 3.6. Let (M_i, ρ_i, μ_i), $i = 1, 2$, be two measure metric spaces. We say that they are *roughly isometric* (or *quasiisometric*) as metric measure spaces if there are two maps $\varphi_k : M_i \to M_j$, $k = (i,j) \in \{(1,2),(2,1)\}$ and a constant A such that for $k' = (j,i)$ we have the following.

1. $\forall x \in M_i,\ \ \rho_i(x, \varphi_{k'} \circ \varphi_k(x)) \leqslant A$.
2. $M_j = \{y \in M_j : \rho_j(y, \varphi_k(M_i)) \leqslant A\}$.
3. $\forall\, x, x' \in M_i,\ \ A^{-1}(\rho_i(x,x') - A) \leqslant \rho_j(\varphi_k(x), \varphi_k(x')) \leqslant A(1 + \rho_i(x,x'))$.
4. $\forall\, x \in M_i,\ \ A^{-1}V(x,1) \leqslant V(\varphi_k(x),1) \leqslant AV(x,1)$.

Condition 3 requires that each of the maps φ_k roughly preserves large enough distances (larger than $2A$, say). Condition 2 requires that each of the maps φ_k is almost surjective, in a quantitative metric sense. The first condition says that the maps φ_k and φ_k' are almost inverse of each other. The last condition concerns volume transport and is obviously specific to the setting of *measure* metric spaces. This definition is nicely symmetric (as an equivalence relation should be!), but is redundant. It is enough to require the existence of one map, say from M_1 to M_2 with the last three properties. The existence of an almost inverse with the desired properties follows from the axiom of choice.

The relevance of rough isometries in the study of Harnack type Dirichlet space lies in the following stability theorem from [32, Theorem 8.3] (although [32] does not explicitly cover the setting of Dirichlet spaces, the same proof applies).

Theorem 3.6. *Let* $(M_i, \mu_i, \mathcal{E}_i, \mathcal{D}(\mathcal{E}_i))$, $i = 1, 2$, *be two regular strictly local Dirichlet spaces satisfying the qualitative conditions* (A1), (A2). *Assume further that these two spaces satisfy the volume doubling property and the* L^2 *Poincaré inequality, uniformly at small scales. If* (M_1, ρ_1, μ_1) *and* (M_2, ρ_2, μ_2) *are roughly isometric as metric measure spaces, then* $(M_1, \mu_1, \mathcal{E}_1, \mathcal{D}(\mathcal{E}_1))$ *is of Harnack type if and only if* $(M_2, \mu_2, \mathcal{E}_2, \mathcal{D}(\mathcal{E}_2))$ *is of Harnack type.*

Example 3.1. In a sense, the following example illustrates in the simplest non-trivial possible way the results of this section. Consider the two-dimensional cubical complex obtained as the subset M of $\mathbb{R}^3$ of those point (x, y, z) with at least one coordinate in $\mathbb{Z}$. In other words, M is the union of the planes $\{x = k\}$, $\{y = k\}$, $\{z = k\}$, $k \in \mathbb{Z}$. It is also the union $M = \bigcup_{\mathbf{k}} Q_{\mathbf{k}}$, where $Q_{\mathbf{k}}$ is the two-dimensional boundary of the unit cube with lower left back corner $\mathbf{k} \in \mathbb{Z}^3$. This space is equipped with its natural measure μ (Lebesgue measure on each of the planes above). To describe the natural Dirichlet form and its domain, we recall that if F is a face on a unite cube Q_k and if a function f in $L^2(F)$ has distributional first order partial derivatives in $L^2(F)$ (i.e., is in the Sobolev space $H^1(F)$), then the trace of f along the one dimensional edges of the face F are well defined, say, as an L^2 function on the edges. Taking into account this remark, we set (the factor of 1/2 is to account for the appearance of each face in exactly two cubes)

$$\mathcal{E}(f, g) = \frac{1}{2} \sum_{\mathbf{k}} \int_{Q_{\mathbf{k}}} \nabla f \cdot \nabla g \, d\mu$$

for all $f, g \in \mathcal{D}(\mathcal{E})$, where $\mathcal{D}(\mathcal{E})$ is the space of those functions $f \in L^2(M)$ which have distributional first order partial derivatives in $L^2(F)$ on each face F of any cube $Q_{\mathbf{k}}$, satisfy $\mathcal{E}(f, f) < \infty$, and have the property that for each pair of faces F_1, F_2 sharing an edge I, the restrictions of $f|_{F_1}$ and $f|_{F_2}$ to the edge I coincide. In the above formula, ∇f refers to the Euclidean gradient of f viewed as a function defined on each of the square faces of the cube $Q_{\mathbf{k}}$. Because of the above-mentioned trace theorem for Sobolev functions, it is easy to see that $(\mathcal{E}, \mathcal{D}(\mathcal{E}))$ is a Dirichlet space. It is local, and one can show (although this is not entirely obvious) that it is regular (see, for example, [82]). The distance ρ associated to this Dirichlet form on M coincides with the natural shortest path distance on this cubical complex. It is not hard to check that

- The uniform small scale doubling property holds.
- The uniform small scale Poincaré inequality holds.

- The metric measure space (M, ρ, μ) is roughly isometric to $\mathbb{R}^3$.

Thus, from Theorem 3.6 it follows that this Dirichlet space is a Harnack type Dirichlet space.

4 Flat Sobolev Inequalities

In the previous sections, we discussed the role of the family of localized Sobolev inequalities (2.11) in Moser's iteration and related techniques. In some sense, the need to consider (2.11) instead of the more classical inequality (2.10) comes from looking at situations that are inhomogeneous either at the level of location or at the level of scales, or both. Because of this one sometimes refers to a global Sobolev inequality that do not require localization as a "flat" Sobolev inequality. For instance, one might ask: What complete n-dimensional Riemannian manifolds satisfy a Sobolev inequality of the form

$$\forall\, f \in \mathcal{C}_c(M),\ \ \|f\|_{2n/(n-2)} \leqslant S\|\nabla f\|_2?$$

It turns out that this inequality is satisfied by a variety of manifolds not having much in common with each others, including manifolds with nonnegative Ricci curvature and maximal volume growth, as well as simply connected manifolds with nonpositive sectional curvature (see [60, Theorem 8.3] and the references therein for this result).

In this section, we discuss such inequalities: how to prove them and what are they good for?

4.1 How to prove a flat Sobolev inequality?

There are many interesting approaches to proving Sobolev inequalities and we will, essentially, discuss only one of them here. One useful aspect of this approach is its robustness. One weakness, among others, is that it never produces best constants.

Definition 4.1. Let (M, g) be a complete Riemannian manifolds. We say that it satisfies an L^p *pseudo-Poincaré inequality* if

$$\forall\, f \in \mathcal{C}_c^\infty(M),\ \ \|f - f_r\|_p \leqslant Ar\|\nabla f\|_p$$

for all $r > 0$, where f_r is a function such that $f_r(x)$ is the average of f over the ball $B(x, r)$.

Theorem 4.1 ([4, Theorem 9.1]**).** *Let (M, g) be a complete Riemannian manifolds satisfying the L^p pseudo-Poincaré inequality. Assume that there exists*

$N > 0$ such that $V(x, r) \geqslant cr^N$ for all $x \in M$ and $r > 0$. Then the inequality

$$\forall\, f \in \mathcal{C}_c^\infty(M),\ \ \int_M |f|^{p(1+1/N)} d\mu \leqslant C(M,p) \left(\int_M |\nabla f|^p d\mu\right) \left(\int_M |f| d\mu\right)^{p/N}$$

holds. If $N > p$, then

$$\forall\, f \in \mathcal{C}_c^\infty(M),\ \ \|f\|_{pN/(N-p)} \leqslant S(M,p) \|\nabla f\|_p.$$

Remark 4.1. The paper [4] shows that a great number of other interesting Sobolev type inequalities follow as a corollary of the above result.

Remark 4.2. The above definition and theorem hold unchanged for $p = 2$ in the context of strictly local regular Dirichlet spaces satisfying the qualitative conditions (A1), (A2).

Remark 4.3. The volume condition $V(x, r) \geqslant cr^N$ is sharp in the sense that it follows from the validity of any of the two stated inequalities.

Remark 4.4. The same result holds if one replaces f_r in the pseudo-Poincaré inequality by $M_r f$ and replaces the volume hypothesis by $\|M_r f\|_\infty \leqslant Cr^{-N}\|f\|_1$. For instance, M_r could be averages over sets different from balls or some more sophisticated operators. As an example, let $M_r = H_{r^2} = e^{r^2 \Delta}$ be the heat semigroup on (M, g) at time $t = r^2$. Then, if one knows that for all $t > 0$, $\|f - H_t f\|_p \leqslant C\sqrt{t}\|\nabla\|_p$ and $\|H_t\|_{1\to\infty} \leqslant Ct^{-N/2}$, then one can conclude that the inequalities stated in the above theorem hold on M. The first of these two hypotheses is always satisfied if $p = 2$.

Example 4.1. Riemannian manifolds with nonnegative Ricci curvature satisfy the pseudo-Poincaré inequality of Definition 4.1 for any $1 \leqslant p \leqslant \infty$. They satisfy the Sobolev inequality

$$\forall\, f \in \mathcal{C}_c^\infty(M),\ \ \|f\|_{pN/(N-p)} \leqslant S(M,p) \|\nabla f\|_p$$

if and only if $V(x, r) \geqslant cr^N$ and $N > p \geqslant 1$ (see [87, Sect. 3.3.5]). On these manifolds, the volume is bounded by $V(x, r) \leqslant C_n r^n$, where n is the topological dimension. Hence $V(x, r) \geqslant cr^N$ for all $r > 0$ is possible only if $N = n$ and $V(x, r) \simeq r^n$.

Example 4.2. Let (M, g) be a connected unimodular Lie group equipped with a left-invariant Riemannian metric. Then the pseudo-Poincaré inequality of Definition 4.1 holds for any $1 \leqslant p \leqslant \infty$ (see [31] or [87, 3.3.4]). The inequality

$$\forall\, f \in \mathcal{C}_c^\infty(M),\ \ \|f\|_{pN/(N-p)} \leqslant S(M,p) \|\nabla f\|_p$$

holds if and only if $V(r) \geqslant cr^N$ for all $r > 0$. For instance, if M is the group of upper-triangular 3 by 3 matrices with 1's on the diagonal (i.e., the Heisenberg group), then for any left-invariant Riemannian metric, $V(x,r) \geqslant cr^N$ for all $r > 0$ and $N \in [3,4]$.

4.2 Flat Sobolev inequalities and semigroups of operators

Sobolev inequalities can be generalized in useful ways in many contexts one of which involves the infinitesimal generator A of a strongly continuous semigroup of operator e^{tA} jointly defined on the spaces $L^p(M,\mu)$, $1 \leqslant p < \infty$. One of the most straightforward results in this context is the following theorem from [26] which extends an earlier result of Varopoulos [100] (see also [103]). For $\alpha > 0$ we set

$$(-A)^{-\alpha/2} = \Gamma(\alpha/2)^{-1} \int_0^\infty t^{-1+\alpha/2} e^{tA} dt.$$

Theorem 4.2. *Fix $p \in (1,\infty)$. Assume that e^{tA} is a bounded holomorphic semigroup of operator on $L^p(M,\mu)$ which extends as an equicontinuous semigroup on both $L^1(M,\mu)$ and $L^\infty(M,\mu)$. Then for any $N > 0$ the following two properties are equivalent.*

- *There exists C_1 such that*

$$\forall\, f \in L^1(M,\mu), \;\; \|e^{tA} f\|_\infty \leqslant C_1 t^{-N/2} \|f\|_1.$$

- *There exists C_2 such that for one pair (equivalently, for all pairs) (α, q) with $0 < \alpha p < N$ and $1/q = 1p - \alpha/N$, we have*

$$\forall\, f \in L^p(M,\mu), \;\; \|(-A)^{-\alpha/2} f\|_q \leqslant C_2 \|f\|_p.$$

Remark 4.5. The first property is known as a form of ultracontractivity (boundedness of e^{tA} from L^1 to L^∞ for all $t > 0$). The second property states that a Sobolev type inequality holds, namely, $\|f\|_q \leqslant C_2 \|(-A)^{\alpha/2} f\|_p$, $f \in \mathrm{Dom}((-A)^{\alpha/2}$.

Remark 4.6. A semigroup e^{tA} is bounded holomorphic on $L^p(X,\mu)$ if

$$t \|A e^{tA} f\|_p \leqslant C \|f\|_p$$

for all $f \in L^p(M,\mu)$ and $t > 0$. This implies that for any $\alpha \in (0,1]$ and f in the domain of $(-A)^{\alpha/2}$

$$\forall\, t > 0, \quad \|f - e^{tA} f\|_p \leqslant C_\alpha t^{\alpha/2} \|(-A)^{\alpha/2} f\|_p.$$

This can be viewed as a form of pseudo-Poincaré inequality.

Theorems such as Theorem 4.2 apply nicely in the context of Dirichlet spaces because the associated semigroups are self-adjoint on $L^2(M,\mu)$ and contract each $L^p(M,\mu)$, $1 \leqslant p \leqslant \infty$. Semigroups of self-adjoint contractions on $L^2(M,\mu)$ are automatically bounded holomorphic on $L^2(M,\mu)$. Moreover, in the regular strictly local Dirichlet space context described earlier, the generator A is related to the form $\mathcal{E}$ and the energy form Γ by

$$\|(-A)^{1/2}f\|_2^2 = \mathcal{E}(f,f) = \int_M d\Gamma(f,f), \quad f \in \mathrm{Dom}((-A)^{1/2}) = \mathcal{D}(\mathcal{E}).$$

For the following result see [15, 100, 103] and also [33].

Theorem 4.3. *Fix $N > 0$. Let $(M,\mu,\mathcal{E},\mathcal{D}(\mathcal{E}))$ be a Dirichlet space with associated semigroup e^{tA}. The following properties are equivalent.*

- *There exists C_1 such that*

$$\forall\, f \in L^1(M,\mu),\ t > 0, \quad \|e^{tA}f\|_\infty \leqslant C_1 t^{-N/2}\|f\|_1.$$

- *For one (equivalently, all) (α,q) with $1 < \alpha < N/2$ and $q = 2N/(N-2\alpha)$ there exists $C(\alpha)$ such that*

$$\forall\, f \in \mathrm{Dom}((-A)^{\alpha/2}), \quad \|f\|_q \leqslant C(\alpha)\|(-A)^{\alpha/2}f\|_2.$$

- *There exists C_2 such that*

$$\forall\, f \in L^1(M,\mu) \cap \mathcal{D}(\mathcal{E}), \quad \|f\|_2^{2(1+2/N)} \leqslant C_2\mathcal{E}(f,f)\|f\|_1^{4/N}.$$

Remark 4.7. The first property is a particular type of ultracontractivity. The second property is a Sobolev type inequality. If $N > 2$, one can take $\alpha = 1$, $q = 2N/(N-2)$ and the inequality takes the form $\|f\|_q \leqslant C_1\mathcal{E}(f,f)$. The third property is a Nash inequality.

Example 4.3. Let (M,g) be an n-dimensional complete Riemannian manifold that is simply connected and has nonpositive sectional curvature. By a simple comparison argument (see, for example, [20, Theorem 6] and the references therein), the heat kernel on M is bounded from above by the Euclidean heat kernel. In particular, for all $t > 0$,

$$\sup_{x,y\in M}\{h(t,x,y)\} \leqslant c_n t^{-n/2}.$$

This implies that for all $t > 0$ we have $\|e^{t\Delta_M}f\|_\infty \leqslant c_n t^{-n/2}\|f\|_1$. Hence the above theorem gives the Sobolev inequality

$$\forall\, f \in \mathcal{C}_c^\infty(M), \quad \|f\|_{2n/(n-2)} \leqslant S_M\|(-\Delta)^{1/2}f\|_2.$$

Of course, $\|(-\Delta)^{1/2} f\|_2 = \|\nabla f\|_2$, so that this inequality can be written as

$$\forall\, f \in \mathcal{C}_c^\infty(M),\;\; \|f\|_{2n/(n-2)} \leqslant S_M \|\nabla f\|_2.$$

There is an open conjecture that this inequality should holds with S_M being the same constant as in the Euclidean n-space.

The first property in Theorem 4.3 obviously calls for a more general formulation. The following general elegant result was obtained by Coulhon [27] (after many attempts by different authors). A smooth positive function Φ defined on $[0,\infty)$ satisfies condition (D) if there exists $\varepsilon \in (0,1)$ such that $\varphi'(s) \geqslant \varepsilon\varphi'(t)$ for all $t > 0$ and $s \in [t, 2t]$, where $\varphi(s) = -\log \Phi(s)$.

Theorem 4.4 ([27]). *Let $(M, \mu, \mathcal{E}, \mathcal{D}(\mathcal{E}))$ be a Dirichlet space with associated semigroup e^{tA}. Let Φ be a positive smooth decreasing function on $[0,\infty)$ satisfying condition* (D)*, and let $\Theta = -\Phi' \circ \Phi^{-1}$. The following properties are equivalent.*

- *There exists a constant $c_1 \in (0,\infty)$ such that*

$$\forall\, f \in L^1(M,\mu),\; t > 0,\;\; \|e^{tA} f\|_\infty \leqslant \Phi(c_1 t)\|f\|_1.$$

- *The exists a constant $C_1 \in (0,\infty)$ such that for all $f \in L^1(M,\mu) \cap \mathcal{D}(\mathcal{E})$ with $\|f\|_1 \leqslant 1$ we have*

$$\Theta(\|f\|_2^2) \leqslant C\mathcal{E}(f,f).$$

We refer the reader to [4, 8, 9, 27] for explicit examples and further results.

4.3 The Rozenblum–Cwikel–Lieb inequality

One of the surprising aspects of the Sobolev inequality

$$\|f\|_{2N/(N-2)}^2 \leqslant S^2 \mathcal{E}(f,f)$$

is how many different equivalent form it takes (hence the title "Sobolev inequalities in disguise" of [4]). Despite the equivalence of this different forms, some appear "stronger" than other. For instance, on one hand, deducing from the above inequality the Nash inequality

$$\|f\|_2^{2(1+2/N)} \leqslant S^2 \mathcal{E}(f,f)\|f\|_1^{4/N}$$

only involves a simple use of Hölder's inequality (and the constant remains the same). On the other hand, recovering the Sobolev inequality from its Nash form involves some more technical arguments. The constant S changes in the process (the two inequalities in $\mathbb{R}^N$ have different best constants) and one needs to assume that $N > 2$.

In 1972, Rozenblum proved a remarkable spectral inequality showing that, in $\mathbb{R}^N$ with $N \geqslant 3$, if V is a nonnegative measurable function and $\mathcal{N}_-(-\Delta - V)$ denotes the number of negative eigenvalues of $-\Delta - V$, then there exists a constant $C(N)$ such that

$$\mathcal{N}_-(-\Delta - V) \leqslant C(N) \int_{\mathbb{R}^N} V(x)^{N/2} dx.$$

Very different proofs were later given by Cwikel and by Lieb, and this inequality is known as the *Rozenblum–Cwikel–Lieb inequality.* We refer the reader to the review of the literature in [70, 84]. The following elegant result is taken from [70] and is based on the technique used in [73] in Euclidean space.

