Universitext

Hervé Le Dret

Nonlinear Elliptic Partial Differential Equations

An Introduction

Springer

Universitext

Universitext

Universitext is a series of textbooks that presents material from a wide variety of mathematical disciplines at master's level and beyond. The books, often well class-tested by their author, may have an informal, personal even experimental approach to their subject matter. Some of the most successful and established books in the series have evolved through several editions, always following the evolution of teaching curricula, to very polished texts.

Thus as research topics trickle down into graduate-level teaching, first textbooks written for new, cutting-edge courses may make their way into *Universitext.*

More information about this series at http://www.springer.com/series/223

Hervé Le Dret

Nonlinear Elliptic Partial Differential Equations

An Introduction

Springer

Hervé Le Dret
Laboratoire Jacques-Louis Lions
Sorbonne Université
Paris, France

ISSN 0172-5939 ISSN 2191-6675 (electronic)
Universitext
ISBN 978-3-319-78389-5 ISBN 978-3-319-78390-1 (eBook)
https://doi.org/10.1007/978-3-319-78390-1

Library of Congress Control Number: 2018938796

Mathematics Subject Classification (2010): 35J15, 35J20, 35J25, 35J47, 35J50, 35J57, 35J60, 35J61, 35J62, 35J66, 35J86, 35J87, 35J88, 35J91, 35J92, 47B33, 47H05, 47J05, 47J20, 47J30, 49J40, 49J45, 49J50

Translation from the French language edition: Équations aux dérivées partielles elliptiques non linéaires by Hervé Le Dret. Copyright © Springer-Verlag Berlin Heidelberg 2013. All Rights Reserved.

Printed on acid-free paper

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

In memory of my wife Catherine and of our son Bryan

Preface

This book initially stems from a graduate class I taught at UPMC[1] in Paris between 2004 and 2007. It was first published in French in 2013. I took the opportunity of an English translation to correct the far too many mistakes that had made it through and that I was able to spot this time, to reorganize some parts here and there, to remove several awkward proofs in favor of more streamlined ones, and to add quite a few additional insights and comments. The bulk of the material is, however, essentially the same as that of the 2013 French version, only slightly augmented.

The goal of the book is to present a selection of mathematical techniques that are geared toward solving semilinear and quasilinear partial differential equations and the associated boundary value problems. These techniques are often put to work on examples, and each time, a series of related exercises is proposed. This selection is not meant to be exhaustive by far, nor does it claim to establish a state of the art in the matter. It is conceived more as a basic toolbox for graduate students learning nonlinear elliptic partial differential equations.

The first chapter is a review of results in real and functional analysis, mostly without proofs, although not always, concerning among others integration theory, distribution theory, Sobolev spaces, variational formulations, and weak topologies. It is designed as a sort of vade mecum. The chapter has an appendix that is not required reading for the sequel, but that is meant to satisfy the reader's natural curiosity regarding the somewhat exotic topological vector spaces that tend to crop up in partial differential equation problems.

Chapter 2 is devoted to giving the proofs of the major fixed point theorems: the Brouwer fixed point theorem and the Schauder fixed point theorem. An application of the Schauder fixed point theorem to the resolution of a semilinear partial differential equation is then given.

The focus of Chap. 3 is on superposition operators, which were already introduced in Chap. 2. We study here their continuity, or lack thereof, between various

[1] Now called Sorbonne Université.

function spaces equipped with various topologies. We introduce the concept of Young measures to deal with the case when there is no continuity.

Chapter 4 presents the Galerkin method for solving nonlinear partial differential equations on two examples. The Galerkin method consists in solving finite dimensional approximated problems first and then in passing to the limit when the dimension tends to infinity. The first example is the same semilinear example already solved with fixed point techniques in Chap. 2. The second example is a totally academic example. It is interesting insofar as its nonlinearity shows similarities with that of the stationary Navier-Stokes equations of fluid mechanics.

Chapter 5 is divided into three separate parts. In the first part, we prove several versions of the maximum principle. The second part is a catalogue of elliptic regularity results, mostly without proofs. In the third part, a combination of maximum principle and elliptic regularity is used on an example to prove existence for a semilinear problem via the method of sub- and super-solutions.

We move to an altogether completely different setting in Chap. 6, where we deal with minimization of functionals in the calculus of variations. This is well adapted to solving quasilinear elliptic problems. We consider the scalar case, for which the central idea is convexity. We also treat the vectorial case, which is associated with systems of equations, where subtler convexity variants come into play: quasiconvexity, polyconvexity, and rank-1-convexity.

Chapter 7 offers another take on the calculus of variations, namely the search for critical points of functionals by topological methods. This approach is better suited for semilinear problems, of which we give several examples.

In Chap. 8, we consider quasilinear problems that are not necessarily associated with a functional of the calculus of variations as in Chap. 6. We introduce monotone and pseudo-monotone operators and solve the accompanying variational inequalities. The example of Leray-Lions operators is discussed in detail.

I would like to thank François Murat, whose lecture notes formed the initial basis of the graduate class from which this book evolved in time.

Paris, France
January 2018

Hervé Le Dret

Contents

Chapter 1
A Brief Review of Real and Functional Analysis

This chapter is meant to provide a very quick review of the real and functional analysis results that will be most frequently used afterwards. Missing proofs can be found in most classical works dealing with these questions. Let us just mention here [2, 11, 17, 27, 28, 61, 62], among many others.

1.1 A Little Topological Uniqueness Trick

This brief section is here only to mention a very simple topological trick that will turn out to be extremely useful in the sequel and that we will actually use a fair number of times. It has to do with uniqueness of limits of subsequences of a given sequence and the convergence of that same sequence. For notational brevity, we will denote sequences by x_n, instead of the more correct $(x_n)_{n\in\mathbb{N}}$.

Lemma 1.1. *Let X be a topological space and x_n a sequence in this space which has the property that there exists $x \in X$ such that, given any subsequence of x_n, we can extract from this subsequence a further subsequence that converges to x. Then the whole x_n sequence converges to x.*

Proof. We argue by contradiction. Assume that the sequence x_n does not converge to x. There thus exists a neighborhood V of x and a subsequence x_{n_m} such that $x_{n_m} \notin V$ for all m. By our hypothesis, we can extract from this first subsequence a second subsequence, $x_{n_{m_l}}$, that converges to x when $l \to +\infty$. Consequently, there exists l_0 such that for all $l \geq l_0$, $x_{n_{m_l}} \in V$, which is a contradiction. □

Lemma 1.1 must be used with convergences associated with a topology, which is not always the case for some popular convergences, for instance almost everywhere convergence, see Remark 1.1 later on.

We next turn to a more substantial section on integration theory.

H. Le Dret, *Nonlinear Elliptic Partial Differential Equations*, Universitext,
https://doi.org/10.1007/978-3-319-78390-1_1

1.2 Integration Theory and the Lebesgue Convergence Theorems

Let (X, μ) be a measure space. More often than not, X will be an open subset of $\mathbb{R}^d$ and μ will be the Lebesgue measure on X, which is defined on the completion of the Borel σ-algebra on X by negligible sets.[1] We will most of the time not distinguish between functions defined on X on the one hand, and μ-almost everywhere equality equivalence classes of functions on X on the other hand, unless it turns out to be necessary to make the distinction between the two.

Let us start with Hölder's inequality.

Theorem 1.1. *Let $p, p' \in [1, +\infty]$ be a pair of Hölder conjugate exponents, that is to say two numbers p and p' such that $\frac{1}{p} + \frac{1}{p'} = 1$ (with the convention $\frac{1}{+\infty} = 0$), $f \in L^p(X, d\mu)$ and $g \in L^{p'}(X, d\mu)$. Then $fg \in L^1(X, d\mu)$ and the following inequality holds:*

$$\|fg\|_{L^1(X,d\mu)} \leq \|f\|_{L^p(X,d\mu)} \|g\|_{L^{p'}(X,d\mu)}.$$

When $p = p' = 2$, this inequality is also known as the Cauchy-Schwarz inequality.

The Lebesgue monotone convergence theorem lies at the heart of all useful integral convergence results, even though it is actually fairly rarely used as such in practice.

Theorem 1.2 (Lebesgue's Monotone Convergence Theorem). *Let f_n be an non-decreasing sequence of measurable functions on X with values in $\overline{\mathbb{R}}_+$. This sequence pointwise converges to a measurable function $f : X \mapsto \overline{\mathbb{R}}_+$ and*

$$\int_X f_n \, d\mu \to \int_X f \, d\mu \text{ when } n \to +\infty.$$

Here $\overline{\mathbb{R}}_+$ denotes the set $[0, +\infty]$ equipped with the order topology and all Lebesgue integrals are $\overline{\mathbb{R}}_+$-valued.

Fatou's lemma is a direct consequence of the monotone convergence theorem. Fatou's lemma can be considered, and perhaps more importantly memorized, as as lower semicontinuity result for the Lebesgue integral with respect to pointwise convergence.

Theorem 1.3 (Fatou's Lemma). *Let f_n be a sequence of measurable functions on X with values in $\overline{\mathbb{R}}_+$. Then*

$$\int_X (\liminf_{n\to\infty} f_n) \, d\mu \leq \liminf_{n\to\infty} \Big(\int_X f_n \, d\mu \Big).$$

[1] A subset of $\mathbb{R}^d$ is negligible if it can be included in open subsets of arbitrarily small Lebesgue measure.

Another way of correctly remembering in which direction the inequality in Fatou's lemma goes, is to keep in mind a simple example for which the inequality is strict. Such an example could be $X = \mathbb{R}$, μ the Lebesgue measure on $\mathbb{R}$ and $f_n = \mathbf{1}_{[n,+\infty[}$. We use the notation $\mathbf{1}_A$ for the *characteristic function* or *indicator function* of a subset A of X. This function is defined on X and takes the value 1 on A, and the value 0 on $X \setminus A$.

An almost direct consequence of Fatou's lemma is the celebrated Lebesgue dominated convergence theorem, which is by far the Lebesgue convergence theorem most commonly used in practice.

Theorem 1.4 (Lebesgue's Dominated Convergence Theorem). *Let f_n be a sequence of measurable functions on X with values in $\mathbb{C}$, such that*

$$f_n(x) \to f(x) \text{ when } n \to +\infty \text{ } \mu\text{-almost everywhere on } X,$$

and such that there exists a function g on X with values in $\overline{\mathbb{R}}_+$ which is integrable and such that

$$|f_n(x)| \leq g(x) \text{ } \mu\text{-almost everywhere on } X,$$

so that each f_n is integrable. Then f is integrable and

$$\int_X |f_n - f| \, d\mu \to 0 \text{ when } n \to +\infty.$$

In particular,

$$\int_X f_n \, d\mu \to \int_X f \, d\mu.$$

Let us note that the dominated convergence theorem is primarily an L^1-convergence result. It also has an L^p-version.

Theorem 1.5. *Let $p \in [1, +\infty[$ and f_n be a sequence of functions of $L^p(X, d\mu)$ such that*

$$f_n(x) \to f(x) \text{ when } n \to +\infty \text{ } \mu\text{-almost everywhere on } X,$$

and such that there exists a function $g \in L^p(X, d\mu)$ satisfying

$$|f_n(x)| \leq g(x) \text{ } \mu\text{-almost everywhere on } X.$$

Then $f \in L^p(X, d\mu)$ and

$$\|f_n - f\|_{L^p(X,d\mu)} \to 0 \text{ when } n \to +\infty.$$

We will make repeated use of the following partial converse of the dominated convergence theorem.

Theorem 1.6. *Let $p \in [1, +\infty[$, f_n a sequence of functions of $L^p(X, d\mu)$ and f another function of $L^p(X, d\mu)$, such that $\|f_n - f\|_{L^p(X,d\mu)} \to 0$ when $n \to +\infty$. Then we can extract a subsequence n_m and find $g \in L^p(X, d\mu)$ such that*

$$f_{n_m}(x) \to f(x) \text{ when } m \to +\infty \ \mu\text{-almost everywhere on } X,$$

and

$$|f_{n_m}(x)| \leq g(x) \ \mu\text{-almost everywhere on } X.$$

This result is slightly less well known that the dominated convergence theorem itself, even though it is very useful in many instances. It is a byproduct of the standard proof that $L^p(X, d\mu)$ is a complete metric space.

Be careful that extracting a subsequence is mandatory in the partial converse of the dominated convergence theorem. This can be seen from the following example. We first consider the integer sequence $\phi \colon m \mapsto \frac{m(m-1)}{2}$. This sequence is strictly increasing for $m \geq 1$ and induces a partition of $\mathbb{N}$. We now define a sequence of functions f_n, $n \geq 1$, on $[0, 1]$ by setting $f_n = \mathbf{1}_{\left[\frac{n-1}{m}-\frac{m-1}{2}, \frac{n}{m}-\frac{m-1}{2}\right]}$ whenever $\phi(m) + 1 \leq n \leq \phi(m + 1)$ for some $m \geq 1$. This is unambiguous due to the partition property: any integer $n \geq 1$ falls into one and only one such interval.

The function f_n thus takes the value 0 except on an interval of length $\frac{1}{m}$ where it takes the value 1. Since $m \geq \sqrt{n} - 1$, we see that $f_n \to 0$ in each $L^p(0, 1)$, $p < +\infty$. On the other hand, the $\frac{1}{m}$-long interval where f_n is equal to 1, sweeps the entirety of $[0, 1]$. Indeed, it is successively equal to $\left[0, \frac{1}{m}\right]$, $\left[\frac{1}{m}, \frac{2}{m}\right]$,...,$\left[1 - \frac{1}{m}, 1\right]$, before starting over with length $\frac{1}{m+1}$ at $\left[0, \frac{1}{m+1}\right]$, and so on. In particular, the sequence $f_n(x)$ does not converge for any value of x since it takes both values 0 and 1 infinitely many times, see Fig. 1.1.

It is however easy to extract a subsequence and find a function g that satisfy the conclusions of Theorem 1.6. Indeed, it is enough to take $n_m = \frac{m(m-1)}{2} + 1$ and $g = 1$.

Remark 1.1. This example shows that almost everywhere convergence is not a topological convergence. Indeed, assume that there existed a topology on say L^1 that induced almost everywhere convergence. From every subsequence, we can extract a further subsequence that converges almost everywhere to 0 by Theorem 1.6. The sequence itself however does not converge almost everywhere to anything. This contradicts Lemma 1.1, therefore, there is no such topology. □

One last result from integration theory that we will use quite often is the Lebesgue points theorem, in the following special case. Let us equip $\mathbb{R}^d$ with the canonical Euclidean distance $d(x, y) = \|x - y\|$ where $\|x\|^2 = \sum_{i=1}^{d} x_i^2$, and let $B(x, r)$ denote the open ball centered at x and of radius r. Now here is an instance where

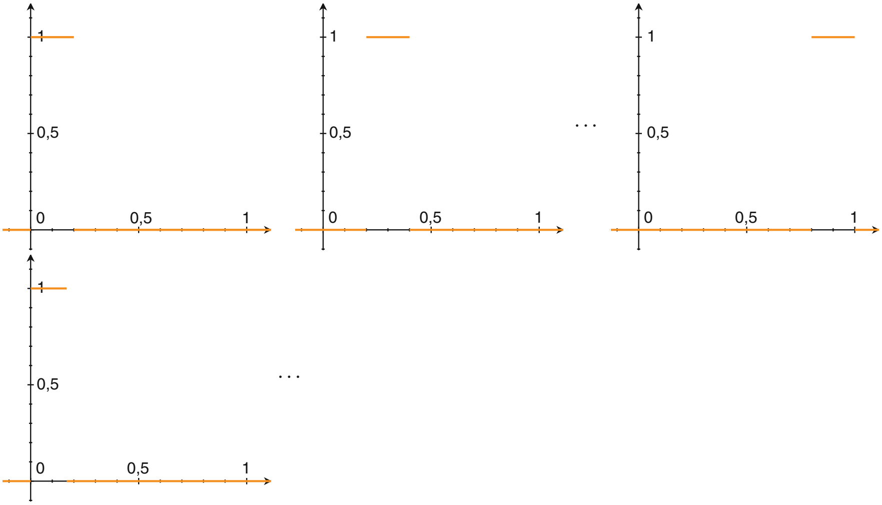


Fig. 1.1 Successive terms in the sequence f_n $(m = 5$ and $m = 6)$

we need to distinguish between function and equivalence class. Here we are talking about actual functions.

Theorem 1.7. *Let f be a locally integrable function on $\mathbb{R}^d$ with respect to the Lebesgue measure. There exists a negligible set N_f such that if $x \notin N_f$, then*

$$\frac{1}{\operatorname{meas}(B(x,r))}\int_{B(x,r)} |f(y) - f(x)|\,dy \to 0 \textit{ when } r \to 0.$$

It follows that

$$\frac{1}{\operatorname{meas}(B(x,r))}\int_{B(x,r)} f(y)\,dy \to f(x) \text{ when } r \to 0 \tag{1.1}$$

for all $x \notin N_f$.

Let us note that the set L_f of points x where the left-hand side quantity of (1.1) converges to some limit when $r \to 0$, only depends on the equivalence class of f for almost everywhere equality, and not on any chosen class representative. Indeed, its definition only involves integrals, which solely depend on the aforementioned equivalence class. Moreover, by Theorem 1.7, the complement of L_f in $\mathbb{R}^d$ is negligible. The set L_f is called the *Lebesgue points set* of (the equivalence class of) f.

Theorem 1.7 thus makes it possible to define a *precise class representative* $\widetilde{f}$ of an equivalence class $f \in L^1_{\text{loc}}(\mathbb{R}^d)$ by setting

$$\widetilde{f}(x) = \begin{cases} \lim\limits_{r\to 0}\Big(\frac{1}{\text{mes}\,(B(x,r))}\int_{B(x,r)} f(y)\,dy\Big) & \text{for } x \in L_f, \\ 0 & \text{for } x \notin L_f, \end{cases}$$

for example, the value 0 for $x \notin L_f$ being arbitrary. When f is the equivalence class of a continuous function, then $L_f = \mathbb{R}^d$ and the precise representative $\widetilde{f}$ is the one and only continuous representative of f.

1.3 Convolution and Mollification

Convolution is a fundamental technique with numerous applications. We will mainly use it as a mollification tool, i.e., a way of smoothing irregular functions defined on the whole of $\mathbb{R}^d$.

Theorem 1.8. *Let f and g be two functions of $L^1(\mathbb{R}^d)$. Then, for almost all x in $\mathbb{R}^d$, the function*

$$y \mapsto f(x-y)g(y)$$

is integrable on $\mathbb{R}^d$ and the function $f \star g$ defined by

$$(f \star g)(x) = \int_{\mathbb{R}^d} f(x-y)g(y)\,dy$$

belongs to $L^1(\mathbb{R}^d)$. This function is called the convolution *of f and g. The convolution operation thus defined is a continuous bilinear mapping from $L^1(\mathbb{R}^d) \times L^1(\mathbb{R}^d)$ into $L^1(\mathbb{R}^d)$, and the following estimate holds:*

$$\|f \star g\|_{L^1(\mathbb{R}^d)} \leq \|f\|_{L^1(\mathbb{R}^d)}\|g\|_{L^1(\mathbb{R}^d)}.$$

Moreover, $f \star g = g \star f$ and $f \star (g \star h) = (f \star g) \star h$, for all f, g and h in $L^1(\mathbb{R}^d)$.

The value of the convolution $f \star g$ at point x can be seen as an average of the values of f around x, weighed by the values of g. Of course, this view of things breaks the symmetry between f and g. Moreover, the "around x" comment is only meaningful when g is somehow concentrated around 0.

The convolution of two functions can be defined more generally when the two functions belong to appropriate $L^p(\mathbb{R}^d)$ spaces. More precisely,

Theorem 1.9. *Let (p, q, r) be a triple of numbers in $[1, +\infty]$ such that*

$$1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q} \text{ with the convention } \frac{1}{+\infty} = 0\cdot \tag{1.2}$$

Let $f \in L^p(\mathbb{R}^d)$ and $g \in L^q(\mathbb{R}^d)$. Then, for almost all x in $\mathbb{R}^d$, the function

$$y \mapsto f(x-y)g(y)$$

is integrable on $\mathbb{R}^d$ and the function $f \star g$ defined by

$$(f \star g)(x) = \int_{\mathbb{R}^d} f(x-y)g(y)\,dy$$

belongs to $L^r(\mathbb{R}^d)$. The mapping $(f, g) \mapsto f \star g$ is a continuous bilinear mapping from $L^p(\mathbb{R}^d) \times L^q(\mathbb{R}^d)$ into $L^r(\mathbb{R}^d)$ with

$$\|f \star g\|_{L^r(\mathbb{R}^d)} \leq \|f\|_{L^p(\mathbb{R}^d)} \|g\|_{L^q(\mathbb{R}^d)}.$$

It is instructive to explore the (p, q) pairs that are admissible for convolution, that to say for which there exists $r \in [1, +\infty]$ satisfying relation (1.2), and the corresponding values of r. Let us notice an interesting particular case, $r = +\infty$, in other words the case when p and q are Hölder conjugates. In this case, not only do we have that $f \star g \in L^\infty(\mathbb{R}^d)$, but in fact $f \star g \in C^0(\mathbb{R}^d)$. In addition, the convolution $f \star g$ tends to 0 at infinity when $1 < p, q < +\infty$, but not necessarily when $p = 1$ and $q = +\infty$.

The definition of convolution given above also works in other contexts, for instance, when $f \in L^1_{\text{loc}}(\mathbb{R}^d)$ and $g \in L^1(\mathbb{R}^d)$ with compact support, see Theorem 1.10 below.

Our primary interest in convolution here comes from the following two results, the conjunction of which makes it possible to approximate nonsmooth functions by smooth functions. This process is called *mollification* or *regularization*. Keep in mind however that convolution has many other applications in analysis.

Theorem 1.10. *Let $k \in \mathbb{N}$. If g belongs to $C^k(\mathbb{R}^d)$ and has compact support, then for all $f \in L^1_{\text{loc}}(\mathbb{R}^d)$, we have $f \star g \in C^k(\mathbb{R}^d)$, and $\partial^\alpha(f \star g) = f \star \partial^\alpha g$ for all multi-indices α such that $|\alpha| \leq k$. If g belongs to $C^\infty(\mathbb{R}^d)$ with compact support, then $f \star g \in C^\infty(\mathbb{R}^d)$.*

The multi-index notation for partial derivatives is as follows. Given a multi-index $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d) \in \mathbb{N}^d$, its length $|\alpha|$ is given by $|\alpha| = \sum_{i=1}^d \alpha_i$ and $\partial^\alpha g = \frac{\partial^{|\alpha|} g}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_d^{\alpha_d}}$ for any function g of class C^k.

If $f \in L^1(\mathbb{R}^d)$, the compact support condition on g can be replaced by an integrability condition.

Theorem 1.11. *Let g be a function of $L^1(\mathbb{R}^d)$ such that $\int_{\mathbb{R}^d} g(y)\,dy = 1$ and $p \in [1, +\infty[$. For all $\varepsilon > 0$, we set*

$$g_\varepsilon(x) = \varepsilon^{-d} g\left(\frac{x}{\varepsilon}\right).$$

Then, for all f in $L^p(\mathbb{R}^d)$, $g_\varepsilon \star f \in L^p(\mathbb{R}^d)$ and

$$\lim_{\varepsilon \to 0} \|g_\varepsilon \star f - f\|_{L^p(\mathbb{R}^d)} = 0.$$

To approximate any function f of $L^p(\mathbb{R}^d)$ in the $L^p(\mathbb{R}^d)$ norm by a sequence of C^∞ functions for the above values of p, it is thus enough to construct one compactly supported C^∞ function, the integral of which is equal to 1. This is actually not too difficult.

Lemma 1.2. *Consider the function θ from $\mathbb{R}$ to $\mathbb{R}$ defined by*

$$\theta(t) = e^{\frac{1}{t-1}} \text{ if } t < 1 \quad \text{and} \quad \theta(t) = 0 \text{ otherwise},$$

and set $g\colon \mathbb{R}^d \to \mathbb{R}$ to be given by

$$g(x) = \frac{1}{\int_{B(0,1)} \theta(|y|^2)\,dy} \theta(|x|^2)$$

for all $x \in \mathbb{R}^d$. Then g is of class C^∞ on $\mathbb{R}^d$, its support is equal to the unit ball $\bar{B}(0,1)$ and its integral over $\mathbb{R}^d$ is equal to 1.

Let us notice that the support of g_ε is $\bar{B}(0,\varepsilon)$. The sequence of functions g_ε concentrates in these smaller and smaller balls when $\varepsilon \to 0$, while keeping their integral equal to 1 due to the multiplicative factor ε^{-d}. Such a sequence g_ε is called a *mollifying sequence* or a *sequence of mollifiers.*

The following remark will make it possible to work not only on the whole of $\mathbb{R}^d$, but also in arbitrary open sets of $\mathbb{R}^d$.

Lemma 1.3. *Let S denote the support of a function $f \in L^1_{\text{loc}}(\mathbb{R}^d)$ and $S_\varepsilon = \{x \in \mathbb{R}^d; d_2(x,S) \leq \varepsilon\}$. Then the support of $g_\varepsilon \star f$ is included in S_ε.*

Proof. Indeed, the support of a locally integrable function is a closed set defined as the complement of the union of all open sets on which f vanishes almost everywhere. Now, if $d_2(x,S) > \varepsilon$, then $B(x,\varepsilon) \cap S = \emptyset$. Consequently, $f = 0$ almost everywhere on $B(x,\varepsilon)$. It follows that

$$g_\varepsilon \star f(x) = \int_{\mathbb{R}^d} g_\varepsilon(x-y) f(y)\,dy = \int_{B(x,\varepsilon)} g_\varepsilon(x-y) f(y)\,dy = 0,$$

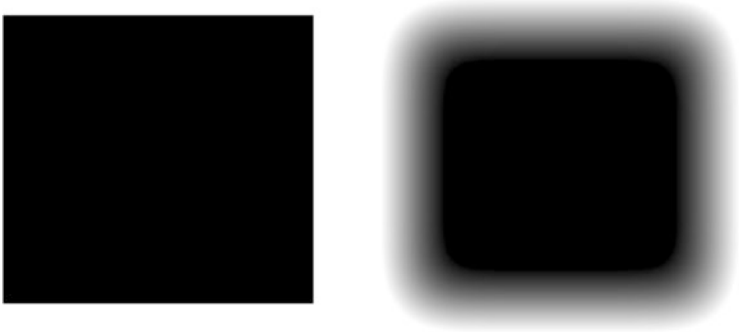


Fig. 1.2 Mollification by convolution

since the support of $y \mapsto g_\varepsilon(x-y)$ is precisely the ball $\bar{B}(x,\varepsilon)$. Hence $g_\varepsilon \star f$ vanishes identically on the complement of S_ε. □

The effect of the convolution by g_ε on the function f, is to "smooth out the rough edges" in the values taken by f, a little bit like blurring a charcoal drawing. This effect is obtained at the price of a slight expansion of the support of f, by a distance of at most ε.

To continue the blurring metaphor, in image processing, a greyscale image is represented by a function from $\mathbb{R}^2$ to [0, 1]. By convention, points colored in black corresponds to points where the function takes the value 0 and points colored in white to points where it takes the value 1. Intermediates values encode various levels of grey. On Fig. 1.2, we see on the left the characteristic function of the outside of the unit square, and on the right, a regularization of this characteristic function by convolution, with a slightly expanded support: the completely black region is smaller than the original unit square.

Let us now consider Ω an arbitrary open subset of $\mathbb{R}^d$. We let $\mathscr{D}(\Omega)$ denote the space of C^∞ functions with compact support in Ω. It is not difficult to prove that there exists an exhaustive sequence of compact subsets in Ω, i.e., a sequence of compacts $K_n \subset \Omega$ such that $K_n \subset \overset{\circ}{K}_{n+1}$ for all n, and $\Omega = \cup_{n\in\mathbb{N}} K_n$.

Theorem 1.12. *The space $\mathscr{D}(\Omega)$ is dense in $L^p(\Omega)$ for all $p \in [1,+\infty[$.*

Proof. We consider an exhaustive sequence of compact subsets K_n in Ω. For all $f \in L^p(\Omega)$, we have that $f\mathbf{1}_{K_n} \to f$ when $n \to +\infty$ in $L^p(\Omega)$ by the dominated convergence theorem. We extend $f\mathbf{1}_{K_n}$ by 0 to the whole of $\mathbb{R}^d$. By Theorem 1.11, we know that $g_\varepsilon \star (f\mathbf{1}_{K_n}) \to f\mathbf{1}_{K_n}$ when $\varepsilon \to 0$ in $L^p(\mathbb{R}^d)$. By Theorem 1.10, these functions are of class C^∞, and by Lemma 1.3, they have support in $(K_n)_\varepsilon$, which is compact. Now, for $\varepsilon < d_2(K_n, \mathbb{R}^d \setminus \Omega)$, it follows that the support of $g_\varepsilon \star (f\mathbf{1}_{K_n})$ is included in Ω, so that its restriction to Ω belongs to $\mathscr{D}(\Omega)$. We conclude by a double limit argument, first taking n large enough, then letting ε tend to 0. □

This result is of course false for $p = +\infty$. The closure of $\mathscr{D}(\Omega)$ in $L^\infty(\Omega)$ is $C_0(\overline{\Omega})$, the set of continuous functions on $\overline{\Omega}$ that vanish on $\partial\Omega$ and tend to 0 at infinity.

1.4 Distribution Theory

Distributions constitute a wide ranging generalization of functions that is relevant in the study of partial differential equations. They are based on the space $\mathscr{D}(\Omega)$, which is a topological vector space.

In day-to-day practice of partial differential equations, it is not necessary to have a complete mastery of the details of the natural topology of $\mathscr{D}(\Omega)$, which is an inductive limit of a sequence of Fréchet spaces. It is more than enough to know what its convergent sequences are. See this Chapter's Appendix for a little more detailed description of this topology.

Proposition 1.1. *A sequence φ_n of functions of $\mathscr{D}(\Omega)$ converges toward $\varphi \in \mathscr{D}(\Omega)$ in the sense of $\mathscr{D}(\Omega)$ if and only if*

i) There exists a compact set $K \subset \Omega$ such that the support of φ_n is included in K for all n,

ii) For all multi-indices α, $\partial^\alpha \varphi_n \to \partial^\alpha \varphi$ uniformly on K when $n \to +\infty$.

As a matter of fact, Proposition 1.1 can be considered as a working definition, which is largely sufficient in practice.

Likewise, there is no need to know the details of the topology of the dual space $\mathscr{D}'(\Omega)$ of $\mathscr{D}(\Omega)$, which is the space of continuous linear forms on $\mathscr{D}(\Omega)$, see the Appendix again. The space $\mathscr{D}'(\Omega)$ is called the *space of distributions on* Ω. We however need to be able to recognize a distribution, i.e., to be able to tell whether a given linear form on $\mathscr{D}(\Omega)$ is continuous or not for this mysterious topology.

Proposition 1.2. *A linear form T on $\mathscr{D}(\Omega)$ is a distribution if and only if for any compact subset K of Ω, there exists an integer n and a constant C such that for all $\varphi \in \mathscr{D}(\Omega)$ with support in K,*

$$|\langle T, \varphi\rangle| \leq C \max_{|\alpha|\leq n, x\in K} |\partial^\alpha \varphi(x)|.$$

Again, this could very well be viewed not as a proposition to be proved, but as a working definition. In general, the integer n and constant C both depend on K. If there exists an integer n in the above estimate that is valid for all K, then the smallest such integer is called the *order* of the distribution T. If no such n exists, the distribution T is said to be of infinite order.

It turns out that the continuity of a linear form on $\mathscr{D}(\Omega)$ is actually equivalent to its sequential continuity. This is not a trivial fact, since the topology of $\mathscr{D}(\Omega)$ is not a metrizable topology.

Proposition 1.3. *A linear form T on $\mathscr{D}(\Omega)$ is a distribution if and only if for all sequences $\varphi_n \in \mathscr{D}(\Omega)$ that converge to φ in the sense of $\mathscr{D}(\Omega)$, there holds*

$$\langle T, \varphi_n \rangle \to \langle T, \varphi \rangle.$$

We can even take $\varphi = 0$ in the above characterization of distributions, by linearity.

The convergence of a sequence of distributions in the space $\mathscr{D}'(\Omega)$, which is also equipped with a natural, but mysterious, topology, is very simply expressed.

Proposition 1.4. *A sequence of distributions T_n converges toward a distribution T in the sense of $\mathscr{D}'(\Omega)$ if and only if for all $\varphi \in \mathscr{D}(\Omega)$, we have*

$$\langle T_n, \varphi \rangle \to \langle T, \varphi \rangle.$$

Distributional convergence is thus just pointwise convergence of linear forms.

Most reasonable function spaces are embedded in the space of distributions. In particular,

Proposition 1.5. *The mapping $\iota \colon L^1_{\text{loc}}(\Omega) \to \mathscr{D}'(\Omega)$ defined by*

$$\forall \varphi \in \mathscr{D}(\Omega), \quad \langle \iota(f), \varphi \rangle = \int_\Omega f\varphi \, dx,$$

is a continuous linear embedding.

Proposition 1.5 thus provides us with a natural way of identifying locally integrable functions—and a fortiori all L^p functions, all continuous functions and so on—with distributions without thinking twice about it. Actually, such an explicit notation as $\iota(f)$ is rarely used, we just write f for the zero order distribution associated with f.

Distributions are indefinitely differentiated by duality.

Proposition 1.6. *Let T be a distribution on Ω. The linear form defined by*

$$\forall \varphi \in \mathscr{D}(\Omega), \quad \Big\langle \frac{\partial T}{\partial x_i}, \varphi \Big\rangle = -\Big\langle T, \frac{\partial \varphi}{\partial x_i} \Big\rangle$$

is a distribution. If T happens to be a class C^1 function, then its partial derivatives in the sense of distributions just defined above are identified with its partial derivatives in the classical sense. Finally, the mapping $T \mapsto \partial T / \partial x_i$ is continuous from $\mathscr{D}'(\Omega)$ into $\mathscr{D}'(\Omega)$.

We can content ourselves with the sequential continuity of the distributional partial derivative operators, which is obvious. Distributions can be multiplied by C^∞ functions, also by duality.

Proposition 1.7. *Let T be a distribution on Ω and $v \in C^\infty(\Omega)$. The linear form defined by*

$$\forall\varphi \in \mathscr{D}(\Omega), \quad \langle vT, \varphi\rangle = \langle T, v\varphi\rangle$$

is a distribution. If T happens to be an L^1_{loc} function, then vT is identified with the pointwise product of v by T in the classical sense.

Owing to the last two propositions, we can apply any linear differential operator with C^∞ coefficients to any distribution, and the result is a distribution. In particular, it makes sense to talk about distributional solutions of a linear partial differential equation with C^∞ coefficients.

A Few Words of Warning A distribution is in general absolutely not a function on Ω (except naturally when it is one in the sense of Proposition 1.5), think about the Dirac mass for example, $\langle\delta, \varphi\rangle = \varphi(0)$. It does not take pointwise nor almost everywhere values in Ω. It cannot be integrated on Ω. Writing such a formula as $\int_\Omega T(x)\varphi(x)\,dx$ instead of a duality bracket $\langle T, \varphi\rangle$ is not acceptable, unless we have checked previously that T is actually in $L^1_{\text{loc}}(\Omega)$, or more rigorously, that T is in the range of the mapping ι of Proposition 1.5. Lastly, convergence in the sense of distributions is not a vaguely magical trick that can be used to justify almost anything.[2] It has a very specific meaning, which is not just hand waving.

Let us point out that, even though a distribution does not take pointwise values, it still retains a local character in the following sense. A distribution on Ω can be restricted to a smaller open set $\omega \subset \Omega$ by dualizing the extension by 0 to Ω of C^∞ functions with compact support in ω. Additionally, a distribution on Ω_1 and another distribution on Ω_2 whose restrictions to $\Omega_1 \cap \Omega_2$ are equal, can be glued together to define a distribution on $\Omega_1 \cup \Omega_2$ (we say that distributions form a *sheaf*), much like functions do.

1.5 Hölder and Sobolev Spaces

A fair amount of the analysis of partial differential equations takes place in Sobolev spaces. Another part takes place in Hölder spaces. More sophisticated function spaces are also used here and there in partial differential equations, but we will not talk about those here. The reader will find various points of view on these function spaces in [2, 11, 15, 17, 36, 49, 57], among many other references in the literature.

Let us talk rapidly about the spaces of Hölder continuous functions. Let Ω be an open subset of $\mathbb{R}^d$. For $0 < \alpha \leq 1$, we say that a real-valued function u on $\overline{\Omega}$ is *Hölder continuous of exponent α* (*Lipschitz continuous* for $\alpha = 1$) if there exists a

[2] By "passing to the limit in the sense of distributions"...

constant C such that, for all pairs of points (x, y) in $\overline{\Omega}$,

$$|u(x) - u(y)| \leq C|x - y|^{\alpha}.$$

We note that such a u is uniformly continuous, so that the above inequality could have been assumed only on Ω with automatic continuous extension to $\partial\Omega$. We set

$$|u|_{C^{0,\alpha}(\overline{\Omega})} = \sup_{\substack{x,y\in\overline{\Omega} \\ x\neq y}} \frac{|u(x) - u(y)|}{|x - y|^{\alpha}},$$

and in the case when $\overline{\Omega}$ is bounded,

$$\|u\|_{C^{0,\alpha}(\overline{\Omega})} = \|u\|_{C^{0}(\overline{\Omega})} + |u|_{C^{0,\alpha}(\overline{\Omega})}.$$

The latter quantity is a norm on the space of Hölder continuous functions $C^{0,\alpha}(\overline{\Omega})$ that makes it into a Banach space. For any integer $k \in \mathbb{N}$, the space $C^{k,\alpha}(\overline{\Omega})$ consists of those functions of class C^k, all the derivatives of which of order k belong to $C^{0,\alpha}(\overline{\Omega})$. It is equipped with the norm

$$\|u\|_{C^{k,\alpha}(\overline{\Omega})} = \|u\|_{C^{k}(\overline{\Omega})} + \max_{|\gamma|=k} |\partial^{\gamma} u|_{C^{0,\alpha}(\overline{\Omega})}$$

for which it is complete.

Let us now turn to Sobolev spaces. There are several different characterizations of these spaces. In the case of Sobolev spaces of nonnegative integer order $k \in \mathbb{N}$, we let

$$W^{k,p}(\Omega) = \{u \in L^p(\Omega);\, \partial^{\alpha} u \in L^p(\Omega) \text{ for all multi-indices } \alpha \text{ such that } |\alpha| \leq k\},$$

with the norm

$$\|u\|_{W^{k,p}(\Omega)} = \Big(\sum_{|\alpha|\leq k} \|\partial^{\alpha} u\|^{p}_{L^p(\Omega)}\Big)^{1/p}$$

for $p \in [1, +\infty[$ and

$$\|u\|_{W^{k,\infty}(\Omega)} = \max_{|\alpha|\leq k} \|\partial^{\alpha} u\|_{L^{\infty}(\Omega)}$$

for $p = +\infty$. The notation $\|\cdot\|_{k,p,\Omega}$ is sometimes used for these norms. The partial derivatives above are naturally meant in the sense of distributions. The $W^{k,p}(\Omega)$ spaces are Banach spaces. For $p = 2$, we also write $W^{k,2}(\Omega) = H^k(\Omega)$, which is a

Hilbert space for the inner product

$$(u|v)_{H^k(\Omega)} = \sum_{|\alpha|\leq k} (\partial^\alpha u|\partial^\alpha v)_{L^2(\Omega)}.$$

The H^k norm is sometimes denoted $\|\cdot\|_{k,\Omega}$. Notice that $W^{0,p}(\Omega) = L^p(\Omega)$ hence the notation $\|\cdot\|_{0,p,\Omega}$ occasionally used for the L^p norm, which becomes $\|\cdot\|_{0,\Omega}$ for L^2.

Clearly, compactly supported C^∞ functions belong to $W^{k,p}(\Omega)$. The closure of $\mathscr{D}(\Omega)$ in $W^{k,p}(\Omega)$ is called $W_0^{k,p}(\Omega)$, for $p < +\infty$, with the notational variant $W_0^{k,2}(\Omega) = H_0^k(\Omega)$. This space is a closed, hence complete, vector subspace of $W^{k,p}(\Omega)$, and is generally a strict subspace thereof for $k \geq 1$. The exceptions to this rule are $\Omega = \mathbb{R}^d$, or $\mathbb{R}^d$ minus a “small” set. The precise meaning of this “smallness” depends on k and p. In particular, $\mathscr{D}(\mathbb{R}^d)$ is dense in $W^{k,p}(\mathbb{R}^d)$ for $p < +\infty$, i.e., $W_0^{k,p}(\mathbb{R}^d) = W^{k,p}(\mathbb{R}^d)$.

By definition, we can approximate any element of $W_0^{k,p}(\Omega)$ in the sense of $W^{k,p}(\Omega)$ by a sequence of functions of $\mathscr{D}(\Omega)$ for any open set Ω. In the same vein, but without compact support,

Theorem 1.13 (Meyers-Serrin Theorem). *The space $C^\infty(\Omega) \cap W^{k,p}(\Omega)$ is dense in $W^{k,p}(\Omega)$, for all $p \in [1, +\infty[$.*

The Meyers-Serrin theorem is proved with a clever use of convolutions. Naturally, functions in $C^\infty(\Omega)$ have no integrability property whatsoever on Ω, hence the intersection with $W^{k,p}(\Omega)$ in the theorem, see [2].

The Meyers-Serrin theorem is valid for any open set, but the question of whether or not functions that are smooth up to the boundary are dense in $W^{k,p}(\Omega)$ depends on the regularity of Ω itself. We start with the definition of a Lipschitz open subset of $\mathbb{R}^d$.

Definition 1.1. An open subset Ω of $\mathbb{R}^d$ is said to be Lipschitz if it is bounded and if its boundary $\partial\Omega$ can be covered by a finite number of open hypercubes C_j, each with an attached system of Cartesian coordinates $y^j = (y_1^j, y_2^j, \dots, y_d^j)$, in such a way that

$$C_j = \{y \in \mathbb{R}^d; |y_i^j| < a_j \text{ for } i = 1, \dots, d\},$$

and for each j, there exists a Lipschitz function $\varphi_j : \mathbb{R}^{d-1} \to \mathbb{R}$ such that

$$\Omega \cap C_j = \{y \in C_j; y_d^j < \varphi_j((y^j)')\},$$

with the notation $\mathbb{R}^{d-1} \ni (y^j)' = (y_1^j, y_2^j, \dots, y_{d-1}^j)$.

Sometimes the open set must be more regular than that, in which case the following definition is also used.

Definition 1.2. A bounded open set $\Omega \subset \mathbb{R}^d$ is said to be of class $C^{k,\alpha}$ if at each point x_0 of $\partial\Omega$, there exists a ball B centered at x_0 and a bijection ψ from B to an open set $D \in \mathbb{R}^d$ such that
i) $\psi(B \cap \Omega) \subset \mathbb{R}^d_+$;
ii) $\psi(B \cap \partial\Omega) \subset \partial\mathbb{R}^d_+$;
iii) $\psi \in C^{k,\alpha}(B; D)$, $\psi^{-1} \in C^{k,\alpha}(D; B)$,
where $\mathbb{R}^d_+ = \{(x_1, x_2, \dots, x_d); x_d \geq 0\}$.

We say that ψ is a $C^{k,\alpha}$-diffeomorphism, in the sense that its inverse is also of class $C^{k,\alpha}$, that locally flattens the boundary. The two definitions of regularity for an open set given above are in a slightly different spirit from each other, but they can naturally be compared, see for example [38]. An open set Ω is said to be of class C^∞, or smooth, if it is of class $C^{k,\alpha}$ for all $k \in \mathbb{N}$.

If Ω is an arbitrary open set, we denote by $C^\infty(\overline{\Omega})$ the space of C^∞ functions on Ω that admit a continuous extension to $\overline{\Omega}$, as well as all their partial derivatives at all orders.

Theorem 1.14. *Let Ω be a Lipschitz open set. Then the space $C^\infty(\overline{\Omega})$ is dense in $W^{1,p}(\Omega)$, for all $p \in [1, +\infty[$.*

The proof uses a localization with a partition of unity, convolution by a mollifying sequence and a translation argument at the boundary, which is where its Lipschitz character intervenes. There are counter-examples to the above density if the Lipschitz hypothesis is omitted. Note that this hypothesis is just a sufficient condition.

We now mention a result of prime importance: the Sobolev embeddings for $\Omega = \mathbb{R}^d$ or Ω smooth. Even though the latter hypothesis may be a little too strong, at least it guarantees that there are no bad surprises…

Theorem 1.15. *Let $k \geq 1$ and $p \in [1, +\infty[$. Then*
i) If $\frac{1}{p} - \frac{k}{d} > 0$, then $W^{k,p}(\Omega) \hookrightarrow L^q(\Omega)$ with $\frac{1}{q} = \frac{1}{p} - \frac{k}{d}$,
ii) If $\frac{1}{p} - \frac{k}{d} = 0$, then $W^{k,p}(\Omega) \hookrightarrow L^q(\Omega)$ for all $q \in [p, +\infty[$ (but not for $q = +\infty$ if $p > 1$),
iii) If $\frac{1}{p} - \frac{k}{d} < 0$, then $W^{k,p}(\Omega) \hookrightarrow L^\infty(\Omega)$. In this case, if $k - \frac{d}{p} > 0$ is not an integer, then $W^{k,p}(\Omega) \hookrightarrow C^{l,\beta}(\overline{\Omega})$ where $l = \lfloor k - \frac{d}{p} \rfloor$ and $\beta = k - \frac{d}{p} - l$ are respectively the integer part and fractional part of $k - \frac{d}{p}$.
All these embeddings are continuous.

Part iii) of the theorem is also known under the name of Morrey's theorem.

If we drop all regularity hypotheses on Ω, the Sobolev embeddings may fail. They however remain true locally, i.e. in any open set compactly included in Ω. In other words, $W^{k,p}(\Omega) \hookrightarrow L^q_{\text{loc}}(\Omega)$, and so on. They remain true globally without regularity hypothesis if we replace $W^{k,p}(\Omega)$ by $W_0^{k,p}(\Omega)$, see [2, 11].

It is useful to rewrite this theorem in the case $k = 1$ and Ω bounded and smooth.

Theorem 1.16. *Let $p \in [1, +\infty[$. Then*
i) If $1 \leq p < d$, then $W^{1,p}(\Omega) \hookrightarrow L^{p^*}(\Omega)$ with $\frac{1}{p^*} = \frac{1}{p} - \frac{1}{d}$, or again $p^* = \frac{dp}{d-p}$,
ii) If $p = d$, then $W^{1,d}(\Omega) \hookrightarrow L^q(\Omega)$ for all $q \in [1, +\infty[$,
iii) If $p > d$, then $W^{1,p}(\Omega) \hookrightarrow C^{0,\beta}(\overline{\Omega})$ where $\beta = 1 - \frac{d}{p}$.

The number p^* is called the *(critical) Sobolev exponent*. Note that if Ω is bounded and smooth, then $W^{1,\infty}(\Omega)$ identifies algebraically and topologically with the space $C^{0,1}(\overline{\Omega})$ of Lipschitz continuous functions on $\overline{\Omega}$.

It is also possible to use the above Sobolev embeddings to obtain that, for $p < d$, $W^{k,p}(\Omega) \hookrightarrow W^{k-1,p^*}(\Omega)$, for example, and so on and so forth. The point is to gain some integrability since $p^* > p$.

Another result of prime importance is a compactness result known as the Rellich-Kondrašov theorem.

Theorem 1.17. *Let Ω be a bounded, smooth open subset of $\mathbb{R}^d$, and $p \in [1, +\infty[$. Then the embeddings*
i) $W^{1,p}(\Omega) \hookrightarrow L^q(\Omega)$ with $1 \leq q < p^*$, if $1 \leq p < d$,
ii) $W^{1,p}(\Omega) \hookrightarrow L^q(\Omega)$ with $1 \leq q < +\infty$, if $p = d$,
iii) $W^{1,p}(\Omega) \hookrightarrow C^{0,\gamma}(\overline{\Omega})$ with $0 \leq \gamma < 1 - \frac{d}{p}$, if $p > d$,
are compact.

In particular, the embedding $W^{1,p}(\Omega) \hookrightarrow L^p(\Omega)$ is compact. An operator is said to be *compact* if it transforms bounded sets into relatively compact sets.

The embeddings remain compact without any regularity hypothesis, but still with Ω bounded, if we replace $L^q(\Omega)$ by $L^q_{\text{loc}}(\Omega)$ and so on in the right-hand side, or $W^{1,p}(\Omega)$ by $W^{1,p}_0(\Omega)$ in the left-hand side, see [2, 11]. They are generally not compact of Ω is unbounded, for instance $\Omega = \mathbb{R}^d$.

Let us now see what we can say about the behavior of functions in $W^{1,p}(\Omega)$ at the boundary of Ω. Remember that functions in $W^{1,p}(\Omega)$ actually are equivalence classes of functions under almost everywhere equality. Case iii) of Theorem 1.16 asserts that if $p > d$, such classes contain a Hölder continuous representative, and thus have a well-defined continuous extension to $\partial\Omega$.

On the other hand, when $p \leq d$, this is no longer true, an equivalence class in $W^{1,p}(\Omega)$ does not need to admit a continuous representative and there is no obvious natural way of ascribing it a reasonable value on the boundary. We will now describe a way of going around this difficulty.

The boundary $\partial\Omega$ of a Lipschitz open set is equipped with a natural $(d-1)$-dimensional measure inherited from Lebesgue's measure on $\mathbb{R}^d$, which we denote by $d\sigma$. This measure allows us to consider L^p spaces on $\partial\Omega$. The boundary also admits an exterior unit normal vector n, which is defined almost everywhere with respect to the surfacic measure $d\sigma$. Both objects can be computed using the functions φ_j of Definition 1.1. The so-called *trace theorem* holds true, see [2, 49].

Theorem 1.18. *Let Ω be Lipschitz open set and $p \in [1, +\infty[$. There exists a unique continuous linear mapping $\gamma_0 \colon W^{1,p}(\Omega) \to L^p(\partial\Omega)$ that extends the restriction mapping $u \mapsto u_{|\partial\Omega}$ defined on the dense subspace $C^1(\overline{\Omega})$ of $W^{1,p}(\Omega)$.*

Of course, the restriction has to be understood as taken on the continuous representative of $u \in C^1(\overline{\Omega})$, when viewed as an element of $W^{1,p}$. The mapping γ_0 thus defined is called the *trace mapping*. It is not onto. Its range is denoted $W^{1-1/p,p}(\partial\Omega)$, or $H^{1/2}(\partial\Omega)$ for $p = 2$. These spaces actually are Sobolev spaces of fractional order, which will not be considered here any further. The kernel of the trace mapping is none other than $W_0^{1,p}(\Omega)$, which qualifies the latter space as being the space of $W^{1,p}$ functions that "vanish on the boundary". Most often, we will just use the case $p = 2$.

From both algebraic and topological points of view, $H^{1/2}(\partial\Omega)$ is isomorphic to the quotient space $H^1(\Omega)/H_0^1(\Omega)$. Since $H_0^1(\Omega)$ is a closed vector subspace of $H^1(\Omega)$, this quotient space is a Banach space for the quotient norm

$$\|\dot{v}\|_{H^1(\Omega)/H_0^1(\Omega)} = \inf_{v \in \dot{v}} \|v\|_{H^1(\Omega)},$$

which can thus be adopted as a norm on $H^{1/2}(\partial\Omega)$ in the form

$$\|g\|_{H^{1/2}(\partial\Omega)} = \inf_{\gamma_0(v)=g} \|v\|_{H^1(\Omega)}.$$

It may not be completely apparent, but this quotient norm is actually a Hilbert norm. There are other equivalent norms on $H^{1/2}(\partial\Omega)$ directly written as integrals over $\partial\Omega$. The space $H^{1/2}(\partial\Omega)$ can also be characterized using the Fourier transform.

The space $H^{1/2}(\partial\Omega)$ is rather delicate to understand intuitively. It is a dense subspace of $L^2(\Omega)$ that, by construction, contains the space of restrictions to $\partial\Omega$ of C^1 functions on $\overline{\Omega}$. It contains unbounded, thus discontinuous functions. For example, if $\Omega =]-e, e[\times]0, 1[$, the function equal to $\ln(|\ln(|x|)|)$ on $[-e, e] \times \{0\}$, 0 elsewhere, is in $H^{1/2}(\partial\Omega)$. On the other hand, functions in $H^{1/2}(\partial\Omega)$ cannot admit discontinuities of the first kind, i.e., jump discontinuities. For instance, the function equal to 0 if $x < 0$, 1 if $x > 0$ on $[-e, e] \times \{0\}$, cannot be the trace of an $H^1(\Omega)$ function (exercise: show it using polar coordinates). See [38, 49] for more details on this space.

Higher order traces are also defined on $W^{k,p}(\Omega)$ for $k > 1$. They extend the first order normal derivative $\frac{\partial u}{\partial n}$ and higher order normal derivatives by continuity. In this respect, it is important to distinguish between $W_0^{k,p}(\Omega)$ et $W^{k,p}(\Omega) \cap W_0^{1,p}(\Omega)$ when $k > 1$. In the first space, all traces up to order $k - 1$ vanish, intuitively, $u = \frac{\partial u}{\partial n} = \cdots = 0$ on $\partial\Omega$, whereas in the second space, only the first trace vanishes.

Let us take note of the following integration by parts formula:

$$\forall u, v \in H^1(\Omega), \quad \int_\Omega \frac{\partial u}{\partial x_i} v \, dx = -\int_\Omega u \frac{\partial v}{\partial x_i} \, dx + \int_{\partial\Omega} \gamma_0(u)\gamma_0(v) n_i \, d\sigma,$$

where n_i is the i-th component of the normal vector n. This formula also goes by various other names, such as Green's formula, and gives rise to many other formulas by repeated application. It is established first for functions in $C^1(\overline{\Omega})$, and then extended by density to $H^1(\Omega)$. As a particular case

$$\forall u \in H^1(\Omega), \quad \int_\Omega \frac{\partial u}{\partial x_i}\,dx = \int_{\partial\Omega} \gamma_0(u) n_i\,d\sigma.$$

We conclude this brief review of Sobolev spaces with Poincaré's inequality.

Theorem 1.19. *Let Ω be an open subset of $\mathbb{R}^d$ that is bounded in one direction. There exists a constant C, which only depends on Ω and p, such that*

$$\forall u \in W_0^{1,p}(\Omega), \quad \|u\|_{L^p(\Omega)} \le C\|\nabla u\|_{L^p(\Omega;\mathbb{R}^d)}.$$

This is again a result first proved for regular functions, then extended by density. It clearly implies that for such an open set, the seminorm $\|\nabla u\|_{L^p(\Omega;\mathbb{R}^d)}$ defines a norm on $W_0^{1,p}(\Omega)$ that is equivalent to the usual $W^{1,p}(\Omega)$ norm. The seminorm is sometimes denoted by $|u|_{1,p,\Omega}$ ($|u|_{1,\Omega}$ for $p = 2$).

1.6 Duality and Weak Convergences in Sobolev Spaces

Let us start with the Lebesgue spaces. For all $p \in [1, +\infty[$, the topological dual space of $L^p(X, d\mu)$ is isometrically identified with $L^{p'}(X, d\mu)$, where the relation $\frac{1}{p} + \frac{1}{p'} = 1$ defines a pair of Hölder conjugate exponents, through the bilinear form $(u, v) \mapsto \int_X uv\,d\mu$. On the other hand, $L^1(X, d\mu)$ is isometrically identified with a strict subspace of $\big(L^\infty(X, d\mu)\big)'$ via the canonical embedding of a space into its bidual, at least for most of the measures μ that can be of interest to us here. Most of the time, for example when X an open set of $\mathbb{R}^d$ and μ is the Lebesgue measure, $L^1(X, d\mu)$ is not a dual space.

We deduce from this that the space $L^p(X, d\mu)$ is reflexive for $1 < p < +\infty$. Consequently, given any bounded sequence u_n in $L^p(X, d\mu)$, we can extract a subsequence u_{n_m} which is weakly convergent to some u. Weak convergence in L^p, $u_n \rightharpoonup u$, means that $\int_X u_n v\,d\mu \to \int_X uv\,d\mu$ for all $v \in L^{p'}(X, d\mu)$.

For $p = +\infty$, X an open set in $\mathbb{R}^d$ and μ the Lebesgue measure, then we can extract from u_n a weakly-star convergent subsequence. This is because L^∞ is the dual of L^1, which is separable. Weak-star convergence in L^∞, $u_n \overset{*}{\rightharpoonup} u$ means that $\int_X u_n v\,d\mu \to \int_X uv\,d\mu$ for all $v \in L^1(X, d\mu)$. Finally, a bounded sequence in $L^1(X, d\mu)$ has in general no weak convergence property without additional hypotheses.[3]

When X is an open subset of $\mathbb{R}^d$ and μ is the Lebesgue measure, these weak and weak-star convergences readily imply convergence in the sense of distributions.