Theorem 4.5 ([70]). *Let $(M, \mu, \mathcal{E}, \mathcal{D}(\mathcal{E}))$ be a Dirichlet space with associated semigroup e^{tA}. Assume that the Sobolev inequality*

$$\forall\, f \in \mathcal{D}(\mathcal{E}), \quad \|f\|^2_{2N/(N-2)} \leqslant S^2 \mathcal{E}(f, f)$$

holds for some $N > 2$. Then for any measurable function $V \geqslant 0$

$$\mathcal{N}_-(-A - V) \leqslant C(N) \int_M V^{N/2} d\mu.$$

In [84], this result is generalized in a number of useful ways. In particular, the following version related to Theorem 4.4 is obtained.

Theorem 4.6 ([84]). *Fix a nonnegative convex function Q on $[0, \infty)$, growing polynomially at infinity and vanishing in a neighborhood of 0. Set*

$$q(u) = \int_0^\infty v^{-1} Q(v) e^{-v/u} dv.$$

Let $(M, \mu, \mathcal{E}, \mathcal{D}(\mathcal{E}))$ be a Dirichlet space with associated semigroup e^{tA}. Assume that

$$\forall\, f \in L^1(M, \mu),\ t > 0, \quad \|e^{tA} f\|_\infty \leqslant \Phi(t) \|f\|_1$$

with Φ continuous, integrable at infinity $\left(\int^\infty \varphi(t) dt < \infty\right)$, and satisfying $\Phi(t) = O(t^{-\alpha})$ at 0 for some $\alpha > 0$. Then for any measurable function V

$$\mathcal{N}_-(-A - V) \leqslant \frac{1}{q(1)} \int_0^\infty \left(\int_M Q(tV(x)) d\mu(x) \right) \frac{\Phi(t)}{t} dt.$$

Remark 4.8. One can take $Q(u) = (u-1)_+$. In this case,

$$\int_0^\infty \left(\int_M Q(tV(x))d\mu(x) \right) \frac{\Phi(t)}{t} dt \leqslant \int_M \left(V(x) \int_{1/V(x)}^\infty \Phi(t)dt \right) d\mu(x)$$

so that if

$$\Psi(u) = \int_u^\infty \Phi(t)dt,$$

then

$$\mathcal{N}_-(-A - V) \leqslant C \int_M V(x)\Psi(1/V(x))d\mu(x).$$

In particular, if $\Phi(t) \simeq t^{-N/2}$, $t > 0$ for some $N > 2$, then $\Psi(u) \simeq u^{-N/2+1}$, $u > 0$, and

$$\mathcal{N}_-(-A - V) \leqslant C \int_M V^{N/2} d\mu.$$

Example 4.4. Let (G, g) be an amenable connected Lie group of topological dimension n equipped with a left-invariant Riemannian metric with Laplace operator Δ. In this case, there are two possible behaviors for the function Φ. If G has polynomial volume growth, then

$$\Phi(t) \simeq \begin{cases} t^{-n/2} & \text{for } t \in (0,1], \\ t^{-N/2} & \text{for } t \in (1,\infty), \end{cases}$$

where N is some integer. If that is not the case, then G has exponential volume growth and

$$\Phi(t) \leqslant C \times \begin{cases} t^{-n/2} & \text{for } t \in (0,1] \\ e^{-ct^{1/3}} & \text{for } t \in (1,\infty) \end{cases}$$

for some $c, C \in (0,\infty)$ (a similar lower bound holds as well).

In the case of polynomial volume growth, application of Theorem 4.6 requires $N > 2$. Assuming that $N > 2$, the function Ψ introduced in the above remark is given by $\Psi(u) \simeq u^{-N/2+1}\mathbf{1}_{u>1} + (1 + u^{-n/2+1})\mathbf{1}_{u\leqslant 1}$. Hence

$$\mathcal{N}_-(-\Delta - V) \leqslant C \Big(\int_{\{V \geqslant 1\}} V(1 + V^{n/2-1})d\mu + \int_{\{V<1\}} V^{N/2} d\mu \Big).$$

In the case of exponential volume growth, one gets

$$\mathcal{N}_-(-\Delta - V) \leqslant C \int_{\{V \geqslant 1\}} V(1 + V^{n/2-1}) d\mu + C \int_{\{V<1\}} e^{-cV^{-1/3}} d\mu.$$

In this case, since the volume growth is exponential, we see that for a smooth positive potential with $V(x) \simeq (1+\rho(e,x))^{-\gamma}$, $\mathcal{N}_-(-\Delta - V)$ is finite if $\gamma > 3$.

4.4 Flat Sobolev inequalities in the finite volume case

Recall that a flat Sobolev inequality of the form $\forall f \in \mathcal{C}_c^\infty(M)$, $\|f\|_{2q} \leqslant S_M \|\nabla f\|_2$, with $q > 1$, on a complete Riemannian manifolds (M, g), implies that the volume grows at least as r^ν with $q = \nu/(\nu - 2)$. In particular, the volume of M cannot be finite. In order to allow for some finite volume manifolds, one needs to consider inequalities of the form

$$\forall f \in \mathcal{C}_c^\infty, \quad \|f\|_{2q}^2 \leqslant a_M \|f\|_2^2 + C_M^2 \|\nabla f\|_2^2. \tag{4.1}$$

If we assume that the volume of M is finite, we can normalize the measure so that $\mu(M) = 1$ and then it is easy to see that the above inequality can hold only if $a_M \geqslant 1$. Moreover, if the global Poincaré inequality $\|f - f_M\|_2 \leqslant A_M \|\nabla f\|_2$ holds for all $f \in \mathcal{C}^\infty(M)$, then (4.1) implies

$$\forall f \in \mathcal{C}_c^\infty, \quad \|f\|_{2q}^2 \leqslant \|f\|_2^2 + S_M^2 \|\nabla f\|_2^2. \tag{4.2}$$

The aim of this section is to point out a beautiful consequence of this inequality obtained by Bakry and Ledoux [5]. We refer the reader to [5] for a complete discussion and detailed references.

Theorem 4.7 ([5, Theorem 2]). *Assume that* (M, g) *is a complete Riemannian manifold with finite volume. Assume that, equipped with its normalized Riemannian measure,* (M, g) *satisfies* (4.2) *for some* $q > 1$ *and* $S_M \in (0, \infty)$. *Then* M *is compact with*

$$\mathrm{Diam}(M) \leqslant \pi \frac{\sqrt{q}}{q-1} S_M.$$

This result is a form of a well-known theorem of Meyers that asserts that an n-dimensional Riemannian manifold whose Ricci curvature is bounded from below by $\mathrm{Ric} \geqslant kg$ with $k > 0$ must be compact with diameter at most $\pi\sqrt{(n-1)/k}$. Indeed, Ilias proved that, on a manifold of dimension n, the hypothesis $\mathrm{Ric} \geqslant kg$ for some $k > 0$ implies the Sobolev inequality (4.2) with $q = n/(n-2)$ and $S_M^2 = 4(n-1)/n(n-2)k$. Hence Meyers' result follows from Ilias' inequality and the above theorem. The upper bound in the theorem is sharp and is attained when M is a sphere.

The above theorem of Bakry and Ledoux is, in fact, obtained in a much more general setting of strictly local Dirichlet spaces (see [5] for a precise description).

4.5 Flat Sobolev inequalities and topology at infinity

We complete this section on flat Sobolev inequalities by pointing out the relevance of the Sobolev inequality in some problems concerning topology. The following result due to Carron [18] is actually closely related to the results concerning the Rozenblum–Cwikel–Lieb inequality.

Theorem 4.8 ([18, Theorem 0.4]). *Let* (M, g) *be a complete Riemannian manifold* (*hence, connected*) *satisfying the Sobolev inequality*

$$\forall\, f \in \mathcal{C}_c^\infty(M), \quad \|f\|_{2\nu/(\nu-2)}^2 \leqslant S_M^2 \int |\nabla f|^2 d\mu$$

for some $\nu > 2$*. Assume that the smallest negative eigenvalue* ric_- *of the Ricci tensor is in* $L^{\nu/2}(M)$*. Then* M *has only finitely many ends. In fact, there exists a constant* $C(\nu)$ *such that the number of ends is bounded by*

$$1 + C(\nu) S_M^2 \int\limits_M |ric_-|^{\nu/2} d\mu.$$

For more sophisticated results in this direction see, for example, [16, 18, 19] and the references therein.

5 Sobolev Inequalities on Graphs

All the ideas and techniques discussed in this paper can be developed and used in the discrete context of graphs, sometimes to great advantage. To a large extend, the context of graph is actually harder to work with than the context of manifolds (and strictly local Dirichlet spaces), but the new difficulties that appear are mostly of a technical nature and can often be overcome. This short section provides pointers to the literature and explains in some detail one of the first applications of Sobolev inequalities on graphs, namely, Varopoulos' solution of Kesten's conjecture regarding random walks on finitely generated groups. We refer to [47] for a short survey and to [98, 105] for a detailed treatment of some aspects.

5.1 Graphs of bounded degree

In what follows, a graph is a pair (V, E), where E is a symmetric subset of $V \times V$ and V is finite or countable. Elements of V are vertices and elements of E are (oriented) edges. For $x, y \in V$ we write $x \sim y$ if $(x, y) \in E$ and we say that x, y are neighbors. A path in V is a sequence of vertices such that consecutive points are neighbors. The length of a path is the number of edges it crosses. The distance $\rho(x, y)$ between two points $x, y \in V$ is the minimal length of a path joining them. The degree $\mu(x)$ of $x \in V$ is the number of $y \in V$ such that $(x, y) \in E$. Throughout the paper, we assume that our graphs are connected, i.e., $\rho(x, y) < \infty$ for all $x, y \in V$ and have uniformly bounded degree, i.e., there exists $D \in [1, \infty)$ such that $\sup\limits_x\{\mu(x)\} = D$. Moreover, we equip V with the measure μ defined by $\mu(A) = D^{-1} \sum\limits_{x\in A} \mu(x)$. A graph is regular if $\mu(x) = D$ for all x. In this case, the measure μ is a counting measure. Let $B(x, r)$ be the (closed) ball of radius r around x, and let $V(x, r) = \mu(B(x, r))$. For a book treatment of various aspects of the study the volume growth in Cayley graphs see [35].

Given a function f on V, we set $df(x, y) = f(y) - f(x)$ and

$$|\nabla f(x)| = \left(\mu(x)^{-1} \sum_{y\sim x} |df(x, y)|^2\right)^{1/2}.$$

Also, set $f_r(x) = V(x, r)^{-1} \sum\limits_{B(x,r)} f(z)\mu(z)$.

We now have all the ingredients to consider whether or not the graph (V, E) satisfies the Sobolev inequality

$$\forall\, f \in \mathcal{C}_c(V), \quad \|f\|_{2q} \leqslant S\|\nabla f\|_2 \tag{5.1}$$

for some $q > 1$, and related inequalities. Here, $\mathcal{C}_c(V)$ is the space of functions with finite support. Moreover, according to our notation, we have

$$\|f\|_{2q}^{2q} = \sum_{x\in V} |f(x)|^{2q}\mu(x) \text{ and } \|\nabla f\|_2^2 = \sum_{x\in V}\sum_{y\sim x} |f(y) - f(x)|^2.$$

In what follows, we concentrate on the simple case of flat Sobolev inequalities because this case is quite interesting and important and does avoid most technical difficulties. For developments paralleling the ideas and results of Sect. 2 we refer the reader to [28, 29, 30, 52, 53, 98] and the references therein.

5.2 Sobolev inequalities and volume growth

We start with the following two theorems.

Theorem 5.1. *Fix $\nu > 0$. For a graph (V,E) as above, the following properties are equivalent.*

- $\forall f \in \mathcal{C}_c(V), \quad \|f\|_2^{(1+2/\nu)} \leqslant N\|\nabla f\|_2 \|f\|_1^{2/\nu}$.
- $\forall f \in \mathcal{C}_c(V)$ *with support in a finite set* Ω, $\|f\|_2 \leqslant C\mu(\Omega)^{1/\nu}\|\nabla f\|_2$.

Moreover, if $\nu > 2$, these properties are equivalent to (5.1) *with $q = \nu/(\nu-2)$. Finally, any of these inequalities implies the existence of $c > 0$ such that*

$$\forall x \in V, r > 0, \quad V(x,r) \geqslant cr^{\nu}.$$

Remark 5.1. The first inequality is a Nash inequality, the second is a Faber–Krahn inequality. For a proof of this theorem see, for example, [4].

The next results gives two Nash inequalities under the volume growth hypothesis that $V(x,r) \geqslant cr^{\nu}$. The first inequality requires no additional hypotheses, whereas the second one depends on the validity of a pseudo-Poincaré inequality. Under that extra hypothesis, the Nash inequality one obtains is, in fact, equivalent to the volume lower bound. Both results are optimal (see [6]).

Theorem 5.2 ([6, 31]). *Fix $\nu > 0$ and assume that a graph (V,E) has volume growth bounded from below:*

$$\forall x \in V, r > 0, \quad V(x,r) \geqslant cr^{\nu}.$$

- *In all the cases,*

$$\forall f \in \mathcal{C}_c(V), \quad \|f\|_2^{(1+1/\gamma)} \leqslant N\|\nabla f\|_2 \|f\|_1^{1/\gamma}, \quad \gamma = \nu/(\nu+1).$$

- *Assume that the pseudo-Poincaré inequality $\forall f \in \mathcal{C}_c(V), \quad \|f - f_r\|_2 \leqslant Cr\|\nabla f\|_2$ holds on (V,E). Then*

$$\forall f \in \mathcal{C}_c(V), \quad \|f\|_2^{(1+2/\nu)} \leqslant N\|\nabla f\|_2 \|f\|_1^{2/\nu}.$$

Proof. First statement. Fix a finite set Ω. For each $x \in \Omega$ let $r(x)$ be the distance between x and $V \setminus (\Omega)$. If f has support in Ω, by a simple use of the Cauchy–Schwarz inequality, for all $x \in \Omega$, $|f(x)|^2 \leqslant r(x)\|\nabla f\|_2^2$. Also $\Omega \supset B(x, r(x)-1)$ for each $x \in \Omega$. Hence, by hypothesis,

$$\mu(\Omega) \geqslant V(x, r(x)-1) \geqslant c(r(x)-1)^{\nu} \geqslant c' r(x)^{\nu}.$$

This yields $|f(x)|^2 \leqslant C\mu(\Omega)^{1/\nu}\|\nabla f\|_2^2$. Summing over Ω, we find

$$\|f\|_2 \leqslant C^{1/2}\mu(\Omega)^{(\nu+1)/2\nu}\|\nabla f\|_2.$$

The desired result follows from Theorem 5.1.

Second statement. Observe that the volume hypothesis yields

$$\|f_r\|_\infty \leqslant c^{-1}r^{-\nu}\|f\|_1.$$

Writing $\|f\|_2^2 = \langle f, f - f_r\rangle + \langle f, f_r\rangle$ and using the hypotheses, we obtain

$$\|f\|_2^2 \leqslant Cr\|f\|_2\|\nabla f\|_2 + c^{-1}r^{-\nu}\|f\|_1^2.$$

Picking $r \simeq (\|f\|_1^2\|f\|_2^{-1}\|\nabla f\|_2^{-1})^{1/(1+\nu)}$, we find

$$\|f\|_2^2 \leqslant C_1\|f\|_2^{\nu/(1+\nu)}\|\nabla f\|_2^{\nu/(1+\nu)}\|f\|_1^{2/(1+\nu)}$$

or

$$\|f\|_2^{(2+\nu)/(1+\nu)} \leqslant C_1\|\nabla f\|_2^{\nu/(1+\nu)}\|f\|_1^{2/(1+\nu)}.$$

Taking the $(1+\nu)/\nu$th power of both sides, we arrive at the desired inequality. □

5.3 Random walks

In the context of graphs, one of the possible natural definitions of the "Laplacian" (and the one we will use) is

$$\Delta_E f(x) = \mu(x)^{-1}\sum_{y\sim x}(f(y) - f(x)) = (K - I)f(x),$$

where I is the identity operator and K is the Markov kernel

$$K(x,y) = \begin{cases} \mu(x)^{-1} & \text{if } y \sim x, \\ 0 & \text{otherwise}, \end{cases}$$

and $Kf(x) = \mu(x)^{-1}\sum\limits_{y\sim x} f(y)$. The random walk interpretation of K is as follows. Think of a particle whose current position at a (discrete) time $t \in \mathbb{N}$, is at $x \in V$ with some probability $\mathbf{p}(t)(\{x\}) = \mathbf{p}(t,x)$. At time $t+1$, the particle picks uniformly one of the neighboring sites and moves there. Hence the probability of the particle to be at a site x at time $t+1$ is

$$\mathbf{p}(t+1,x) = \sum_{y\sim x}\mathbf{p}(t,y)\mu(y)^{-1} = \mathbf{p}(t)K(x),$$

where the action of K on a measure $\mathbf{p}$ is defined naturally by $\mathbf{p}K(f) = \mathbf{p}(Kf)$. It follows immediately that the operator K is a self-adjoint contraction on $L^2(V, \mu)$ and the function

$$u(t, x) = \mu(x)^{-1}\mathbf{p}(t, x)$$

is a solution of discrete time discrete space heat equation

$$u(t + 1, \cdot) - u(t, \cdot) = \Delta_E u(t, \cdot).$$

In this context, the heat kernel $h(t, x, y)$ is obtained by setting

$$h(t, x, y) = u_x(t, y) = \mu(y)^{-1}\mathbf{p}_x(t, y), \quad \mathbf{p}_x(0, y) = \delta_x(y).$$

It is a symmetric function of x, y, and for any f with finite support on V

$$u(t, x) = \sum_{y\in V} h(t, x, y)f(y)\mu(y)$$

is a solution of the heat equation with the initial value f. Finally, by definition,

$$\mathbf{p}_x(t, y) = h(t, x, y)\mu(y)$$

is the probability that our particle is at y at (discrete) time t given that it started at x at time 0.

The idea of applying Sobolev type inequalities in this context was introduced by Varopoulos [101] and produced a remarkable breakthrough in the study of random walks on graphs and finitely generated groups. The book [105] gives a detailed treatment of many aspects of the resulting developments. The following theorem is the most basic result (see [15, 101, 103, 105]).

Theorem 5.3. *Fix $\nu > 0$. Let (V, E) be a connected graph with bounded degree as above. The following properties are equivalent.*

- $\forall\, f \in \mathcal{C}_c(V), \quad \|f\|_2^{(1+2/\nu)} \leqslant N\|\nabla f\|_2\|f\|_1^{2/\nu}.$
- $\forall\, t \in \mathbb{N},\ x, y \in V, \quad h(t, x, y) \leqslant C(1 + t)^{-\nu/2}.$

Example 5.1. A rather interesting family of examples is as follows. Assume that the graph (V, E) has no loops (i.e., is a tree) and there exists $\nu > 0$ such that $V(x, r) \simeq r^\nu$. Such a tree must have many leaves (vertices of degree 1). For examples of such trees see [6]. Applying Theorems 5.2 and 5.3, we obtain the estimate $h(t, x, y) \leqslant C(1 + t)^{-\nu/(1+\nu)}$. As is proved in [7], this estimate is optimal in the sense that

$$h(2t, x, x) \simeq (1 + t)^{-\nu/(\nu+1)}.$$

Much more generally, the following assertion similar to Theorem 4.4 holds as well.

Theorem 5.4 ([27]). *Let* (V, E) *be as above. Let* Φ *be a positive smooth decreasing function on* $[0, \infty)$ *satisfying condition* (D)*, and let* $\Theta = -\Phi' \circ \Phi^{-1}$*. The following properties are equivalent.*

- *There exists a constant* $c_1 \in (0, \infty)$ *such that*

$$\forall\, t \in \mathbb{N},\ x, y \in V,\ \ h(t, x, y) \leqslant \Phi(c_1 t).$$

- *The exists a constant* $C_1 \in (0, \infty)$ *such that for all* $f \mathcal{C}_c(V)$ *with* $\|f\|_1 \leqslant 1$

$$\Theta(\|f\|_2^2) \leqslant C \|\nabla f\|_2^2.$$

Example 5.2. A case of interest is when $\Phi(t) = ce^{-t^\gamma}$ for some $\gamma \in (0, 1)$. Then $-\Phi'(t) = ct^{\gamma-1}e^{-t^\gamma}$, $\Phi^{-1}(s) = (c+\log 1/s)^{1/\gamma}$, and $\Theta(s) = s(c+\log 1/s)^{1-1/\gamma}$.

5.4 Cayley graphs

A *Cayley graph* is a graph (V, E) as above, where $V = G$ is a finitely generated group equipped with a finite generating set S and $(x, y) \in V \times V$ is in E if and only if $y = xs$ with $s \in S \cup S^{-1}$. Hence one can assume that S is symmetric, i.e., $S = S^{-1}$. These graphs are regular of degree $D = \#S$, and thus the measure μ used earlier is just a counting measure. Denote by e the identity element in G.

The random walk on a Cayley graph can be described as follows. Let $\xi_1, \xi_2, \dots$ be independent uniform picks in the finite symmetric generating set S. Then for $t \in \mathbb{N}$ and $x, y \in G$, $\mathbf{p}_x(t, y)$ is the probability that the product $X_t = x\xi_1 \cdots \xi_t$ is equal to y. It is oblivious that $\mathbf{p}_x(t, y) = \mathbf{p}_e(t, x^{-1}y)$ (left-invariance). For general finitely generated groups the study of such random walks originated in H. Kesten's thesis. Later, Kesten considered the natural question of when such a random walk is recurrent. Recall that recurrence here means that, with probability 1, the walk returns infinitely often to its starting point. A walk that is not recurrent is called transient and has the property that, with positive probability, it never returns to it starting point. By a celebrated result of Polya, the random walk on the integer lattices $\mathbb{Z}^n$ is recurrent if $n = 0, 1, 2$ and is transient otherwise. One of Kesten's questions about the recurrence of random walks can be formulated as follows: What are the groups that admit recurrent random walks (with generating support). For a long time, the conjectural answer known as Kesten's conjecture was that the only groups that admit recurrent random walks are the finite extensions of $\mathbb{Z}^n$, $n = 0, 1, 2$ (i.e., those groups that contain $\{0\}$ or $\mathbb{Z}$ or $\mathbb{Z}^2$ with finite index).

A basic result around this question (see, for example, [105]) is that recurrence is equivalent to

$$\sum_{t=1}^{\infty} \mathbf{p}_e(t,e) = \infty.$$

Indeed, $\sum_{t=1}^{\infty} \mathbf{p}_x(t,y)$ can be understood as the mean number of returns to y starting from x. Thus, the question is really a question about the behavior of the associated heat kernel $h(t,x,x)$.

Theorem 5.5 ([31]). *Fix $p \in [1,\infty]$. Let (V,E) be the Cayley graph associated to a finitely generated group G equipped with a finite symmetric generating set S. Then the pseudo-Poincaré inequality*

$$\|f - f_r\|_p \leqslant Cr\|\nabla f\|_p \tag{5.2}$$

holds, as well as the Poncaré type inequality

$$\sum_B |f - f_B|^p \leqslant Cr^p \frac{V(2r)}{V(r)} \sum_{2B} |\nabla f|^p, \tag{5.3}$$

where $B = B(e,r)$, $2B = B(e,2r)$, $V(r) = \#B(e,r)$, and f_B is the average of f over B.

Remark 5.2. The paper [31] treats mostly the case $p = 1$ (and the case $p = 2$, briefly, towards the end), partly because the other cases are obvious variations on the same argument. The inequality (5.2) with $p = 1$ is contained in [31, p. 296]. The inequality (5.3) with $p = 1$ and $p = 2$ is contained in [31, pp. 308–310] because, on a Cayley graph and under an invariant choice of paths, the constants $K(x,n)$ and $K_2(x,n)$ appearing in [31] can be of order $nV(2n)/V(n)$ and $n^2V(2n)/V(n)$ respectively. Below, we give a complete proof of the case $p = 2$, emphasizing the great similarity between these two inequalities.

Proof. We treat the case $p = 2$ (other cases are similar except for $p = \infty$ which is trivial and has little content). The crucial observation is that for any set $A \subset G$

$$\sum_{x\in A} \sum_{y\in B(e,s)\cap x^{-1}A} |f(xy) - f(x)|^2 \leqslant (\#S)s^2V(s) \sum_{A_{s/2}} |\nabla f|^2.$$

Here, $A_\tau = \{z \in G : \rho(z,A) \leqslant \tau\}$. To prove this inequality, for each y denote by γ_y a fixed path of minimal length from e to y and use the Cauchy–Schwarz inequality to get

$$|f(xy) - f(x)|^2 \leqslant (\#S)|y| \sum_{z\in\gamma_y} |\nabla f|(xz)^2,$$

where $|y| = \rho(e,y)$ is the graph distance between e and y (i.e., the length of y). Note that $\rho(xz,A) \leqslant \min\{\rho(e,z),\rho(z,y)\} \leqslant |y|/2 \leqslant s/2$. Moreover, and this is the crucial point of the argument, y being fixed, a given vertex $\xi = xz$ can appear for at most $|y|$ different points x. Hence

$$\sum_{x\in A}\sum_{y\in B(e,s)\cap x^{-1}A} |f(xy)-f(x)|^2 \leqslant (\#S)s^2V(s)\sum_{A_{s/2}}|\nabla f|^2.$$

Taking $A = G$, $r = s$ and dividing both sides by $V(r)$, we obtain the pseudo-Poincaré inequality $\|f_r - f\|_2 \leqslant (\#S)^{1/2}r\|\nabla f\|_2$. Taking $A = B(e,r)$, $s = 2r$ and dividing both sides by $V(r)$, we find

$$\sum_B |f-f_B|^2 \leqslant Cr^2\frac{V(2r)}{V(r)}\sum_{2B}|\nabla f|^2.$$

□

Theorem 5.6. *Let (V,E) be the Cayley graph associated to a finitely generated group G equipped with a finite symmetric generating set S. Assume that $V(r) \geqslant cr^\nu$, $r > 0$. Then there are constants N and C such that*

$$\forall\, f\in \mathcal{C}_c(G),\quad \|f\|_2^{(1+2/\nu)} \leqslant N\|\nabla f\|_2\|f\|_1^{2/\nu}$$

and

$$\forall\, t\in\mathbb{N},\ x,y\in G,\quad h(t,x,y)\leqslant C(1+t)^{-\nu/2}.$$

In addition, if the doubling volume property $V(2r)\leqslant DV(r)$ holds, then the scale invariant Poincaré inequalities

$$\forall\, B = B(x,r),\ \sum_B |f-f_B|^p \leqslant P_p r^p\sum_B |\nabla f|^p$$

are satisfied for all $p\in[1,\infty]$.

Proof. The first two properties are equivalent and follow from Theorems 5.2, 5.3, and 5.5. The last statement follows from Theorem 5.5 and a well-known, but somewhat subtle argument to get rid of the doubling of the ball over which one integrates the gradient (see, for example, [87, Sect. 5.3]). □

Remark 5.3. The statement that $V(r) \geqslant cr^\nu$ implies $h(t,x,x) \leqslant C_\varepsilon(1+t)^{-(\nu-\varepsilon)/2}$, $\varepsilon > 0$, was first proved by Varopoulos [104, 102] by different, but related methods.

Returning to Kesten's conjecture, let us observe that the above theorem implies that if a finitely generated group G satisfies $V(r)\geqslant cr^\nu$ with $\nu > 2$, then

$$\sum_{t=1}^{\infty} h(t,e,e) = \sum_{t=1}^{\infty}\mathbf{p}_e(t,e) < \infty, \tag{5.4}$$

i.e., the random walks on the Cayley graphs of G are transient (it is easy to see that different generating sets S always yield comparable growth functions

V). This means that a group carrying a recurrent random walk must have a volume growth function satisfying

$$\forall \varepsilon > 0, \ \liminf_{r\to\infty} r^{-(2+\varepsilon)} V(r) < \infty.$$

By the celebrated theorem of Gromov [54] (and its extension in [99]), the condition

$$\exists\, A > 0, \quad \liminf_{r\to\infty} r^{-A} V(r) < \infty \tag{5.5}$$

implies that G contains a nilpotent subgroup of finite index. Since a subgroup of finite index in G has volume growth comparable to that of G and, by a theorem due to Bass, nilpotent groups have volume growth of type r^ν for some integer ν (see, for example, [35]), we see that a group carrying a recurrent walk must contain a nilpotent subgroup of finite index and volume growth of type r^0 or r^1 or r^2. It is easy to check that this means that G is a finite extension of $\{0\}$ or $\mathbb{Z}$ or $\mathbb{Z}^2$, as desired.

Theorem 5.7 (solution of Kesten's conjecture, [104]). *If a finitely generated group G admits a finite symmetric generating set S such that the associated random walk is recurrent, then G is a finite extension of $\{0\}$ or $\mathbb{Z}$ or $\mathbb{Z}^2$.*

Remark 5.4. In a recent preprint [67], Kleiner gave a new proof of Gromov's theorem on groups of polynomial volume growth. His argument is quite significant since it avoids the use of the Montgomery–Zippin–Yamabe structure theory of locally compact groups (and of the solution of Hilbert fifth problem). It is also very significant from the viewpoint of the present paper and in relation to Theorem 5.7, as we will explain. The proof of Theorem 5.7 is based on two main results: the theorem of Gromov on groups of polynomial growth (albeit, only in the "small growth" case (5.4)) and Varopoulos' result that links volume growth to the decay of the probability of return of a random walk as expressed in Theorem 5.6. Until Kleiner's work on Gromov's theorem, these two corner stones of the proof of Theorem 5.7 appeared to be rather unrelated. However, it is remarkable that one of the key ingredients of Kleiner's proof is the Poincaré inequality (5.3). Recall that, in Theorem 2.4, we stated a result of Colding and Minicozzi to the effect that, on complete manifolds, the Poincaré inequality and the doubling property imply the *finite dimensionality* of the spaces of harmonic functions of polynomial growth. One of Kleiner's main ideas in [67] is to show that, because one has (5.3), the Colding–Minicozzi finite dimensionality results for harmonic functions of polynomial growth does hold for Cayley graphs under the (weak) polynomial volume growth hypothesis (5.5). This makes Theorem 5.5 central for each of the two main ingredients of the proof of Kesten's conjecture.

We complete with what can be seen as a generalization of Theorem 5.7 which involves Sobolev's inequalities. Because of the relation between the Sobolev inequality and the Nash inequality and the decay of the probability of return in Theorem 5.3, it is possible to formulate Theorem 5.7 in an equivalent way as follows: a Cayley graph always satisfies the inequality $\|f\|_2^{5/3} \leqslant N\|\nabla f\|_2\|f\|_1^{2/3}$ unless the group is a finite extension of a nilpotent group of growth degree at most 2. More generally, the following assertion holds.

Theorem 5.8. *Fix a positive integer ν. On the Cayley graph of a finitely generated group G, the Nash inequality*

$$\forall\, f \in \mathcal{C}_c(G), \quad \|f\|_2^{1+2/\nu} \leqslant N\|\nabla f\|_2\|f\|_1^{2/\nu}$$

always holds for some constant $N \in (0,\infty)$ (depending on ν, G and the Cayley graph structure) unless G is a finite extension of a nilpotent group of volume growth degree at most $\nu - 1$.

Similarly, the Sobolev inequality

$$\forall\, f \in \mathcal{C}_c(G), \quad \|f\|_{p\nu/(\nu-p)} \leqslant S\|\nabla f\|_p$$

always holds for all $\nu > p \geqslant 1$ and some constant $S \in (0,\infty)$ (depending on p, ν, G and the Cayley graph structure) unless G is a finite extension of a nilpotent group of volume growth degree at most $\nu - 1$.

Acknowledgement. Research partially supported by NSF (grant no. DMS 0603886).

References

1. Alexopoulos, G., Lohoué, N.: Sobolev inequalities and harmonic functions of polynomial growth. J. London Math. Soc. (2) **48**, no. 3, 452–464 (1993)
2. Aronson, D.G., Serrin, J.: Local behavior of solutions of quasilinear parabolic equations. Arch. Ration. Mech. Anal. **25**, 81–122 (1967)
3. Aronson, D.G., Serrin, J.: A maximum principle for nonlinear parabolic equations. Ann. Sc. Norm. Super. Pisa (3) **21**, 291–305 (1967)
4. Bakry, D., Coulhon, T., Ledoux, M., Saloff-Coste, L.: Sobolev inequalities in disguise. Indiana Univ. Math. J. **44**, no. 4, 1033–1074 (1995)
5. Bakry, D., Ledoux, M.: Sobolev inequalities and Myers's diameter theorem for an abstract Markov generator. Duke Math. J. **85**, no. 1, 253–270 (1996)
6. Barlow, M., Coulhon, T., Grigor'yan, A.: Manifolds and graphs with slow heat kernel decay. Invent. Math. **144**, no. 3, 609–649 (2001)
7. Barlow, M.T., Coulhon, T., Kumagai, T.: Characterization of sub-Gaussian heat kernel estimates on strongly recurrent graphs. Commun. Pure Appl. Math. **58**, no. 12, 1642–1677 (2005)
8. Bendikov, A., Coulhon, T., Saloff-Coste, L.: Ultracontractivity and embedding into L^∞. Math. Ann. **337**, no. 4, 817–853 (2007)
9. Bendikov, A., Maheux, P.: Nash type inequalities for fractional powers of nonnegative self-adjoint operators. Trans. Am. Math. Soc. **359**, no. 7, 3085–3097 (electronic) (2007)

10. Biroli, M., Mosco, U.: A Saint-Venant type principle for Dirichlet forms on discontinuous media. Ann. Mat. Pura Appl. (4) **169**, 125–181 (1995)
11. Biroli, M., Mosco, U.: Sobolev and isoperimetric inequalities for Dirichlet forms on homogeneous spaces. Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl. **6**, no. 1, 37–44 (1995)
12. Biroli, M., Mosco, U.: Sobolev inequalities on homogeneous spaces. Potential Anal. **4**, no. 4, 311–324 (1995)
13. Bougerol, P., Élie, L.: Existence of positive harmonic functions on groups and on covering manifolds. Ann. Inst. H. Poincaré Probab. Statist. **31**, no. 1, 59–80 (1995)
14. Buser, P.: A note on the isoperimetric constant. Ann. Sci. École Norm. Sup. (4) **15**, no. 2, 213–230 (1982)
15. Carlen, E.A., Kusuoka, S., Stroock, D.W.: Upper bounds for symmetric Markov transition functions. Ann. Inst. H. Poincaré Probab. Statist. **23**, no. 2, suppl., 245–287 (1987)
16. Carron, G.: Formes harmoniques L^2 sur les variétés non-compactes. Rend. Mat. Appl. (7) **21**, no. 1–4, 87–119 (2001)
17. Carron, G.: Inégalités isopérimétriques de Faber-Krahn et conséquences. In: Actes de la Table Ronde de Géométrie Différentielle (Luminy, 1992) Sémin. Congr. Vol. 1, pp. 205–232. Soc. Math. France, Paris (1996)
18. Carron, G.: Une suite exacte en L^2-cohomologie. Duke Math. J. **95**, no. 2, 343–372 (1998)
19. Carron, G.: L^2-cohomologie et inégalités de Sobolev. Math. Ann. **314**, no. 4, 613–639 (1999)
20. Chavel, I.: Eigenvalues in Riemannian geometry Academic Press Inc., Orlando, FL (1984)
21. Chavel, I.: Riemannian Geometry–a Modern Introduction. Cambridge Univ. Press, Cambridge (1993)
22. Cheeger, J., Gromov, M., Taylor, M.: Finite propagation speed, kernel estimates for functions of the Laplace operator, and the geometry of complete Riemannian manifolds. J. Differ. Geom. **17**, no. 1, 15–53 (1982)
23. Cheng, S.Y., Li, P., Yau, Sh.T.: On the upper estimate of the heat kernel of a complete Riemannian manifold. Am. J. Math. **103**, no. 5, 1021–1063 (1981)
24. Colding, T.H., Minicozzi, W.P.II: Harmonic functions on manifolds. Ann. Math. (2) **146**, no. 3, 725–747 (1997)
25. Colding, T.H., Minicozzi, W.P.II: Weyl type bounds for harmonic functions. Invent. Math. **131**, no. 2, 257–298 (1998)
26. Coulhon, T.: Dimension à l'infini d'un semi-groupe analytique. Bull. Sci. Math. **114**, no. 4, 485–500 (1990)
27. Coulhon, T.: Ultracontractivity and Nash type inequalities. J. Funct. Anal. **141**, no. 2, 510–539 (1996)
28. Coulhon, T.: Analysis on infinite graphs with regular volume growth. In: Random Walks and Discrete Potential Theory (Cortona, 1997) Sympos. Math. Vol. XXXIX, pp. 165–187. Cambridge Univ. Press, Cambridge (1999)
29. Coulhon, T., Grigor'yan, A.: Pointwise estimates for transition probabilities of random walks on infinite graphs. In: Fractals in Graz 2001, pp. 119–134. Birkhäuser, Basel (2003)
30. Coulhon, T., Grigor'yan, A., Zucca, F.: The discrete integral maximum principle and its applications. Tohoku Math. J. (2) **57**, no. 4, 559–587 (2005)
31. Coulhon, T., Saloff-Coste, L.: Isopérimétrie pour les groupes et les variétés. Rev. Mat. Iberoam. **9**, no. 2, 293–314 (1993)
32. Coulhon, T., Saloff-Coste, L.: Variétés riemanniennes isométriques à l'infini. Rev. Mat. Iberoam. **11**, no. 3, 687–726 (1995)