[3] There are several charaterizations of weakly compact subsets of L^1.

More details and more generality on weak and weak-star topologies to be found in Sect. 1.7.

Let us now consider the duality of Sobolev spaces. First of all, the spaces $W^{1,p}(\Omega)$, $1 < p < +\infty$, are reflexive. We can thus extract a weakly convergence subsequence from any bounded sequence. Unfortunately for us, the dual space of $W^{1,p}(\Omega)$ is not so easily identified with another concrete function space. This is hardly a problem, since we have the following characterization of weak convergence in $W^{1,p}(\Omega)$.

Proposition 1.8. *A sequence u_n weakly converges to u in $W^{1,p}(\Omega)$, if and only if there exist $g_i \in L^p(\Omega)$ such that $u_n \rightharpoonup u$ weakly in $L^p(\Omega)$ and $\frac{\partial u_n}{\partial x_i} \rightharpoonup g_i$ weakly $L^p(\Omega)$, $i = 1, \dots, d$. In this case, $g_i = \frac{\partial u}{\partial x_i}$.*

Proof. Let u_n be such that $u_n \rightharpoonup u$ weakly in $W^{1,p}(\Omega)$. Given any continuous linear form ℓ on $W^{1,p}(\Omega)$, we thus have $\ell(u_n) \to \ell(u)$. Let us take any $v \in L^{p'}(\Omega)$. The linear forms $\ell(u) = \int_\Omega uv\,dx$ and $\ell_i(u) = \int_\Omega \frac{\partial u}{\partial x_i} v\,dx$ are continuous on $W^{1,p}(\Omega)$ by Hölder's inequality. Consequently, $u_n \rightharpoonup u$ in $L^p(\Omega)$ and $\frac{\partial u_n}{\partial x_i} \rightharpoonup \frac{\partial u}{\partial x_i}$ in $L^p(\Omega)$.

Conversely, let u_n be such that $u_n \rightharpoonup u$ weakly in $L^p(\Omega)$ and $\frac{\partial u_n}{\partial x_i} \rightharpoonup g_i$ weakly in $L^p(\Omega)$ for $i = 1, \dots, d$. On the one hand, this implies that $u_n \to u$ and $\frac{\partial u_n}{\partial x_i} \to g_i$ in $\mathscr{D}'(\Omega)$. Taking partial derivatives in the sense of distributions is a continuous operation, see Proposition 1.6, therefore $g_i = \frac{\partial u}{\partial x_i}$. On the other hand, it follows from the hypothesis that u_n is bounded in $W^{1,p}(\Omega)$. Since $1 < p < +\infty$, we can extract a subsequence u_{n_m} that converges weakly in the reflexive space $W^{1,p}(\Omega)$ to some $v \in W^{1,p}(\Omega)$. Of course, $u_{n_m} \rightharpoonup u$ in L^p already, so that $v = u$. We conclude by uniqueness of the limit, cf. Lemma 1.1. □

Due to Rellich's theorem, we have in addition that $u_n \to u$ strongly in $L^p_{\text{loc}}(\Omega)$. There is an analogous characterization with stars in the case $p = +\infty$.

The case of $H^1(\Omega)$ is special, since this is a Hilbert space. It is therefore isometric to its dual space by means of its inner product, by Riesz's theorem. We can thus say that $u_n \rightharpoonup u$ weakly in $H^1(\Omega)$ if and only if, for all v in $H^1(\Omega)$,

$$\int_\Omega (u_n v + \nabla u_n \cdot \nabla v)\,dx \to \int_\Omega (uv + \nabla u \cdot \nabla v)\,dx.$$

However, all things considered, Proposition 1.8 is often the most practical.

The duality of $W_0^{1,p}(\Omega)$ is simpler in a sense. Indeed, $\mathscr{D}(\Omega)$ is dense in $W_0^{1,p}(\Omega)$ by definition. It follows that any continuous linear form on $W_0^{1,p}(\Omega)$ defines a distribution which is unique. In other words, there is a canonical embedding of $(W_0^{1,p}(\Omega))'$ into $\mathscr{D}'(\Omega)$. We let $W^{-1,p'}(\Omega)$ denote the image of this canonical embedding.[4] Clearly, a distribution T belongs to $W^{-1,p'}(\Omega)$ if and only if there

[4] This is a negative order Sobolev space, which we have not introduced here.

exists a constant C such that

$$\forall \varphi \in \mathscr{D}(\Omega), \quad |\langle T, \varphi \rangle| \leq C|\varphi|_{1,p,\Omega},$$

because it extends as a continuous linear form to $W_0^{1,p}(\Omega)$, due to the Poincaré inequality. The norm of T in $W^{-1,p'}(\Omega)$, denoted $\|T\|_{-1,p',\Omega}$, is the infimum of all constants C that can appear in the above estimate. It is also the dual norm in the usual abstract sense.

The space $W^{-1,p'}(\Omega)$ can also be characterized in terms of first partial derivatives of $L^{p'}(\Omega) = W^{0,p'}(\Omega)$ functions, which sort of explains the notation. When $p = 2$, the notation becomes $H^{-1}(\Omega)$.

A Word of Caution the space $H_0^1(\Omega)$ is a Hilbert space in its own right with the gradient inner product, due to the Poincaré inequality. It can thus be identified with its dual space via this inner product. This identification, which says that for any continuous linear form ℓ sur $H_0^1(\Omega)$, there exists a unique $v \in H_0^1(\Omega)$ such that $\ell(u) = \int_\Omega \nabla u \cdot \nabla v \, dx$ for all $u \in H_0^1(\Omega)$, is just as legitimate as the previous one. The two identifications $(H_0^1(\Omega))' \simeq H^{-1}(\Omega)$ and $(H_0^1(\Omega))' \simeq H_0^1(\Omega)$ are nonetheless quite different from each other.

For instance, the second identification is not compatible with the identification of the dual space of $L^2(\Omega)$ with itself using its inner product, an identification that is not really open to debate. On the other hand, the identification $(H_0^1(\Omega))' \simeq H^{-1}(\Omega)$ is compatible with it, as well as with the identification of $L^2(\Omega)$ with a subspace of $\mathscr{D}'(\Omega)$.

Indeed, if $T \in L^2(\Omega)$, then

$$\forall \varphi \in \mathscr{D}(\Omega), \quad |\langle T, \varphi \rangle| = \left| \int_\Omega T\varphi \, dx \right| \leq \|T\|_{0,\Omega} \|\varphi\|_{0,\Omega} \leq C\|T\|_{0,\Omega} |\varphi|_{1,\Omega},$$

first by the Cauchy-Schwarz inequality, then by the Poincaré inequality. This estimate shows that $L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$ canonically,[5] with $\|T\|_{-1,\Omega} \leq C\|T\|_{0,\Omega}$, where the constant C is the Poincaré inequality constant. In fact, this embedding is nothing else but the transpose of the canonical embedding of $H_0^1(\Omega)$ into $L^2(\Omega)$. When we use the $H^{-1}(\Omega)$ identification, we find ourselves in the nice diagram

$$\mathscr{D}(\Omega) \hookrightarrow H_0^1(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow H^{-1}(\Omega) \hookrightarrow \mathscr{D}'(\Omega)$$

where all embeddings are continuous, dense and canonical.

This is why in practice $H^{-1}(\Omega)$ is often preferred. Nonetheless, every once in a while, it is more advantageous to use the identification $(H_0^1(\Omega))' \simeq H_0^1(\Omega)$ of Riesz's theorem, which has more to do with variational formulations, see Sect. 1.8 below.

[5] Whereas there is absolutely no canonical embedding of $L^2(\Omega)$ into $H_0^1(\Omega)$!

More generally, when we have to deal with two Hilbert spaces $V \hookrightarrow H$ with a continuous and dense embedding, in order to identify their dual spaces in a compatible way, the same scheme is most often used:

$$V \hookrightarrow H \simeq H' \hookrightarrow V'.$$

Finally a word about the duality of trace spaces, for which we set $H^{-1/2}(\partial\Omega) = (H^{1/2}(\partial\Omega))'$ with a similar identification.

1.7 The Weak and Weak-Star Topologies

We revisit here the definitions and properties of weak and weak-star topologies from a more abstract angle. There will be more detail in an even more general setting in Appendix. Let E be a real Banach space. The topology induced by the norm of E is called the strong topology and E' denotes the topological dual of E, the vector space of all linear forms on E that are continuous for the strong topology. The dual space is also a Banach space for the dual norm $\|\ell\|_{E'} = \sup_{\|x\|_E \le 1} |\ell(x)|$.

The weak topology on E, $\sigma(E, E')$, is the coarsest topology, that is to say the topology with the least amount of open sets, that keeps all strongly continuous linear forms, i.e., all the elements of E', continuous. This is a projective topology in the sense of Definition 1.5, see Appendix. It is a separated topology.

The convergence for the weak topology of E is denoted with the $\rightharpoonup$ sign. It follows from the considerations of Appendix, that $x_n \rightharpoonup x$ if and only if for all $\ell \in E'$, $\ell(x_n) \to \ell(x)$. Moreover, a weakly convergent sequence is bounded—this follows from the Banach-Steinhaus theorem—and $\|x\|_E \le \liminf \|x_n\|_E$. It is also quite clear that if $x_n \rightharpoonup x$ in E and $\ell_n \to \ell$ in E', then $\ell_n(x_n) \to \ell(x)$. Indeed, $\ell_n(x_n) - \ell(x) = \ell_n(x_n) - \ell(x_n) + \ell(x_n) - \ell(x)$, and $|\ell_n(x_n) - \ell(x_n)| \le \|\ell_n - \ell\|_{E'} \|x_n\|_E \to 0$, by the boundedness of x_n.

Let us note that the topological dual of E for the weak topology is algebraically the same as the original one, E', which means that we have obviously not added any new continuous linear form by weakening the topology of E.

A neighborhood basis for 0 for the weak topology is available. The elements of this basis are of the form $\{x \in E; |\ell_i(x)| < \varepsilon, i = 1, \dots, k\}$ with $\ell_i \in E'$, k an integer and $\varepsilon > 0$. It follows that in the infinite dimensional case, a nonempty weak open set is never bounded. In fact, it even contains an infinite dimensional affine space. For example, the neighborhood of 0, $\{x \in E; |\ell_i(x)| < 1, i = 1, \dots, k\}$, contains the space $\cap_{i=1}^k \ker \ell_i$, which is infinite dimensional as a finite intersection of hyperplanes. In particular, any ball has empty weak interior in infinite dimension.

It is not too difficult to show that when E is infinite dimensional, the weak topology does not have a countable basis of neighborhoods of 0. It is therefore not metrizable. On the other hand, in finite dimension, both weak and strong topologies clearly coincide.

Since the weak topology has less open sets than the strong topology, there are a priori more weakly convergent sequences than strongly convergent sequences. More precisely, $x_n \to x$ implies that $x_n \rightharpoonup x$. Conversely, there are less weakly continuous mappings from E with values in another topological space than there are strongly continuous ones. More precisely, if $f\colon E \to F$ is continuous for the weak topology on E, then it is continuous for the strong topology. The converse is generally not true, which is one of the difficulties encountered in nonlinear partial differential equations problems.

A Banach space E is said to be uniformly convex if its norm is such that for all $\varepsilon > 0$, there exists $\delta > 0$ such that if $\max(\|x\|_E, \|y\|_E) \leq 1$ and $\|x - y\|_E > \varepsilon$, then $\left\|\frac{x+y}{2}\right\|_E < 1 - \delta$. A uniformly convex Banach space is reflexive, and has the following very useful property:

Proposition 1.9. *Let E be a uniformly convex Banach space and x_n be a sequence in E such that $x_n \rightharpoonup x$ weakly in E and $\|x_n\|_E \to \|x\|_E$. Then, $x_n \to x$ strongly in E.*

The L^p spaces are uniformly convex for $1 < p < +\infty$, see [11].

A fundamental result concerning the weak topology is the characterization of closed convex sets. The result directly follows from the Hahn-Banach theorem in the second geometrical form, see [11]. It will be very useful when we start talking about questions in the calculus of variations in Chap. 6.

Theorem 1.20. *A convex set C in E is weakly closed if and only if it is strongly closed.*

Proof. Since the weak topology is coarser than the strong topology, every weakly closed set is strongly closed. Therefore, every weakly closed convex set is a strongly closed convex set. The heart of the theorem is in the converse statement.

Let thus C be a strongly closed convex set and let $x_0 \notin C$. According to the geometrical form of the Hahn-Banach theorem, there exists $\ell \in E'$ and $\alpha \in \mathbb{R}$ such that for all $y \in C$, $\ell(x_0) < \alpha < \ell(y)$. The set $V = \{x \in E; \ell(x) < \alpha\}$ is a weak open set, since it is the inverse image of an open set of $\mathbb{R}$ by a weakly continuous mapping, $x_0 \in V$ and $V \cap C = \emptyset$, see Fig. 1.3. Consequently, the complement of C is weakly open and C is weakly closed. □

Notice that this implies that a (strongly or weakly) closed convex set is the intersection of all closed half-spaces that contain it. Notice also that the Hahn-Banach theorem implies immediately that the weak topology is separated, as already said above.

Theorem 1.20 has the following consequence, also known as Mazur's lemma. It ensues from considering the convex hull of a sequence x_n and its strong and weak closures, which are equal.

Corollary 1.1 (Mazur). *Let x_n be a sequence in E that converges weakly to $x \in E$. Then, there exists a sequence of convex combinations of the x_n that converges strongly to x.*

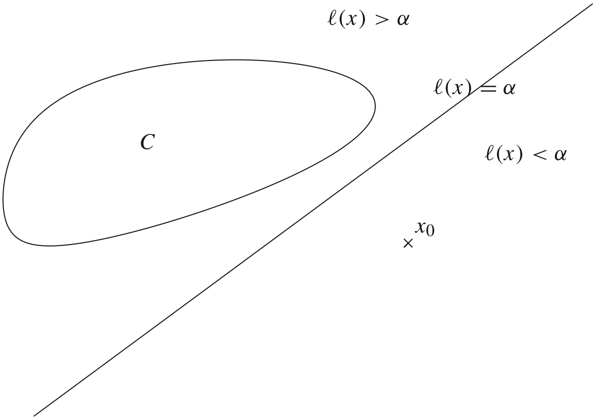


Fig. 1.3 Seeing it is believing it! Point x_0 belongs to a weak open set, itself contained in the complement of C

For example, if we have a sequence on the unit sphere that converges weakly to 0, then we can find a sequence of linear combinations of the values of that initial sequence the norms of which go to 0.

We have mentioned before that the dual space E' of a Banach space E is also a Banach space. It is thus also equipped with the weak topology $\sigma(E', E'')$. This is however not necessarily the most interesting weak sort of topology on E'. The weak-star topology, $\sigma(E', E)$, has better properties in some, but not all, respects. The weak-star topology is the coarsest topology that keeps all the linear forms on E' induced by elements of E continuous, that is to say all mappings of the form $\ell \mapsto \ell(x)$ for a certain $x \in E$. This is also a projective topology and it is also separated. Convergence for this topology is denoted $\overset{*}{\rightharpoonup}$. Naturally, $\ell_n \overset{*}{\rightharpoonup} \ell$ if and only if for all $x \in E$, $\ell_n(x) \to \ell(x)$, a weakly-star convergent sequence is bounded and $\|\ell\|_{E'} \leq \liminf \|\ell_n\|_{E'}$. As was the case for the weak topology, it is fairly clear that if $\ell_n \overset{*}{\rightharpoonup} \ell$ in E' and $x_n \to x$ in E, then $\ell_n(x_n) \to \ell(x)$. In general however, if $\ell_n \overset{*}{\rightharpoonup} \ell$ in E' and $x_n \rightharpoonup x$ in E, $\ell_n(x_n) \not\to \ell(x)$. It should be noted that a general Banach space E has no reason to be a dual space, thus there is no weak-star topology on a Banach space that has no predual.

As a rule, the weak-star topology is coarser than the weak topology on E', and strictly so except when the canonical embedding of E into E'' is an isomorphism, i.e., when E is reflexive. It has less open sets, hence in return, more compact sets than the weak topology. In particular, the following Banach-Alaoglu theorem holds, see [11].

Theorem 1.21. *The closed unit ball of E' is weakly-$*$ compact.*

Corollary 1.2. *If E is reflexive, then the closed unit ball of E is weakly compact.*

Proof. Indeed, in this case $E = (E')'$ and the weak and weak-$*$ topologies on E coincide. □

Remark 1.2. This property is actually a necessary and sufficient condition for reflexivity. □

More generally, we see that

Corollary 1.3. *If E is reflexive and K is a closed bounded convex subset of E, then K is weakly compact.*

Proof. The set K is weakly closed by Theorem 1.20. Since it is bounded, there exists $\lambda \in \mathbb{R}_+$ such that $K \subset \lambda \bar{B}_E$, where $\bar{B}_E$ is the closed unit ball of E. It is thus a weakly closed subset of a weak compact, hence is itself weakly compact. □

We have implicitly used the fairly obvious fact that a scaling is weakly continuous. We can also make statements about sequences, with additional hypotheses of separability. Let us quickly mention a few results in this direction.

Proposition 1.10. *If E is separable, then the restriction of the weak-star topology of E' to its unit ball is metrizable.*

This is a slightly surprising result since in infinite dimension, neither the weak topology as already mentioned before, nor the weak-star topology are metrizable. A simple way of seeing this is to take a sequence $x_k \in E$ such that $x_k \rightharpoonup 0$ but $\|x_k\|_E = 1$.[6] We introduce the double-indexed sequence $x_{k,n} = nx_k$. For n fixed, we have $x_{k,n} \rightharpoonup 0$ when $k \to +\infty$. If the topology was metrizable for a distance d, we would thus have $d(x_{k,n}, 0) \to 0$ when $k \to +\infty$. For all n, we could then choose $k(n)$ such that $d(x_{k(n),n}, 0) \leq \frac{1}{n}$, that is to say $x_{k(n),n} \rightharpoonup 0$. Now, $\|x_{k,n}\|_E = n$. Consequently, the sequence $x_{k(n),n}$ is not bounded and it certainly cannot converge weakly to anything.

This example shows above all that the familiar double approximation argument is in general false for the weak or weak-star topologies, and that one should be wary of it! Except of course if it is known in advance that the sequence remains in a bounded set of the dual space of a separable Banach space.

Corollary 1.4. *If E is separable and ℓ_n is a bounded sequence in E', then there exists a subsequence $\ell_{n'}$ that is weakly-$*$ convergent.*

This result is used very often for instance in $E' = L^\infty(\Omega)$, which is the dual space of the notoriously separable Banach space $E = L^1(\Omega)$, and in the Sobolev spaces built on $L^\infty(\Omega)$.

Corollary 1.5. *If E is reflexive and x_n is a bounded sequence in E, there exists a weakly convergent subsequence $x_{n'}$.*

[6] Such a sequence exists in most reasonable infinite dimensional Banach spaces, the opposite property being rather pathological, even though it happens too. Indeed, the strong and weak topologies are always distinct, and the latter is never metrizable, but it may happen, very rarely, that they have the same convergent sequences. This is the case of the space $\ell^1(\mathbb{N})$, for instance.

There is no need for separability here, since it is enough to work on the closure of the vector space spanned by the elements of the sequence, which is separable. The fact that a closed vector subspace of a reflexive space is reflexive is also used.

A few remarks to close this section. First of all, a strongly closed convex subset of E' is not necessarily weakly-star closed. A simple example of this is $C^0([0,1])$, which is a strongly closed convex subset of $L^\infty(0,1)$ that is not weakly-star closed. It is enough to pointwise approximate the characteristic function of an interval with a sequence of continuous, $[0,1]$-valued functions to see that. Naturally, $C^0([0,1])$ is weakly closed in $L^\infty(0,1)$, which goes to show that the weak topology of $L^\infty(0,1)$, which is induced by its dual space $(L^\infty(0,1))'$, is a rather strange object.

At the other extreme, we remark that the unit ball of $L^1(0,1)$ is not weakly sequentially compact. Consider the sequence $u_n(x) = n\mathbf{1}_{[0,1/n]}(x)$. Assume that there is a subsequence $u_{n'}$ that weakly converges to some u in $L^1(0,1)$. This means that for all $v \in L^\infty(0,1)$, $\int_0^1 u_{n'}v\,dx \to \int_0^1 uv\,dx$. In particular, for $v = 1$, we obtain

$$\int_0^1 u\,dx = \lim_{n'\to+\infty} n' \int_0^{1/n'} dx = 1,$$

on the one hand, and for $v = \mathbf{1}_{[a,b]}$ with $a > 0$

$$\int_a^b u\,dx = \lim_{n'\to+\infty} n' \int_0^{1/n'} \mathbf{1}_{[a,b]}(x)\,dx = 0,$$

on the other hand, since $[0,1/n'] \cap [a,b] = \emptyset$ as soon as $n' > 1/a$. The second equality implies that $u = 0$, which contradicts the first equality. We say in this case that the sequence u_n is not equi-integrable (a necessary condition of weak compactness in $L^1(0,1)$). We can sum up the respective qualities and defects of the various weak topologies in the following table:

	Weak topology on E	Weak-star topology on E'
For	Strongly closed convexes are closed	Compact unit ball
Against	Unit ball not always compact	Strongly closed convexes not always closed

We see that everything goes for the best in the best of all worlds when E is reflexive. A very common example is of course the case of Hilbert spaces.

1.8 Variational Formulations and Their Interpretation

Let us start with the basic result in the context of linear elliptic PDEs, namely the Lax-Milgram theorem, see [11] for instance.

Theorem 1.22 (Lax-Milgram Theorem). *Let V be a Hilbert space, ℓ a continuous linear form on V and a be a continuous bilinear form on V such that there exists $\alpha > 0$ with*

$$\forall v \in V, \quad a(v, v) \geq \alpha \|v\|^2 \text{ (we say that } a \text{ is } V\text{-elliptic).}$$

Then the variational problem: Find $u \in V$ such that

$$\forall v \in V, \quad a(u, v) = \ell(v),$$

has one and only one solution. Moreover, the mapping $\ell \mapsto u$ is linear continuous from V' into V.

This is a fairly simple abstract Hilbert space result. In order to deduce from it existence and uniqueness results for PDE boundary value problems, we need to *interpret* such variational problems. Let us give two simple examples.

Let us first consider the homogeneous Dirichlet problem for the Laplace equation Let Ω be an open bounded subset of $\mathbb{R}^d$. We are looking for a function u such that $-\Delta u = f$ in Ω and $u = 0$ on $\partial\Omega$ in some sense, where the right-hand side f is a given function.

We associate to this boundary value problem the following variational problem: take $V = H_0^1(\Omega)$ which incorporates the homogeneous Dirichlet condition, $a(u, v) = \int_\Omega \nabla u \cdot \nabla v \, dx$ and $\ell(v) = \langle f, v \rangle$ with a given $f \in H^{-1}(\Omega)$. The Lax-Milgram theorem applies, V-ellipticity is an immediate consequence of the Poincaré inequality. In which way does the function u thus found abstractly satisfy the original boundary value problem? This is the point of interpretation.

Concerning the Dirichlet condition, $u = 0$ sur $\partial\Omega$, we have seen that it makes no sense for u in $H^1(\Omega)$. On the other hand, belonging to the subspace $H_0^1(\Omega)$ is an advantageous replacement. Indeed, we have also seen that when Ω is sufficiently regular, then $\gamma_0(u) = 0$. In addition, if u itself is regular, the vanishing of the trace is equivalent to u being actually 0 in the classical sense on $\partial\Omega$. We thus have a natural extension of the homogeneous Dirichlet condition to the case of $H^1(\Omega)$ functions.

Concerning the PDE, we first notice that $H^1(\Omega) \subset \mathscr{D}'(\Omega)$, therefore $-\Delta u$ makes sense as a distribution. In fact, it is easy to see that $-\Delta$ is a continuous operator from $H^1(\Omega)$ into $H^{-1}(\Omega)$. Indeed, using the Einstein repeated indices summation convention,

$$\langle -\Delta u, \varphi \rangle = -\langle \partial_{ii} u, \varphi \rangle = \langle \partial_i u, \partial_i \varphi \rangle = \int_\Omega \nabla u \cdot \nabla \varphi \, dx = a(u, \varphi),$$

for all $\varphi \in \mathscr{D}(\Omega)$, simply by definition of distributional derivatives and the identification of square integrable functions with distributions using the integral. Consequently,

$$|\langle -\Delta u, \varphi\rangle| \leq \|\nabla u\|_{0,\Omega}\|\nabla\varphi\|_{0,\Omega} \leq \|u\|_{1,\Omega}\|\nabla\varphi\|_{0,\Omega},$$

by the Cauchy-Schwarz inequality. We thus see that $u \mapsto -\Delta u$ continuous from $H^1(\Omega)$ to $H^{-1}(\Omega)$.

Since $\mathscr{D}(\Omega) \subset H_0^1(\Omega)$, we can moreover apply the variational formulation using any $\varphi \in \mathscr{D}(\Omega)$ as a test-function. This tells us that $a(u,\varphi) = \ell(\varphi)$ and therefore that

$$\langle -\Delta u, \varphi\rangle = \langle f, \varphi\rangle,$$

for all $\varphi \in \mathscr{D}(\Omega)$. This in turn means that $-\Delta u = f$ in the sense of $\mathscr{D}'(\Omega)$, and actually in the sense of $H^{-1}(\Omega)$, to which both distributions belong. We have just *interpreted* this particular variational problem in terms of a boundary value problem.

Conversely, let us assume that someone has handed us a function $u \in H_0^1(\Omega)$ such that $-\Delta u = f$ in the sense of $\mathscr{D}'(\Omega)$. By going through the previous calculations backwards, we see that this implies that

$$\forall\varphi \in \mathscr{D}(\Omega), \quad a(u,\varphi) = \ell(\varphi).$$

Now, by definition, $\mathscr{D}(\Omega)$ is dense in $H_0^1(\Omega)$. For all $v \in H_0^1(\Omega)$, there exists a sequence $\varphi_n \in \mathscr{D}(\Omega)$ such that $\varphi_n \to v$ in $H_0^1(\Omega)$. Since a and ℓ are continuous, we can pass to the limit in the above equality and obtain that u is a solution of the variational problem. But this solution is unique and given by the Lax-Milgram theorem. In other words, the boundary value problem with the PDE in the sense of distributions has no other solution in $H_0^1(\Omega)$ than the variational solution u.

A second, slightly more delicate example is the non homogeneous Neumann problem, $-\Delta u + u = f$ in Ω and $\frac{\partial u}{\partial n} = g$ on $\partial\Omega$, where f and g are given. We assume here that Ω is Lipschitz.

We have a corresponding variational problem with:

$$V = H^1(\Omega), \quad a(u,v) = \int_\Omega (\nabla u \cdot \nabla v + uv)\,dx$$

and

$$\ell(v) = \int_\Omega fv\,dx + \langle g, \gamma_0(v)\rangle_{H^{-1/2}(\partial\Omega),H^{1/2}(\partial\Omega)},$$

if we take $f \in L^2(\Omega)$ and $g \in H^{-1/2}(\partial\Omega)$. The Lax-Milgram theorem applies trivially.

Let us interpret this problem. We start by obtaining the PDE in the open set by taking test-functions in $\mathscr{D}(\Omega)$. The computations are similar to the previous ones and yield

$$\langle -\Delta u + u, \varphi\rangle = \langle f, \varphi\rangle,$$

for all $\varphi \in \mathscr{D}(\Omega)$, i.e., $-\Delta u + u = f$ in the sense of $\mathscr{D}'(\Omega)$. Indeed, $\gamma_0(\varphi) = 0$. By rewriting $-\Delta u = f - u$, we see that $-\Delta u \in L^2(\Omega)$ and the PDE also holds in the sense of $L^2(\Omega)$. Now, and this is the main difference with the Dirichlet problem, $\mathscr{D}(\Omega)$ is not dense in $H^1(\Omega)$ and we are far at this point from having exploited the entirety of the variational problem.