33. Davies, E.B.: Heat Kernels and Spectral Theory. Cambridge Univ. Press, Cambridge (1989)
34. DeGiorgi, E.: Sulla differenziabilità e l'analiticità delle estremali degli integrali multipli regolari. Mem. Accad. Sci. Torino. Cl. Sci. Fis. Mat. Nat. (3) **3**, 25–43 (1957)
35. de la Harpe, P.: Topics in Geometric Group Theory. University of Chicago Press, Chicago, IL (2000).
36. Deny, J.: Méthodes hilbertiennes en théorie du potentiel. In: Potential Theory (C.I.M.E., I Ciclo, Stresa, 1969), pp. 121–201. Edizioni Cremonese, Rome (1970)
37. Duong, T.X., Ouhabaz, ElM., Sikora, A.: Plancherel type estimates and sharp spectral multipliers. J. Funct. Anal. **196**, no. 2, 443–485 (2002)
38. Duong, T.X., Ouhabaz, ElM., Sikora, A.: Spectral multipliers for self-adjoint operators. In: Geometric Analysis and Applications (Canberra (2000)),pp. 56–66. Austral. Nat. Univ., Canberra (2001)
39. Franchi, B., Gutiérrez, C.E., Wheeden, R.L.: Weighted Sobolev–Poincaré inequalities for Grushin type operators. Commun. Partial Differ. Equ. **19**, no. 3-4, 523–604 (1994)
40. Fukushima, M., Ōshima, Yō., Takeda, M.: Dirichlet Forms and Symmetric Markov Processes. Walter de Gruyter, Berlin (1994)
41. Gallot, S., Hulin, D., Lafontaine, J.: Riemannian Geometry. Springer-Verlag, Berlin (1990)
42. Gilbarg, D., Trudinger, N.S.: Elliptic Partial Differential Equations of Second Order. Springer-Verlag, Berlin (2001)
43. Grigor'yan, A.: The heat equation on noncompact Riemannian manifolds (Russian). Mat. Sb. **182**, no. 1, 55–87 (1991); English transl.: Math. USSR, Sb. **72**, no. 1, 47–77 (1992)
44. Grigor'yan, A.: Heat kernel upper bounds on a complete non-compact manifold. Rev. Mat. Iberoam. **10**, no. 2, 395–452 (1994)
45. Grigor'yan, A.: Analytic and geometric background of recurrence and non-explosion of the Brownian motion on Riemannian manifolds. Bull. Am. Math. Soc. (N.S.) **36**, no. 2, 135–249 (1999)
46. Grigor'yan, A.: Estimates of heat kernels on Riemannian manifolds. In: Spectral Theory and Geometry (Edinburgh (1998)), pp. 140–225. Cambridge Univ. Press, Cambridge (1999)
47. Grigor'yan, A.: Heat kernels on manifolds, graphs and fractals. In: European Congress of Mathematics, Vol. I (Barcelona (2000)), pp. 393–406. Birkhäuser, Basel (2001)
48. Grigor'yan, A.: Heat kernels on weighted manifolds and applications. In: The Ubiquitous Heat Kernel, pp. 93–191. Am. Math. Soc., Providence, RI (2006)
49. Grigor'yan, A., Saloff-Coste, L.: Heat kernel on connected sums of Riemannian manifolds. Math. Res. Lett. **6**, no. 3-4, 307–321 (1999)
50. Grigor'yan, A., Saloff-Coste, L.: Stability results for Harnack inequalities. Ann. Inst. Fourier (Grenoble) **55**, no. 3, 825–890 (2005)
51. Grigor'yan, A., Saloff-Coste, L.: Heat kernel on manifolds with ends. (2007)
52. Grigor'yan, A., Telcs. A.: Sub-Gaussian estimates of heat kernels on infinite graphs. Duke Math. J. **109**, no. 3, 451–510 (2001)
53. Grigor'yan, A., Telcs. A.: Harnack inequalities and sub-Gaussian estimates for random walks. Math. Ann. **324**, no. 3, 521–556 (2002)
54. Gromov, M.: Groups of polynomial growth and expanding maps. Inst. Hautes Études Sci. Publ. Math. **53**, 53–73 (1981)
55. Guivarc'h, Y.: Croissance polynomiale et périodes des fonctions harmoniques. Bull. Soc. Math. France **101**, 333–379 (1973)
56. Gutiérrez, C.E., Wheeden, R.: Mean value and Harnack inequalities for degenerate parabolic equations. Colloq. Math. **60/61**, no. 1, 157–194 (1990)

57. Gutiérrez, C.E., Wheeden, R.: Harnack's inequality for degenerate parabolic equations. Commun. Partial Differ. Equ. **16**, no. 4-5, 745–770 (1991)
58. Gutiérrez, C.E., Wheeden, R.: Bounds for the fundamental solution of degenerate parabolic equations. Commun. Partial Differ. Equ. **17**, no. 7-8, 1287–1307 (1992)
59. Hajłasz, P., Koskela, P.: Sobolev met Poincaré. Mem. Am. Math. Soc. no. 688 (2000)
60. Hebey, E.: Nonlinear Analysis on Manifolds: Sobolev Spaces and Inequalities. Courant Lect. Notes **5**. Am. Math. Soc., Providence, RI (1999)
61. Hebisch, W., Saloff-Coste, L.: On the relation between elliptic and parabolic Harnack inequalities. Ann. Inst. Fourier (Grenoble) **51**, no. 5, 1437–1481 (2001)
62. Heinonen, J.: Lectures on analysis on metric spaces. Universitext. Springer-Verlag, New York (2001)
63. Heinonen, J.: Nonsmooth calculus. Bull. Am. Math. Soc. (N.S.) **44**, no. 2, 163–232 (electronic), (2007)
64. Kanai, M.: Rough isometries, and combinatorial approximations of geometries of noncompact Riemannian manifolds. J. Math. Soc. Japan **37**, no. 3, 391–413 (1985)
65. Kanai, M.: Analytic inequalities, and rough isometries between noncompact Riemannian manifolds. In: Curvature and Topology of Riemannian Manifolds (Katata, (1985), pp. 122–137. Springer, Berlin (1986)
66. Kanai, M.: Rough isometries and the parabolicity of Riemannian manifolds. J. Math. Soc. Japan **38**, no. 2, 227–238 (1986)
67. Kleiner, B.: A new proof of Gromov's theorem on groups of polynomial growth. (2007) arXiv:0710.4593v4
68. Kuchment, P., Pinchover, Y.: Liouville theorems and spectral edge behavior on abelian coverings of compact manifolds. Trans. Am. Math. Soc. **359**, no. 12, 5777–5815 (electronic) (2007)
69. Ladyzhenskaya, O.A., Solonnikov, V.A., Uraltseva, N.N.: Linear and Quasilinear Equations of Parabolic Type Am. Math. Soc., Providence, RI (1968)
70. Levin, D., Solomyak, M.: The Rozenblum–Lieb–Cwikel inequality for Markov generators. J. Anal. Math. **71**, 173–193 (1997)
71. Li, P.: Harmonic functions on complete Riemannian manifolds. In: Tsing Hua Lectures on Geometry and Analysis (Hsinchu, (1990-1991), pp. 265–268. Int. Press, Cambridge, MA (1997)
72. Li, P.: Harmonic sections of polynomial growth. Math. Res. Lett. **4**, no. 1, 35–44 (1997)
73. Li, P., Yau, Sh.T.: On the Schrödinger equation and the eigenvalue problem. Commun. Math. Phys. **88**, no. 3, 309–318 (1983)
74. Li, P., Yau, Sh.T.: On the parabolic kernel of the Schrödinger operator. Acta Math. **156**, no. 3-4, 153–201 (1986)
75. Maz'ya, V.G.: Sobolev Spaces. Springer-Verlag, Berlin–Tokyo (1985)
76. Morrey, Ch.B., Jr.: Multiple Integrals in the Calculus of Variations. Springer-Verlag, New York (1966)
77. Moser, J.: On Harnack's theorem for elliptic differential equations. Commun. Pure Appl. Math. **14**, 577–591 (1961)
78. Moser, J.: A Harnack inequality for parabolic differential equations. Commun. Pure Appl. Math. **17**, 101–134 (1964)
79. Moser, J.: Correction to: "A Harnack inequality for parabolic differential equations." Commun. Pure Appl. Math. **20**, 231–236 (1967)
80. Moser, J.: On a pointwise estimate for parabolic differential equations. Commun. Pure Appl. Math. **24**, 727–740 (1971)
81. Nash, J.: Continuity of solutions of parabolic and elliptic equations. Am. J. Math. **80**, 931–954 (1958)

82. Pivarski, M., Saloff-Coste, L.: Small time heat kernel behavior on riemannian complexe.
83. Porper, F.O., Eidel'man, S.D.: Two-sided estimates of fundamental solutions of second-order parabolic equations, and some applications (Russian). Uspekhi Mat. Nauk **39**, no. 3, 107–156 (1984); English transl.: Russ. Math. Surv. **39**, no. 3, 119–178 (1984)
84. Rozenblum, G., Solomyak, M.: The Cwikel–Lieb–Rozenblum estimates for generators of positive semigroups and semigroups dominated by positive semigroups (Russian). Algebra Anal. **9**, no. 6, 214–236 (1997); English transl.: St. Petersbg. Math. J. **9**, no. 6, 1195–1211 (1998)
85. Saloff-Coste, L.: A note on Poincaré, Sobolev, and Harnack inequalities. Int. Math. Res. Not. **2**, 27–38 (1992)
86. Saloff-Coste, L.: Parabolic Harnack inequality for divergence form second order differential operators. Potential Anal. **4**, no. 4, 429–467 (1995)
87. Saloff-Coste, L.: Aspects of Sobolev-Type Inequalities. Cambridge Univ. Press, Cambridge (2002)
88. Saloff-Coste, L.: Analysis on Riemannian co-compact covers. In: Surveys in Differential Geometry. Vol. 9, pp. 351–384. Int. Press, Somerville, MA (2004)
89. Semmes, S.: Some Novel Types of Fractal Geometry. Clarendon Press, Oxford (2001)
90. Serrin, J.: Local behavior of solutions of quasi-linear equations. Acta Math. **111**, 247–302 (1964)
91. Sikora, A.: Sharp pointwise estimates on heat kernels. Q. J. Math., Oxf. II. Ser. **47**, no. 187, 371–382 (1996)
92. Sikora, A.: Riesz transform, Gaussian bounds and the method of wave equation. Math. Z. **247**, no. 3, 643–662 (2004)
93. Sobolev, S.L.: On a theorem of functional analysis (Russian). Mat. Sb. **46**, 471–497 (1938); English transl.: Am. Math. Soc., Transl., II. Ser. **34**, 39–68 (1963)
94. Sturm, K.T.: Analysis on local Dirichlet spaces. III. The parabolic Harnack inequality. J. Math. Pures Appl. (9) **75**, no. 3, 273–297 (1996)
95. Sturm, K.T.: Analysis on local Dirichlet spaces. I. Recurrence, conservativeness and L^p-Liouville properties. J. Reine Angew. Math. **456**, 173–196 (1994)
96. Sturm, K.T.: Analysis on local Dirichlet spaces. II. Upper Gaussian estimates for the fundamental solutions of parabolic equations. Osaka J. Math. **32**, no. 2, 275–312 (1995)
97. Sturm, K.T.: The geometric aspect of Dirichlet forms. In: New Directions in Dirichlet Forms, pp. 233–277. Am. Math. Soc., Providence, RI (1998)
98. Telcs, A.: The Art of Random Walks. Springer-Verlag, Berlin (2006)
99. van den Dries, L., Wilkie, A.J.: Gromov's theorem on groups of polynomial growth and elementary logic. J. Algebra **89**, no. 2, 349–374 (1984)
100. Varopoulos, N.Th.: Hardy-Littlewood theory for semigroups. J. Funct. Anal. **63**, no. 2, 240–260 (1985)
101. Varopoulos, N.Th.: Isoperimetric inequalities and Markov chains. J. Funct. Anal. **63**, no. 2, 215–239 (1985)
102. Varopoulos, N.Th.: Convolution powers on locally compact groups. Bull. Sci. Math. (2) **111**, no. 4, 333–342 (1987)
103. Varopoulos, N.Th., Saloff-Coste, L., Coulhon, T.: Analysis and Geometry on Groups. Cambridge Univ. Press, Cambridge (1992)
104. Varopoulos, N.Th.: Théorie du potentiel sur des groupes et des variétés. C. R. Acad. Sci., Paris Sér. I Math. **302**, no. 6, 203–205 (1986)
105. Woess, W.: Random Walks on Infinite Graphs and Groups. Cambridge Univ. Press, Cambridge (2000)

A Universality Property of Sobolev Spaces in Metric Measure Spaces

Nageswari Shanmugalingam

Abstract Current research on analysis in metric measure spaces has used alternative notions of Sobolev functions on metric measure spaces. We show that, under some mild geometric assumptions on the metric measure space, all these notions give the same class of functions.

1 Introduction

Motivated by the goal of understanding geometry of singular (non-Riemannian) spaces, the study of analysis on metric measure spaces in recent years has seen significant development. The metric spaces obtained as Gromov–Hausdorff limits of Riemannian manifolds may be non-Riemannian in nature, and the study of smooth analysis may be insufficient in understanding the geometry of such singular spaces. Furthermore, analysis on fractal type sets would be incomplete if one is restricted to studying only smooth functions. Thus, the current research on metric spaces utilizes notions of first order Sobolev space and an appropriate formulation of Poincaré inequalities. In metric spaces where this machinery is available, some of the current research considers variational problems, partial differential equations, potential theory, and other aspects of analysis.

The theory of Sobolev spaces of functions in Euclidean domains is now well understood (see, for example, a standard reference [20]). Currently, there exist various approaches to defining Sobolev type spaces of functions in metric measure spaces. The geometric construction using the notion of upper gradients from [11] was first studied independently in [5] and [22]; this construction yields the classical first order Sobolev space $W^{1,p}(\Omega)$ for $1 \leqslant p < \infty$ when Ω is an Euclidean domain or a Riemannian manifold, and under additional

Nageswari Shanmugalingam
University of Cincinnati, Cincinnati, OH 45221-0025, USA, e-mail: nages@math.uc.edu

V. Maz'ya (ed.), *Sobolev Spaces in Mathematics I, International Mathematical Series.*
doi: 10.1007/978-0-387-85648-3,

assumptions such as Poincaré inequality, allows for a comprehensive first-order theory of variational methods in potential theory and partial differential equations [16]. These Sobolev type spaces are called Newtonian spaces. A second notion of Sobolev type spaces, due to Hajłasz [8], gives an almost everywhere pointwise control of Sobolev type functions in terms of a function that plays the role of a maximal function of the gradient. A third avenue of development of first order calculus in metric measure spaces is due to Korevaar and Schoen [17]. The approaches of [8, 9, 17] yield notions of gradients that are not local; i.e., from the fact that a Sobolev function is constant on a Borel set one cannot conclude that the corresponding notion of gradient is zero almost everywhere on that set. However, the Newtonian space, where the notion of gradient is the concept of upper gradient, has this locality property. Another concept of Sobolev type space is obtained via a probabilistic approach using a symmetric bilinear form called a Dirichlet form. This concept yields a version of the Sobolev space $W^{1,2}$ which is naturally endowed with a Hilbert-space structure. The standard reference for this approach is [6], and has been developed in the setting of fractals in [1, 14, 15]. The literature on Dirichlet forms on self-similar sets is rapidly expanding, and we cannot hope to provide a bibliography which does justice to the field here. Concurrently, the abstract theory of Dirichlet forms on general metric spaces has been studied, for example, in [21, 3, 4, 25, 26, 27, 12] and the references therein.

By examining the analytic properties satisfied by the above various notions of Sobolev spaces in metric measure spaces, Gol'dshtein and Troyanov [7] developed the foundational theory of axiomatic Sobolev spaces. Of the different approaches to the notion of Sobolev functions in metric measure spaces, in this paper we focus on the Newtonian spaces of [22], the axiomatic Sobolev spaces of Gol'dshtein and Troyanov [7], and the Dirichlet forms developed in [6, 12]. Under suitable conditions such as strong locality of the Sobolev type spaces, doubling property of the measure, and the satisfaction of a Poincaré inequality, we show that these three spaces are isomorphic. For an expository description of connections between the Hajłasz–Sobolev spaces, the Sobolev spaces of Korevaar–Schoen, and the Newtonian spaces we refer the interested reader to [24].

This paper is structured as follows. Section 2 provides the background material and basic concepts used in the main theorems. There are two main results in this paper. The first, demonstrating that under certain conditions that are standard in analysis of Dirichlet forms, the domain of such a Dirichlet form, called the Dirichlet domain, coincides with the Newtonian space $N^{1,2}(X)$. Section 3 begins with a description of Dirichlet forms, and then provides a proof of Theorem 3.7. The final section gives a definition of the axiomatic Sobolev spaces of Gol'dshtein–Troyanov and proves that under strong locality assumptions these spaces coincide with the Newtonian spaces (Theorem 4.3).

2 Background

In this paper, (X, d, μ) denotes a set X equipped with a metric d and a Borel measure μ such that nonempty open sets have positive measure and bounded sets have finite measure. We also assume that X is complete and μ is doubling, i.e., there is a constant $C > 0$ such that whenever $B(x,r) = \{y \in X \,:\, d(x,y) < r\}$ is a ball in X,

$$\mu(B(x,2r)) \leqslant C\,\mu(B(x,r)).$$

It is easy to see that such a metric space X is proper, i.e., closed and bounded subsets of X are compact.

Heinonen and Koskela [11] proposed the following substitute for gradients in metric spaces.

Definition 2.1. Given a function $u : X \to [-\infty, \infty]$, we say that a nonnegative Borel measurable function ρ on X is an *upper gradient* of u if whenever γ is a compact rectifiable path in X,

$$|u(x) - u(y)| \leqslant \int\limits_{\gamma} \rho\, ds,$$

where x, y denote the endpoints of γ. If $u(x)$ or $u(y)$ is not finite, the above inequality is interpreted to mean that

$$\int\limits_{\gamma} \rho\, ds = \infty.$$

We say that ρ is a *p-weak upper gradient* of u for some $1 \leqslant p < \infty$ if there is a nonnegative Borel measurable function $\rho_0 \in L^p(X)$ such that whenever γ does not satisfy the above inequality,

$$\int\limits_{\gamma} \rho_0\, ds = \infty.$$

The uniform convexity of $L^p(X)$ when $1 < p < \infty$ implies that if u has a p-weak upper gradient $\rho \in L^p(X)$, then u has a *minimal p-weak upper gradient* $\rho_u \in L^p(X)$ (see [22]). If u is constant on a Borel set A and $\rho \in L^p(X)$ is a p-weak upper gradient of u, then $\rho\,\chi_{X\setminus A}$ is also a p-weak upper gradient of u; for a proof of this fact we refer the interested reader to [23]. Such a property is called the *strong locality* property.

For functions $u \in L^p(X)$ with upper gradients $\rho \in L^p(X)$ we define

$$\|u\|_{N^{1,p}(X)} = \|u\|_{L^p(X)} + \inf_{\rho} \|\rho\|_{L^p(X)},$$

where the infimum is taken over all upper gradients of u. The function space $\widehat{N}^{1,p}(X)$ is the collection of all functions u on X such that $u \in L^p(X)$ and u has an upper gradient in $L^p(X)$. It is easy to see that $\widehat{N}^{1,p}(X)$ is a vector space with a lattice structure. The Newtonian space $N^{1,p}(X) = \widehat{N}^{1,p}(X)/\sim$, where the equivalence relation $\sim$ is given by the rule $u \sim v$ if $\|u-v\|_{N^{1,p}(X)} = 0$.

Observe that if X is a domain in an Euclidean space, equipped with the Euclidean metric and the standard Lebesgue measure, then $N^{1,p}(X) = W^{1,p}(X)$ (see, for example, [22]).

Definition 2.2. We say that X supports a $(1,p)$-Poincaré inequality for Newtonian functions if there are constants $C > 0$ and $\tau \geqslant 1$ such that for all $u \in N^{1,p}(X)$, upper gradients ρ of u, and balls $B(x,r) \subset X$

$$\fint_{B(x,r)} |u - u_{B(x,r)}|\, d\mu \leqslant Cr\Big(\fint_{B(x,\tau r)} \rho^p \, d\mu\Big)^{1/p}. \tag{2.1}$$

Here,

$$\fint_{B(x,r)} u\, d\mu = u_{B(x,r)} = \mu(B(x,r))^{-1} \int_{B(x,r)} u\, d\mu.$$

The following deep result, due to Cheeger [5], was proved for a Sobolev type space $H^{1,p}(X)$ that appears on the surface to be different from the Newtonian space. However, when $1 < p < \infty$, it can be seen that $H^{1,p}(X) = N^{1,p}(X)$; we direct the interested reader to [22].

Proposition 2.3 ([5, Sect. 6])**.** *If the measure on X is doubling and supports a $(1,p)$-Poincaré inequality for Newtonian functions with $p > 1$, and u is a Lipschitz function on X, then*

$$\operatorname{Lip} u(x) := \limsup_{y \to x} \frac{|u(y) - u(x)|}{d(y,x)}$$

is the unique (up to sets of μ-measure zero) minimal p-weak upper gradient of u.

The following result, due to Keith, is a key tool in proving that the three candidate Sobolev spaces are isomorphic under certain conditions.

Proposition 2.4 ([13, Theorem 2])**.** *Let X be a complete metric measure space equipped with a doubling measure, and let $p \geqslant 1$. The following are equivalent.*

1. *X supports a $(1,p)$-Poincaré inequality for all measurable functions.*
2. *X supports a $(1,p)$-Poincaré inequality for all compactly supported Lipschitz functions with compactly supported continuous upper gradients.*
3. *X supports a $(1,p)$-Poincaré inequality for all Newtonian functions.*

In proving one of the two focal theorems of this paper (Theorem 3.7), we need the auxiliary candidate space called the Korevaar–Schoen space, first given in [17] and also studied in [18]. The Korevaar–Schoen space $KS^{1,2}(X)$ is defined to be the collection of all functions $f \in L^2(X)$ for which the approximating energies

$$e_\varepsilon^2(x;f) = \fint_{B(x,\varepsilon)} \frac{d_Y(f(x),f(y))^2}{\varepsilon^2}\, d\mu_X(y)$$

converge to a finite energy

$$E_{KS}(f) := \sup_{\text{balls } B} \limsup_{\varepsilon\to 0} \int_B e_\varepsilon^2(x;f)\, d\mu_X(x) < \infty.$$

3 Dirichlet Forms and $N^{1,2}(X)$

We follow the spirit of [12] in defining Dirichlet forms. Recall that $L^2(X) = L^2(X,\mu)$ is a Hilbert space endowed with the inner product

$$(u,v) := \int_X uv\, d\mu.$$

Definition 3.1. A symmetric bilinear form $\mathcal{E} : L^2(X)\times L^2(X) \to \mathbb{R}^+ \cup \{\infty\}$ is a *Dirichlet form* if the following conditions are satisfied:

1. *Quadratic contraction property* If $\mathcal{E}(u,u) < \infty$ and $\varphi : \mathbb{R} \to \mathbb{R}$ is L-Lipschitz with $\varphi(0) = 0$, then
$$\mathcal{E}(\varphi\circ u, \varphi\circ u) \leqslant L^2\mathcal{E}(u,u). \tag{3.1}$$
2. *Closedness* If $\{u_n\}$ is a Cauchy sequence in $L^2(X)$ such that $\mathcal{E}(u_n) < \infty$ for all n and $\mathcal{E}(u_n-u_m, u_n-u_m) \to 0$ as $n,m \to \infty$, then $\mathcal{E}(u_n-u, u_n-u) \to 0$ as $n \to \infty$. Here, u denotes the L^2-limit of the Cauchy sequence $\{u_n\}$.
3. *Density condition* The domain $D(\mathcal{E})$ of the bilinear form, consisting of all functions u in $L^2(X)$ for which $\mathcal{E}(u,u) < \infty$, is dense in $L^2(X)$ with respect to its norm and is dense in the space $C_0(X)$ of continuous functions with compact support with respect to the supremum norm.
4. For every $u,v \in L^2(X)$, we have the *Minkowski inequality*
$$\sqrt{\mathcal{E}(u+v,u+v)} \leqslant \sqrt{\mathcal{E}(u,u)} + \sqrt{\mathcal{E}(v,v)}. \tag{3.2}$$

The approach of [6] begins with a Markovian condition which is weaker than the quadratic contraction property. However, as much of the current research on Dirichlet forms also require them to satisfy the quadratic contraction property, we include it as part of the definition.

It can be seen by the quadratic contraction property that $D(\mathcal{E})$ is a vector space, and indeed is a normed space when equipped with the norm

$$\|u\|_{\mathcal{E}} := \|u\|_{L^2(X)} + \sqrt{\mathcal{E}(u,u)}.$$

The closability condition implies that if $\{u_n\}$ is a sequence which is Cauchy in this norm with L^2-limit u, then $\mathcal{E}(u,u) = \lim_{n\to\infty} \mathcal{E}(u_n,u_n)$ is finite. Moreover, as a consequence of the closedness condition and the Minkowski inequality, the normed space $(D(\mathcal{E}), \|\cdot\|_{\mathcal{E}})$ is a Hilbert space equipped with the norm $\|\cdot\|_{\mathcal{E}}$.

Using the fact that $L^2(X)$ is a Hilbert space, Beurling and Deny [2] construct a unique signed Radon measure-valued symmetric bilinear form associated with $\mathcal{E}$:

$$\eta : D(\mathcal{E}) \times D(\mathcal{E}) \to \mathcal{M}(X)$$

(where $\mathcal{M}(X)$ is the collection of all finite signed Radon measures on X) which satisfies

$$\mathcal{E}(u,v) = \int_X d\eta(u,v).$$

The measure $\eta(u,u)$ plays the role of the (square of the norm of the) pointwise derivative of u in the general setting, satisfying the quadratic contraction property

$$d\eta(\varphi\circ u, \varphi\circ u) \leqslant L^2 d\eta(u,u)$$

whenever φ is an L-Lipschitz function on $\mathbb{R}$, and for all Borel sets $E \subset X$

$$|\eta(u,v)(E)| \leqslant \sqrt{\eta(u,u)(E)\,\eta(v,v)(E)}. \tag{3.3}$$

If $\mathcal{E}$ is strongly local, then for all $u,v \in D(\mathcal{E})$ the singular (jump) part of $\eta(u,v)$ with respect to μ is zero.

For more discussion about Dirichlet forms, we refer the reader to [12, 2, 25, 26, 21] and the text [6].

Definition 3.2. We say that $\mathcal{E}$ is *local* if whenever $u,v \in D(\mathcal{E})$ such that the support of u and the support of v are disjoint compact sets, then $\mathcal{E}(u,v) = 0$. We say that $\mathcal{E}$ is *strongly local* if whenever $U \subset X$ is an open set and $u \in D(\mathcal{E})$ is constant on U, then for all $v \in D(\mathcal{E})$, $\eta(u,v)$ is supported in $X \setminus U$.

If X is a Euclidean domain, the form

$$\mathcal{E}(u,v) = \int_X \langle \nabla u, \nabla v\rangle \, dx$$

is a strongly local Dirichlet form in the above sense. Bilinear Dirichlet forms on certain fractals and more general self-similar sets were constructed by Barlow–Bass [1] and Kigami [14, 15]; these are not strongly local in general. If X is a metric measure space equipped with a doubling measure supporting a $(1, 2)$-Poincaré inequality for Newtonian functions, then by a deep Rademacher type theorem of [5], it follows that $N^{1,2}(X) = D(\mathcal{E})$ for some strongly local Dirichlet form on X. Sturm [27] proved that the Korevaar–Schoen space $KS^{1,2}(X)$ (a definition of this space is provided at the end of Sect. 2) is a Dirichlet domain corresponding to a local Dirichlet form.

In considering Dirichlet forms, traditionally the underlying space is only a Hausdorff topological space equipped with a Borel measure, and the Poincaré inequality for such a Dirichlet form $\mathcal{E}$ is typically expressed using the so-called *intrinsic metric* associated with $\mathcal{E}$.

A subspace Γ of $D(\mathcal{E}) \cap C_0(X)$ is a *μ-separating core* if it is dense in $C_0(X)$ with respect to the supremum norm, is dense in $D(\mathcal{E})$ with respect to the norm $\|\cdot\|_{\mathcal{E}}$, and has the following separation property: for every pair of distinct points x, y in X there is a function $\varphi \in \Gamma$ such that $\varphi(x) \neq \varphi(y)$ and $d\eta(\varphi, \varphi) \leqslant d\mu$. The intrinsic metric $d_{\mathcal{E}}$ is defined as follows:

$$d_{\mathcal{E}}(x, y) = \sup\{\varphi(x) - \varphi(y) \ : \ \varphi \in \Gamma, d\eta(\varphi, \varphi) \leqslant d\mu\}, \qquad x, y \in X. \quad (3.4)$$

Following [12, 3], we make the standing assumption that the topology induced by $d_{\mathcal{E}}$ coincides with the underlying topology on X and that the measure μ is doubling for balls in the new metric, i.e., there exists a constant $C \geqslant 1$ such that

$$\mu(2B_{\mathcal{E}}) \leqslant C\mu(B_{\mathcal{E}})$$

for every ball $B_{\mathcal{E}}$ in X. A change in the underlying metric results in a change in the corresponding Newtonian spaces. Hence from now on we also assume that $d_{\mathcal{E}} = d$ and every 1-Lipschitz function φ on X is in the μ-separating core Γ with $d\eta(\varphi, \varphi) \leqslant d\mu$. Given a Lipschitz function φ on X, let

$$\operatorname{Lip}(u, x) := \lim_{r \to 0^+} \sup_{y,z \in B(x,r), y \neq z} \frac{|\varphi(y) - \varphi(z)|}{d(y, z)}. \quad (3.5)$$

Standard measure theoretical arguments show that $\operatorname{Lip}(u, \cdot)$ is a bounded μ-measurable function on X.

Lemma 3.3. *Let $\mathcal{E}$ be a strongly local Dirichlet form such that every compactly supported 1-Lipschitz function φ on X is in $D(\mathcal{E})$ with $d\eta(\varphi, \varphi) \leqslant d\mu$. Then whenever u is a compactly supported Lipschitz function on X, $u \in D(\mathcal{E})$ and for μ-almost every $x \in X$,*

$$d\eta(u, u)(x) \leqslant \operatorname{Lip}(u, x)^2 \, d\mu(x).$$

Proof. Fix $x \in X$ and $r > 0$. Let

$$L_u(x;r) := \sup_{y,z\in B(x,r),y\neq z} \frac{|\varphi(y)-\varphi(z)|}{d(y,z)}.$$

Then $\operatorname{Lip}(u,x) = \lim_{r\to 0^+} L_u(x;r)$. Restricting u to the ball $B(x,r)$, it is easy to see that this restriction is $L_u(x;r)$-Lipschitz continuous on this ball. We can find a Lipschitz extension φ of the restriction of u to X (for example, via a McShane construction [10]) such that φ is $L_u(x;r)$-Lipschitz on X. By modifying φ outside $B(x,r)$ (by multiplying φ by a compactly supported Lipschitz function that is 1 on $B(x,r)$ if necessary), we can ensure that φ is compactly supported. Hence $L_u(x;r)^{-1}\varphi \in D(\mathcal{E})$ with $d\eta(L_u(x;r)^{-1}\varphi, L_u(x;r)^{-1}\varphi) \leqslant d\mu$ by hypothesis. Now, by the quadratic contraction property, $d\eta(\varphi,\varphi) \leqslant L_u(x;r)^2\, d\mu$.

Since $\mathcal{E}$ is strongly local, by (3.3) we have $d\eta(\varphi,\varphi) = d\eta(u,u)$ on $B(x,r)$. Hence $d\eta(u,u) \leqslant L_u(x;r)^2\, d\mu$ on $B(x,r)$. Now, letting $r\to 0$ completes the proof for $x\in X$ that are Lebesgue points for $\operatorname{Lip}(u,\cdot)$. □

Definition 3.4. We say that X supports a *weak* $(1,2)$-*Poincaré inequality* for the form $\mathcal{E}$ if there are constants $\tau \geqslant 1$ and $C > 0$ such that for all $u\in D(\mathcal{E})$ and balls $B(x,r)\subset X$

$$⨍_{B(x,r)} |u - u_{B(x,r)}|\, d\mu \leqslant Cr\sqrt{\frac{\eta(u,u)(B(x,\tau r))}{\mu(B(x,\tau r))}}. \tag{3.6}$$

Remark 3.5. The paper [19] studies metric space-valued Sobolev type spaces and Dirichlet domains, and demonstrates that if a strongly local Dirichlet form satisfies a Poincaré inequality and for all Lipschitz functions φ on X, $d\eta(\varphi,\varphi)(x) \approx \operatorname{Lip}(\varphi,x)\, d\mu(x)$, then the corresponding Dirichlet domain embeds isomorphically into the Korevaar–Schoen space $K^{1,2}(X)$ (see [19, Proposition 4.2]). The condition that $d\eta(\varphi,\varphi)(x) \approx \operatorname{Lip}(\varphi,x)\, d\mu(x)$ for all Lipschitz functions φ on X was used there only to ensure that the underlying metric used in constructing the Korevaar–Schoen space is bi-Lipschitz equivalent to the intrinsic metric of the Dirichlet form. We need this result of [19], but as we consider only the intrinsic metric on the space ($d_{\mathcal{E}} = d$), this condition on the measures η will not be needed by us. The main theorem of this section, Theorem 3.7, is an improvement of Proposition 4.2 of [19].

Recall that X is said to be quasiconvex if there is a constant $C > 0$ such that whenever $x,y\in X$ there is a curve γ connecting x to y with length no more than $Cd(x,y)$.

Lemma 3.6. *Let (X,d,μ) be a complete metric measure space such that μ is doubling and supports the following Poincaré inequality: for every Lipschitz function φ on X, whenever $B(x,r)$ is a ball in X,*

$$\fint_{B(x,r)} |u - u_{B(x,r)}|\, d\mu \leqslant Cr\Big(\fint_{B(x,\tau r)} \mathrm{Lip}(u,y)^2\, d\mu(y)\Big)^{1/2}.$$

Then X is quasiconvex.

Proof. The proof of this lemma is along the lines of a proof of Semmes on quasiconvexity (see [5, Appendix]). For $\varepsilon > 0$ and $x, y \in X$ we say that $x \sim_\varepsilon y$ if there is a finite collection of points $x_1, \ldots, x_k$ in X such that $d(x_1, x) < \varepsilon, d(x_k, y) < \varepsilon$, and for $j = 1, \ldots, k-1$, $d(x_j, x_{j+1}) < \varepsilon$. An easy topological argument shows that the equivalence relation $\sim_\varepsilon$ partitions X into pairwise disjoint closed sets. If U_1 and U_2 are such sets, then if both are nonempty, the distance between them is at least ε, and hence these sets are open as well. Hence the characteristic function $u = \chi_{U_1}$ of U_1 is $1/\varepsilon$-Lipschitz, with $\mathrm{Lip}(u, \cdot) = 0$. Since U_2 is nonempty and open, for sufficiently large balls B in X the integral

$$\fint_B |u - u_B|\, d\mu$$

is positive, thus violating the above version of Poincaré inequality. Hence every $x, y \in X$ have $x \sim_\varepsilon y$.

For $\varepsilon > 0$ and $x \in X$, let $\rho_\varepsilon : X \to \mathbb{R}$ be the function given by

$$\rho_\varepsilon(y) = \inf_{(x_1,\ldots,x_k)} d(x, x_1) + d(x_k, y) + \sum_{j=1}^{k-1} d(x_j, x_{j+1}).$$

If $y_1, y_2 \in X$ such that $d(y_1, y_2) < \varepsilon$, then

$$|\rho_\varepsilon(y_1) - \rho_\varepsilon(y_2)| \leqslant d(y_1, y_2),$$

i.e., $\mathrm{Lip}(\rho_\varepsilon, y) \leqslant 1$ on X. Hence an employment of the above version of Poincaré inequality together with the telescoping argument given in [5, Sect. 4] shows that $(\rho_\varepsilon)_\varepsilon$ is a bounded sequence of continuous functions in $L^1(B(x, R))$ for all $R > 0$. Since ρ_ε increases as ε decreases, $\lim_{\varepsilon\to\infty} \rho_\varepsilon$ exists, and by the above boundedness converges to a finite-valued function ρ_∞ on X. The final part of the argument of [5, Appendix] now completes the proof. □

The main theorem of this section is the following result.

Theorem 3.7. *Let $\mathcal{E}$ be a strongly local Dirichlet form on X such that the intrinsic metric d induced by $\mathcal{E}$ is compatible with the topology of X. Suppose that X supports a Borel measure μ that is doubling (with respect to the metric d). If every compactly supported 1-Lipschitz function on X is in $D(\mathcal{E})$ with $d\eta(\varphi, \varphi) \leqslant d\mu$ and $\mathcal{E}$ satisfies a weak $(1, 2)$-Poincaré inequality, then X supports a weak $(1, 2)$-Poincaré inequality for Newtonian functions and $D(\mathcal{E}) = N^{1,2}(X)$ as a Banach space isomorphism.*

Proof. By Proposition 2.4, it suffices to prove the Poincaré inequality for compactly supported Lipschitz functions u with compactly supported Lipschitz upper gradients ρ.