The interpretation of the Neumann boundary conditions requires the introduction of a new kind of trace. This trace is in a certain sense dual to the previously introduced trace γ_0. Let $H(\Delta, \Omega)$ denote the space of functions in $H^1(\Omega)$, the Laplacian of which in the sense of distributions is in $L^2(\Omega)$. We equip this space with the norm

$$\|v\|_{H(\Delta,\Omega)} = \big(\|v\|^2_{H^1(\Omega)} + \|\Delta v\|^2_{L^2(\Omega)}\big)^{1/2},$$

which makes it a Hilbert space. We have just seen that the solution u of the variational problem is an element of $H(\Delta, \Omega)$.

Proposition 1.11. *There exists a continuous linear mapping γ_1 from $H(\Delta, \Omega)$ into $H^{-1/2}(\partial\Omega)$, called the* normal trace*, such that $\gamma_1(v) = \frac{\partial v}{\partial n}$ for all $v \in C^2(\overline{\Omega})$. It is given by*

$$\forall g \in H^{1/2}(\partial\Omega),\ \langle \gamma_1(v), g\rangle_{H^{-1/2}(\partial\Omega), H^{1/2}(\partial\Omega)} = \int_\Omega \nabla w \cdot \nabla v\,dx + \int_\Omega w\Delta v\,dx, \tag{1.3}$$

where $w \in H^1(\Omega)$ is any function such that $\gamma_0(w) = g$.

Proof. Let us call $\lambda(v, w)$ the right-hand side of (1.3), which is well defined for all $v \in H(\Delta, \Omega)$ and $w \in H^1(\Omega)$. We first show that $\lambda(v, w)$ only depends on $\gamma_0(w)$. By linearity, it is enough to show that it vanishes when $\gamma_0(w) = 0$, that is to say for $w \in H^1_0(\Omega)$. Now, if $\varphi \in \mathscr{D}(\Omega)$, the definition of partial derivatives in the sense of distributions leads to the following formula:

$$\int_\Omega \varphi\Delta v\,dx = \langle \Delta v, \varphi\rangle = -\langle \nabla v, \nabla\varphi\rangle = -\int_\Omega \nabla\varphi \cdot \nabla v\,dx,$$

since $v \in H(\Delta, \Omega)$. Therefore, $\lambda(v, \varphi) = 0$. Also due to the fact that $v \in H(\Delta, \Omega)$, the mapping $w \mapsto \lambda(v, w)$ is continuous on $H^1(\Omega)$. Since $H^1_0(\Omega)$ is the closure of $\mathscr{D}(\Omega)$ in $H^1(\Omega)$, it follows that $\lambda(v, w) = 0$ for all $w \in H^1_0(\Omega)$ by density.

We thus see that the linear form $w \mapsto \lambda(v, w)$ is actually well defined on the quotient space modulo $H^1_0(\Omega)$, which means that it is in fact a linear form on $H^{1/2}(\partial\Omega)$. Let us call it $\gamma_1(v)$. We now show that it is continuous. For all $g \in H^{1/2}(\partial\Omega)$, there holds

$$|\gamma_1(v)(g)| \leq \int_\Omega \|\nabla w\|\|\nabla v\|\,dx + \int_\Omega |w||\Delta v|\,dx \leq \|w\|_{H^1(\Omega)}\|v\|_{H(\Delta,\Omega)},$$

for all w such that $\gamma_0(w) = g$, by the Cauchy-Schwarz inequality. Taking the infimum of the right-hand side with respect to w, we thus obtain that

$$|\gamma_1(v)(g)| \leq \|g\|_{H^{1/2}(\partial\Omega)}\|v\|_{H(\Delta,\Omega)},$$

using the quotient norm for $H^{1/2}$, which shows on the one hand that $\gamma_1(v) \in H^{-1/2}(\partial\Omega)$ and on the other hand that the mapping γ_1, which is clearly linear, is continuous from $H(\Delta,\Omega)$ into $H^{-1/2}(\partial\Omega)$, with norm less than 1.

If we now take v in $C^2(\overline{\Omega})$, the Green formula, which is valid for $v \in H(\Delta,\Omega)$, shows that in this case $\gamma_1(v) = \frac{\partial v}{\partial n}$. □

Let us return to the interpretation of the variational problem. We now take an arbitrary test-function v in $H^1(\Omega)$ and write

$$\int_\Omega \nabla u \cdot \nabla v\,dx + \int_\Omega v\Delta u\,dx + \int_\Omega (u - \Delta u)v\,dx \\ = \int_\Omega fv\,dx + \langle g, \gamma_0(v)\rangle_{H^{-1/2}(\partial\Omega),H^{1/2}(\partial\Omega)}.$$

We have already established that $u - \Delta u = f$ in the previous step, and noticed that $u \in H(\Delta,\Omega)$. It follows that

$$\langle \gamma_1(u), \gamma_0(v)\rangle_{H^{-1/2}(\partial\Omega),H^{1/2}(\partial\Omega)} = \langle g, \gamma_0(v)\rangle_{H^{-1/2}(\partial\Omega),H^{1/2}(\partial\Omega)},$$

for all $v \in H^1(\Omega)$. The trace mapping γ_0 is onto $H^{1/2}(\partial\Omega)$, therefore, the Neumann condition $\gamma_1(u) = g$ is satisfied in the sense of $H^{-1/2}(\partial\Omega)$. Naturally, if we can otherwise establish that u is more regular than just $H^1(\Omega)$,[7] then we obtain the classical form of the Neumann condition.

Just as was the case for the Dirichlet problem, we can walk back through the computations starting from $-\Delta u + u = f$ and $\gamma_1(u) = g$ with $u \in H^1(\Omega)$, to recover the unique solution of the variational problem.

In this section, we have always considered that a variational formulation was given and that we just had to check that it solved the right boundary value problem. What we have not discussed is how to find such an appropriate variational formulation starting from a given boundary value problem. A general rule of thumb

[7] See the *elliptic regularity* results in Chap. 5.

is to assume that every function is regular enough, multiply the PDE by a regular test-function and integrate the result by parts as many times as possible, all the while pretending the formulas hold true, in such a way as to arrive to something that looks formally like a variational formulation. At this point, one has to precisely identify the right Hilbert space, the bilinear form, the linear form, and check that the Lax-Milgram theorem applies.

Once this is done, it is necessary to go back and rigorously interpret the variational problem as in the two examples above, to make sure that the somewhat shaky construction that led to this variational problem, actually led to the right variational problem.

1.9 Some Spectral Theory

Even though spectral theory is not a central theme of the present work, we will nonetheless have to deal with the eigenvalues and eigenfunctions of an elliptic operator every once in a while. Of particular interest will be those of $-\Delta$. So we give here an extremely abbreviated review of the basic concepts of spectral theory. A more in-depth study can be found in [19] for example.

Let A be a continuous linear operator from a Hilbert space H into itself. The *spectrum* of A, $\sigma(A)$, is the set of scalars λ such that $A - \lambda\mathrm{Id}$ is not invertible. We say that λ is an *eigenvalue* of A if $\ker(A - \lambda\mathrm{Id}) \neq \{0\}$, that is to say if $A - \lambda\mathrm{Id}$ is not injective. In this case, every nonzero element of this kernel is called an *eigenvector* associated with the eigenvalue λ. Naturally, in a finite dimensional space, the spectrum only contains eigenvalues. This is not the case in an infinite dimensional space where, in general, it is entirely possible for $A-\lambda\mathrm{Id}$ to be injective without being invertible. Note at this point that it would be generally advisable to complexify the Hilbert space H and talk of the spectrum as a subset of $\mathbb{C}$. We will not do this here because we will only have to deal with situations in which the spectrum is real anyway.

Given a continuous linear operator A on H, it is easy to see that there exists a unique continuous linear operator A^* on H such that

$$\forall (x, y) \in H \times H,\ (Ax|y)_H = (x|A^*y)_H.$$

The operator A^* is called the *adjoint* of A. When A is such that $A^* = A$, we say that A is *self-adjoint*.

The basic spectral theorem for our purposes is a generalisation of the well-known orthogonal diagonalization theorem for symmetric endomorphisms of a Euclidean space, or equivalently of symmetric matrices, to the Hilbert space setting.

Theorem 1.23. *Let A be a compact self-adjoint operator in a Hilbert space H, which is separable and infinite dimensional.*

i) The spectrum of A is the union of $\{0\}$ *and either a sequence* $(\lambda_j)_{j\in\mathbb{N}^*}$ *of real, nonzero eigenvalues that tends to* 0*, or a finite number of nonzero eigenvalues.*

ii) For any nonzero eigenvalue λ, the eigenspace $\ker(A-\lambda\mathrm{Id})$ *is finite dimensional.*

iii) There exists a Hilbert basis $(e_j)_{j\in\mathbb{N}^*}$ of H composed of eigenvectors and

$$\forall x\in H,\quad Ax=\sum_{j\in\mathbb{N}^*}\tilde{\lambda}_j(x|e_j)_H e_j \text{ and } \|Ax\|_H^2=\sum_{j\in\mathbb{N}^*}\tilde{\lambda}_j^2(x|e_j)_H^2,$$

where the family $(\tilde{\lambda}_j)_{j\in\mathbb{N}^*}$ consists either of the sequence of nonzero eigenvalues, or of the finite family of nonzero eigenvalues, counting multiplicities, in both cases possibly completed by the zero eigenvalue. Moreover, $\|A\|_{\mathscr{L}(H)}=\max_{j\in\mathbb{N}^*}|\tilde{\lambda}_j|$.

Let us notice that 0 is always in the spectrum, but is not necessarily an eigenvalue. When it is not an eigenvalue, then we necessarily have a sequence of nonzero eigenvalues with finite dimensional eigenspaces. On the other hand, when there are only a finite number of nonzero eigenvalues, then 0 is necessarily an eigenvalue with infinite dimensional eigenspace. Indeed, in this case, A is of finite rank. So, point i) could also be formulated as "the union of $\{0\}$ and a sequence $(\lambda_j)_{j\in\mathbb{N}^*}$ of real eigenvalues that tends to 0", but we have preferred to distinguish between 0 and nonzero eigenvalues because they do not behave in the same way.

Let us apply all this to the operator $-\Delta$. More precisely, we consider the eigenvalue problem

$$-\Delta\phi=\lambda\phi, \tag{1.4}$$

with $\phi\in H_0^1(\Omega)$ nonzero, thus with a homogeneous Dirichlet condition. In this context, the eigenvector ϕ is rather called an *eigenfunction* associated with the eigenvalue λ. The following result holds:

Theorem 1.24. *Let Ω be a bounded open subset of $\mathbb{R}^d$. There exists a sequence of eigenvalues $(\lambda_j)_{j\in\mathbb{N}^*}$ and eigenfunctions $(\phi_j)_{j\in\mathbb{N}^*}$ such that* $0<\lambda_1<\lambda_2\le\cdots\le\lambda_j\le\cdots$*, with $\lambda_j\to+\infty$ when $j\to+\infty$, $\phi_j\in H_0^1(\Omega)$ and $-\Delta\phi_j=\lambda_j\phi_j$. The family of eigenfunctions is a Hilbert basis of $L^2(\Omega)$ and is orthogonal and complete in $H_0^1(\Omega)$ equipped with the inner product associated with the seminorm.*

To prove this, we apply the general spectral theorem to $H=L^2(\Omega)$ and $A=(-\Delta)^{-1}$ defined via the Lax-Milgram theorem: $Au=v$ if and only if $-\Delta v=u$ and $v\in H_0^1(\Omega)$. This operator is obviously self-adjoint on H, due to the variational formulation. It is compact due to Rellich's theorem, which is why we assumed that Ω is bounded. It does not admit 0 as an eigenvalue. The eigenvalues of $-\Delta$ are the inverses of the eigenvalues of $(-\Delta)^{-1}$ given by Theorem 1.23.

Theorem 1.24 is a little more specific than the general theorem in that it indicates that the eigenvalues are all strictly positive. This fact follows immediately from the variational formulation. It also says that the family of eigenfunctions is orthogonal and complete in H_0^1, which also follows from the variational formulation. It moreover asserts that $\lambda_1 < \lambda_2$, which is a roundabout way of saying that the first eigenvalue λ_1 is simple: $\dim(\ker(-\Delta - \lambda_1 \mathrm{Id})) = 1$. The other eigenvalues can be multiple eigenvalues, for example when Ω possesses geometrical symmetries. In the same spirit, the following finer result holds:

Proposition 1.12. *There exists* $\phi_1 \in \ker(-\Delta - \lambda_1 \mathrm{Id})$ *such that* $\phi_1(x) > 0$ *for all* x *in* Ω.

In other words, the eigenfunctions associated with the first eigenvalue do not vanish in Ω and we can thus choose one that is strictly positive in Ω.[8] The result is a consequence of the Krein-Rutman theorem, see [63]. As a rule, eigenfunctions associated with the other eigenvalues do vanish on so-called nodal sets and change sign.

Appendix: The Topologies of $\mathscr{D}$ and $\mathscr{D}'$

In the literature of applied partial differential equations, it is customary not to dwell too much on the description of the topologies of $\mathscr{D}(\Omega)$ and $\mathscr{D}'(\Omega)$. It is true that there is no need to know them in detail in order to work efficiently. The sequential properties sketched above are amply sufficient. There is thus no real drawback in not reading the rest of the present section.

On the other hand, it is quite legitimate to be curious with regard to these topologies, without being willing to read through the theory of abstract topological vector spaces, which is rather imposing, see [8, 63, 68]. In effect, we are dealing with topological vector spaces that are not normed spaces, but are naturally equipped with much more sophisticated topologies.

A topological vector space on $\mathbb{R}$, or more generally on a topological field $\mathbb{K}$, is a $\mathbb{K}$-vector space E equipped with a topology which is such that the addition is continuous from $E \times E$ to E and the scalar multiplication is continuous from $\mathbb{K} \times E$ to E, both product spaces being equipped with their product topology. We will stick with $\mathbb{K} = \mathbb{R}$ from now on.

Let us start with the concept of *Fréchet space*. A seminorm on a vector space on $\mathbb{R}$ is a mapping with nonnegative values that is absolutely homogeneous and satisfies the triangle inequality, i.e., is subadditive.

Let E be an $\mathbb{R}$-vector space and $(p_n)_{n\in\mathbb{N}}$ a countable, nondecreasing family of seminorms on E such that for all $u \neq 0$ in E, there exists $n \in \mathbb{N}$ with $p_n(u) > 0$. For all $n \in \mathbb{N}$ and $\alpha > 0$, we let

$$V_{n,\alpha}(u) = \{v \in E;\ p_n(v - u) < \alpha\}.$$

[8] Elliptic regularity, see Chap. 5, implies that the eigenfunctions are smooth.

We define a family $\mathscr{O}$ of subsets of E by saying that

$$U \in \mathscr{O} \text{ if and only if } \forall u \in U, \exists n \in \mathbb{N}, \exists \alpha \in \mathbb{R}_+^*, V_{n,\alpha}(u) \subset U.$$

This is reminiscent of the way a topology arises from a distance in a metric space, and the sets $V_{n,\alpha}(u)$ are meant to evoke neighborhoods of u, except that we do not have one distance, but a whole family of seminorms.

Proposition 1.13. *The family $\mathscr{O}$ is a topology on E which makes it a topological vector space. This topology is said to be* generated by the family of seminorms. *This topology is metrizable and the mapping $E \times E \to \mathbb{R}_+$,*

$$d(u,v) = \sum_{n=0}^{\infty} 2^{-n} \min(1, p_n(u-v)), \tag{1.5}$$

is a distance that also generates the topology.

Proof. Let us check that the axioms of a topology are satisfied. Trivially, $E \in \mathscr{O}$ and $\emptyset \in \mathscr{O}$, since in the latter case, the condition to be met is empty.

Let $(U_i)_{i=1,\dots,k}$ be a finite family of elements of $\mathscr{O}$ and $U = \cap_{i=1}^k U_i$.[9] Let us take $u \in U$. By definition, for all $i = 1, \dots, k$, there exists $n_i \in \mathbb{N}$, $\alpha_i > 0$ such that $V_{n_i,\alpha_i}(u) \subset U_i$. We set $n = \max\{n_i\} \in \mathbb{N}$ and $\alpha = \min\{\alpha_i\} > 0$. Now the sequence p_n is nondecreasing, so that the inequalities

$$p_{n_i}(v-u) \leq p_n(v-u) < \alpha \leq \alpha_i,$$

show that $V_{n,\alpha}(u) \subset V_{n_i,\alpha_i}(u) \subset U_i$ for all i. Consequently, $V_{n,\alpha}(u) \subset U$, which implies that $U \in \mathscr{O}$.

Let now $(U_\lambda)_{\lambda \in \Lambda}$ be an arbitrary family of elements of $\mathscr{O}$ and $U = \cup_{\lambda \in \Lambda} U_\lambda$. Let us take $u \in U$. By definition, there exists $\lambda \in \Lambda$ such that $u \in U_\lambda$. We can thus choose an n and an α such that $V_{n,\alpha}(u) \subset U_\lambda$. Trivially, $U_\lambda \subset U$ and thus $U \in \mathscr{O}$.

We are thus assured that we are dealing with a topology. Let us check that this topology is a topological vector space topology, i.e., that the vector space operations are continuous.

For the addition, let us be given v_1 and v_2 in E, and let $u = v_1 + v_2$. We consider an open set U that contains u. There exist n and α such that $V_{n,\alpha}(u) \subset U$. We claim that $V_{n,\alpha/2}(v_1) + V_{n,\alpha/2}(v_2) \subset V_{n,\alpha}(u)$, which implies the continuity of the addition at point (v_1, v_2). Indeed, if $w_1 \in V_{n,\alpha/2}(v_1)$ and $w_2 \in V_{n,\alpha/2}(v_2)$, then

$$\begin{aligned} p_n(w_1 + w_2 - u) &= p_n(w_1 - v_1 + w_2 - v_2) \\ &\leq p_n(w_1 - v_1) + p_n(w_2 - v_2) < \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha \end{aligned}$$

by the triangle inequality.

[9] The empty family has zero element, it is also finite, and its intersection is E.

For the scalar multiplication, we note first that for all λ, μ, u, and v,

$$\mu v - \lambda u = \mu(v - u) + (\mu - \lambda)u.$$

Let U be an open set containing λu and $V_{n,\alpha}(\lambda u)$ be an open neighborhood of λu that is included in U. We wish to find $\eta > 0$ and $\beta > 0$ such that if $(\mu, v) \in]\lambda - \eta, \lambda + \eta[\times V_{n,\beta}(u)$, then $\mu v \in V_{n,\alpha}(\lambda u)$.

It follows from the above decomposition that

$$p_n(\mu v - \lambda u) \leq |\mu| p_n(v - u) + |\mu - \lambda| p_n(u)$$

by absolute homogeneity of the seminorm and the triangle inequality. There are two cases. First of all, if $p_n(u) > 0$, an obvious choice for η is then $\eta = \frac{\alpha}{2p_n(u)}$, which yields

$$|\mu - \lambda| p_n(u) < \frac{\alpha}{2}.$$

If $p_n(u) = 0$, then we take $\eta = 1$, and the above estimate still holds true. Then, since $|\mu| < |\lambda| + \eta$, we can take $\beta = \frac{\alpha}{2(|\lambda|+\eta)}$ and obtain

$$|\mu| p_n(v - u) < \frac{\alpha}{2},$$

thus establishing the continuity of scalar multiplication.

Let us finally show that this topological vector space topology is metrizable. It is not difficult to check that d defined by formula (1.5) is a distance on E. This is where the hypothesis that for all $u \neq 0$, there exists n such that $p_n(u) > 0$, comes into play, in order to ensure that $d(v, w) = 0$ if and only if $v = w$. To conclude, we must show that for all $u \in E$, every open set of $\mathscr{O}$ containing u contains an open ball of d also containing u, and conversely. Since all the quantities involved are translation invariant, it is enough to prove this for $u = 0$. We denote by $B(0, \alpha)$ the open ball of d centered at 0 and of radius α.

Let us thus be given an open set U of $\mathscr{O}$ that contains 0. By definition, there exists $n \in \mathbb{N}$ and $\alpha > 0$ such that $V_{n,\alpha}(0) \subset U$. We want to find $\beta > 0$ such that $B(0, \beta) \subset V_{n,\alpha}(0)$. We notice that it is enough to consider the case $\alpha < 1$, since $V_{n,\alpha}(0) \subset V_{n,\alpha'}(0)$ as soon as $\alpha \leq \alpha'$. We take $\beta = 2^{-n}\alpha$. If $v \in B(0, \beta)$, then

$$2^{-n} \min(1, p_n(v)) \leq d(0, v) < 2^{-n}\alpha.$$

Since $\alpha < 1$, the minimum in the left-hand side is not 1 and it follows that $p_n(v) < \alpha$, or in other words $v \in V_{n,\alpha}(0)$, hence $B(0, \beta) \subset V_{n,\alpha}(0)$.

Conversely, let us be given $\beta > 0$ and consider the ball $B(0, \beta)$. For all $n \in \mathbb{N}$, there holds

$$d(0, v) = \sum_{k=0}^{n} 2^{-k} \min(1, p_k(v)) + \sum_{k=n+1}^{\infty} 2^{-k} \min(1, p_k(v)).$$

Now $\min(1, p_k(v)) \leq 1$, therefore

$$\sum_{k=n+1}^{\infty} 2^{-k} \min(1, p_k(v)) \leq 2^{-n}.$$

We pick n large enough so that $2^{-n} < \beta/2$. Since $\min(1, p_k(v)) \leq p_k(v)$ and the sequence of seminorms is nondecreasing, we have

$$\sum_{k=0}^{n} 2^{-k} \min(1, p_k(v)) \leq 2p_n(v).$$

We then choose $\alpha = \beta/4$. It follows that if $v \in V_{n,\alpha}(0)$, then

$$d(0, v) \leq 2p_n(v) + \beta/2 < \beta,$$

that is to say $V_{n,\alpha}(0) \subset B(0, \beta)$. □

More generally, the topological vector spaces whose topology is generated by an arbitrary family of seminorms, i.e., not necessarily a countable family, are called *locally convex topological vector spaces*. It is (almost[10]) clear that the countability of the seminorm family plays no role in the fact that we are dealing with a topological vector space. The countability only comes into play for metrizability. An equivalent characterization is the existence of a basis of convex neighborhoods of 0, hence the name. The equivalence uses the gauge or Minkowski functional of a convex set, see Chap. 2. All the topological vector spaces considered here fall into this category.

Definition 1.3. We say that a vector space E equipped with a countable family of seminorms as above is a *Fréchet space* if it is complete for the distance (1.5).

Remark 1.3. Fréchet spaces provide an example of the usefulness of the concept of metrizable space and the subtle distinction made with metric spaces: their topology is naturally defined using neighborhoods. It so happens that there is a distance that generates the same topology, but this distance is not especially natural. Besides,

[10] Well, we have used the nondecreasing character of the family in places to speed things up. This has to be worked around in the general case. Note that the condition on $p(u) > 0$ is not necessary in general, a locally convex topology does not need to be separated.

there are many other distances that are equivalent to it. When we work in a Fréchet space, we rarely use the distance explicitly. To the contrary, seminorms and associated neighborhoods will be used. Note that a Fréchet space being metrizable and complete, it is a Baire space. □

Definition 1.4. A subset A of a topological vector space E is said to be *bounded* if for any neighborhood V of 0, there exists a scalar λ such that $A \subset \lambda V$ (we say that A is absorbed by any neighborhood of 0).

Caution The concept of bounded subset just introduced is not a metric concept, but a topological vector space concept. As a matter of fact, the distance d defined above in the case of a countable family of seminorms is itself bounded. Clearly, the diameter of E is less than 2, which is not what we have in mind for a bounded subset of a topological vector space. The two concepts however coincide in the case of a normed vector space when we consider the distance that is canonically associated with the norm, $d(u, v) = \|u - v\|$.

Bounded subsets of E are easily characterized in terms of the seminorms used to define the topology.

Proposition 1.14. *Let E be a topological vector space the topology of which is generated by a family of seminorms $(p_n)_{n\in\mathbb{N}}$. A subset A of E is bounded if and only if for all n, there exists a constant λ_n such that for all $u \in A$, $p_n(u) < \lambda_n$.*

Proof. Let A be a bounded subset of E. For all n, $V_{n,1}(0)$ is a neighborhood of 0. Thus, by definition, for all n, there exists λ_n such that $A \subset \lambda_n V_{n,1}(0) = V_{n,\lambda_n}(0)$, which means that for all $u \in A$, $p_n(u) < \lambda_n$.

Conversely, let A satisfy the condition of the Proposition. Let U be a neighborhood of 0. It contains a neighborhood of the form $V_{n,\alpha}(0)$, for some n and α. By hypothesis, for all $u \in A$, $p_n(u) < \lambda_n$. In particular $\lambda_n > 0$. Consequently, $p(\alpha u/\lambda_n) < \alpha$, by absolute homogeneity. This says that $\frac{\alpha}{\lambda_n}u \in V_{n,\alpha}(0)$, which in turn readily implies that $A \subset \frac{\lambda_n}{\alpha} V_{n,\alpha}(0) \subset \frac{\lambda_n}{\alpha} U$. □

A subset of E is thus bounded if and only if it is bounded for every seminorm. Here too, the countability of the family plays no role in the above characterization of bounded subsets.

If the sequence of seminorms is stationary, i.e., if there exists $n_0 \in \mathbb{N}$ such that $p_n = p_{n_0}$ for all $n \geq n_0$, then it is easily checked that p_{n_0} is a norm that generates the topology. We have thus only introduced a possibly new object compared to normed spaces, or Banach spaces in the complete case, if the sequence p_n is not stationary. The latter condition is however not sufficient to ensure that a given Fréchet space is not normable.

Proposition 1.15. *Let E be a topological vector space, the topology of which is generated by a nondecreasing family of seminorms $(p_n)_{n\in\mathbb{N}}$ such that for all n, there exists $m > n$ such that p_m is not equivalent to p_n. Then the topology of E is not normable.*

Proof. We have $p_n \leq p_m$, therefore saying that p_m is not equivalent to p_n means that $\sup\{p_m(u); u \in E, p_n(u) < 1\} = +\infty$. If a topological vector space is normable, it has a bounded set with nonempty interior, namely the unit ball of a norm that generates the topology. We are thus going to show that every bounded set of E has empty interior.

Let thus A be a bounded subset of E and λ_n, $n \in \mathbb{N}$, the scalars that express its boundedness in terms of the seminorms. Let U be an open set included in A and assume that U is nonempty. It thus contains a neighborhood $V_{n,\alpha}(u)$ for some triple (n, α, u) with $\alpha > 0$. We take $m > n$ as above. For all $v \in V_{n,\alpha}(u)$, we see that $w = \frac{v-u}{\alpha} \in V_{n,1}(0)$, and conversely if $w \in V_{n,1}(0)$, then $v = u + \alpha w \in V_{n,\alpha}(u)$. By hypothesis, there exists $w \in V_{n,1}(0)$ such that $p_m(w) \geq \frac{\lambda_m + p_m(u)}{\alpha}$. It follows that $p_m(v) \geq \alpha p_m(w) - p_m(u) \geq \lambda_m$ by the triangle inequality, with $v \in V_{n,\alpha}(u) \subset U \subset A$. This is a contradiction, therefore U is empty. □

Conversely, it is not hard to see that if all seminorms are equivalent starting from a certain rank, then the topology is normable by one of these equivalent seminorms.

Since the topology of a Fréchet space is metrizable, it can also be worked out in terms of convergent sequences. These sequences are very easily described.

Proposition 1.16. *Let E be a topological vector space with a countable family of seminorms as above. A sequence u_k tends to u in the sense of E if and only if, $p_n(u_k - u) \to 0$ when $k \to +\infty$, for all $n \in \mathbb{N}$.*

Proof. A sequence u_k tends to u if and only if, for any neighborhood V of u, there exists k_0 such that $u_k \in V$ for all $k \geq k_0$. This happens if and only if for all n and $\alpha > 0$, there exists k_0 such that $u_k \in V_{n,\alpha}(u)$, that is to say $p_n(u_k - u) < \alpha$, for all $k \geq k_0$. □

Of course, for such a sequence $d(u_k, u) \to 0$ and conversely, which is a simple exercise when we do not know yet that both topologies coincide.

Let us stop here with generalities about Fréchet spaces and introduce our main example in the context of distributions.

Proposition 1.17. *Let Ω be an open subset of $\mathbb{R}^d$ and K a compact subset of Ω with nonempty interior. The space*

$$\mathscr{D}_K(\Omega) = \{\varphi \in C^\infty(\Omega); \operatorname{supp} u \subset K\}$$

of indefinitely differentiable functions with support in K, equipped with the family of seminorms

$$p_n(\varphi) = \max_{|\gamma| \leq n, x \in K} |\partial^\gamma \varphi(x)|, \tag{1.6}$$

is a Fréchet space.