Since $\mathcal{E}$ satisfies a Poincaré inequality, by Lemma 3.3, whenever $B(x,r)$ is a ball in X,

$$\fint_{B(x,r)} |u-u_B|\,d\mu \leqslant Cr\Big(\fint_{B(x,\tau r)} \operatorname{Lip}(u,y)^2\,d\mu(y)\Big)^{1/2}.$$

Let ρ be a compactly supported Lipschitz upper gradient of u. By quasiconvexity of X (see Lemma 3.6), whenever $z,y\in B(x,r)$, there is a curve γ connecting z and y in $B(x,Cr)$ with length

$$\ell(\gamma)\leqslant C_Q d(z,y).$$

Hence

$$|u(z)-u(y)|\leqslant \int_\gamma \rho\,ds \leqslant \ell(\gamma)\rho(w_\gamma)\leqslant C_Q d(z,y)\rho(w_\gamma)$$

for some w_γ in the trajectory of γ. Hence

$$|u(z)-u(y)|/d(z,y)\leqslant C_Q\rho(w_\gamma).$$

Therefore,

$$L_u(x;r)\leqslant C_Q \sup_{w\in B(x,Cr)} \rho(w).$$

Letting $r\to 0$ and noting that ρ is continuous, we obtain

$$\operatorname{Lip}(u,x)\leqslant C_Q\rho(x).$$

It now follows that

$$\fint_{B(x,r)} |u-u_B|\,d\mu \leqslant CC_Q r\Big(\fint_{B(x,\tau r)} \rho^2\,d\mu\Big)^{1/2}.$$

Since $\mathcal{E}$ supports a weak $(1,2)$-Poincaré inequality, by [19, Proposition 4.2] (see Remark 3.5 above), $D(\mathcal{E})$ embeds via a Banach space isomorphism into the Korevaar–Schoen space $KS^{1,2}(X)$. Since X supports a $(1,2)$-Poincaré inequality for Newtonian functions, $KS^{1,2}(X)=N^{1,2}(X)$ as a Banach space isomorphism. Therefore, $D(\mathcal{E})$ is embedded isomorphically in $N^{1,2}(X)$.

It suffices now to prove that $N^{1,2}(X)$ embeds into $D(\mathcal{E})$. For $\varepsilon>0$ we can cover X by countably many balls $B_i=B(x_i,\varepsilon)$ such that

$$\sup_{x\in X}\sum_{i\in\mathbb{N}}\chi_{5B_i}(x)\leqslant C$$

and obtain a partition of unity φ_i subordinate to this cover; $0 \leqslant \varphi_i \leqslant 1$, φ_i is supported on B_i, $\sum_{i\in\mathbb{N}} \varphi_i = 1$, and φ_i is C/ε-Lipschitz (see [10] or [18]). As in the proof of [18, Lemma 4.6], we obtain a discrete convolution of $u \in N^{1,2}(X) = KS^{1,2}(X)$ as follows:

$$u_\varepsilon(x) = \sum_{i\in\mathbb{N}} u_{B_i}\varphi_i(x).$$

By Lemma 4.6 of [18] and Hölder's inequality, whenever $y, z \in X$ with $d(y,z) < \varepsilon$,

$$\frac{|u_\varepsilon(y) - u_\varepsilon(z)|}{d(y,z)} \leqslant C\Big(\fint_{B(y,2\varepsilon)} e^2_{5\varepsilon}(w;u)\, d\mu(w)\Big)^{1/2},$$

and hence if $x \in X$ and $r < \varepsilon/2$, we have

$$\mathrm{Lip}_{u_\varepsilon}(x;r) \leqslant C\Big(\fint_{B(x,4\varepsilon)} e^2_{5\varepsilon}(w;u)\, d\mu(w)\Big)^{1/2}.$$

It follows that

$$\mathrm{Lip}(u_\varepsilon, x) \leqslant C\Big(\fint_{B(x,4\varepsilon)} e^2_{5\varepsilon}(w;u)\, d\mu(w)\Big)^{1/2},$$

and hence, by Lemma 3.3,

$$\eta(u_\varepsilon,u_\varepsilon)(B_i) \leqslant C\int_{B_i} \fint_{B(x,4\varepsilon)} e^2_{5\varepsilon}(w;u)\, d\mu(w)\, d\mu(x) \leqslant C\int_{B_i} e^2_{5\varepsilon}(x;u)\, d\mu(x).$$

Hence, by the bounded overlap property of the cover B_i,

$$\eta(u_\varepsilon,u_\varepsilon)(X) \leqslant C\int_X e^2_{5\varepsilon}(x;u)\, d\mu(x),$$

i.e., for sufficiently small $\varepsilon > 0$

$$\mathcal{E}(u_\varepsilon,u_\varepsilon) \leqslant CE_{KS}(u).$$

By the fact that X supports $(1,2)$-Poincaré inequality for Newtonian functions, as in [18] we see that $u_\varepsilon \to u$ in $KS^{1,2}(X) = N^{1,2}(X)$. Thus, by the closability property of the Dirichlet form, we see that

$$\mathcal{E}(u,u) \leqslant C\|u\|^2_{N^{1,2}(X)},$$

i.e., $N^{1,2}(X) \subset D(\mathcal{E})$. □

4 Axiomatic Sobolev Spaces and $N^{1,p}(X)$

The focus of this section is the axiomatic theory of Gol'dshtein–Troyanov [7]. The starting point for this theory is the assumption of a class of "derivatives" in the metric space setting, called a *D structure*, which associates with each $u \in L^p_{loc}(X)$ a class $D[u]$ of nonnegative measurable functions on X such that the following five axioms are satisfied:

1. *Nontriviality* If u is a nonnegative L-Lipschitz function supported on a set $A \subset X$, then $L\chi_A \in D[u]$.
2. *Upper linearity* If $\rho_1 \in D[u_1]$, $\rho_2 \in D[u_2]$, and $\alpha, \beta \in \mathbb{R}$, and $g \geqslant |\alpha|\rho_1 + |\beta|\rho_2$, then $g \in D[\alpha u_1 + \beta u_2]$.
3. *Leibnitz rule* If $g \in D[u]$ and φ is a bounded L-Lipschitz function on X, then $\|\varphi\|_{L^\infty(X)} g + L|u| \in D[\varphi u]$.
4. *Lattice property* If $\rho_1 \in D[u_1]$, $\rho_2 \in D[u_2]$, then
$$\max\{\rho_1, \rho_2\} \in D[\max\{u_1, u_2\}] \cap D[\min\{u_1, u_2\}].$$
5. *Completeness* If $g_i \in D[u_i]$ for $i \in \mathbb{N}$, $u_i \to u$ in $L^p(X)$, and $g_i \to g$ in $L^p(X)$, then $g \in D[u]$.

The axiomatic Sobolev space $\mathcal{L}^{1,p}(X)$ consists of all $u \in L^p(X)$ for which there exists $g \in D[u] \cap L^p(X)$. This space, equipped with the norm

$$\|u\|_{\mathcal{L}^{1,p}(X)} = \|u\|_{L^p(X)} + \inf_{g \in D[u]} \|g\|_{L^p(X)}$$

is a Banach space by the axioms above.

It was shown in [7] that if $1 < p < \infty$, then for every $u \in \mathcal{L}^{1,p}(X)$ there is a unique function $g_u \in D[u]$, called the *minimal pseudogradient* of u, such that

$$\inf_{g \in D[u]} \|g\|_{L^p(X)} = \|g_u\|_{L^p(X)}.$$

Definition 4.1. We say that the D structure supports a $(1,p)$-Poincaré inequality if there are constants $C > 0$ and $\tau \geqslant 1$ such that whenever $u \in L^p_{loc}(X)$ and $g \in D[u]$, for all balls $B(x,r)$ in X

$$\fint_{B(x,r)} |u - u_{B(x,r)}|\, d\mu \leqslant Cr\Big(\fint_{B(x,\tau r)} g^p \, d\mu\Big)^{1/p}.$$

We say that the D structure is *strongly local* if whenever $u_1, u_2 \in \mathcal{L}^{1,p}(X)$ and $u_1 = u_2$ on a Borel set $A \subset X$, then for all $g_1 \in D[u_1]$ and $g_2 \in D[u_2]$

$$g_1\chi_A + g_2\chi_{X\setminus A} \in D[u_2].$$

The axiomatic regularity theory developed in [28] would not be applicable to solutions to variational problems based on the axiomatic Sobolev space theory without the assumption of the support of Poincaré inequality and strict locality. Hence the hypotheses of the following main theorem of this section are reasonable.

Lemma 4.2. *If the D structure is strongly local and* $1 < p < \infty$*, then whenever* $u \in \mathcal{L}^{1,p}(X)$ *and* $g \in D[u]$*,* $g \geqslant g_u$ μ*-a.e. in* X *(*g_u *is the minimal pseudogradient of* u*).*

The proof is a direct application of the strong locality of the D structure, and is left to the reader.

Theorem 4.3. *Suppose that* $\mathcal{L}^{1,p}(X)$ *is an axiomatic Sobolev space on a complete metric measure space with doubling measure. If the associated D structure supports a* $(1,p)$*-Poincaré inequality and is strongly local, then* X *supports a* $(1,p)$*-Poincaré inequality for Newtonian functions, and* $\mathcal{L}^{1,p}(X) = N^{1,p}(X)$ *as a Banach space isomorphism.*

Proof. First we show that, under the strict locality assumption, whenever $\varphi : X \to \mathbb{R}$ is a Lipschitz function on X, $\mathrm{Lip}(\varphi,\cdot) \in D[\varphi]$. Without loss of generality, we assume that φ is compactly supported, and hence, by the nontriviality axiom, $\varphi \in \mathcal{L}^{1,p}(X)$. Fix $r > 0$. For $x \in X$ let $\psi_{x,r}$ be a Lipschitz extension (given by McShane for example, see [10]) of $\varphi|_{B(x,r)}$ to X. Then the global Lipschitz constant of $\psi_{x,r}$ is given as follows:

$$L_{\psi_{x,r}} = \sup_{y,z\in X: y\neq z} \frac{|\psi_{x,r}(y) - \psi_{x,r}(z)|}{d(y,z)} = \sup_{y,z\in B(x,r): y\neq z} \frac{|\varphi(y) - \varphi(z)|}{d(y,z)}$$
$$= \mathrm{Lip}_\varphi(x;r).$$

By the nontriviality axiom, $L_{\psi_{x,r}} \in D[\psi_{x,r}]$. Note that

$$\mathrm{Lip}(\varphi,x) = \limsup_{r\to 0^+} \mathrm{Lip}_\varphi(x;r).$$

Since $\varphi = \psi_{x,r}$ on the open ball $B(x,r)$, by the strong locality property, if g_φ is the minimal pseudogradient of φ, then $g_\varphi \leqslant L_{\psi_{x,r}}$ on $B(x,r)$ (see the lemma above).

Let Z be the collection of all non-Lebesgue points of g_φ. Then $\mu(Z) = 0$ and for all $x \in X \setminus Z$

$$\frac{1}{\mu(B(x,r))} \int\limits_{B(x,r)} g_\varphi \, d\mu \leqslant \mathrm{Lip}_\varphi(x;r)$$

and hence

$$g_\varphi(x) = \lim_{r\to 0^+} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} g_\varphi \, d\mu \leqslant \limsup_{r\to 0^+} \mathrm{Lip}_\varphi(x;r) = \mathrm{Lip}(\varphi, x),$$

i.e., $g_\varphi \leqslant \mathrm{Lip}(\varphi,\cdot)$ almost everywhere on X. Hence $\mathrm{Lip}(\varphi,\cdot) \in D[\varphi]$ by the upper linearity axiom.

Now, if the axiomatic space satisfies a $(1,p)$-Poincaré inequality, then whenever φ is a Lipschitz function on X, for all balls $B \subset X$

$$\frac{1}{\mu(B)} \int_B |\varphi - \varphi_B| \, d\mu \leqslant C \,\mathrm{rad}(B) \left(\fint_{\tau B} \mathrm{Lip}(\varphi, y)^p \, d\mu(y) \right)^{1/p},$$

and then, as in the proof of Theorem 3.7, we conclude that X supports a $(1,p)$-Poincaré inequality for Newtonian functions and $N^{1,p}(X) = \mathcal{L}^{1,p}(X)$, thus concluding the proof. □

Acknowledgement. We wish to thank Marc Troyanov and Sergey Timoshin for valuable discussions on the topic of axiomatic Sobolev spaces.

References

1. Barlow, M., Bass, R.: Brownian motion and harmonic analysis on Sierpinski carpets. Canad. J. Math. **51**, 673–744 (1999)
2. Beurling, A., Deny, J.: Dirichlet spaces. Proc. Nat. Acad. Sci. U.S.A. **45**, 208–215 (1959)
3. Biroli, M., Mosco, U.: A Saint-Venant type principle for Dirichlet forms on discontinuous media. Ann. Mat. Pura Appl. (4) **169**, 125–181 (1995)
4. Biroli, M., Mosco, U.: Sobolev and isoperimetric inequalities for Dirichlet forms on homogeneous spaces. Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl. **6**, 37–44 (1995)
5. Cheeger, J.: Differentiability of Lipschitz functions on metric measure spaces. Geom. Funct. Anal. **9**, 428–517 (1999)
6. Fukushima, M., Oshima, Y., Takeda, M.: Dirichlet Forms and Symmetric Markov Processes. Walter de Gruyter, Berlin (1994)
7. Gol'dshtein, V.M., Troyanov, M.: Axiomatic theory of Sobolev spaces. Exposition. Math. **19**, 289–336 (2001)
8. Hajłasz, P.: Sobolev spaces on an arbitrary metric space. Potential Anal. **5**, 403–415 (1996)
9. Hajłasz, P., Koskela, P.: Sobolev met Poincaré. Mem. Am. Math. Soc. **688** (2000)
10. Heinonen, J.: Lectures on Analysis on Metric Spaces. Springer, New York (2000)
11. Heinonen, J., Koskela, P.: Quasiconformal maps in metric spaces with controlled geometry. Acta Math. **181**, 1–61 (1998)
12. Jost, J.: Generalized Dirichlet forms and harmonic maps. Calc. Var. Partial Differ. Equ. **5**, 1–19 (1997)
13. Keith, S.: Modulus and the Poincaré inequality on metric measure spaces. Math. Z. **245**, 255–292 (2003)

14. Kigami, J.: A harmonic calculus on the Sierpiński spaces. Japan J. Appl. Math. **6**, 259–290 (1989)
15. Kigami, J.: Harmonic calculus on p.c.f. self-similar sets. Trans. Am. Math. Soc. **335**, 721–755 (1993)
16. Kinnunen, J., Shanmugalingam, N.: Regularity of quasi-minimizers on metric spaces. Manuscr. Math. **105**, 401–423 (2001)
17. Korevaar, J N., Schoen, R.M.: Sobolev spaces and harmonic maps for metric space targets. Commun. Anal. Geom. **1**, 561–659 (1993)
18. Koskela, P., MacManus, P.: Quasiconformal mappings and Sobolev spaces. Stud. Math. **131**, 1–17 (1998)
19. Koskela, P., Shanmugalingam, N., Tyson, J.T.: Dirichlet forms, Poincaré inequalities, and the Sobolev spaces of Korevaar–Schoen. Potential Anal. **21**, 241–262 (2004)
20. Maz'ya, V.G.: Sobolev Spaces. Springer, Berlin-Tokyo (1985)
21. Mosco, U.: Composite media and asymptotic Dirichlet forms J. Funct. Anal. **123**, 368–421 (1994)
22. Shanmugalingam, N.: Newtonian spaces: an extension of Sobolev spaces to metric measure spaces. Rev. Mat. Iberoam. **16**, 243–280 (2000)
23. Shanmugalingam, N.: Harmonic functions on metric spaces. Ill. J. Math. **45**, 1021–1050 (2001)
24. Shanmugalingam, N.: Sobolev type spaces on metric measure spaces. Potential theory in Mat., Adv. Stud. Pure Math. **44**, 77–90 (2006)
25. Sturm, K.T.: Monotone approximation of energy functionals for mappings into metric spaces - II. Potential Anal. **11**, 359–386 (1999)
26. Sturm, K.T.: Analysis on local Dirichlet spaces I. Recurrence, conservativeness and L^p-Liouville properties. J. Reine Angew. Math. **456**, 173–196 (1994)
27. Sturm, K.T.: Diffusion processes and heat kernels on metric spaces. Ann. Probab. **26**, 1–55 (1998)
28. Timoshin, S.: Axiomatic regularity on metric spaces. Michigan Math. J. (2008) [To appear]

Cocompact Imbeddings and Structure of Weakly Convergent Sequences

Kiril Tintarev

Abstract The concentration compactness method is a powerful technique for establishing the existence of minimizers for inequalities and critical points of functionals. We give a functional-analytic formulation for the method in a Banach space. The key object is a dislocation space, i.e., a triple (X, F, D), where F is a convex functional defining a norm on a Banach space X and D is a group of isometries on X. Bounded sequences in dislocation spaces admit a decomposition into an asymptotic sum of "profiles" $w^{(n)} \in X$ dislocated by the actions of D. This decomposition allows to extend the weak convergence argument from variational problems with compactness to problems, where X is *cocompactly* (relatively to D) imbedded into a Banach space Y. We prove a general statement on the existence of minimizers in cocompact imbeddings that applies, in particular, to Sobolev imbeddings which lack compactness (an unbounded domain, a critical exponent) including the subelliptic Sobolev spaces and spaces over Riemannian manifolds.

1 Introduction

Minimizers for an inequality in a functional space often do not exist or cannot be obtained by a straightforward compactness reasoning since, in general, one may expect that a minimization sequence converges only weakly. The concentration compactness method presented in the celebrated papers by P.-L.Lions [8]–[11]) uses a detailed structural information about the minimization sequences in order to verify the convergence in problems that a priori lack compactness. The core idea of the concentration compactness is that if the problem possesses a noncompact invariance group G, lack of convergence can

Kyril Tintarev
Uppsala University,P.O.Box 480, SE-751 06 Uppsala, Sweden,
e-mail: kyril.tintarev@math.uu.se

V. Maz'ya (ed.), *Sobolev Spaces in Mathematics I, International Mathematical Series.*
doi: 10.1007/978-0-387-85648-3,

be attributed to the action of G, and thus a given sequence becomes convergent only after the terms ("profiles"), dislocated by the transformations, are "factored out." Elaborations of the original classification of weak convergent sequences by Lions into tight, vanishing and dichotomous, which are often called *splitting lemmas*, were given for specific cases by Struwe [14], Brezis and Coron [4], Lions [12], and numerous authors afterwards. The "splitting lemmas," which were originally established for critical sequences of specific functionals in specific functional spaces, were later summarized by the author in a structural statement that holds in the general Hilbert space (see [15] and references therein) by using the asymptotic orthogonality of dislocated profiles: if $g_k \in D$, where D is a fixed group of unitary operators, $u_k \rightharpoonup w$, $g_k u_k \rightharpoonup w_2$, and $v_k = u_k - w_1 - g_k^{-1} w_2$, then $u_k = w_1 + g_k^{-1} w_2 + v_k$ is the asymptotically orthogonal sum in the sense that the scalar product of any two terms of the sum converges to zero. Furthermore, this construction may be iterated. Under general conditions, the subtraction of all dislocated profiles of a bounded sequence (nonzero weak limits of sequences $g_k u_k$ with different sequences g_k) amounts to a sequence that weakly converges to zero under all dislocations (the D-weak convergence). In fact, this construction is useful only to an extent that the D-weak convergence is meaningful. One may say that the Hilbert space is cocompactly (relatively to D) imbedded into a Banach space Y if the D-weak convergence in X implies the convergence in Y. For example, subcritical Sobolev imbeddings on complete Riemannian manifolds are cocompact with respect to the action of any subgroup of the isometry group of a manifold if the manifold itself is cocompact with respect to this subgroup.

In the present paper, we give a tentative formulation of this framework for Banach spaces, where one can no longer rely on the notion of asymptotic orthogonality. Its natural counterpart is asymptotic additivity or subadditivity of energy functionals with respect to dislocated profiles [it makes sense indeed to call a functional with such an additivity property an *energy*, indicating that it is asymptotically additive over asymptotically separate (for example, with asymptotically disjoint supports) clusters of the physical system that it models]. Such an asymptotic additivity is realized, in particular, in Brezis–Lieb lemma [3]

Many applications of the concentration compactness method, such as the existence of minimizers in isoperimetric problems or the compactness of imbeddings of subspaces of functions with symmetries, are realized already on the functional-analytic level, with immediate applications to Sobolev spaces $W^{m,p}$ over Riemannian (and sub-Riemannian) manifolds and their flask subdomains.

In Sect. 2 we prove the main structural theorem. Section 3 deals with functional-analytic statements on the existence of minimizers in isoperimetric problems. Section 4 extends the results of two previous cases to noninvariant subspaces, and in Sect. 5 some compactness results are given.

2 Dislocation Space and Weak Convergence Decomposition

In this section, we prove a structural theorem for bounded sequences in a class of Banach spaces associated with convex functionals.

Lemma 2.1. *Let X be a vector space, and let $F \in C^1(X)$ be an even non-negative convex function with $F^{-1}(0) = \{0\}$. Then the map $\lambda : X \to [0, \infty)$,*

$$\lambda(u) = \inf\{\lambda > 0 : F(\lambda^{-1}u) \leqslant 1\}, \tag{2.1}$$

is a norm on X and $\lambda = \|u\|$ for any $u \in X \setminus \{0\}$ is a unique solution of $F(\lambda^{-1}u) = 1$.

Proof. The homogeneity of $\lambda(u)$ is immediate from definition. If $u = 0$, then $F(\lambda^{-1}u) = 0$ for all $\lambda > 0$ and thus $\lambda(0) = 0$. Since $F^{-1}(0) = \{0\}$, for every $u \in X \setminus \{0\}$ the even convex function $t \in \mathbb{R}_+ \mapsto F(tu)$ is strictly monotone and unbounded from above. In particular, $\lambda(u) > 0$, whenever $u \neq 0$. Furthermore, by strict monotonicity, $F(\lambda^{-1}u) = 1$ has a unique solution λ_1 and, since $F(\lambda^{-1}u) > 1$ for $\lambda < \lambda_1$, the infimum in (2.1) is attained at $\lambda_1 = \lambda(u)$. It remains to prove the triangle inequality. By the convexity of F, we have

$$\begin{aligned} F\left(\frac{u+v}{\lambda(u)+\lambda(v)}\right) &= F\left(\frac{\lambda(u)}{\lambda(u)+\lambda(v)}\frac{u}{\lambda(u)} + \frac{\lambda(v)}{\lambda(u)+\lambda(v)}\frac{v}{\lambda(v)}\right) \\ &\leqslant \frac{\lambda(u)}{\lambda(u)+\lambda(v)}F\left(\frac{u}{\lambda(u)}\right) + \frac{\lambda(v)}{\lambda(u)+\lambda(v)}F\left(\frac{v}{\lambda(v)}\right) \\ &= \frac{\lambda(u)}{\lambda(u)+\lambda(v)} + \frac{\lambda(v)}{\lambda(u)+\lambda(v)} = 1. \end{aligned}$$

The lemma is proved. □

Definition 2.2. A *dislocation space* is a triple (X,F,D), where the pair (X, F) is the same as in Lemma 2.1, $F \in C^1(X)$ is uniformly continuous on bounded sets, a Banach space X is separable and reflexive, and D is a group of linear operators on X, closed with respect to the strong (elementwise) convergence, satisfying $F \circ g = F$ for all $g \in D$ and such that

$$g_k \in D,\ g_k \not\rightharpoonup 0,\ u_k \rightharpoonup 0 \Rightarrow g_k u_k \rightharpoonup 0 \text{ on a subsequence.} \tag{2.2}$$

Moreover, if sequences $\{g_k^{(n)}\}_{k\in\mathbb{N}} \subset D$, $n = 1, \dots, M$, $M \in \mathbb{N}$, satisfy

$$g_k^{(m)^{-1}} g_k^{(n)} \rightharpoonup 0, \quad m \neq n, \tag{2.3}$$

and $u_k \in X$ is a bounded sequence such that ${g_k^{(n)}}^{-1} u_k \rightharpoonup w^{(n)}$, $n = 1, \ldots, M$, then

$$\liminf F(u_k) \geqslant \sum_{n=1}^{M} F(w^{(n)}) \tag{2.4}$$

and

$$F\Big(\sum_{n=1}^{M} g_k^{(n)} w^{(n)} \Big) \to \sum_{n=1}^{M} F(w^{(n)}). \tag{2.5}$$

Remark 2.3. It is easy to see that the conditions (2.4) and (2.5) are satisfied if F satisfies the Brezis–Lieb property

$$u_k \rightharpoonup u \Rightarrow F(u_k) - F(u) - F(u_k - u) \to 0.$$

In particular, if X is a Hilbert space and $F(u) = \|u\|^2$, then

$$\|u_k\|^2 - \|u_k - u\|^2 - \|u\|^2 = 2(u_k, u) - 2\|u\|^2 \to 0.$$

When $F(u) = \int \varphi(u) d\mu$ with φ from a class of functions on a measure space that includes $\varphi(t) = |t|^p$, $p \in (1, \infty)$, Brezis–Lieb property was verified in [3] under the additional condition $u_k \to 0$ a.e., although, since L^2 is a Hilbert space, this condition redundant when $p = 2$.

Examples of dislocation spaces

1. (H, F, D), where H is a separable Hilbert space; $F(u) = \|u\|^2$; and the group D of unitary operators satisfies (2.2), in particular, as in any of the examples below with $p = 2$. This case is elaborated in [15].

2. $(W^{1,p}(M), \|\cdot\|^p, D)$, where M is a complete (sub-) Riemannian manifold, $W^{1,p}(M)$, $p > 1$, is a Sobolev space associated with the p-(sub-)Laplacian and $D = \{u \mapsto u \circ \eta\}_{\eta \in \mathrm{Iso}\,(M)}$. In particular, $(W^{1,p}(\mathbb{R}^N), \|\cdot\|^p, D)$ with the group of shifts $D = \{u \mapsto u(\cdot + y)\}_{y \in \mathbb{R}^N}\}$.

3. $(\mathcal{D}^{1,p}(G), \|\sqrt{L(u)}\|_p^p, D')$, where G is a Carnot group of homogeneous dimension Q with invariant subelliptic Lagrangian $L(u) = \sum_i |du(X_i)|$, X_i are generators of the correspondent Lie algebra, $1 < p < Q$, and D' is a product group of the actions of left group shifts and of the group of dilation actions $u \mapsto t^{\frac{Q-p}{p}} u \circ \delta_t$, $t > 0$, where $\delta_t : G \to G$, $t \in (0, \infty)$ are homogeneous dilations on G. In particular, $(\mathcal{D}^{1,p}(\mathbb{R}^N), \|\nabla \cdot\|_p^p, D')$, $1 < p < N$, where D' is a product group of Euclidean shifts and of the group of dilation actions $u \mapsto t^{\frac{N-p}{p}} u(t\cdot)$, is a dislocation space.

Definition 2.4. Let X be a Banach space, and let D be a group of linear isometries on X. A sequence $u_k \in X$ *converges* D*-weakly* to $u \in X$ (denoted as $u_k \overset{D}{\rightharpoonup} u$) if for any sequence $g_k \in D$ we have $g_k(u_k - u) \rightharpoonup 0$.

Lemma 2.5. *Let* (X, F, D) *be a dislocation space. If* $F(u_k) \to 0$ *then* $u_k \overset{D}{\rightharpoonup} 0$.

Proof. Since $F(u_k) \leqslant 1$, for all k sufficiently large $\|u_k\| \leqslant 1$. On every weakly convergent renumbered subsequence of u_k, $F(\text{w-lim}\, u_k) \leqslant \lim F(u_k) = 0$ and, consequently, $u_k \rightharpoonup 0$. Since D preserves F, the same conclusion applies to $g_k u$ for any sequence $g_k \in D$. □

Theorem 2.6. *Let* (X, F, D) *be a dislocation space. If* $u_k \in X$ *is a bounded sequence, then there exists a set* $\mathbb{N}_0 \subset \mathbb{N}$, $w^{(n)} \in X$, *sequences* $\{g_k^{(n)}\}_{k \in \mathbb{N}} \subset D$ *with* $g_k^{(1)} = \mathrm{id}$ *satisfying* (2.3), $n \in \mathbb{N}_0$, *such that for a renumbered subsequence*

$$w^{(n)} = \text{w-lim}\, {g_k^{(n)}}^{-1} u_k, \tag{2.6}$$

$$\sum_{n \in \mathbb{N}_0} F(w^{(n)}) \leqslant \limsup F(u_k), \tag{2.7}$$

$$u_k - \sum_{n \in \mathbb{N}_0} g_k^{(n)} w^{(n)} \overset{D}{\rightharpoonup} 0, \tag{2.8}$$

where the series $\sum_{n \in \mathbb{N}_0} g_k^{(n)} w^{(n)}$ *converges uniformly in* k *in the sense that*

$$\sup_{k \in \mathbb{N}} F\Big(\sum_{n \geqslant m} g_k^{(n)} w^{(n)} \Big) \to 0 \text{ as } m \to \infty. \tag{2.9}$$

Proof. 1. Once (2.3) is proved and $w^{(n)}$ satisfy (2.6), the inequality (2.4) in Definition 2.2 holds for every $M \in \mathbb{N}$ and thus the series in (2.7) converges.

2. Observe that if $u_k \overset{D}{\rightharpoonup} 0$, then the theorem is verified with $\mathbb{N}_0 = \varnothing$. Otherwise, we consider expressions of the form

$$w^{(1)} =: \text{w-lim}\, {g_k^{(1)}}^{-1} u_k. \tag{2.10}$$

The sequence u_k is bounded, D is a set of isometries, so the sequence in (2.10) is bounded and thus, for any choice of $g_k \in D$, it has a weakly convergent subsequence. Since we assume that u_k does not converge D-weakly to zero, there exists necessarily a renumbered sequence $g_k^{(1)}$ that yields a nonzero limit in (2.10).

Let

$$v_k^{(1)} = u_k - g_k^{(1)} w^{(1)}.$$

By (2.10),

$$g_k^{(1)^{-1}} v_k^{(1)} = g_k^{(1)^{-1}} (u_k - w^{(1)}) \rightharpoonup 0. \tag{2.11}$$

If $v_k^{(1)} \overset{D}{\rightharpoonup} 0$, the theorem is verified with $\mathbb{N}_0 = \{1\}$. Otherwise (we repeat the argument above), there exists necessarily a sequence $g_k^{(2)} \in D$ and $w^{(2)} \neq 0$ such that, on a renumbered subsequence,

$$g_k^{(2)^{-1}} v_k^{(1)} \rightharpoonup w^{(2)}.$$

We set

$$v_k^{(2)} = v_k^{(1)} - g_k^{(2)} w^{(2)}.$$

Then we obtain an obvious analog of (2.11):

$$g_k^{(2)^{-1}} v_k^{(2)} = g_k^{(2)^{-1}} (v_k^{(1)} - w^{(2)}) \rightharpoonup 0. \tag{2.12}$$

If we assume that

$$g_k^{(1)^{-1}} g_k^{(2)} \not\rightharpoonup 0,$$

then, by (2.12) and (2.2),

$$g_k^{(1)^{-1}} (v_k^{(1)} - g_k^{(2)} w^{(2)}) \rightharpoonup 0,$$

which, by (2.11), yields

$$g_k^{(1)^{-1}} g_k^{(2)} w^{(2)} \rightharpoonup 0. \tag{2.13}$$

We now use (2.2) again to replace in (2.13) $g_k^{(1)^{-1}}$ with $g_k^{(2)^{-1}}$, which results in

$$w^{(2)} \rightharpoonup 0, \tag{2.14}$$

which cannot be true since we assumed that $w^{(2)} \neq 0$. From this a contradiction follows:

$$g_k^{(1)^{-1}} g_k^{(2)} \rightharpoonup 0. \tag{2.15}$$

Then

$$g_k^{(2)^{-1}} g_k^{(1)} \rightharpoonup 0.$$

Indeed, if this were false, then from (2.2) and (2.15) we have on a subsequence

$$\mathrm{id} = g_k^{(2)^{-1}} g_k^{(1)} g_k^{(1)^{-1}} g_k^{(2)} \rightharpoonup 0,$$

which is obviously false.

We define recursively:

$$v_k^{(n)} = v_k^{(n-1)} - g_k^{(n)} w^{(n)} = u_k - g_k^{(1)} w^{(1)} - \cdots - g_k^{(n)} w^{(n)}, \tag{2.16}$$

where

$$w^{(n)} = \text{w-lim}\, {g_k^{(n)}}^{-1} v_k^{(n-1)}$$

calculated on a successively renumbered subsequence. We subordinate the choice of $g_k^{(n)}$ and thus the extraction of this subsequence for every given n to the following requirements. For every $n \in \mathbb{N}$ we set

$$W_n = \{w \in H \setminus \{0\} : \exists g_j \in D, \{k_j\} \subset \mathbb{N} : g_j^{-1} v_{k_j}^{(n)} \rightharpoonup w\}$$

and

$$t_n = \sup_{w \in W_n} F(w).$$

Note that $t_n < \infty$ since all the operators involved at all the steps leading to the definition of W_n have uniform bounds.

If $t_n = 0$ for some n, the theorem is proved with $\mathbb{N}_0 = \{1, \ldots, n-1\}$. Otherwise, we choose $w^{(n+1)} \in W_n$ such that

$$F(w^{(n+1)}) \geqslant \frac{1}{2} t_n \tag{2.17}$$

and the sequence $g_k^{(n+1)}$ is chosen so that on a subsequence that we renumber

$${g_k^{(n+1)}}^{-1} v_k^{(n)} \rightharpoonup w^{(n+1)}. \tag{2.18}$$

As above, for $n = 1$ we have

$${g_k^{(p)}}^{-1} g_k^{(q)} \rightharpoonup 0 \text{ whenever } p \neq q, p, q \leqslant n. \tag{2.19}$$

This allows us to deduce immediately (2.6) from (2.18), as well as (2.7). From (2.4) and(2.17) if follows that

$$\sum_{n \geqslant 2} t_n \leqslant 2F(u_k).$$

Let φ_i, $i \in \mathbb{N}$, be a normalized basis for X^*. By the definition of W_n,

$$\limsup_k \sum_i 2^{-i} \sup_{g \in D} \langle g v_k^{(n)}, \varphi_i \rangle^2 \leqslant 4 t_n^2, \quad n \in \mathbb{N}.$$

Let $k(n)$ be such that

$$\sum_i 2^{-i} \sup_{g \in D} \langle g v_{k(n)}^{(n)}, g \varphi_i \rangle^2 \leqslant 8 t_n^2, \quad n \in \mathbb{N}. \tag{2.20}$$

This implies that

$$\sup_{g\in D}\langle gv_{k(n)}^{(n)},\varphi\rangle \to 0$$

for any φ that is a linear combination of φ_i, and an elementary density argument extends this relation to any $\varphi \in X^*$, so that

$$v_{k(n)}^{(n)} \overset{D}{\rightharpoonup} 0$$

as $n \to \infty$. Instead of $k(n)$ selected for each n from the index set of a renumbered subsequence of u_k (that was produced by successive extractions), we now use the correspondent index (preserving the notation $k(n)$) from the original enumeration of u_k. (This change of enumeration affects also the terms $g_{k(n)}^{(j)}$, $j = 1,\dots,n$, in the definition (2.16) of $v_{k(n)}^{(n)}$.) Then

$$v_{k(n)}^{(n)} = u_{k(n)} - \sum_{j\leqslant n} g_{k(n)}^{(j)} w^{(j)} \overset{D}{\rightharpoonup} 0.$$

Since the final extraction is a subsequence of the sequence in (2.19), we obtain (2.3).

Note that $k(n)$ can be chosen in (2.20) arbitrarily large and, in particular, such that the series $\sum\limits_j g_{k(n)}^{(j)} w^{(j)}$ is uniformly convergent in the sense of (2.9) due to (2.7) and (2.3), and therefore (2.8) follows. Indeed, one can always choose a subsequence of $g_k^{(m+1)}$ such that, by (2.5),

$$\left| F\Big(\sum_{n=1}^{m+1} g_k^{(n)} w^{(n)}\Big) - F\Big(\sum_{n=1}^{m} g_k^{(n)} w^{(n)}\Big) - F(w^{(m+1)})\right| \leqslant 2^{-k-m}.$$

Finally, if $w^{(1)} = \text{w-lim}\, u_k \neq 0$, we could have chosen $g_k^{(1)} = \text{id}$ at the first step. If $\text{w-lim}\, u_k = 0$, we renumber terms in the expansion by $n = 2, 3, \dots$ and set $g_k^{(1)} = \text{id}$, $w^{(1)} = 0$. □

3 Cocompactness and Minimizers

In this section, we give a functional-analytic formalization of the minimization reasoning of Lions ([8]) in cocompactly imbedded dislocation spaces.

Definition 3.1. A continuous imbedding of a Banach space X into a Banach space Y is *cocompact* relatively to a group D of isometric linear operators on X if every D-weakly convergent sequence $u_k \in X$ converges in Y.

Note that it does not follow from this definition that the quotient X/D is compactly imbedded into Y. If $D = \{\text{id}\}$, the cocompact imbedding becomes compact.

Examples of cocompact imbeddings

1. $W^{1,p}(\mathbb{R}^N)$, is cocompactly imbedded into $L^q(\mathbb{R}^N)$ relatively to the group of lattice shifts $u \mapsto u(\cdot + y)$, $y \in \mathbb{Z}^N$, when $p < q < \frac{Np}{N-p}$ for $N > p$ or $q > p$ for $N \leqslant p$.
2. Let M be a complete N-dimensional Riemannian manifold, cocompact with respect to a subgroup G of its isometry group Iso (M), i.e., there exists a compact set $V \subset M$ such that $\bigcup_{\eta \in G} \eta V = M$. Then $W^{1,p}(M)$ with the invariant norm $\|u\|^p = \int (|du|^p + |u|^p) d\mu$ is cocompactly imbedded into $L^p(M)$ for the same values of p as above, relatively to the group $\{u \mapsto u \circ \eta\}_{\eta \in G}$.
3. Let G be a Carnot group of homogeneous dimension Q. Then $\mathcal{D}^{1,p}(G)$, $p < Q$, is cocompactly imbedded into $L^{p^*}(G)$, where $p^* = \frac{pQ}{Q-p}$, relatively to a product group of left shifts and discrete dilation action $u \mapsto 2^{\frac{Q-p}{p} j} u \circ \delta_{2^j}$, $j \in \mathbb{Z}$. In particular, $\mathcal{D}^{1,p}(\mathbb{R}^N)$ is cocompactly imbedded into $L^{p^*}(\mathbb{R}^N)$ for $p < N$.