Proof. The family of seminorms is clearly nondecreasing. Moreover, p_0 is a norm, so that $p_0(\varphi) > 0$ as soon as $\varphi \neq 0$. The only difficulty is the completeness.

Let thus $\varphi_k \in \mathscr{D}_K(\Omega)$ be a Cauchy sequence. We notice that p_n actually is the norm on $C^n_K(\Omega)$. Therefore, if φ_k is Cauchy in $\mathscr{D}_K(\Omega)$, it is *a fortiori* Cauchy in $C^n_K(\Omega)$ for all n. Now $C^n_K(\Omega)$ is complete, and $C^{n+1}_K(\Omega) \hookrightarrow C^n_K(\Omega)$. It follows that φ_k converge in $C^n_K(\Omega)$ toward some φ, which is the same for all n. Hence, $\varphi \in \mathscr{D}_K(\Omega)$, and according to Proposition 1.16, the sequence φ_k converges to φ in $\mathscr{D}_K(\Omega)$. □

Let us remark in passing that Proposition 1.16 translates in this particular case into the fact that a sequence converges in $\mathscr{D}_K(\Omega)$ if and only if all its partial derivatives at all orders converge uniformly on K.

The space $\mathscr{D}_K(\Omega)$ is not normable because its family of seminorms satisfies the hypothesis of Proposition 1.15. There is no norm that can generate its Fréchet space topology, at least we have not worked for nothing. This can also be seen as a consequence of the next proposition, which can be slightly surprising at first when we only know about infinite dimensional normed spaces.

Proposition 1.18. *Bounded closed subsets of $\mathscr{D}_K(\Omega)$ are compact.*

Proof. Let A be a bounded subset of $\mathscr{D}_K(\Omega)$. For all n, there thus exists λ_n such that

$$\forall \varphi \in A, \quad p_n(\varphi) \leq \lambda_n.$$

By the mean value inequality, this implies that $\partial^\gamma A$ is an equicontinuous family in $C^0_K(\Omega)$ for all multi-indices $|\gamma| \leq n-1$, and that $\max_K |\partial^\gamma \varphi| \leq \lambda_n$ for all $\varphi \in A$. Since K is compact, we can apply Ascoli's theorem to deduce that all these sets are relatively compact in $C^0_K(\Omega)$.

Let us now take a sequence in A. Using the above remark, we extract a subsequence that converges in all $C^n_K(\Omega)$, $n \in \mathbb{N}$, by the diagonal argument. The set A is thus relatively compact. □

Of course, in a separated topological vector space, all compact subsets are bounded and closed. A space that has the property of Proposition 1.18 and is reflexive is called a *Montel space*. Because of the Riesz theorem, an infinite dimensional normed space is not a Montel space. Now, since K has nonempty interior, $\mathscr{D}_K(\Omega)$ is patently infinite dimensional, hence is not normable, as we already noticed before.[11]

Be careful that this does not mean that $\mathscr{D}_K(\Omega)$ is locally compact. In fact, a more general version of Riesz's theorem states that a separated topological vector space is locally compact if and only if it is finite dimensional, see [63]. Simply, here the closed bounded sets, which are the compact sets of $\mathscr{D}_K(\Omega)$, all have empty interior and no nonempty open set is relatively compact.

[11] When K has empty interior, $\mathscr{D}_K(\Omega) = \{0\}$.

It should finally be kept in mind that since $\mathscr{D}_K(\Omega)$ is infinite dimensional, it can be equipped with several different reasonable topologies. The topology we have described up to now is called the strong topology of $\mathscr{D}_K(\Omega)$.

Let us now talk about the dual space of $\mathscr{D}_K(\Omega)$, denoted $\mathscr{D}'_K(\Omega)$.

Proposition 1.19. *A linear form T on $\mathscr{D}_K(\Omega)$ is continuous if and only if there exists $n \in \mathbb{N}$ and $C \in \mathbb{R}$ such that*

$$\forall \varphi \in \mathscr{D}_K(\Omega), \quad |\langle T, \varphi\rangle| \leq C p_n(\varphi), \tag{1.7}$$

and if and only if, for any sequence $\varphi_k \to \varphi$ in $\mathscr{D}_K(\Omega)$, $\langle T, \varphi_k\rangle \to \langle T, \varphi\rangle$.

Proof. The second characterization is trivial since $\mathscr{D}_K(\Omega)$ is metrizable. For the first characterization, we start with noticing that it is enough to prove the continuity of T at 0 by linearity. Let T be a linear form that satisfies (1.7). Since $\varphi_k \to 0$ implies that $p_n(\varphi_k) \to 0$ for all n, we clearly have $\langle T, \varphi_k\rangle \to 0$, hence T is continuous.

Conversely, let us take $T \in \mathscr{D}'_K(\Omega)$. This is a continuous mapping from $\mathscr{D}_K(\Omega)$ into $\mathbb{R}$, therefore the preimage of any open set of $\mathbb{R}$ is an open set of $\mathscr{D}_K(\Omega)$. In particular, since $0 \in T^{-1}(]-1,1[)$, there exists $n \in \mathbb{N}$ and $\alpha > 0$ such that $V_{n,\alpha}(0) \subset T^{-1}(]-1,1[)$. In other words, this means that if $p_n(\varphi) < \alpha$, then $|\langle T, \varphi\rangle| < 1$. Now, if $\varphi \neq 0$, then $p_n\big(\frac{\alpha\varphi}{2p_n(\varphi)}\big) < \alpha$ by absolute homogeneity. It follows that for all nonzero φ, $\big|\big\langle T, \frac{\alpha\varphi}{2p_n(\varphi)}\big\rangle\big| < 1$. This implies that $|\langle T, \varphi\rangle| \leq \frac{2}{\alpha} p_n(\varphi)$ for all φ, including $\varphi = 0$ for which the latter inequality obviously holds. □

Note the little trick of switching from strict inequalities to a non-strict one in the end in order to accommodate the case $\varphi = 0$. With a little more work, $\frac{2}{\alpha}$ can be replaced by $\frac{1}{\alpha}$.

Which topology are we going to use on $\mathscr{D}'_K(\Omega)$? Once again, there are several choices. We will only be interested here in the weak-star topology. Just like in the case of the dual space of a normed vector space, cf. Sect. 1.7, this is the coarsest topology, that is to say the one with as little open sets as possible, which makes all the mappings of the form $T \mapsto \langle T, \varphi\rangle$ continuous, with φ arbitrary in $\mathscr{D}_K(\Omega)$.

Let us focus for a moment on this concept of "the coarsest topology having this or that property," from an abstract viewpoint.

Proposition 1.20. *Let X be a set and $\mathscr{A} \subset \mathscr{P}(X)$ be a set of subsets of X. There exists a unique topology on X which is the coarsest of all topologies containing $\mathscr{A}$. This topology is called the* topology generated by $\mathscr{A}$. *It consists in all arbitrary unions of finite intersections of elements of $\mathscr{A}$.*

Proof. A topology on X is an element of $\mathscr{P}(\mathscr{P}(X))$ with special properties. It thus makes sense to consider families of topologies on X. It is then fairly obvious that the intersection of such a nonempty family of topologies is a topology on X. Now $\mathscr{A}$ is also an element of $\mathscr{P}(\mathscr{P}(X))$, and the discrete topology $\mathscr{P}(X)$

contains $\mathscr{A}$. Consequently, the family of topologies containing $\mathscr{A}$ is nonempty, and its intersection is the smallest possible topology containing $\mathscr{A}$.

Let us describe this topology more explicitly. If it contains $\mathscr{A}$, it must contain all finite intersections of elements of $\mathscr{A}$, due to closure under finite intersections. Due to closure under arbitrary unions, it must then contain arbitrary unions of such finite intersections. It is thus enough to show that the set of arbitrary unions of finite intersections of elements of $\mathscr{A}$ is a topology.

Let $\mathscr{O}$ be this set. It obviously contains $\emptyset$ and X, and is closed under arbitrary unions. The only (small) difficulty is closure under finite intersections. It is enough to consider the case of two elements U_1 and U_2 of $\mathscr{O}$, the general case follows by induction on the number of elements in the family.

Let us thus be given U_1 and U_2 such that there are two sets of indices Λ_1 and Λ_2, and for each $\lambda \in \Lambda_i$, an integer p_λ such that we can write[12]

$$U_1 = \bigcup_{\lambda\in\Lambda_1} \Big(\bigcap_{k=1}^{p_\lambda} A_{\lambda,k}\Big), \quad U_2 = \bigcup_{\mu\in\Lambda_2} \Big(\bigcap_{l=1}^{p_\mu} A_{\mu,l}\Big),$$

where $A_{\lambda,k}$ and $A_{\mu,l}$ belong to $\mathscr{A}$. We want to show that $U_1 \cap U_2 \in \mathscr{O}$. Let us set $\Lambda = \Lambda_1 \times \Lambda_2$ and

$$V = \bigcup_{(\lambda,\mu)\in\Lambda} \Big(\Big(\bigcap_{k=1}^{p_\lambda} A_{\lambda,k}\Big) \bigcap \Big(\bigcap_{l=1}^{p_\mu} A_{\mu,l}\Big)\Big),$$

so that $V \in \mathscr{O}$. Let $x \in U_1 \cap U_2$. There exist $\lambda \in \Lambda_1$ and $\mu \in \Lambda_2$ such that $x \in \bigcap_{k=1}^{p_\lambda} A_{\lambda,k}$ and $x \in \bigcap_{l=1}^{p_\mu} A_{\mu,l}$. In other words, $x \in \big(\bigcap_{k=1}^{p_\lambda} A_{\lambda,k}\big) \bigcap \big(\bigcap_{l=1}^{p_\mu} A_{\mu,l}\big)$. This shows that $U_1 \cap U_2 \subset V$.

Conversely, let us take $x \in V$. Then, there exists $(\lambda, \mu) \in \Lambda$ such that $x \in \big(\bigcap_{k=1}^{p_\lambda} A_{\lambda,k}\big) \bigcap \big(\bigcap_{l=1}^{p_\mu} A_{\mu,l}\big)$, that is to say $x \in \bigcap_{k=1}^{p_\lambda} A_{\lambda,k}$ and $x \in \bigcap_{l=1}^{p_\mu} A_{\mu,l}$, that is to say $x \in U_1$ and $x \in U_2$. This shows that $V \subset U_1 \cap U_2$. With the previous inclusion, it follows that $V = U_1 \cap U_2$, hence the intersection of two elements of $\mathscr{O}$ belongs to $\mathscr{O}$. We conclude by induction on the number of elements of $\mathscr{O}$ to be intersected with each other. □

Definition 1.5. Let X be a set, $(X_\lambda)_{\lambda\in\Lambda}$ a family of topological spaces and for each λ, a mapping $f_\lambda \colon X \to X_\lambda$. The coarsest topology on X that makes all the mappings f_λ continuous is called the *projective topology* or *initial topology* with respect to the family $(X_\lambda, f_\lambda)_{\lambda\in\Lambda}$.

This topology exists and is unique. Indeed, it is simply the topology generated by the family of sets $f_\lambda^{-1}(U_\lambda)$ where λ range over Λ and for each such λ, U_λ range over the open sets of X_λ. A basis for the topology—that is to say a family of sets that generate the open sets by arbitrary unions—is given by sets of the form

[12] We do not consider empty families with 0 element, since they pose no problem.

$\bigcap_{k=1}^{p} f_{\lambda_k}^{-1}(U_{\lambda_k})$ where U_{λ_k} is an open set of X_{λ_k}, according to Proposition 1.20. It follows that a mapping $f\colon Y \to X$ from a topological space Y into X equipped with the projective topology is continuous if and only if for all $\lambda \in \Lambda$, $f_\lambda \circ f$ is continuous from Y to X_λ. Finally, why look for the coarsest topology in this case? This is because it lives on X which is the domain of all f_λ. Adding sets to a topology on X can only improve the continuity status of these mappings, so the challenge is really to remove as many of them as possible.

The convergent sequences for this topology are also very simple.

Proposition 1.21. *Let x_n be a sequence in X equipped with the projective topology with respect to the family $(X_\lambda, f_\lambda)_{\lambda\in\Lambda}$. Then $x_n \to x$ in X if and only if $f_\lambda(x_n) \to f_\lambda(x)$ in X_λ, for all $\lambda \in \Lambda$.*

Proof. Let us assume that $x_n \to x$. Since each f_λ is continuous, it follows that $f_\lambda(x_n) \to f_\lambda(x)$.

Conversely, let us assume that $f_\lambda(x_n) \to f_\lambda(x)$ for all $\lambda \in \Lambda$. Let us take a neighborhood of x for the projective topology, of the form $\bigcap_{k=1}^{p} f_{\lambda_k}^{-1}(U_{\lambda_k})$. By hypothesis, for all $1 \le k \le p$, there exists an integer n_k such that $f_{\lambda_k}(x_n) \in U_{\lambda_k}$ for all $n \ge n_k$. Let us set $n_0 = \max\{n_1, \dots, n_p\}$. We thus see that $x_n \in \bigcap_{k=1}^{p} f_{\lambda_k}^{-1}(U_{\lambda_k})$ for all $n \ge n_0$. This shows that $x_n \to x$ for the projective topology. □

Let us apply all this to $\mathscr{D}'_K(\Omega)$. The weak-star topology is nothing but the projective topology with respect to $\Lambda = \mathscr{D}_K(\Omega)$, $\lambda = \varphi$, $X_\varphi = \mathbb{R}$ and $f_\varphi(T) = \langle T, \varphi\rangle$. According to Proposition 1.20, a neighborhood basis for 0 is given by sets of the form $\bigcap_{k=1}^{n}\{T \in \mathscr{D}'_K(\Omega); |\langle T, \varphi_k\rangle| < \varepsilon\}$. This neighborhood basis makes it very easy to check that we are dealing with a topological vector space topology, i.e., that the vector space operations are continuous. Furthermore, we see that this neighborhood basis is associated with the seminorms $p(T) = \max_{k\le n} |\langle T, \varphi_k\rangle|$, hence it is a locally convex topology. A sequence T_n converges to T for the weak-star topology if and only if $\langle T_n, \varphi\rangle \to \langle T, \varphi\rangle$ for all $\varphi \in \mathscr{D}_K(\Omega)$.[13]

Let us now deal with the space $\mathscr{D}(\Omega)$, a little faster. First of all, $\mathscr{D}(\Omega)$ is actually a vector space. Indeed, $\operatorname{supp}(\varphi+\psi) \subset \operatorname{supp}(\varphi)\cup\operatorname{supp}(\psi)$ and $\operatorname{supp}(\lambda\varphi) \subset \operatorname{supp}(\varphi)$ which are compact subsets of Ω.

We have to start again with some abstraction.

Proposition 1.22. *Let X be a set and $(\mathscr{O}_\lambda)_{\lambda\in\Lambda}$ a nonempty family of topologies on X. There exists a unique topology $\mathscr{O}$ on X which is the finest of all the topologies included in each $\mathscr{O}_\lambda$.*

Proof. It is enough to take $\mathscr{O} = \bigcap_{\lambda\in\Lambda} \mathscr{O}_\lambda$ which is obviously a topology, hence by construction the largest in the sense of inclusion that is contained in each $\mathscr{O}_\lambda$. □

A set is thus an open set of $\mathscr{O}$ if and only if it is an open set of $\mathscr{O}_\lambda$ for all $\lambda \in \Lambda$.

[13] This explains why this topology is sometimes called the pointwise convergence topology.

Definition 1.6. Let X be a set, $(X_\lambda)_{\lambda\in\Lambda}$ a family of topological spaces and for each λ, a mapping $f_\lambda\colon X_\lambda \to X$. The finest topology on X that makes all the f_λ continuous is called the *inductive topology* or *final topology* with respect to the family $(X_\lambda, f_\lambda)_{\lambda\in\Lambda}$.

This topology is well defined. Indeed, let

$$\mathscr{O}_\lambda = \{U \subset X;\ f_\lambda^{-1}(U) \text{ is an open set of } X_\lambda\}.$$

This is clearly a topology on X and the finest for which f_λ is continuous. We just take the intersection of these topologies for all $\lambda \in \Lambda$. It also appears that a mapping f from X into a topological space Y is continuous for the inductive topology if and only if all the mappings $f \circ f_\lambda$ are continuous from X_λ to Y. Here too, why look for the finest topology in this case? This is because it lives on X which is the codomain of all f_λ. Removing sets from a topology on X can only improve the continuity status of these mappings. The situation is actually opposite to that of the projective topology, with all arrows reversed.

Let us apply this to the case of $\mathscr{D}(\Omega)$. We make use of the fact that the open set Ω has an exhaustive sequence of compact subsets $K_n \subset \mathring{K}_{n+1}$, $\bigcup_{n\in\mathbb{N}} K_n = \Omega$. Let $\iota_n\colon \mathscr{D}_{K_n}(\Omega) \to \mathscr{D}(\Omega)$ be the canonical embedding simply given by inclusion. Since any compact subset of Ω is included in one of the K_n, it follows that $\mathscr{D}(\Omega) = \bigcup_n \mathscr{D}_{K_n}(\Omega)$. We equip $\mathscr{D}(\Omega)$ with the inductive topology associated with the family ι_n, after having checked that it does not depend on a specific choice of exhaustive sequence of compacts.[14] The topology $\mathscr{O}_n$ associated with ι_n is a Fréchet topology, hence a locally convex topological vector space topology. Since an intersection of convex sets is convex, the inductive topology on $\mathscr{D}(\Omega)$ is also a locally convex topological vector space topology.

As a matter of fact, we also have another family of embeddings $\iota_{nm}\colon \mathscr{D}_{K_n}(\Omega) \to \mathscr{D}_{K_m}(\Omega)$ for $n \le m$ that commute with the original embeddings, since the compact sets K_n are ordered under inclusion. Moreover, the topology induced on $\mathscr{D}_{K_n}(\Omega)$ by that of $\mathscr{D}_{K_m}(\Omega)$ when $m \ge n$ coincides with the topology of $\mathscr{D}_{K_n}(\Omega)$, because the seminorms coincide. In this case, we talk about a *strict inductive limit topology*, denoted by

$$\mathscr{D}(\Omega) = \varinjlim \mathscr{D}_{K_n}(\Omega).$$

This is thus the finest topology such that all embeddings ι_n are continuous. An open set U in $\mathscr{D}(\Omega)$ is a subset of $\mathscr{D}(\Omega)$ such that $U \cap \mathscr{D}_{K_n}(\Omega)$ is open in $\mathscr{D}_{K_n}(\Omega)$ for all n, which means that

$$U \text{ is open } \Leftrightarrow \forall\varphi \in U, \forall n \text{ such that } \operatorname{supp}\varphi \subset K_n;\ \exists p_n, \alpha_n,\ V_{K_n,p_n,\alpha_n}(\varphi) \subset U,$$

[14] Which is important, if slightly tedious.

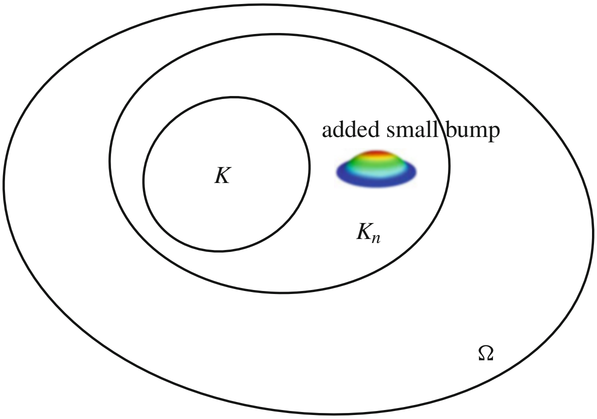


Fig. 1.4 All $\mathscr{D}_K(\Omega)$ have empty interior in $\mathscr{D}(\Omega)$

with an obvious notation $V_{K_n,p,\alpha}(\varphi)$ for the neighborhood basis of $\mathscr{D}_{K_n}(\Omega)$.

We note that the sequence of topologies $\mathscr{O}_n$ that are intersected, is decreasing for inclusion. In a sense, we are imposing more and more restrictions on the sets considered as n increases. Let us also note that a nonempty open set necessarily contains functions with arbitrarily large support. In fact, all $\mathscr{D}_K(\Omega)$ have empty interior in $\mathscr{D}(\Omega)$. Indeed, let K be a compact subset of Ω and U an open set such that $U \cap \mathscr{D}_K(\Omega)$ is nonempty, containing some function φ. Let n be such that $K \subset \mathring{K}_n$. Since $U \cap \mathscr{D}_{K_n}(\Omega)$ is a nonempty open set of $\mathscr{D}_{K_n}(\Omega)$, it contains a neighborhood $V_{K_n,p_n,\alpha_n}(\varphi)$. Now, this neighborhood is not included in $\mathscr{D}_K(\Omega)$ because it contains functions whose support is strictly larger than K. To see this, add to φ a small "bump" whose support is located in $K_n \setminus K$, see Fig. 1.4. Consequently, $U \not\subset \mathscr{D}_K(\Omega)$.

We also see that $\mathscr{D}_K(\Omega)$ is closed in $\mathscr{D}(\Omega)$. Take $\varphi \in \mathscr{D}(\Omega) \setminus \mathscr{D}_K(\Omega)$, and let $x \notin K$ be such that $\varphi(x) \neq 0$. Choosing n such that $\varphi \in \mathscr{D}_{K_n}(\Omega)$, it is quite obvious that $V_{K_n,p_0,|\varphi(x)|/2}(\varphi) \subset \mathscr{D}(\Omega) \setminus \mathscr{D}_K(\Omega)$ since no element of this neighborhood can vanish at point x.

The convergence of a sequence in $\mathscr{D}(\Omega)$ is actually given by Proposition 1.1. Indeed, let us take a sequence φ_k that satisfies conditions i) and ii) of the Proposition. By condition i), there exists n_0 such that all φ_k have support in K_{n_0}. Condition ii) then asserts that $\varphi_k \to \varphi$ in $\mathscr{D}_{K_{n_0}}(\Omega)$. The continuity of ι_{n_0} in turns implies that $\varphi_k \to \varphi$ in $\mathscr{D}(\Omega)$ for the inductive limit topology.

Conversely, assume that $\varphi_k \to \varphi$ in $\mathscr{D}(\Omega)$. Let us prove that condition i) holds. Once it is established, condition ii) is trivial. It is enough to consider the case $\varphi = 0$. Indeed, $\operatorname{supp}(\varphi_k - \varphi + \varphi) \subset \operatorname{supp}(\varphi_k - \varphi) \cup \operatorname{supp}(\varphi)$ which is a compact subset of Ω. Let thus U be an open set that contains 0. It follows that there exists k_0 such that $\varphi_k \in U$ for all $k \geq k_0$.

We argue by contradiction with an especially well picked open set U. Let us thus assume that condition i) is not satisfied. Then for all m, there exists k_m such that

$\operatorname{supp}\varphi_{k_m} \not\subset K_m$ which implies that there exists $x_m \in \Omega \setminus K_m$ with $\varphi_{k_m}(x_m) \neq 0$. Let $\ell(m) = \min\{\ell; x_m \in K_\ell\}$. We set

$$p(\psi) = \sum_{m=0}^{+\infty} 2 \max_{x \in K_{\ell(m)} \setminus K_m} \left| \frac{\psi(x)}{\varphi_{k_m}(x_m)} \right|.$$

We first observe that this quantity is well-defined on $\mathscr{D}(\Omega)$, because for all ψ with compact support, there is only a finite number of nonzero terms. It is in fact fairly clear that this is a seminorm on $\mathscr{D}(\Omega)$. Let us now choose

$$U = \{\psi \in \mathscr{D}(\Omega); p(\psi) < 1\}.$$

For any n, there is only a fixed finite number of nonzero terms in the sum on any $\mathscr{D}_{K_n}(\Omega)$, hence there is a constant C_n such that $p \le C_n p_0$ on $\mathscr{D}_{K_n}(\Omega)$. It follows that p is continuous on $\mathscr{D}(\Omega)$ and that U is open for the inductive topology of $\mathscr{D}(\Omega)$. Of course $0 \in U$, but on the other hand, $p(\varphi_{k_n}) \ge 2$, which implies that $\varphi_{k_n} \notin U$, contradiction.

A few final words on the space $\mathscr{D}(\Omega)$. It is not normable because a strict inductive limit of a sequence of Montel spaces is a Montel space. In fact, its topology is not metrizable. Of course, this is a question of uncountable neighborhood bases, but we can also see it simply by using a variant of the function g of Lemma 1.2 in one dimension. Let us set

$$\varphi_{k,n}(x) = \begin{cases} e^{\frac{k}{(x-\frac{1}{n})^2 - \frac{1}{4n^2}}} & \text{for } \frac{1}{2n} < x < \frac{3}{2n}, \\ 0 & \text{otherwise,} \end{cases}$$

with $k > 0$ and $n \ge 1$. By construction, $\varphi_{k,n} \in \mathscr{D}(]0,2[)$. Moreover, for n fixed, $\varphi_{k,n} \to 0$ in $\mathscr{D}(]0,2[)$ when $k \to +\infty$ because of the exponential term that turns up as a factor in all the derivatives. If the topology was metrizable, the usual double limit argument would enable us to find a sequence $k(n)$ such that $\varphi_{k(n),n} \to 0$ in $\mathscr{D}(]0,2[)$ when $n \to +\infty$. This is obviously not the case, because such a sequence cannot satisfy the support condition i) since $\operatorname{supp}\varphi_{k(n),n} = [\frac{1}{2n}, \frac{3}{2n}]$.

Another amusing way of showing that $\mathscr{D}(\Omega)$ is not metrizable is to call on completeness. There is a general theory of so-called *uniform structures* which makes it possible to extend the concept of completeness to non metrizable spaces by replacing Cauchy sequences with more general objects called Cauchy filters, see [63]. Now it so happens that a strict inductive limit of Fréchet spaces is complete in this generalized sense.[15] Since $\mathscr{D}(\Omega) = \cup_{n\in\mathbb{N}}\mathscr{D}_{K_n}(\Omega)$ and each $\mathscr{D}_{K_n}(\Omega)$ is closed with empty interior, we are thus faced with a complete space which is not a Baire space. It can then certainly not be metrizable.

[15] In the sense that every Cauchy filter converges.

Let us at last talk rapidly about the space of distributions $\mathscr{D}'(\Omega)$, the dual space of $\mathscr{D}(\Omega)$, from the topological point of view. Dually to what was seen above, we have restriction mappings $r_n\colon \mathscr{D}^*(\Omega) \to \mathscr{D}^*_{K_n}(\Omega)$ defined by $\langle r_n T, \varphi\rangle = \langle T, \iota_n \varphi\rangle$ (we are taking here the *algebraic* duals, without continuity condition). The definition of inductive limit topology implies that T is continuous if and only if $r_n T$ is continuous for all n, that is to say according to Proposition 1.19, the condition of Proposition 1.2. Indeed, T is continuous if and only if for any open set ω of $\mathbb{R}$, $T^{-1}(\omega)$ is open, or again if and only if $T^{-1}(\omega) \cap \mathscr{D}_{K_n}(\Omega)$ is open in $\mathscr{D}_{K_n}(\Omega)$ for all n.

Concerning the condition of Proposition 1.3, let us consider a linear form T such that $\langle T, \varphi_k\rangle \to \langle T, \varphi\rangle$ as soon as $\varphi_k \to \varphi$. In particular, this convergence holds for all sequences supported in K_n. Since $\mathscr{D}_{K_n}(\Omega)$ is metrizable, we conclude that $r_n T$ is continuous for all n, which implies that T itself is continuous, i.e., a distribution.

We equip $\mathscr{D}'(\Omega)$ with the projective topology associated with the restriction mappings r_n and spaces $\mathscr{D}'_{K_n}(\Omega)$ equipped with their weak-star topology. The compacts K_n are ordered by inclusion, and we then talk of a *projective limit topology*, which is denoted

$$\mathscr{D}'(\Omega) = \varprojlim \mathscr{D}'_{K_n}(\Omega).$$

This is once more a locally convex topological vector space topology.