The Euclidean case of the statements above, with the group of $\mathbb{R}^N$-shifts, and, in the limit Sobolev case, with the continuous dilation group, is due to Lieb [7] and Lions ([8, 10]). The proof in the case of a manifold and of discrete dilations can be found in [15] for $p = 2$. The general case can be proved in a similar way.

In what follows, we assume that the following condition holds:

(**A**) (X, F, D) is a dislocation space, (Y, G) is the same as in Lemma 2.1, $G \in C(X)$, X is continuously imbedded into Y, and D has an extension into Y such that $G \circ g = G$ for all $g \in D$. Moreover, G satisfies the Brezis–Lieb property

$$G(u_k) - G(u) - G(u - u_k) \to 0 \text{ whenever } u_k \rightharpoonup u \text{ in } X. \tag{3.1}$$

Lemma 3.2. *Let (Y, G, D) satisfy assumption* (**A**). *Then the function G is continuous in Y.*

Proof. Let $u_k \to u$ in Y. By Lemma 2.1, $G(u_k - u) \to 0$. By (3.1),

$$\lim G(u_k) - G(u) = \lim G(u_k - u) = 0.$$

The lemma is proved. □

Let

$$c_t = \inf_{u \in X : G(u) = t} F(u), \qquad t > 0. \tag{3.2}$$

Proposition 3.3. *Assume that* (**A**) *and the following condition hold:*

(**B**) *D contains a subsequence $g_k \rightharpoonup 0$.*

Then for any $\tau \in [0, t]$,

$$c_t \leqslant c_\tau + c_{t-\tau}. \tag{3.3}$$

Proof. Fix $\varepsilon > 0$. Let $v, w \in X$ satisfy $G(v) = \tau$, $F(v) \leqslant c_\tau + \varepsilon/2$ and $G(w) = t - \tau$, $F(w) \leqslant c_{t-\tau} + \varepsilon/2$ respectively. Let $g_k \rightharpoonup 0$, and let $u_k = v + g_k w$. Then $G(u_k) \to G(v) + G(w) = t$ by (3.1). On the other hand, $F(u_k) \to F(v) + F(w) \leqslant c_\tau + c_{t-\tau} + \varepsilon$ by (2.5). Since ε is arbitrary, this implies (3.3). □

We show the existence of constrained minima under assumptions of the strict inequality in (3.3) and cocompactness. This is a functional-analytic formalization of analogous results due to Lions.

Lemma 3.4. *Assume that* (**A**) *holds. Let the embedding of X into Y be cocompact. Then for all $a, b > 0$*

$$\inf_{G(au)>b} F(u) > 0. \tag{3.4}$$

Proof. Assume that there is a sequence $u_k \in X$ such that $F(u_k) \to 0$, while $G(au_k) > b$. By Lemma 2.5, $au_k \overset{D}{\rightharpoonup} 0$ and, by the cocompactness of imbedding, $au_k \to 0$ in Y. By Lemma 3.2, G is continuous. Therefore, $G(au_k) \to 0$ and we arrive at a contradiction. □

Theorem 3.5. *Assume that* (**A**) *and* (**B**) *hold. Then for every minimizing sequence u_k for* (3.2), *$t > 0$, there exists a sequence $g_k \in D$ such that $g_k u_k$ converges D-weakly to a point of minimum if and only if for every $\tau \in (0, t)$*

$$c_t < c_\tau + c_{t-\tau}. \tag{3.5}$$

Proof. Note that $c_t > 0$ by (3.4).

Sufficiency. Assume that (3.5) holds. Let $u_k \in X$ be a minimizing sequence, i.e., $F(u_k) = c_t$ and $G(u_k) \to t$. If $u_k \overset{D}{\rightharpoonup} 0$, then $u_k \to 0$ in Y by cocompactness and $G(u_k) \to 0$ by Lemma 3.2, which contradicts $c_t > 0$. Consequently, there exists a sequence $g_k \in D$ such that, on a renamed subsequence, $g_k u_k \rightharpoonup w^{(1)} \neq 0$ in X. By the invariance of the problem, $g_k u_k$ is also a minimizing sequence that we now rename as u_k. Let $g_k^{(n)}$, $w^{(n)}$ be as provided by Theorem 2.6. By (2.7), we have

$$\sum_n F(w^{(n)}) \leqslant c(t),$$

and from the iteration of (3.1) and cocompactness of imbedding it follows that

$$\sum_n G(w^{(n)}) = t.$$

Let $G(w^{(n)}) = \tau_n$. Then

$$\sum_n c_{\tau_n} \leqslant c_t,$$

which, by (3.5), is false unless all but one of the values τ_n is zero. Since $\tau_1 \neq 0$, we conclude that $u_k - w^{(1)} \overset{D}{\rightharpoonup} 0$. By cocompactness, $u_k \to w^{(1)}$ in Y and, by the continuity of G, $G(w^{(1)}) = t$. By the weak lower semicontinuity, $F(u) \leqslant c_t$. Since c_t is the infimum over functions with $G(u) = t$, $w^{(1)}$ is necessarily a minimizer.

Necessity. Assume that (3.5) does not hold for some $0 < \tau < t$. By (3.3), this implies $c_\tau + c_{t-\tau} = c_t$. Let $v_n, w_n \in X$ satisfy respectively $G(v_n) = \tau$, $F(v_n) \leqslant c_\tau + 1/n$ and $G(w_n) = t - \tau$, $F(w_n) \leqslant c_{t-\tau} + 1/n$, $n \in \mathbb{N}$. Let $g_k \rightharpoonup 0$, and let $u_{nk} = v_n + g_k w_n$. Then for every n there exists k_n such that for all $k \geqslant k_n$ we have $\sup_{k\geqslant k_n} |G(u_{nk}) - t| \to 0$ by (3.1) and $\sup_{k\geqslant k_n} |F(u_{nk}) - c_t| \to 0$ by (2.5) as $n \to \infty$. Without loss of generality, $v_n \rightharpoonup v \neq 0$ and $w_n \rightharpoonup w \neq 0$ (if one of v_n and w_n is D-weakly convergent to zero, then $\tau = 0$ or $\tau = t$). Let ψ_j, $j \in \mathbb{N}$ be a basis for X^*. Then

$$\sum_{j\in\mathbb{N}} \frac{|\langle \psi_j, g_{k'_n} w_n\rangle|^2}{2^j} \to 0$$

if $k'_n \geqslant k_n$ are sufficiently large. This implies $g_{k'_n} w_n \rightharpoonup 0$. A similar argument allows us to select a further subsequence such that $g_{k''_n}^{-1} v_n \rightharpoonup 0$. Consequently, w-lim$(v_n + g_{k''_n} w_n) = v \neq 0$, while w-lim$(g_{k''_n}^{-1} v_n + w_n) = w \neq 0$. Thus, we have constructed a minimization sequence that is not D-weakly convergent. □

Note that the proof of sufficiency does not require assumption (**B**).

Theorem 3.6. *Let* (X, F, D) *be a dislocation space. Assume that* (3.5), (**A**), *and* (**B**) *hold. Let* $f, g : X \to \mathbb{R}$ *be nonnegative weakly continuous functions, at least one of them is positive for* $u \neq 0$, *and let*

$$c'_t = \inf_{G(u)+g(u)=t} (F(u) - f(u)), \quad t > 0. \tag{3.6}$$

If for every $\tau \in (0, t)$

$$c'_t < c'_\tau + c_{t-\tau}, \tag{3.7}$$

then every minimizing sequence for (3.6) *converges* D*-weakly to a point of minimum.*

Using an argument repetitive of that in Proposition 3.3, we see that $c'_t \leqslant c'_\tau + c_{t-\tau}$ for any $\tau \in (0,t)$, $t > 0$, so the role of the condition (3.7) is similar to that of (3.5).

Proof. The proof is analogous to that of Theorem 3.5 and of the similar statement in [8]. Let $u_k \in X$ be a minimizing sequence, i.e., $(F-f)(u_k) = c'_t$ and $(G+g)(u_k) \to t$. Let $g_k^{(n)}$, $w^{(n)}$ be as provided by Theorem 2.6. By the iteration of (3.1), taking into account cocompactness, we have

$$g(w^{(1)}) + \sum_n G(w^{(n)}) = t.$$

Let $G(w^{(1)}) + g(w^{(1)}) = \tau_1$, and let $G(w^{(n)}) = \tau_n$, $n \geqslant 2$, so that $\sum \tau_n \leqslant t$. By (2.7), we have

$$\sum_n F(w^{(n)}) - f(w^{(1)}) \leqslant c'(t),$$

which implies

$$c'_{\tau_1} + \sum_{n \geqslant 2} c_{\tau_n} \leqslant c'_t.$$

This contradicts (3.5) and (3.5) unless all but one of the values τ_n is zero. Assume that $\tau_m = 1$ for some $m \geqslant 2$. Then $c_t \leqslant c'_t$, which is false (the opposite strict inequality follows by substituting the minimizer of (3.2) into (3.6)). Consequently, $u_k - w^{(1)} \overset{D}{\rightharpoonup} 0$, $(G+g)(w^{(1)}) = t$, and $(F-f)(w^{(1)}) \leqslant c'_t$, so $w^{(1)}$ is necessarily a minimizer. □

4 Flask Subspaces

Theorem 2.6 can be extended to certain subspaces of a dislocation space which are not D-invariant.

Definition 4.1. Let (X, F, D) be a dislocation space. A subspace X_0 of X is a *flask subspace* if $g_k u_k \rightharpoonup u$, $g_k \in D$, $u_k \in X_0$, implies that $gu \in X_0$ for some $g \in D$.

Proposition 4.2. *Let (X, F, D) be a dislocation space with a flask subspace X_0, and let $u_k \in X_0$ be a bounded sequence. Then Theorem 2.6 holds with $w^{(n)} \in X_0$.*

Proof. Since X_0 is a flask subspace, $g_n w^{(n)} \in X_0$ for some $g_n \in D$, $n \in \mathbb{N}$. Set $\widetilde{w}^{(n)} = g_n w^{(n)}$ and $\widetilde{g}_k^{(n)} = g_k^{(n)} g_n^{-1}$, so that (2.8) holds with $\widetilde{w}^{(n)}$ and $\widetilde{g}_k^{(n)}$. It is easy to see that sequences $\widetilde{g}_k^{(n)}$ satisfy (2.3). □

Corollary 4.3. *Proposition* 3.3*, Lemma* 3.4*, and Theorems* 3.5 *and* 3.6 *remain valid if* X *in* (3.2) *and* (3.6) *is replaced by a flask subspace* X_0*.*

Flask subspaces are a functional-analytic generalization of $H_0^1(\Omega)$ with a flask domain $\Omega \subset \mathbb{R}^N$ in the sense of del Pino–Felmer [13].

Proposition 4.4. *Let* M *be a complete Riemannian manifold, cocompact with respect to a subgroup* G *of its isometry group* $\mathrm{Iso}\,(M)$*. Let* $\Omega \subset M$ *be an open set with a piecewise smooth boundary. If for every sequence* $\eta_k \in G$ *there exists* $\eta \in \mathrm{Iso}\,(M)$ *such that*

$$\liminf \eta_k(\Omega) \subset \eta(\Omega), \tag{4.1}$$

then $W_0^{1,p}(\Omega)$*,* $p > 1$*, is a flask subspace of* $W^{1,p}(M)$ *relatively to the group* $\{u \mapsto u \circ \eta\}_{\eta \in G}$*.*

Proof. First observe that for arbitrary functions if $u_k(x) \to u(x)$ and $u(x) \neq 0$, then necessarily $u_k(x) \neq 0$ for all k sufficiently large. In other words,

$$\{u \neq 0\} \subset \liminf\{u_k \neq 0\}.$$

If $u_k \circ \eta_k \rightharpoonup u$ in $W^{1,p}(M)$, then $u_k \circ \eta_k$ converges almost everywhere as well, and from (4.1) we conclude that for some $\eta \in \mathrm{Iso}\,(M)$, $u = 0$ a.e. on $M \setminus \eta(\Omega)$. In order to apply the Hedberg trace theorem [2] (to regularized u), it remains to note that $u = 0$ on $M \setminus (\overline{\eta(\Omega)}$ and, since $\partial\Omega$ is sufficiently smooth, $u = 0$ on $\eta(\partial\Omega)$ as well, which yields $u \in W_0^{1,p}(\eta(\Omega))$. ☐

5 Compact Imbeddings

This section deals with abstract analogs of sufficient conditions for the compactness of Sobolev imbeddings on unbounded domains (see, for example, [1, 5]).

Proposition 5.1. *Let* (X, F, D) *be a dislocation space, cocompactly imbedded into a Banach space* Y*. Assume that* (**B**) *holds. Let* X_0 *be a subspace of* X*. If for every sequence* $u_k \in X_0$

$$\{g_k\} \subset D, \quad g_k \rightharpoonup 0 \Rightarrow g_k u_k \rightharpoonup 0, \tag{5.1}$$

then the imbedding of X_0 *into* Y *is compact.*

Proof. By (**B**), a sequence $g_k \rightharpoonup 0$ exists. Then, since the sequence $g_k u_k$ is bounded and g_k are isometries, u_k is a bounded sequence. Without loss of generality, it suffices to assume that u_k has the form (2.8). Then (5.1) implies

$u_k - \text{w-lim}\, u_k \overset{D}{\rightharpoonup} 0$. Since the imbedding of X into Y is cocompact, this implies $u_k - \text{w-lim}\, u_k \to 0$ in Y. □

Corollary 5.2. *Let M be a sub-Riemannian manifold of homogeneous dimension Q cocompact with respect to* $\text{Iso}\,(M)$*. If $\Omega \subset M$ is an open set and for any sequence $\eta_k \in \text{Iso}\,(M)$ such that, for some $x_0 \in M$, $\eta_k(x_0)$ has no convergent subsequence,*

$$\liminf \eta_k(\Omega) \text{ has measure zero,}$$

then $W_0^{1,p}(\Omega)$ is compactly imbedded into $L^q(\Omega)$, $1 < p < q < p^*$.

Proof. Since the imbedding in question is cocompact, the statement follows from Proposition 5.1 once we observe that the operator sequence $u \mapsto u \circ \eta_k$, with η_k as above, is weakly convergent to zero. Indeed, if it does not, then, necessarily, there exists a compact set $V \subset M$ such that, for a renamed subsequence, $\bigcup_k \eta_k V$ is a bounded set. Since η_k are isometries, this yields, by Arzela–Ascoli theorem, that a subsequence of η_k is convergent uniformly on compact sets and, in particular, $\eta_k(x_0)$ is convergent, and we arrive at a contradiction. □

The following statement generalizes the well-known compactness for subspaces of radial functions (see, for example, [6]).

Theorem 5.3. *Let (X, F, D) be a dislocation space, cocompactly imbedded into a Banach space Y. Let C be the group of linear automorphisms of X that preserves F such that for every $c \in C \setminus \{\text{id}\}$ and every sequence $g_k \in D$, $g_k \rightharpoonup 0$, we have*

$$g_k^{-1} c g_k \rightharpoonup 0. \tag{5.2}$$

Let

$$X_C = \{u \in X : cu = u, c \in C\}.$$

Then the imbedding of the subspace X_C into Y is compact.

Proof. Let u_k be a bounded sequence in X_C. Consider its expansion (2.8). Then for any $c \in C$ we have $c^{-1} u_k = u_k$ and, consequently,

$$u_k - \sum_n c g_k^{(n)} w^{(n)} \tag{5.3}$$

Assume that there is at least one term $w^{(n)} \neq 0$ with $n \geqslant 2$, say, with $n = 2$. Then, by (5.2),

$$g_k^{(2)^{-1}} c g_k^{(2)} \rightharpoonup 0, \quad c \in C \setminus \{\text{id}\},$$

for every $c, c' \in C$, $c' \neq c$

$$(c'g_k^{(2)})^{-1}cg_k^{(2)} \rightharpoonup 0;$$

furthermore,

$$(cg_k^{(2)})^{-1}u_k \rightharpoonup w^{(2)}, \quad c \in C.$$

Let $M \in \mathbb{N}$, and let C_M be any subset of C with M elements. By (2.4),

$$F(u_k) \geqslant \sum_{c \in C_M} F(w^{(2)}) = MF(w^{(2)}).$$

Since M is arbitrary and the left-hand side is bounded, we arrive at a contradiction. Consequently, $u_k \stackrel{D}{\rightharpoonup} w^{(1)}$. Since the imbedding of X_C into Y is cocompact, $u_k \to w^{(1)}$ in Y. □

Acknowledgement. The topic of this paper was suggested, at different times, by Moshe Marcus and Adimurthi. The author thanks Vladimir Maz'ya for a discussion concerning the use of Hedberg trace theorem in the context of Proposition 4.4. The paper was written as the author was visiting Technion – Haifa Institute of Technology, and he thanks the mathematics faculty there for their warm hospitality.

References

1. Adams, R.A.: Compact imbedding theorems for quasibounded domains. Trans. Am. Math. Soc. **148**, 445–459 (1970)
2. Adams, D.R., Hedberg, L.I.: Function Spaces and Potential Theory. Springer-Verlag (1995)
3. Brezis, H., Lieb, E.: A relation between pointwise convergence of functions and convergence of functionals. Proc. Am. Math. Soc. **88**, 486–490 (1983)
4. Brezis, H., Coron, J.M.: Convergence of solutions of H-systems or how to blow bubbles. Arch. Ration. Mech. Anal. **89**, 21–56 (1985)
5. Clark, C.: An embedding theorem for function spaces. Pacific J. Math. **19**, 243–251 (1966)
6. Esteban, M., Lions, P.-L.: A compactness lemma. Nonlinear Anal., Theory Methods Appl. **7**, 381–385 (1983)
7. Lieb, E.: On the lowest eigenvalue of the Laplacian for the intersection of two domains. Invent. Math. **74**, 441–448 (1983)
8. Lions, P.-L.: The concentration-compactness principle in the calculus of variations. The locally compact case, part 1. Ann. Inst. H. Poincaré, Anal. non Linéaire **1**, 109–1453 (1984)
9. Lions, P.-L.: The concentration-compactness principle in the calculus of variations. The locally compact case, part 2. Ann. Inst. H. Poincaré, Anal. non Linéaire **1**, 223–283 (1984)
10. Lions, P.-L.: The concentration-compactness principle in the calculus of variations. The Limit Case. Rev. Mat. Iberoam. Part 1. **1.1**, 145–201 (1985)
11. Lions, P.-L.: The concentration-compactness principle in the calculus of variations. The Limit Case. Rev. Mat. Iberoam. Part 2. **1.2**, 45–121 (1985)

12. Lions, P.-L.: Solutions of Hartree–Fock equations for Coulomb systems. Commun. Math. Phys. **109**, 33–97 (1987)
13. del Pino, M., Felmer, P.: Least energy solutions for elliptic equations in unbounded domains. Proc. R. Soc. Edinb., Sect. A, Math. **126**, 195–208 (1996)
14. Struwe, M.: A global compactness result for elliptic boundary value problems involving limiting nonlinearities. Math. Z. **187**, 511–517 (1984)
15. Tintarev, K., Fieseler, K.-H.: Concentration Compactness: Functional-Analytic Grounds and Applications. Imperial College Press (2007)

Index