The open sets of the $\mathscr{D}'(\Omega)$ are of very little practical interest for applications to partial differential equations. They are however easy to describe since this is a projective topology. Indeed an open set must be of the form $U = \bigcup_{\lambda\in\Lambda}\bigl(\bigcap_{i=1}^{k_\lambda} r_{n_i}^{-1}(U_i)\bigr)$ where U_i is an open set of the weak-star topology of $\mathscr{D}'_{K_{n_i}}(\Omega)$. Each open set U_i is itself of the form $V = \bigcup_{\text{arbitrary}}\bigl(\bigcap_{\text{finite}}\{T \in \mathscr{D}'_K(\Omega); |\langle T, \varphi\rangle - a| < \varepsilon\}\bigr)$ with $\varphi \in \mathscr{D}_K(\Omega)$, $a \in \mathbb{R}$ and $\varepsilon > 0$, where we have dropped all multiple indices for legibility. Now

$$r_n^{-1}\bigl(\{T \in \mathscr{D}'_{K_n}(\Omega); |\langle T, \varphi\rangle - a| < \varepsilon\}\bigr) = \{T \in \mathscr{D}'(\Omega); |\langle T, \varphi\rangle - a| < \varepsilon\}\bigr),$$

since there is no harm in identifying a function $\varphi \in \mathscr{D}_{K_n}(\Omega)$ with its extension by 0 to Ω, which is an element of $\mathscr{D}(\Omega)$. We have seen earlier in Proposition 1.20 that a finite intersection of arbitrary unions of finite intersections of sets is itself an arbitrary union of finite intersections. Therefore, it turns out that any open set of $\mathscr{D}'(\Omega)$ is of the form $U = \bigcup_{\text{arbitrary}}\bigl(\bigcap_{\text{finite}}\{T \in \mathscr{D}'(\Omega); |\langle T, \varphi\rangle - a| < \varepsilon\}\bigr)$ with $\varphi \in \mathscr{D}(\Omega)$, $a \in \mathbb{R}$ and $\varepsilon > 0$. We thus see that the projective limit topology is also the weak-star topology on the dual space of $\mathscr{D}(\Omega)$.

On the other hand, the convergence of sequences of distributions is of paramount interest for the applications we have in mind. Due to the general properties of projective topologies, a sequence of distributions T_k converges to T if and only if $r_n T_k \to r_n T$ in $\mathscr{D}'_{K_n}(\Omega)$ for all n. Proposition 1.4 follows right away, given what we know about convergence in $\mathscr{D}'_{K_n}(\Omega)$.

We notice that the topology of $\mathscr{D}'(\Omega)$ is not a metrizable topology either. Indeed, let $T_{k,n} = \frac{1}{k}\delta^{(n)} \in \mathscr{D}'(\mathbb{R})$, where $\delta^{(n)}$ is the nth derivative of the Dirac mass. For n fixed, obviously $T_{k,n} \to 0$ when $k \to +\infty$. Let $k(n)$ be any sequence of integers tending to infinity. By a theorem of Borel, there exists a function $\varphi \in \mathscr{D}(\mathbb{R})$ such that $\varphi^{(n)}(0) = k(n)$ for all n. Therefore $\langle T_{k(n),n}, \varphi\rangle = 1 \not\to 0$ so that $T_{k(n),n}$ does not converge to 0 in the sense $\mathscr{D}'(\mathbb{R})$. Now, if the topology was metrizable, we could find such a sequence $k(n)$ for which this convergence to 0 would hold.

We have thus more or less covered all the practical properties of distributions.

Chapter 2
Fixed Point Theorems and Applications

If f is a mapping from a set E into itself, any element x of E such that $f(x) = x$ is called a *fixed point* of f. Many problems, including nonlinear partial differential equations problems, may be recast as problems of finding a fixed point of a certain mapping in a certain space. We will see several examples of this a little later on. It is therefore interesting to have fixed point theorems at our disposal that are as general as possible.

Let us first mention the Banach fixed point theorem for a strict contraction in a complete metric space. This is a relatively elementary result that is not very useful for the applications we have in mind, but we state it anyway.

Theorem 2.1. *Let (E, d) be a complete metric space, $T : E \to E$ a strict contraction, i.e., a mapping such that there exists a constant $k < 1$ with*

$$\forall x, y \in E, \quad d(T(x), T(y)) \leq kd(x, y).$$

Then T admits a unique fixed point $x^* = T(x^*) \in E$. Moreover, for all $x_0 \in E$, the sequence of iterates $x_m = T^m(x_0)$ converges to x^* when $m \to +\infty$.

This theorem, or variants thereof, is nonetheless useful in the context of evolution ordinary or partial differential equations to establish the Picard-Lindelöf theorem, but we will not be concerned with this context here.

2.1 Brouwer's Fixed Point Theorem

Brouwer's theorem is the basic fixed-point theorem in finite dimension. Let $\bar{B}^d = \{x \in \mathbb{R}^d, \|x\| \leq 1\}$ denote the closed unit ball of $\mathbb{R}^d$ equipped with the standard Euclidean norm, and $S^{d-1} = \partial\bar{B}^d$ the unit sphere, which is the ball's boundary. Brouwer's fixed point theorem asserts that:

H. Le Dret, *Nonlinear Elliptic Partial Differential Equations*, Universitext,
https://doi.org/10.1007/978-3-319-78390-1_2

Theorem 2.2. *Every continuous mapping from $\bar{B}^d$ into $\bar{B}^d$ admits at least one fixed point.*

We note an amusing "physical"[1] illustration of Brouwer's theorem. If we take a disk cut out of paper and set on a table, crumple it up without tearing it and put it back on the table so that it does not stick out of its original position, then at least one point in the crumpled paper ends up exactly on the same vertical as its original precrumpled position. The continuous mapping to which we apply Brouwer's theorem is simply that which sends each point of the disk to its projection on the table after crumpling.

Brouwer's theorem is a nontrivial result, except in the $d = 1$ case where it follows readily from a connectedness argument. There are many different proofs in the general case, which all call for more or less elementary ideas. We give below a proof that feels as elementary as possible.[2] Other accessible proofs can be found in [30, 40, 53], but there are still more proofs, for instance of combinatoric nature or using algebraic topology.

We start with the no-retraction theorem in the C^1 case. A retraction from a topological space to one of its subsets is a continuous mapping from the space to the subset, the restriction of which to the subset is the identity mapping. So, a retraction continuously sends the whole space to a subset without moving any of the points of the subset.

Theorem 2.3. *There is no mapping $f : \bar{B}^d \to S^{d-1}$ of class C^1 such that $f_{|S^{d-1}} = \mathrm{Id}$.*

Proof. We argue by contradiction. Let f be a retraction of $\bar{B}^d$ onto S^{d-1} of class C^1. For all $t \in [0, 1]$, we set $f_t(x) = (1-t)x + tf(x)$. Since the ball is convex, f_t is a mapping from $\bar{B}^d$ into $\bar{B}^d$. Moreover, f is a retraction, so the restriction of f_t to S^{d-1} is the identity. For $t = 1$, we have $f_1 = f$, for $t = 0$, $f_0 = \mathrm{Id}$ and for small t, f_t is a small perturbation of the identity. In a first step, we are going to show that f_t is a diffeomorphism of the open ball onto itself for small t.

Let $c = \max_{\bar{B}^d} \|\nabla f\|$, where ∇f denotes the Jacobian matrix, i.e., the $d \times d$ matrix of partial derivatives of the components of f, and where we have taken the matrix norm induced by the standard Euclidean norm on $\mathbb{R}^d$. By the mean value inequality, it follows that for all $x, y \in \bar{B}^d$,

$$\|f(x) - f(y)\| \leq c\|x - y\|.$$

Now since

$$\|f_t(x) - f_t(y)\| \geq (1-t)\|x - y\| - t\|f(x) - f(y)\| \geq \big((1-t) - ct\big)\|x - y\|,$$

[1] To some extent. In the same vein, there is also a stirred coffee cup example.

[2] Keeping in mind that this is pretty subjective.

we see that f_t is injective as long as $0 \le t < 1/(1+c)$. Since f_t is the identity on S^{d-1}, we conclude that $f_t(B^d) \subset B^d$, where B^d denotes the open ball, for these values of t.

There also holds,

$$\nabla f_t = (1-t)I + t\nabla f = (1-t)\Big(I + \frac{t}{1-t}\nabla f\Big),$$

with, for $0 \le t < 1/(1+c)$,

$$\frac{t}{1-t}\|\nabla f\| \le \frac{ct}{1-t} < 1.$$

Consequently, ∇f_t is everywhere invertible for those same values of t. The inverse function theorem then implies that f_t is a local C^1-diffeomorphism on B^d. In particular, the image of B^d by f_t is an open set, still for these values of t.

Let us now show that f_t is also surjective for $0 \le t < 1/(1+c)$. It is enough to show that $f_t(B^d) = B^d$, since $f_t(S^{d-1}) = S^{d-1}$. We argue by contradiction. Let us assume that $f_t(B^d) \ne B^d$. We have seen that $f_t(B^d) \subset B^d$, thus there exists $y_1 \in B^d \setminus f_t(B^d)$. Let us pick a point $y_0 = f_t(x_0) \in f_t(B^d)$ and set

$$\bar\lambda = \inf\{\lambda \in \mathbb{R}_+;\, y_\lambda = (1-\lambda)y_0 + \lambda y_1 \notin f_t(B^d)\}.$$

We know that $f_t(B^d)$ is an open set, therefore $\bar\lambda > 0$. We also have $\bar\lambda \le 1$ due to the choice of y_1. The sequence $y_{\bar\lambda-1/k}$ thus belongs to $f_t(B^d)$ for k large enough, which means that there exists a sequence $x_k \in B^d$ such that $f_t(x_k) = y_{\bar\lambda-1/k} \to y_{\bar\lambda}$. We may as well assume that x_k tends to some $\bar x \in \bar B^d$, which implies that $f_t(\bar x) = y_{\bar\lambda}$. This implies that $\bar x \notin S^{d-1}$ for otherwise, we would have $y_{\bar\lambda} = \bar x \in S^{d-1}$, whereas the segment $\{y_\lambda, \lambda \in [0,1]\}$ is included in B^d by convexity. Thus $\bar x \in B^d$, $\bar\lambda < 1$ and f_t is a local diffeomorphism at $\bar x$. The image of f_t is thus a neighborhood of $y_{\bar\lambda}$. It follows that there exists $1 > \mu > \bar\lambda$ such that $y_\lambda \in f_t(B^d)$ for all $\lambda \in [\bar\lambda, \mu[$. This contradicts the definition of $\bar\lambda$ as an infimum, see Fig. 2.1.

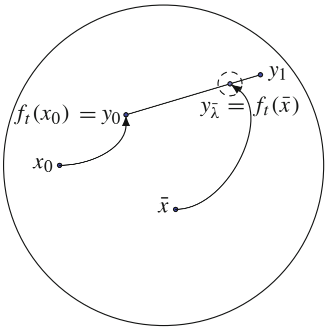


Fig. 2.1 The mapping f_t is a local diffeomorphism at $\bar x$

Up to now, we have shown that for $0 \leq t < 1/(1+c)$, f_t is a C^1-diffeomorphism from B^d onto B^d. Let us set

$$V(t) = \int_{B^d} \det \nabla f_t(x)\,dx.$$

It is clear by continuity that for t in a neighborhood of 0, $\det \nabla f_t > 0$ on B^d. It is therefore equal to the Jacobian of f_t and by the change of variable formula,

$$V(t) = \int_{B^d} dx = \text{volume}\, B^d,$$

for t nonnegative and small enough. This value thus does not depend on t on this neighborhood of 0. Now it is also clear from the form of ∇f_t that the function $t \mapsto V_t$ is a polynomial of degree at most d in the variable t. It must thus be a constant polynomial, hence $V(1) = V(0) = \text{volume}\, B^d > 0$. Now $f_1 = f$ and f is a retraction, thus $\det \nabla f_1 = 0$, otherwise the image of f would not have an empty interior by the inverse function theorem. It follows that $V(1) = 0$, a contradiction. □

When considered as a function in the two variables t and x, $(t, x) \mapsto f(t, x) = f_t(x)$ is a *homotopy* between Id and f. We now are in a position to establish Brouwer's theorem, i.e., Theorem 2.2.

Proof of Brouwer's Theorem. We argue once more by contradiction. Let g be a continuous mapping from $\bar{B}^d$ into $\bar{B}^d$ that has no fixed point. By compactness, there exists $\alpha > 0$ such that $\|x - g(x)\| > \alpha$ for all $x \in \bar{B}^d$. The Stone-Weierstrass theorem gives us a polynomial mapping $h\colon \bar{B}^d \to \mathbb{R}^d$ such that $\max_{\bar{B}^d} \|g(x) - h(x)\| \leq \alpha/2$. We note that

$$\|x - h(x)\| = \|x - g(x) + g(x) - h(x)\| > \alpha - \frac{\alpha}{2} = \frac{\alpha}{2}, \tag{2.1}$$

for all $x \in \bar{B}^d$. Of course, h has no reason to be $\bar{B}^d$-valued, but $h_\alpha = (1 + \frac{\alpha}{2})^{-1} h$ is and has no fixed point in $\bar{B}^d$. Indeed, any such fixed point would be such that $h(x) = x + \frac{\alpha}{2} x$, so that $\|x - h(x)\| \leq \frac{\alpha}{2}$, thus contradicting inequality (2.1).

We have now replaced a continuous mapping g from $\bar{B}^d$ into $\bar{B}^d$ without any fixed point by a C^1 mapping[3] from $\bar{B}^d$ into $\bar{B}^d$ without any fixed point. From now on, we assume without loss of generality that g is C^1.

The mapping $x \mapsto \|x - g(x)\|^{-1}$ is thus of class C^1 on $\bar{B}^d$ as the composition of C^1 mappings, and the same goes for $x \mapsto u(x) = \|x - g(x)\|^{-1}(x - g(x))$, which is furthermore S^{d-1}-valued. We consider the straight line going through x in

[3] In fact, a polynomial, hence C^∞ mapping.

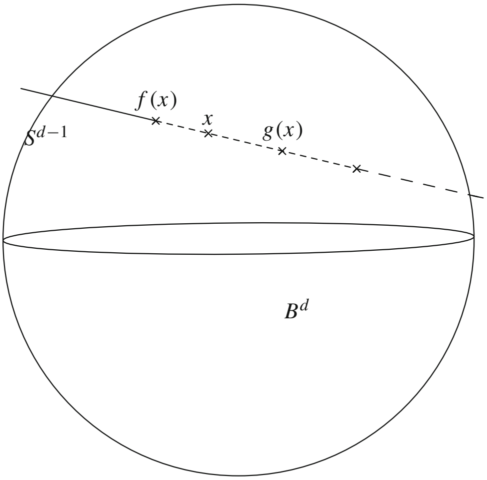


Fig. 2.2 Constructing a retraction f from a mapping g with no fixed point

the direction of $u(x)$. This straight line intersects the sphere S^{d-1} at two points.[4] We call $f(x)$ the intersection point on the side of x, opposite to $g(x)$.

It is clear by construction that f is S^{d-1}-valued and that if $x \in S^{d-1}$, then $f(x) = x$. Let us check that f is of class C^1. By definition of f, there exists a real number $t(x)$ such that $f(x) = x + t(x)u(x)$. This number is obtained by solving the second degree equation expressing that $\|f(x)\|^2 = 1$ and by taking its positive root. An elementary computation gives

$$t(x) = -x \cdot u(x) + \sqrt{1 - \|x\|^2 + (x \cdot u(x))^2}.$$

Let us notice that the number under the radical sign is always strictly positive. Therefore the mapping $x \mapsto t(x)$ is of class C^1, and so is f, which is thus a C^1 retraction of the ball on the sphere. This is impossible. □

See Fig. 2.2 for the geometrical construction of the C^1 retraction f starting from g.

Remark 2.1. It is amusing to realize that this proof of Brouwer's theorem relies, among other things, on the inverse function theorem, that can itself be shown by using the Banach fixed point theorem. Let also note that the multivariate change of variable formula, another crucial ingredient in this proof, is not so elementary a result, see for example [40] for a complete proof. □

Using Brouwer's theorem, we can now generalize the no-retraction theorem with a more topologically satisfying formulation.

[4] Indeed, if $x \in S^{d-1}$, $u(x)$ is not a tangent vector by construction.

Theorem 2.4. *There is no continuous mapping* $f\colon \bar{B}^d \to S^{d-1}$ *such that* $f_{|S^{d-1}} = \mathrm{Id}$.

Proof. Let f be such a retraction. We set $g(x) = -f(x)$. Then $g \in C^0(\bar{B}^d; \bar{B}^d)$ has a fixed point x_0, i.e., $x_0 = -f(x_0)$. Since f takes its values in the unit sphere, $x_0 \in S^{d-1}$. Since f is a retraction, $f(x_0) = x_0$, and therefore $x_0 = 0$, which is definitely not on the sphere. □

We finally see that the no-retraction theorem and Brouwer's theorem are two equivalent results.

The closed unit ball of $\mathbb{R}^d$ is not the only topological space that has the fixed point property. The following is an easy consequence of Brouwer's theorem.

Theorem 2.5. *Let* K *be a compact set homeomorphic to the closed unit ball of* $\mathbb{R}^d$*. Every continuous mapping from* K *into* K *admits at least one fixed point.*

Proof. Let g be a continuous mapping from K into K and h a homeomorphism from K to the closed unit ball. The mapping $h \circ g \circ h^{-1}$ is continuous from $\bar{B}^d$ into $\bar{B}^d$. It thus has a fixed point $y = h \circ g \circ h^{-1}(y) \in \bar{B}^d$. Hence, $h^{-1}(y) \in K$ is a fixed point of g. □

Let us give a useful consequence of this result.

Theorem 2.6. *Let* C *be a nonempty convex compact subset of* $\mathbb{R}^d$*. Every continuous mapping from* C *into* C *admits at least one fixed point.*

Proof. We are going to show that a nonempty convex compact subset C of $\mathbb{R}^d$ is either homeomorphic to the closed unit ball of $\mathbb{R}^n$ for some $n \leq d$ and apply the previous theorem, or is reduced to one point, in which case there is only one mapping from C to C that cannot help but have a fixed point.

Let thus C be such a set. We assume first that $\overset{\circ}{C} \neq \emptyset$. By translation, which is a homeomorphism, we can always assume that $0 \in \overset{\circ}{C}$ and therefore, there is an open ball $B(0, r)$, $r > 0$, included in C. We introduce the gauge function of the convex set C. The gauge function is the mapping $\mathbb{R}^d \to \mathbb{R}_+$ defined by

$$j(x) = \inf\{t > 0; x/t \in C\},$$

see Fig. 2.3.

It follows from the fact that $B(0, r) \subset C$ that $j(x) \leq \|x\|/r$ for all $x \in \mathbb{R}^d$, which confirms that j is real-valued.[5] Likewise, since C is compact, it is included in a ball $B(0, R)$ for some R, which implies that $\|x\|/R \leq j(x)$. The following properties are easily checked:

i) if $x \in C$ then $j(x) \leq 1$,

ii) $j(\lambda x) = \lambda j(x)$ for all $x \in \mathbb{R}^d$ and $\lambda \geq 0$,

iii) $j(x + y) \leq j(x) + j(y)$ for all $x, y \in \mathbb{R}^d$.

[5] In the sense that the infimum is not $+\infty$.

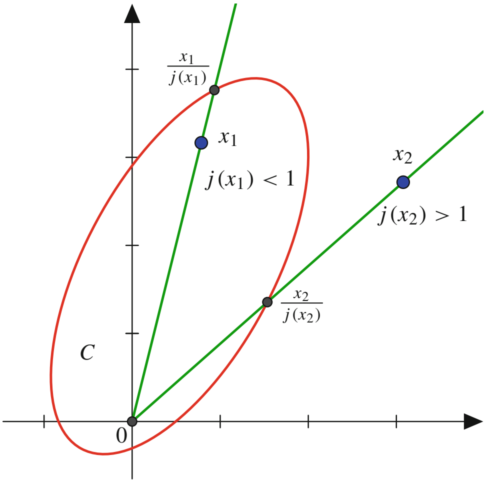


Fig. 2.3 The gauge function

Property i) is fairly obvious, since $x/1 \in C$. For property ii), we see that

$$j(\lambda x) = \inf\{t > 0; (\lambda x)/t \in C\} = \inf\{t > 0; x/(t/\lambda) \in C\}$$
$$= \lambda \inf\{(t/\lambda) > 0; x/(t/\lambda) \in C\} = \lambda j(x),$$

when $\lambda > 0$. Of course, $j(0x) = j(0) = 0 = 0j(x)$.

Property iii) is the one that makes use of the convexity of C. Given x and y, we take $t_1 > 0$ and $t_2 > 0$ such that x/t_1 and y/t_2 belong to C. It follows that

$$\frac{x+y}{t_1+t_2} = \Bigl(\frac{t_1}{t_1+t_2}\Bigr)\frac{x}{t_1} + \Bigl(\frac{t_2}{t_1+t_2}\Bigr)\frac{y}{t_2} \in C,$$

since $\frac{t_1}{t_1+t_2}$ and $\frac{t_2}{t_1+t_2}$ are in $[0, 1]$ and add up to 1. Consequently, by definition of j, $j(x + y) \leq t_1 + t_2$. Property iii) results from taking the infimums in the right-hand side of the latter inequality.

Conversely to i), we notice that if $j(x) \leq 1$, then $x \in C$. In effect, there then exists a sequence $t_n \to j(x)$ and $y_n \in C$ such that $x = t_n y_n + (1 - t_n)0$. If there exists an integer n_0 such that $t_{n_0} \leq 1$, then x is a convex combination of elements of C, thus is in C. If $t_n > 1$ for all n, then $t_n \to j(x) = 1$ and $y_n \to x$. Since C is closed, this yields $x \in C$.

We deduce from the subadditivity of j that it is a continuous function from $\mathbb{R}^d$ to $\mathbb{R}$. Indeed, for all $x, y \in \mathbb{R}^d$,

$$-j(-y) \leq j(x + y) - j(x) \leq j(y),$$

(write $x = x + y + (-y)$ for the inequality on the left), and thus

$$|j(x + y) - j(x)| \leq \max(|j(-y)|, |j(y)|) \leq \|y\|/r.$$

At this point, we define two mappings g and h from $\mathbb{R}^d$ to $\mathbb{R}^d$ by setting

$$g(x) = \begin{cases} \dfrac{j(x)}{\|x\|}x & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases} \quad \text{and} \quad h(y) = \begin{cases} \dfrac{\|y\|}{j(y)}y & \text{if } y \neq 0, \\ 0 & \text{if } y = 0. \end{cases}$$

It is easy to check that g and h are inverse of each other. Furthermore, they are continuous outside of 0 due to the continuity of j. Moreover, we have $\|g(x)\| \leq j(x) \leq \|x\|/r$, so that g is also continuous at 0 and as $\|h(y)\| \leq \|y\|^2/j(y) \leq R\|y\|$, the same goes for h. Consequently, g and h form a pair of inverse homeomorphisms of $\mathbb{R}^d$.

To conclude this part of the proof, we notice that $g(C) = \bar{B}^d$. Indeed, if $x \in C$, then $j(x) \leq 1$ and $\|g(x)\| \leq 1$, i.e., $g(C) \subset \bar{B}^d$. Conversely, let $y \in \bar{B}^d$. We have $j(h(y)) = \frac{\|y\|}{j(y)} j(y) \leq 1$, so that $h(y) \in C$. This means that $h(\bar{B}^d) \subset C$, hence by composition by g, $\bar{B}^d \subset g(C)$. This completes the case $\overset{\circ}{C} \neq \emptyset$.

We now assume that $\overset{\circ}{C} = \emptyset$. We argue by decreasing induction on the space dimension. If C contained a linearly independent family of d vectors, then it would contain the simplex generated by these vectors and the zero vector, and would not have empty interior. Therefore, $\overset{\circ}{C} = \emptyset$ implies that C is contained in an affine hyperplane of $\mathbb{R}^d$, that is to say a space of dimension $d - 1$.

We then have the following alternative: either the interior of C as a subset of this hyperplane is nonempty, in which case we apply the preceding step, or this interior is empty and we start over and decrease the dimension by one. After at most $d - 1$ steps, we have thus established that C is homeomorphic to a ball of dimension less or equal to d or that C is included in a straight line and has empty interior in this straight line. In the latter case, C is reduced to one point. □

Remark 2.2. i) We can give a much shorter proof by considering a closed ball B that contains C and taking the mapping $g: B \to B$, $g = f \circ P$, where P is the orthogonal projection on C, a continuous mapping. By Brouwer's theorem, g has a fixed point $x \in B$, such that $x = f(P(x))$, from which it follows that $x \in C$ and therefore $P(x) = x$.

ii) In general, there is no reason for a Brouwer fixed point to be unique, see Fig. 2.4.

iii) There exist compact subsets of $\mathbb{R}^d$ that do not have the fixed point property. For example, Brouwer's theorem is clearly false in a circular annulus, even though there is no retraction of the circular annulus on its boundary. As a matter of fact, more generally, if X is a compact manifold with boundary, there is no retraction of X onto ∂X. The no-retraction and fixed point properties are thus not related in general.

iv) Completely aside of the present discussion, we observe that the gauge function is defined for any convex subset of any real vector space. If C is *circled*, meaning here that if $x \in C$ then $-x \in C$, then its gauge j is a seminorm on the vector space in question, if it is real-valued. This is the link mentioned in the Appendix in Chap. 1 between the definitions of locally convex topological vectors

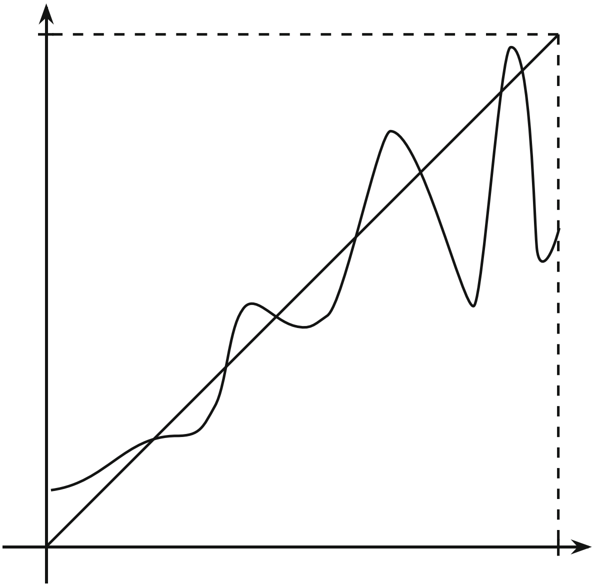


Fig. 2.4 Non uniqueness of a Brouwer fixed point

spaces via a family of seminorms on the one hand, and via a family of convex neighborhoods of 0 on the other hand. □

We now give another result which is also equivalent to Brouwer's theorem and to the no-retraction theorem in the ball, see [48].

Theorem 2.7. *Let E be a Euclidean space (thus finite dimensional) and $P\colon E \to E$ a continuous mapping such that there exists $\rho > 0$ for which at every point x of the sphere of radius ρ centered at 0, there holds $P(x) \cdot x \geq 0$. Then there exists at least one point x_0, $\|x_0\| \leq \rho$, such that $P(x_0) = 0$.*

Proof. Let us assume that P does not vanish on the closed ball $\bar{B}(0, \rho)$, in other words that for all $\|x\| \leq \rho$, $\|P(x)\| > 0$. The function $g\colon \bar{B}(0, \rho) \to \partial\bar{B}(0, \rho)$ defined by

$$g(x) = -\frac{\rho}{\|P(x)\|} P(x)$$

is thus continuous. By Brouwer's theorem, this mapping has a fixed point x^*, which is such that

$$g(x^*) = x^* = -\frac{\rho}{\|P(x^*)\|} P(x^*).$$

On the one hand, it follows that $\|x^*\| = \rho$ and on the other hand that $\rho^2 = \|x^*\|^2 = g(x^*) \cdot x^* = -\frac{\rho}{\|P(x^*)\|} P(x^*) \cdot x^* \leq 0$, which contradicts the hypothesis $\rho > 0$. □

Remark 2.3. i) The hypothesis $P(x) \cdot x \geq 0$ can be interpreted geometrically without resorting to the Euclidean structure as the fact that the function P points outwards on the boundary of the convex set $\bar{B}(0, \rho)$. Changing P in $-P$, we can assume as well that it points inwards.

ii) Theorem 2.7 implies the no-retraction theorem for the ball, because we can apply it to a retraction, which must thus vanish at some point in the ball. It is therefore equivalent to Brouwer's theorem.

iii) The stirred cup of coffee physical illustration belongs more here than with Brouwer's theorem: at any instant, there is at least one point at the surface of the coffee that has zero horizontal speed.[6] □

2.2 The Schauder Fixed Point Theorems

The previous results make crucial use of the finite dimension of the ambient space, for example through the use of the change of variable formula or of the compactness of the closed ball. In infinite dimension, Brouwer's theorem is no longer true. Here is a counter-example. We consider the space of square summable sequences $l^2 = \{(x_i)_{i\in\mathbb{N}}, \sum_{i=0}^{\infty} x_i^2 < +\infty\}$. When equipped with the norm $\|x\|_{l^2} = (\sum_{i=0}^{\infty} x_i^2)^{1/2}$, this is an infinite dimensional Hilbert space. Let $\bar{B}$ be its closed unit ball and S its unit sphere. Then the mapping

$$T \colon \bar{B} \to \bar{B}, T(x) = \Big(\sqrt{1 - \|x\|_{l^2}^2}, x_0, x_1, x_2, \dots\Big)$$

is continuous but has no fixed point. In effect, T takes its values in the unit sphere, thus any potential fixed point should simultaneously satisfy $\|x\|_{l^2} = 1$, then $x_0 = 0$ and $x_{i+1} = x_i$ for all $i \geq 0$, hence $x = 0$.

Likewise, the mapping $R \colon \bar{B} \to S$ resulting from the same construction as in the finite dimensional case, i.e.,

$$\begin{cases} R(x)_0 = x_0 + \dfrac{1 - \|x\|_{l^2}^2}{\|x - T(x)\|_{l^2}^2}\left(x_0 - \sqrt{1 - \|x\|_{l^2}^2}\right), \\[2ex] R(x)_i = x_i + \dfrac{1 - \|x\|_{l^2}^2}{\|x - T(x)\|_{l^2}^2}\left(x_i - x_{i-1}\right), i \geq 1, \end{cases}$$

is a retraction of $\bar{B}$ on S.

The problem actually comes from a lack of compactness: infinite dimensional spaces are not locally compact. In order to go around this difficulty, we need a property of approximation of compact sets in a normed vector space by finite

[6] This is more striking for an inviscid coffee that does not adhere to the side of the cup...

dimensional sets, which will make it possible to fall back on the latter case, assuming some compactness.

Lemma 2.1. *Let E be a normed vector space and K a compact subset of E. For all $\varepsilon > 0$, there exists a finite dimensional vector subspace F_ε of E and a continuous mapping g_ε from K into F_ε such that for all $x \in K$, $\|x - g_\varepsilon(x)\|_E < \varepsilon$. Moreover, $g_\varepsilon(K) \subset \operatorname{conv} K$.*

Here $\operatorname{conv} K$ denotes the convex hull of K. This is the set of all convex combinations of elements of K, or equivalently the smallest convex set containing K. The lemma says that any compact set can be approximated arbitrarily closely by a finite dimensional subspace of E via a continuous mapping g_ε, the image of which is included in the convex hull of K.

Proof. The set K is compact, therefore for all $\varepsilon > 0$, there exist a finite number of points x_i of K, $i = 1, \dots, p$, such that K is covered by the open balls of center x_i and radius ε, i.e., $K \subset \cup_{i=1}^p B(x_i, \varepsilon)$. For each i, we define a nonnegative function on E by

$$\delta_i(x) = (\varepsilon - \|x - x_i\|_E)_+.$$

It is clear that $\delta_i \in C^0(E; \mathbb{R}_+)$ because it is a composition of continuous mappings and that it is strictly positive in the open ball $B(x_i, \varepsilon)$ and 0 elsewhere. Furthermore,

$$\forall x \in K, \quad \sum_{i=1}^p \delta_i(x) > 0.$$

Indeed, for all x in K, there exists an index j such that $x \in B(x_j, \varepsilon)$ by the covering property. For this particular j, we thus have $\delta_j(x) > 0$ and consequently, $\sum_{i=1}^p \delta_i(x) \geq \delta_j(x) > 0$.[7] Let us then set

$$g_\varepsilon(x) = \frac{\sum_{i=1}^p \delta_i(x) x_i}{\sum_{i=1}^p \delta_i(x)}.$$

It follows that $g_\varepsilon \in C^0(K; F_\varepsilon)$, where F_ε is the vector subspace of E generated by the points x_i, which implies that $\dim F_\varepsilon \leq p$. Furthermore, it is clear by construction that $g_\varepsilon(K) \subset \operatorname{conv} K$.

Let us now check the approximation property. For all x in K and all i, we may write $x_i = x + h_i$, where h_i is such that $\|h_i\|_E < \varepsilon$ if and only if $\delta_i(x) > 0$. Therefore

$$g_\varepsilon(x) = x + \frac{\sum_{i=1}^p \delta_i(x) h_i}{\sum_{i=1}^p \delta_i(x)},$$

[7] In fact, this sum is even bounded below on K by some constant $\delta > 0$.

and

$$\left\| \frac{\sum_{i=1}^{p} \delta_i(x) h_i}{\sum_{i=1}^{p} \delta_i(x)} \right\|_E \leq \frac{\sum_{i=1}^{p} \delta_i(x) \|h_i\|_E}{\sum_{i=1}^{p} \delta_i(x)} < \varepsilon$$

because the only nonzero terms in the sum in the numerator of the second term in the inequalities, are those for which $\|h_i\|_E < \varepsilon$. □

We now are in a position to establish a first version of the Schauder fixed point theorems.

Theorem 2.8. *Let E be a normed vector space, C a compact convex subset of E and T a continuous mapping from C into C. Then T admits at least one fixed point.*

Proof. According to Lemma 2.1, for all integers $n \geq 1$, there exist a finite dimensional vector subspace F_n of E and a continuous mapping g_n from C into F_n such that

$$\forall x \in C, \quad \|x - g_n(x)\|_E < \frac{1}{n} \text{ and } g_n(C) \subset \operatorname{conv} C = C.$$

Let us denote by $\overline{\operatorname{conv}}\, A$ the closed convex hull of a subset A of E, that is to say the closure of the convex hull of A.[8] We set $K_n = \overline{\operatorname{conv}}\, g_n(C)$. This is a convex subset of F_n, which is furthermore compact as a closed subset of the compact set C.

We now consider the mapping $T_n\colon K_n \to K_n$ defined by $T_n(x) = g_n(T(x))$. This mapping is a composition of continuous mappings, hence it is continuous. Consequently, we can apply Theorem 2.6 (remember that K_n is finite dimensional), and T_n has a fixed point $x_n \in K_n \subset C$. Now C is metric compact, so that we can extract from x_n a subsequence $x_{n'}$ that converges to a certain $x \in C$.

The mapping T is continuous, thus $T(x_{n'}) \to T(x)$ when $n' \to +\infty$. Then, by the triangle inequality, we see that

$$\begin{aligned}\|x - T(x)\|_E \leq \|x - x_{n'}\|_E &+ \|x_{n'} - T_{n'}(x_{n'})\|_E \\ &+ \|T_{n'}(x_{n'}) - T(x_{n'})\|_E + \|T(x_{n'}) - T(x)\|_E.\end{aligned}$$

The first and last terms in the right-hand side were just shown to tend to 0. The second term is identically 0 by the fixed point property for $T_{n'}$. Finally,

$$\|T_{n'}(x_{n'}) - T(x_{n'})\|_E = \|g_{n'}(T(x_{n'})) - T(x_{n'})\|_E < \frac{1}{n'} \to 0 \text{ when } n' \to +\infty,$$

by construction of $g_{n'}$. It follows that $x = T(x)$ is a fixed point of T. □

[8] Or equivalently, the smallest closed convex set containing A.

In the case of a Banach space, there is another version of the Schauder theorem which is often used.

Theorem 2.9. *Let E be a Banach space, C a closed convex subset of E and T a continuous mapping from C into C such that $T(C)$ is relatively compact. Then T admits at least one fixed point.*

Proof. Let $C' = \overline{\operatorname{conv}}\, T(C)$. This is a convex subset of C. Indeed, $T(C) \subset C$, so that $\operatorname{conv} T(C) \subset C$ because C is convex, and $\overline{\operatorname{conv}}\, T(C) \subset C$ because C is closed. Moreover, C' is compact as the closed convex hull of a relatively compact subset of a Banach space, cf. the next lemma.

We then apply the first Schauder theorem to the restriction of T to C'. □

Remark 2.4. i) In the sequel, any mention of the Schauder fixed point theorem will indifferently refer to Theorems 2.8 or 2.9, which are clearly equivalent to each other in a Banach space. In practice, we always work in Banach spaces.

ii) When applying Schauder's theorem to nonlinear boundary value problems, we have a certain amount of freedom. We must first reformulate the problem as a fixed point problem for a certain mapping T. Then we have to choose a space E on which T is continuous, then a closed convex C invariant by T, which is either compact or such that $T(C)$ is relatively compact. Let us note that the latter property can sometimes be proved by just showing that for all sequences $x_n \in C$, there exists a subsequence such that $T(x_{n'})$ converges in E, without necessarily proving that the subsequence $x_{n'}$ itself converges. Anyway, this not even always the case. □

In the course of the above proof, we have used a compactness result about closed convex hulls that deserves a proof of its own.

Lemma 2.2. *Let E be a Banach space and A a relatively compact subset of E. Then $\overline{\operatorname{conv}}\, A$ is compact.*

Proof. By the relative compactness of A, for all $\varepsilon > 0$, there exist a finite number of points $x_1, \dots, x_k$ of A such that the open balls centered at x_i and of radius $\varepsilon/2$ cover A, i.e.,

$$A \subset \bigcup_{i=1}^{k} B(x_i, \varepsilon/2). \tag{2.2}$$

We set $C = \operatorname{conv}\{x_1, \dots, x_k\}$. This is a bounded convex set of dimension less than $k - 1$, it is therefore relatively compact.[9] Thus there exist a finite number of points $y_1, \dots, y_m$ of $C \subset \operatorname{conv} A$ such that

$$C \subset \bigcup_{j=1}^{m} B(y_j, \varepsilon/2). \tag{2.3}$$

[9] Actually, it is compact.

Let us now take $z \in \operatorname{conv} A$. There exist a finite number of points $z_l \in A$, $l = 1, \dots, p$, and scalars $\lambda_l \in [0,1]$ with $\sum_{l=1}^{p} \lambda_l = 1$, such that $z = \sum_{l=1}^{p} \lambda_l z_l$. By the first covering (2.2), for each value of l, we can write

$$z_l = x_{k_l} + r_{k_l} \text{ for a certain index } k_l\text{, with } \|r_{k_l}\|_E \leq \varepsilon/2.$$

This yields

$$z = \sum_{l=1}^{p} \lambda_l x_{k_l} + \sum_{l=1}^{p} \lambda_l r_{k_l}.$$

Now, $\sum_{l=1}^{p} \lambda_l x_{k_l} \in C$, thus by the second covering (2.3), we can also write

$$\sum_{l=1}^{p} \lambda_l x_{k_l} = y_j + s_j \text{ for a certain index } j\text{, with } \|s_j\|_E \leq \varepsilon/2.$$

Consequently,

$$z = y_j + \Big(s_j + \sum_{l=1}^{p} \lambda_l r_{k_l}\Big),$$

with

$$\Big\| s_j + \sum_{l=1}^{p} \lambda_l r_{k_l} \Big\|_E \leq \varepsilon,$$

by the triangle inequality. In other words, we have shown that for all $\varepsilon > 0$, there exist a finite number of points $y_1, \dots, y_m$ of $\operatorname{conv} A$ such that

$$\operatorname{conv} A \subset \bigcup_{j=1}^{m} B(y_j, \varepsilon),$$

a property that is characteristic of relatively compact subsets of a complete metric space. □

Remark 2.5. If E is finite dimensional and if $K \subset E$ is compact, then $\operatorname{conv} K$ is also compact. In effect, there is a theorem by Carathéodory which states that if $A \subset E$, then $\operatorname{conv} A = \{x \in E, x = \sum_{i=1}^{\dim E+1} \lambda_i x_i, \lambda_i \geq 0, \sum_{i=1}^{\dim E+1} \lambda_i = 1, x_i \in A\}$. Therefore, if K is a compact subset of E, then $\operatorname{conv} K$ is the image of the compact set

$$K^{\dim E+1} \times \Big\{\lambda_i \geq 0, \sum_{i=1}^{\dim E+1} \lambda_i = 1\Big\}$$

by the continuous mapping $\big((x_i), (\lambda_i)\big) \mapsto \sum_{i=1}^{\dim E+1} \lambda_i x_i$.

This is no longer true in the infinite dimensional case. Here is a counterexample. Let us consider the space l^2 with its canonical Hilbert basis $(e_i)_{i\in\mathbb{N}}$. Let $K = \{e_i/i\}_{i\in\mathbb{N}^*} \cup \{0\}$, clearly a compact set. We consider the sequence

$$x_k = \sum_{i=1}^{k-1} \frac{e_i}{i2^i} + \big(1 - \frac{1}{2^k}\big)\frac{e_k}{k} \in \operatorname{conv} K.$$

Then $x_k \to \sum_{i=1}^{\infty} \frac{e_i}{i2^i} \notin \operatorname{conv} K$ when $k \to +\infty$. Therefore $\operatorname{conv} K$ is not closed, hence not compact. □

Let us mention that it is not necessary for E to be a normed space. In fact, the Tychonov fixed point theorem states that

Theorem 2.10. *Let E be a separated locally convex topological vector space, C a convex, compact subset of E and T a continuous mapping from C into C. Then T admits at least one fixed point.*

Proof. This result can be established using topological degree arguments, see [64]. □

Remark 2.6. We could think of using the Tychonov fixed point theorem in the following situation. Let E be a reflexive Banach space equipped with the weak topology. Then every bounded convex set is compact, and this part of the hypotheses comes for free. On the other hand, there will probably be difficulties in showing that a given nonlinear mapping T is continuous for the weak topology. We will see more precisely in Chap. 3 what kind of problems are likely to arise in this context. This limits the scope of applicability of this theorem, at least in such situations. □

We close this section with a few more fixed point theorems that may be useful. The first result is called the Schaefer fixed point theorem and is a consequence of the Schauder fixed point theorem.

Theorem 2.11. *Let E be a Banach space and T a continuous compact mapping from E into E which satisfies the following condition: there exists $R \geq 0$ such that if $x = tT(x)$ for some $t \in [0,1[$ then necessarily $\|x\|_E \leq R$. Then T admits at least one fixed point.*

Proof. We consider the mapping $T^*\colon \bar{B}(0, R+1) \to \bar{B}(0, R+1)$ defined by

$$T^*(x) = \begin{cases} T(x) & \text{if } \|T(x)\|_E \leq R+1, \\ (R+1)\frac{T(x)}{\|T(x)\|_E} & \text{if } \|T(x)\|_E > R+1. \end{cases}$$

The mapping T^* is the composition of T with a continuous mapping from E to E, it is thus continuous and compact. Therefore, $T^*(\bar{B}(0, R+1))$ is relatively compact and T^* admits a fixed point $x^* \in \bar{B}(0, R+1)$ by Schauder's theorem, second version.

Let us show that x^* is also a fixed point of T. If $\|T(x^*)\|_E \leq R+1$, then it clearly is a fixed point of T. If $\|T(x^*)\|_E > R+1$, then on the one hand $\|x^*\|_E = \|T^*(x^*)\|_E = R+1$, and on the other hand $x^* = tT(x^*)$ with $t = (R+1)/\|T(x^*)\|_E \in [0,1[$. It follows from the hypothesis that $\|x^*\|_E \leq R$, a contradiction, and the second case cannot occur. □

Remark 2.7. Replacing $R+1$ by $R+1/n$ with arbitrary $n \in \mathbb{N}^*$, we see that we can ensure that there is a fixed point such that $\|x^*\|_E \leq R$. □

There is a much more general version due to Leray and Schauder, using the Leray-Schauder topological degree theory, see [59].

Theorem 2.12. *Let E be a Banach space and T a continuous compact mapping from $[0,1] \times E$ into E that satisfies the following condition: $T(0,x) = 0$ and there exists $R \geq 0$ such that $x = T(t,x)$ with $t \in [0,1]$ implies that $\|x\|_E \leq R$. Then, for all $t \in [0,1]$, $T(t,.)$ admits a fixed point $x^*(t)$ which depends continuously on t.*

Let us finally mention,

Theorem 2.13. *Let E be a Banach space and T a continuous compact mapping from E into E such that there exists $R \geq 0$ with $T(\partial B(0,R)) \subset \bar{B}(0,R)$. Then T admits at least one fixed point.*

Proof. Similar to that of Theorem 2.11. □

Remark 2.8. The Leray and Schauder theorem is surprising. Consider the case of $(t,x) \mapsto tT(x)$. If we are able to somehow prove that, if each element in the family tT has a fixed point, then this fixed point must remain in a ball that is independent of t, then this (plus some compactness) implies the very existence of these fixed points. In other words, an a priori estimate of potential solutions, even before we know whether they exist or not, is enough to ensure that they actually exist. Let us note that the expression "a priori estimate" is often improperly used when estimating solutions the existence of which has already been established. □

2.3 Solving a Model Problem Using a Fixed Point Method

We are now going to illustrate how fixed point theorems can be used to solve nonlinear elliptic PDE problems. For this, we will consider the following very simple model problem. Let Ω be a bounded open subset of $\mathbb{R}^d$ and f be a function in $C^0(\mathbb{R}) \cap L^\infty(\mathbb{R})$. The problem consists in finding a function $u \in H_0^1(\Omega)$ such that $-\Delta u = f(u)$ in the sense of $\mathscr{D}'(\Omega)$. We will see later on how to make proper sense of this problem.

In order to rewrite it in the form of a fixed point problem, we must start with a linear existence and uniqueness result.

Proposition 2.1. *Let $g \in H^{-1}(\Omega)$. There exists a unique $v \in H_0^1(\Omega)$ such that $-\Delta v = g$ in the sense of $\mathscr{D}'(\Omega)$. This function v is also the unique solution of the* *variational problem:*

$$\forall w \in H_0^1(\Omega), \quad \int_\Omega \nabla v \cdot \nabla w \, dx = \langle g, w \rangle. \tag{2.4}$$

Moreover, the mapping $g \mapsto (-\Delta)^{-1} g = v$ is continuous from $H^{-1}(\Omega)$ into $H_0^1(\Omega)$.

Proof. See Chap. 1, Sect. 1.8, for the existence and uniqueness.

The continuity of $(-\Delta)^{-1}$ stems directly from the variational formulation with $w = v$ and the Poincaré inequality. Or again just as directly from the Lax-Milgram Theorem 1.22. □

Corollary 2.1. *The mapping $(-\Delta)^{-1}$ is continuous from $L^2(\Omega)$ into $H_0^1(\Omega)$.*

Proof. Indeed, $L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$. □

The right-hand side of the model problem contains a term of the form $f(u)$, the meaning of which has not been made precise yet. This is one of the objects of the Carathéodory Theorem 2.14 below. We first need a simple, almost obvious lemma. Let $\sim$ denote the almost everywhere equality equivalence relation for measurable functions.

Lemma 2.3. *Let Ω be an open set of $\mathbb{R}^d$ and $f \in C^0(\mathbb{R})$. For all pairs of measurable functions u_1 and u_2 on Ω, $f \circ u_i$ are measurable and if $u_1 \sim u_2$ then $f \circ u_1 \sim f \circ u_2$.*

Proof. Note first that if a function u is measurable $f \circ u$ then is also measurable, since f is continuous. Let us assume that $u_1 \sim u_2$, i.e., $u_1 = u_2$ almost everywhere in Ω. There thus exists a negligible set N such that if $x \notin N$ then $u_1(x) = u_2(x)$, hence $f(u_1(x)) = f(u_2(x))$, that is to say $f \circ u_1 \sim f \circ u_2$. □

We have just seen that the mapping $u \mapsto f \circ u$ is well-defined on the quotient space of measurable functions under almost everywhere equality. The Carathéodory theorem then states that

Theorem 2.14. *Let Ω be an open bounded subset of $\mathbb{R}^d$ and $f \in C^0(\mathbb{R})$ be such that $|f(t)| \leq a + b|t|$ for some $a, b \geq 0$. For any equivalence class u of measurable functions on Ω, we define the equivalence class $f(u) = f \circ u$ as in Lemma 2.3. Then the mapping $\tilde{f} : u \mapsto f \circ u$ is well-defined on $L^2(\Omega)$ with values in $L^2(\Omega)$ and is continuous for the strong topology.*

Proof. If $u \in L^2(\Omega)$, then

$$\int_\Omega |f(u)|^2 \, dx \leq 2a^2 \text{meas}\, \Omega + 2b^2 \|u\|_{L^2(\Omega)}^2 < +\infty,$$

so that $\tilde{f}(u) \in L^2(\Omega)$.

Let us show that the mapping thus defined is continuous from $L^2(\Omega)$ with the strong topology to $L^2(\Omega)$ also with the strong topology. Let thus u_n be a sequence in $L^2(\Omega)$ that converges to a limit u, and $u_{n'}$ be a subsequence of u_n. Due to Theorem 1.6, we can extract a further subsequence $u_{n''}$ that converges almost everywhere and such that there exists a function $g \in L^2(\Omega)$ with $|u_{n''}(x)| \leq g(x)$ almost everywhere.

We thus have $|f(u_{n''}(x)) - f(u(x))|^2 \to 0$ almost everywhere since f is continuous, and $|f(u_{n''}) - f(u)|^2 \leq 4a^2 + 4b^2g^2 + 2|f(u)|^2$ almost everywhere. The right-hand side of this estimate is a function in $L^1(\Omega)$ that does not depend on n''. We can thus apply the (direct) Lebesgue dominated convergence theorem to deduce that $\int_\Omega |f(u_{n''}) - f(u)|^2\,dx \to 0$, i.e., that $\tilde{f}(u_{n''}) \to \tilde{f}(u)$ in $L^2(\Omega)$ strong.

What we have shown is that, from any subsequence $\tilde{f}(u_{n'})$, we can extract a subsequence that converges to $\tilde{f}(u)$ in $L^2(\Omega)$ strong. The uniqueness of this limit implies that the whole sequence $\tilde{f}(u_n)$ converges to $\tilde{f}(u)$, cf. Lemma 1.1. Consequently, $\tilde{f}$ is continuous. □

Remark 2.9. i) The Ω bounded hypothesis is not necessary here. It is clearly enough that meas $\Omega < +\infty$. Likewise, the hypothesis that Ω is an open set of $\mathbb{R}^d$ equipped with the Lebesgue measure can obviously be considerably generalized.

ii) The Carathéodory theorem is actually more general than this. For example, let A be a Borel subset of $\mathbb{R}^d$ and $f\colon A \times \mathbb{R} \to \mathbb{R}$ a function such that

$$\begin{cases} f(\cdot, s) \text{ is measurable on } A \text{ for all } s \in \mathbb{R}, \\ f(x, \cdot) \text{ is continuous on } \mathbb{R} \text{ for almost all } x \in A, \end{cases}$$

(such a function is called a Carathéodory function). We assume that there exist exponents $1 \leq p, q < +\infty$, a function $a \in L^q(A)$ and a constant $b \geq 0$ such that

$$|f(x,s)| \leq a(x) + b|s|^{p/q} \quad \text{for almost all } x \text{ and for all } s.$$

Then the mapping $u \mapsto \tilde{f}(u)$ defined by $\tilde{f}(u)(x) = f(x, u(x))$ is continuous from $L^p(A)$ strong to $L^q(A)$ strong. The proof is very close to the previous one: measurability of $\tilde{f}(u)$ is established by approximating u almost everywhere by a sequence of simple functions, then the continuity from L^p to L^q follows from the partial converse of the dominated convergence theorem and then the dominated convergence theorem itself. □

We now are in a position to tackle the model problem.

Theorem 2.15. *Let Ω be a bounded open subset of $\mathbb{R}^d$ and $f \in C^0(\mathbb{R}) \cap L^\infty(\mathbb{R})$. Then there exists at least one solution $u \in H_0^1(\Omega)$ of the nonlinear partial differential equation $-\Delta u = f(u)$ in the sense of $\mathscr{D}'(\Omega)$.*

Proof. We give two proofs, each one using one variant of the Schauder theorem.

First Proof. We base everything on the Banach space $E = L^2(\Omega)$. According to Theorem 2.14, if $v \in E$ then $f(v) \in E$.[10] Let us set $T(v) = (-\Delta)^{-1}(f(v))$. Then $T\colon E \to E$ is continuous. Indeed, the mapping T is the composition of the following continuous mappings:

$$\begin{tikzcd}
L^2(\Omega) \arrow[r, "\tilde{f}"] & L^2(\Omega) \arrow[r, "(-\Delta)^{-1}"] & H_0^1(\Omega) \arrow[r, "\text{embedding}"] & L^2(\Omega) \\
v \arrow[r, mapsto] & f(v) \arrow[r, mapsto] & T(v) \arrow[r, mapsto] & T(v)
\end{tikzcd}$$

Let us check that any fixed point of T is a solution to our problem. Let thus $u \in L^2(\Omega)$ be such that $T(u) = u$. We first see that $u \in H_0^1(\Omega)$, because $T(u) = (-\Delta)^{-1}(f(u))$, on the one hand. On the other hand, by definition of the operator $(-\Delta)^{-1}$, $-\Delta T(u) = f(u)$ in the sense of $\mathscr{D}'(\Omega)$ and thus u solves the model problem, and conversely.

In order to apply Schauder's theorem, we must choose a compact, convex subset of $L^2(\Omega)$ that is invariant by T. We take here $C = \{v \in H_0^1(\Omega);\ \|v\|_{H_0^1(\Omega)} \leq M\}$ where M is a constant to be chosen later on. Here we use $\|v\|_{H_0^1(\Omega)} = \|\nabla v\|_{L^2(\Omega)}$, which is an equivalent norm on $H_0^1(\Omega)$ due to the Poincaré inequality. This set is a ball for a norm, hence it is convex. By Rellich's theorem, the embedding of $H_0^1(\Omega)$ into $L^2(\Omega)$ is compact, thus C, which is bounded in $H_0^1(\Omega)$, is relatively compact in E.

Furthermore, C is closed in E. Indeed, if we consider a sequence $v_n \in C$ that converges to $v \in L^2(\Omega)$ in the sense of $L^2(\Omega)$, then v_n is bounded in $H_0^1(\Omega)$ and we can thus extract a subsequence $v_{n'}$ that converges weakly to an element of $H_0^1(\Omega)$, which has no other choice than to be v. Hence $v \in H_0^1(\Omega)$. In addition, the weak sequential lower semicontinuity of the norm implies that $\|v\|_{H_0^1(\Omega)} \leq \liminf_{n' \to +\infty} \|v_{n'}\|_{H_0^1(\Omega)} \leq M$, i.e., that $v \in C$. Consequently, C is compact in E.

We now choose the constant M in such a way that $T(C) \subset C$. This is an *estimation* question. According to Proposition 2.1, $T(v)$ is the unique solution of the variational problem:

$$\forall w \in H_0^1(\Omega), \qquad \int_\Omega \nabla T(v) \cdot \nabla w \, dx = \int_\Omega f(v) w \, dx.$$

Taking $w = T(v)$ as a test-function, we obtain

$$\|\nabla T(v)\|_{L^2(\Omega)}^2 = \int_\Omega f(v) T(v) \, dx \leq \|f\|_{L^\infty(\mathbb{R})} \int_\Omega |T(v)| \, dx,$$

[10] Actually, here $f(v) \in L^\infty(\Omega)$.

since $|f(v)T(v)| \leq \|f\|_{L^\infty(\mathbb{R})}|T(v)|$.[11] It follows that

$$\|\nabla T(v)\|^2_{L^2(\Omega)} \leq \|f\|_{L^\infty(\mathbb{R})}(\text{meas}\,\Omega)^{1/2}\|T(v)\|_{L^2(\Omega)},$$

by the Cauchy-Schwarz inequality. The Poincaré inequality implies that there exists a constant C_Ω such that for all z in $H^1_0(\Omega)$, $\|z\|_{L^2(\Omega)} \leq C_\Omega\|\nabla z\|_{L^2(\Omega)}$. Since $T(v) \in H^1_0(\Omega)$, we thus see that for all v in E,

$$\|\nabla T(v)\|_{L^2(\Omega)} \leq C_\Omega\|f\|_{L^\infty(\mathbb{R})}(\text{meas}\,\Omega)^{1/2}.$$

To make sure that $T(C) \subset C$, it is thus enough to set $M = C_\Omega\|f\|_{L^\infty(\mathbb{R})}(\text{meas}\,\Omega)^{1/2}$, since in this case $T(E) \subset C$.

The hypotheses of the first version of Schauder's theorem are satisfied, consequently, there exists at least one solution to the model problem.

Second Proof. This time, we take $E = H^1_0(\Omega)$ and still let $T(v) = (-\Delta)^{-1}(f(v))$. Then $T\colon E \to E$ is continuous as the composition of the continuous mappings:

$$\begin{array}{ccccccc} & \text{embedding} & & \tilde f & & (-\Delta)^{-1} & \\ H^1_0(\Omega) & \longrightarrow & L^2(\Omega) & \longrightarrow & L^2(\Omega) & \longrightarrow & H^1_0(\Omega) \\ v & \longmapsto & v & \longmapsto & f(v) & \longmapsto & T(v) \end{array}$$

By Rellich's theorem, the first embedding is compact, therefore T transforms bounded subsets of E into relatively compact subsets of E, since the image of a compact set by a continuous mapping is compact.

We take the same convex set $C = \{v \in H^1_0(\Omega);\ \|v\|_{H^1_0(\Omega)} \leq M\}$ still with $M = C_\Omega\|f\|_{L^\infty(\mathbb{R})}(\text{meas}\,\Omega)^{1/2}$. By the same estimate as before, we thus have $T(C) \subset C$. The set C is closed ball of E, hence closed and convex. It is moreover bounded, so that $T(C)$ is relatively compact. The hypotheses of the second version of Schauder's theorem are satisfied. □

Remark 2.10. i) It is noteworthy that the ingredients used in the two proofs are essentially the same, but the order in which they are invoked is not the same.

ii) We assumed that f is bounded. If f is not bounded, solutions may exist, but this is not always the case. Let us give an example. Let $\lambda_1 > 0$ be the first eigenvalue of the $-\Delta$ operator on Ω with homogeneous Dirichlet condition on the boundary. By Proposition 1.12, we can choose an associated eigenfunction $\phi_1 \in H^1_0(\Omega)$ such that $\phi_1 > 0$ in Ω. We consider the function $f(t) = 1 + \lambda_1 t$. Then $-\Delta u = f(u)$ has no solution u in $H^1_0(\Omega)$. In effect, any solution should in particular satisfy

$$\int_\Omega \nabla u \cdot \nabla\phi_1\,dx = \int_\Omega \phi_1\,dx + \lambda_1\int_\Omega u\phi_1\,dx.$$

[11] This is where the hypothesis f bounded is used.

But ϕ_1 is an eigenfunction associated with the eigenvalue λ_1, so we also have

$$\int_\Omega \nabla\phi_1 \cdot \nabla u\, dx = \lambda_1 \int_\Omega \phi_1 u\, dx.$$

Consequently, $\int_\Omega \phi_1\, dx = 0$, which is impossible. □

Let us close this chapter with a uniqueness result for the model problem.

Theorem 2.16. *In addition to the previous hypotheses, we assume that f is nonincreasing. Then the solution u of the model problem is unique.*

Proof. Let u_1 and u_2 be two solutions. According to Theorem 2.15, they both satisfy

$$\forall v_1 \in H_0^1(\Omega), \quad \int_\Omega \nabla u_1 \cdot \nabla v_1\, dx = \int_\Omega f(u_1)v_1\, dx,$$
$$\forall v_2 \in H_0^1(\Omega), \quad \int_\Omega \nabla u_2 \cdot \nabla v_2\, dx = \int_\Omega f(u_2)v_2\, dx.$$

We take $v_1 = u_1 - u_2$ and $v_2 = u_2 - u_1$ and add the two equations. We obtain

$$\int_\Omega |\nabla(u_1 - u_2)|^2\, dx = \int_\Omega (f(u_1) - f(u_2))(u_1 - u_2)\, dx \le 0.$$

This is nonpositive because the right-hand side integrand is nonpositive due to f being nonincreasing. Consequently, $u_1 = u_2$ by the Poincaré inequality. □

Of course, in general, there is no reason for uniqueness to hold.

2.4 Exercises of Chap. 2

1. Show Theorem 2.13.

2. Show that every bounded open convex set of a Banach space is homeomorphic to the open ball of this Banach space. Deduce from this a fixed point theorem analogous to Theorem 2.13 using such a convex set.

3. Let E be a Banach space and $T: E \to E$ a continuous, compact mapping such that there exists $R > 0$ with

$$\|x - T(x)\|_E^2 \ge \|T(x)\|_E^2 - \|x\|_E^2$$

for all x such that $\|x\|_E \ge R$. Show that T admits a fixed point. (*Hint: note that if $0 \le t < 1$, the mapping tT can have no fixed point outside of the ball of radius R.*) Deduce from this that if T continuous, compact is such that $\|T(x)\|_E \le a\|x\|_E + b$ with $0 \le a < 1$, then T admits a fixed point.

4. Let $d \geq 3$, Ω a bounded open subset of $\mathbb{R}^d$ and $f\colon \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ a continuous mapping such that

$$|f(s,\xi)| \leq a + b|s|^{\frac{d}{d-2}} + c\|\xi\|$$

with $a, b, c \geq 0$.

4.1. Show that the mapping $u \mapsto f(u, \nabla u)$ is well defined and continuous from $H_0^1(\Omega)$ into $L^2(\Omega)$ with their respective strong topologies. (*Hint: use the Sobolev embeddings.*)

4.2. Assume now that $b = c = 0$. Show that the problem: $u \in H_0^1(\Omega)$, $-\Delta u = f(u, \nabla u)$ admits at least one solution. (*Hint: show that the* $(-\Delta)^{-1}$ *operator is compact from* $L^2(\Omega)$ to $H_0^1(\Omega)$ *by proving that if* $g_n \rightharpoonup g$ *in* $L^2(\Omega)$*, then* $u_n = (-\Delta)^{-1}g_n \rightharpoonup (-\Delta)^{-1}g = u$ *in* $H_0^1(\Omega)$ *and* $\|\nabla u_n\|_{L^2} \to \|\nabla u\|_{L^2}$*, using the variational formulation.*)

5. Let V and H be two Hilbert spaces such that $V \hookrightarrow H$ with a continuous, compact embedding. Let a be a bilinear form V, continuous and V-elliptic and $F\colon H \to V'$ be a continuous mapping such that there exists $R > 0$ with $F(H) \subset B_{V'}(0, R)$. Show that the variational problem: Find $u \in V$ such that

$$\forall v \in V, \quad a(u, v) = \langle F(u), v\rangle,$$

admits at least one solution.

6. An application of the previous exercise. Let Ω be a bounded open subset of $\mathbb{R}^d$ and A a $d \times d$ matrix-valued function on Ω, whose coefficients a_{ij} are in $L^\infty(\Omega)$ and such that there exists $\alpha > 0$ with

$$\sum_{i,j=1}^{d} a_{ij}(x)\xi_i\xi_j \geq \alpha\|\xi\|^2$$

for all $\xi \in \mathbb{R}^d$ and almost all $x \in \Omega$. Let us also be given $d + 1$ functions from $\mathbb{R}$ to $\mathbb{R}$ denoted f and g_i, $i = 1, \dots, d$, continuous and bounded.

Show that the boundary value problem

$$\begin{cases} -\mathrm{div}\,(A\nabla u) = f(u) + \displaystyle\sum_{i=1}^{d} \partial_i(g_i(u)) \text{ in } \Omega, \\ u = 0 \text{ on } \partial\Omega, \end{cases}$$

admits at least one solution.

Chapter 3
Superposition Operators

In the previous chapter, when studying the model problem $-\Delta u = f(u)$, we have already met operators of the form $u \mapsto f(u)$ where f is a mapping from $\mathbb{R}$ to $\mathbb{R}$, and u belongs to some function space defined over an open subset of $\mathbb{R}^d$. This type of operator is called a *superposition operator* or *Nemytsky operator*. Let us now study these operators in more detail in various functional contexts.

3.1 Superposition Operators in $L^p(\Omega)$

We have seen that under reasonable technical hypotheses, these operators continuously map $L^p(\Omega)$ equipped with the strong topology into $L^q(\Omega)$ also equipped with the strong topology, by Carathéodory's theorem. This no longer holds true for the weak topologies. In this section, we still use the notation $\tilde{f}$ for the mapping $u \mapsto f \circ u$. We focus on the case $p = q$ for brevity.

Proposition 3.1. *Let Ω and f be as in the Carathéodory theorem. The mapping $\tilde{f}$ is sequentially continuous from $L^p(\Omega)$ weak to $L^p(\Omega)$ weak for $p < +\infty$, and from $L^\infty(\Omega)$ weak-star to $L^\infty(\Omega)$ weak-star if and only if f is affine.*

Proof. Let Q be a cube included in Ω. By a change of coordinates, we can always assume that $Q =]0, 1[^d$. Let a and b be two arbitrary real numbers and θ a number between 0 and 1. We define a sequence of functions on $L^p(\Omega)$ by

$$u_n(x) = \begin{cases} 0 & \text{if } x \notin Q, \\ v_n(x_1) & \text{otherwise,} \end{cases}$$

H. Le Dret, *Nonlinear Elliptic Partial Differential Equations*, Universitext,
https://doi.org/10.1007/978-3-319-78390-1_3

where

$$v_n(t) = \begin{cases} a & \text{if } \frac{\lfloor nt\rfloor}{n} \leq t \leq \frac{\lfloor nt\rfloor+\theta}{n}, \\ b & \text{if } \frac{\lfloor nt\rfloor+\theta}{n} < t < \frac{\lfloor nt\rfloor+1}{n}, \end{cases}$$

where $\lfloor s\rfloor$ denotes the integer part of s. The restrictions to the cube Q of the functions u_n take values that oscillate between a and b increasingly rapidly as n grows. In fact, the functions v_n take the value a on those intervals that are of the form $\left[\frac{k}{n}, \frac{k+\theta}{n}\right]$ and the value b on those intervals that are of the form $\left]\frac{k+\theta}{n}, \frac{k+1}{n}\right[$, $k \in \mathbb{Z}$. The sequence $f \circ u_n$ has similar properties since

$$f \circ u_n(x) = \begin{cases} f(0) & \text{if } x \notin Q, \\ w_n(x_1) & \text{otherwise,} \end{cases}$$

where

$$w_n(t) = \begin{cases} f(a) & \text{if } \frac{\lfloor nt\rfloor}{n} \leq t \leq \frac{\lfloor nt\rfloor+\theta}{n}, \\ f(b) & \text{if } \frac{\lfloor nt\rfloor+\theta}{n} < t < \frac{\lfloor nt\rfloor+1}{n}. \end{cases}$$

Both sequences u_n and $f \circ u_n$ are bounded in $L^p(\Omega)$. We can thus extract a subsequence n' such that each one of them is weakly convergent, or weakly-star convergent in the case $p = +\infty$. We thus have $u_{n'} \rightharpoonup u$ and $f \circ u_{n'} \rightharpoonup g$ (the star is understood for $p = +\infty$) for some u and g, and our goal now is to identify u and g.

First of all, it is clear that

$$u(x) = \begin{cases} 0 & \text{if } x \notin Q, \\ v(x_1) & \text{otherwise,} \end{cases}$$

where v is the weak or weak-star limit in $L^p(0, 1)$ of the sequence $v_{n'}$ restricted to $[0, 1]$. Note that the space of L^p functions vanishing outside Q and depending only on x_1 is a closed subspace of $L^p(\Omega)$ isometric to $L^p(0, 1)$. We can thus work in one dimension of space only.

Let us consider any subinterval $[t_1, t_2]$ of $[0, 1]$. From the weak or weak-star convergence of the sequence $v_{n'}$, we deduce that

$$\int_{t_1}^{t_2} v_{n'}(t)\,dt \longrightarrow \int_{t_1}^{t_2} v(t)\,dt \text{ when } n' \to +\infty.$$

Let us directly compute the limit of the left-hand side by slicing up the interval in parts that coincide with the oscillations of $v_{n'}$,

$$\int_{t_1}^{t_2} v_{n'}(t)\,dt = \int_{t_1}^{\frac{\lfloor n't_1\rfloor+1}{n'}} v_{n'}(t)\,dt + \sum_{k=\lfloor n't_1\rfloor+1}^{\lfloor n't_2\rfloor-1} \int_{\frac{k}{n'}}^{\frac{k+1}{n'}} v_{n'}(t)\,dt + \int_{\frac{\lfloor n't_2\rfloor}{n'}}^{t_2} v_{n'}(t)\,dt.$$

By construction of v_n, we see that

$$\int_{\frac{k}{n'}}^{\frac{k+1}{n'}} v_{n'}(t)\,dt = \frac{\theta a + (1-\theta)b}{n'}.$$

It is furthermore easy to bound the two integrals on the ends from above

$$\max\left\{\left|\int_{t_1}^{\frac{\lfloor n't_1\rfloor+1}{n'}} v_{n'}(t)\,dt\right|, \left|\int_{\frac{\lfloor n't_2\rfloor}{n'}}^{t_2} v_{n'}(t)\,dt\right|\right\} \leq \frac{\max\{|a|,|b|\}}{n'}.$$

We thus obtain

$$\begin{aligned}\int_{t_1}^{t_2} v_{n'}(t)\,dt &= \frac{\lfloor n't_2\rfloor - \lfloor n't_1\rfloor - 1}{n'}(\theta a + (1-\theta)b)\\ &\qquad + \int_{t_1}^{\frac{\lfloor n't_1\rfloor+1}{n'}} v_{n'}(t)\,dt + \int_{\frac{\lfloor n't_2\rfloor}{n'}}^{t_2} v_{n'}(t)\,dt\\ &= \frac{\lfloor n't_2\rfloor - \lfloor n't_1\rfloor - 1}{n'}(\theta a + (1-\theta)b) + r_{n'},\end{aligned}$$

with $|r_{n'}| \leq 2\frac{\max\{|a|,|b|\}}{n'}$. Now clearly, $\frac{\lfloor n't_2\rfloor - \lfloor n't_1\rfloor - 1}{n'} \to t_2 - t_1$ when $n' \to +\infty$, which shows that

$$\int_{t_1}^{t_2} v_{n'}(t)\,dt \longrightarrow (t_2 - t_1)(\theta a + (1-\theta)b).$$

Consequently,

$$\frac{1}{t_2 - t_1}\int_{t_1}^{t_2} v(t)\,dt = \theta a + (1-\theta)b.$$

We let t_2 tend to t_1 and apply the Lebesgue points theorem to conclude that v is almost everywhere equal to the constant function $t \mapsto \theta a + (1-\theta)b$, see Fig. 3.1.

The same argument applied to the sequence $w_{n'}$ shows that

$$g(x) = \begin{cases} f(0) & \text{if } x \notin Q,\\ \theta f(a) + (1-\theta)f(b) & \text{otherwise.}\end{cases}$$

A necessary condition for weak or weak-star sequential continuity, thus a fortiori weak or weak-star continuity of $\tilde{f}$ is therefore that $\tilde{f}(u) = g$, or in other words that

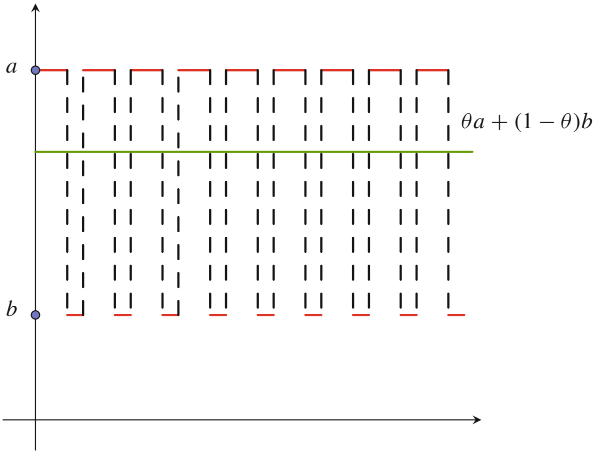


Fig. 3.1 A term of the oscillating sequence v_n and its constant weak limit

for all a, b and θ,

$$f(\theta a + (1-\theta)b) = \theta f(a) + (1-\theta) f(b).$$

Thus, f must be affine. The converse is clear. □

Remark 3.1. i) The proof shows that not only do subsequences weakly converge, but actually the whole sequence weakly converges by Lemma 1.1.

ii) On the other hand, this is an example of a sequence that does not admit any almost everywhere convergent subsequence, whenever $a \neq b$. Let us take $a = 0$, $b = 1$ and $\theta = \frac{1}{2}$. If there was an almost everywhere convergent subsequence, its almost everywhere limit would be $\{0, 1\}$-valued. By the dominated convergence theorem, it would also converge strongly in all $L^p(0, 1)$, $p < +\infty$. This contradicts its weak convergence to the constant function $t \mapsto 1/2$. In fact, it follows from the previous considerations, that any subsequence may converge at most on a negligible set.

iii) The same proof works between L^p and L^q. It shows that the superposition operator is weakly continuous between L^p and L^q if and only if f is affine when $p \geq q$, and if and only if f is constant when $p < q$.

iv) We have used the weak limit of a sequence of functions that oscillate between two values. This idea, suitably generalized, has a number of other applications. Let us note that if $f \in L^\infty(\mathbb{R})$ is T-periodic, then it can be shown in a very similar way that the sequence $f_n(t) = f(nt)$ converges in $L^\infty(\mathbb{R})$ weak-star to the constant function equal to the average value of f, $\frac{1}{T}\int_0^T f(t)\,dt$.

v) We see in this example that weak convergence generally does not get along well with nonlinearity. This is part of what makes nonlinear PDE problems more delicate than linear ones. □

3.2 Young Measures

The link between weak convergence in Lebesgue spaces and nonlinear mappings can be described in a finer manner by objects that are called *Young measures*, see [69]. Indeed, if we know that in general $f(u_n)$ does not converge weakly to $f(u)$ when u_n tends weakly to u, as we have just seen, we still would like to determine the weak limit of $f(u_n)$ when f is not affine.[1] This is the object of the next theorem.

Theorem 3.1. *Let Ω be an open subset of $\mathbb{R}^d$ and u_n a bounded sequence in $L^\infty(\Omega)$. Then there exists a subsequence $u_{n'}$ and for almost all $x \in \Omega$ a Borel probability measure ν_x on $\mathbb{R}$ such that, for all $f \in C^0(\mathbb{R})$, there holds*

$$f(u_{n'}) \overset{*}{\rightharpoonup} \bar{f} \textit{ in } L^\infty(\Omega), \tag{3.1}$$

where

$$\bar{f}(x) = \int_{\mathbb{R}} f(y)\,d\nu_x(y) \textit{ a.e. in } \Omega. \tag{3.2}$$

Proof. We first extract a subsequence that will work. By hypothesis, there is a number M such that $\|u_n\|_{L^\infty(\Omega)} \le M$ for all n. Let $(p_k)_{k\in\mathbb{N}}$ be a dense countable family in $C^0([-M,M])$. Such a family exists due to the Weierstrass theorem, for example the polynomials with rational coefficients. For each integer k, we trivially have $\|p_k(u_n)\|_{L^\infty(\Omega)} \le \|p_k\|_{C^0([-M,M])}$. Thus, by the diagonal argument, there exists a subsequence $u_{n'}$ and for all $k \in \mathbb{N}$ a certain $\bar{p}_k \in L^\infty(\Omega)$ such that

$$p_k(u_{n'}) \overset{*}{\rightharpoonup} \bar{p}_k.$$

To go from the countable family to the whole space, we remark that if $f \in C^0([-M,M])$, then by the same argument as before, there exists a subsequence n'' of the sequence n' such that $f(u_{n''}) \overset{*}{\rightharpoonup} \bar{f}$ in $L^\infty(\Omega)$ for a certain $\bar{f}$. Let us show that in fact, the whole sequence $f(u_{n'})$ converges and thus that $\bar{f}$ is entirely determined by the first subsequence n'.

Let us take $g \in L^1(\Omega)$ nonzero, $\varepsilon > 0$ and by density of the family, choose an integer k in such a way that $\|f - p_k\|_{C^0([-M,M])} \le \frac{\varepsilon}{4\|g\|_{L^1(\Omega)}}$. There holds

$$\begin{aligned}\int_\Omega (f(u_{n'}) - f(u_{m'}))g\,dx &= \int_\Omega (f(u_{n'}) - p_k(u_{n'}))g\,dx \\ &\quad + \int_\Omega (p_k(u_{n'}) - p_k(u_{m'}))g\,dx + \int_\Omega (p_k(u_{m'}) - f(u_{m'}))g\,dx,\end{aligned}$$

[1] As of now, we stop distinguishing between function f and superposition operator $\tilde{f}$ for a lighter notation.

so that

$$\Big|\int_\Omega (f(u_{n'}) - f(u_{m'}))g\,dx\Big| \le \Big|\int_\Omega (p_k(u_{n'}) - p_k(u_{m'}))g\,dx\Big| + \frac{\varepsilon}{2},$$

from which it follows that the sequence $\int_\Omega f(u_{n'})g\,dx$ is Cauchy in $\mathbb{R}$. Since it contains a subsequence indexed by n'' that converges to $\int_\Omega \bar f g\,dx$, it is convergent with that same limit and we obtain (3.1).

The mapping $f \mapsto \bar f$ thus defined is obviously linear. It is moreover continuous, with norm 1 from $C^0([-M, M])$ to $L^\infty(\Omega)$. As a matter of fact,

$$\|\bar f\|_{L^\infty(\Omega)} = \sup_{\|g\|_{L^1(\Omega)}\le 1} \int_\Omega \bar f g\,dx = \sup_{\|g\|_{L^1(\Omega)}\le 1} \lim_{n'\to\infty} \int_\Omega f(u_{n'})g\,dx.$$

Now for such functions g, Hölder's inequality implies that

$$\int_\Omega f(u_{n'})g\,dx \le \|f(u_{n'})\|_{L^\infty(\Omega)} \le \|f\|_{C^0([-M,M])}.$$

Therefore

$$\|\bar f\|_{L^\infty(\Omega)} \le \|f\|_{C^0([-M,M])}.$$

Another way of seeing this consists in saying that $\|f(u_{n'})\|_{L^\infty(\Omega)} \le \|f\|_{C^0([-M,M])}$ and that since $f(u_{n'}) \overset{*}{\rightharpoonup} \bar f$, we have $\|\bar f\|_{L^\infty(\Omega)} \le \liminf \|f(u_{n'})\|_{L^\infty(\Omega)}$. So the norm of the linear map $f \mapsto \bar f$ is less than 1, with equality achieved for $f = 1$.

Let us now establish the representation formula (3.2). According to the Lebesgue points theorem (see Theorem 1.7 of Chap. 1), for each $k \in \mathbb{N}$, there exists a negligible set $N_k \subset \Omega$ such that if $x \notin N_k$, then

$$\frac{1}{\text{meas}\, B(x,\rho)} \int_{B(x,\rho)} \bar p_k(y)\,dy \to \bar p_k(x) \text{ when } \rho \to 0.$$

Let $V = \text{vect}\{(p_k)_{k\in\mathbb{N}}\}$, which is a dense vector subspace of $C^0([-M, M])$. We set $N = \cup_{k\in\mathbb{N}} N_k$. It is still a negligible set. Every element q of V is by definition a linear combination of the p_k, and by linearity the corresponding weak limit $\bar q$ is the same linear combination of $\bar p_k$. Consequently, we see that for all $x \notin N$ and all $q \in V$, there holds

$$\frac{1}{\text{meas}\, B(x,\rho)} \int_{B(x,\rho)} \bar q(y)\,dy \to \bar q(x) \text{ when } \rho \to 0.$$

For all $x \notin N$, we thus introduce a linear form ℓ_x on V by

$$\ell_x(q) = \lim_{\rho\to 0}\Bigl(\frac{1}{\text{meas } B(x,\rho)}\int_{B(x,\rho)} \bar{q}(y)\,dy\Bigr).$$

This linear form is continuous on V because

$$\frac{1}{\text{meas } B(x,\rho)}\Bigl|\int_{B(x,\rho)} \bar{q}(y)\,dy\Bigr| \leq \|\bar{q}\|_{L^\infty(\Omega)} \leq \|q\|_{C^0([-M,M])},$$

so that passing to the limit when $\rho \to 0$,

$$|\ell_x(q)| \leq \|q\|_{C^0([-M,M])}.$$

It thus has a unique continuous extension, still denoted ℓ_x, to $C^0([-M, M])$. For all $f \in C^0([-M, M])$, we thus have a mapping $x \mapsto \ell_x(f)$ defined almost everywhere on Ω. By construction, for all $q \in V$, $\bar{q} = \ell_x(q)$, therefore

$$\bar{f} - \ell_x(f) = \bar{f} - \bar{q} + \bar{q} - \ell_x(q) + \ell_x(q - f),$$

thus

$$\|\bar{f} - \ell_x(f)\|_{L^\infty(\Omega)} \leq 2\|f - q\|_{C^0([-M,M])}.$$

It follows that $\bar{f} = \ell_x(f)$ almost everywhere for all f in $C^0([-M, M])$, by density of V.

We now show that for all $x \notin N$, the linear form ℓ_x is positive. We have to prove that if $f \geq 0$, then $\ell_x(f) \geq 0$ for all $x \in N$. It could be tempting to say that $\bar{f} = \ell_x(f)$ almost everywhere and that $\bar{f} \geq 0$ if $f \geq 0$ to conclude, but this does not work, because this particular "almost everywhere" depends on f. Moreover, we need the result for all such f, which form an uncountable set.

We go around this difficulty as follows. If $f \geq 0$, then for all $p \in \mathbb{N}^*$, $f + \frac{1}{p} \geq \frac{1}{p}$, and we can find $q_p \in V$ such that $|f + \frac{1}{p} - q_p| \leq \frac{1}{2p}$. We thus have $q_p \geq 0$ and $q_p \to f$ in $C^0([-M, M])$ when $p \to +\infty$. Naturally, a weak-star limit of a sequence of nonnegative functions is nonnegative, since for all $g \in L^1(\Omega)$, $g \geq 0$, then $0 \leq \int_\Omega q_p(u_{n'})g\,dx \to \int_\Omega \bar{q}_p g\,dx$. So we have that $\bar{q}_p \geq 0$ and consequently, $\ell_x(q_p) \geq 0$ for all $x \notin N$ by construction. It follows that $0 \leq \lim_{p\to+\infty} \ell_x(q_p) = \ell_x(f)$ for all such x.

The Riesz representation theorem, see for example [61], tells us that ℓ_x is represented by a positive Radon measure ν_x on $[-M, M]$, as follows

$$\forall f \in C^0([-M, M]), \quad \ell_x(f) = \int_{-M}^{M} f(y)\,d\nu_x(y),$$

(a Radon measure is a Borel measure that is finite on compact sets). Extending ν_x by 0 outside of $[-M, M]$, we finally obtain

$$\forall f \in C^0(\mathbb{R}), \quad \bar{f}(x) = \int_{\mathbb{R}} f(y)\, d\nu_x(y) \text{ a.e. in } \Omega.$$

To conclude the proof, we apply the above formula to $f = 1$, for which $\bar{f} = 1$, hence $1 = \int_{\mathbb{R}} d\nu_x(y)$ et ν_x is a probability measure. □

The family of probability measures ν_x parametrized by the points x of Ω is called the *family of Young measures* associated with the subsequence $u_{n'}$. They are also sometimes called a *parametrized measure*. The family describes the asymptotic distribution of the values taken by the functions $u_{n'}$. It clearly depends on the subsequence.[2] In the example of the functions u_n of Proposition 3.1, the Young measures do not depend on the subsequence, nor on x, and we have $\nu_x = \theta\delta_a + (1-\theta)\delta_b$. This means that the sequence u_n in question takes the value a an average time θ and the value b an average time $1-\theta$.

For an example of Young measure with actual dependence on x, consider $u_n + \sin x$, where $\nu_x = \theta\delta_{a+\sin x} + (1-\theta)\delta_{b+\sin x}$. In general, even if the weak limit of the sequence is a constant function, the associated Young measures depend on x, take for instance $u_n(x) = \sin x \sin nx$.

Remark 3.2. Taking $f(y) = y$, we obtain in particular that

$$u_{n'} \overset{*}{\rightharpoonup} u = \int_{\mathbb{R}} y\, d\nu_x(y) \text{ in } L^\infty(\Omega), \tag{3.3}$$

thus the Young measures also obviously encode the weak-star limit of the sequence $u_{n'}$ itself. They however contain a lot more information on the asymptotic behavior of the sequence than just its weak-star limit, which smoothes out oscillations, cf. the above $u_n(x) = \sin x \sin nx$ example. □

Remark 3.3. We can wonder why take the rather back road of the Lebesgue points theorem instead of just letting $\ell_x(q) = \bar{q}(x)$. The reason is to have a choice of representative of $\bar{q} \in V$ which ensures the continuity and positivity of ℓ_x. Indeed, as already said before, ν_x are defined for almost all x once and for all, but the equality $\bar{f} = \int_{\mathbb{R}} f\, d\nu_x$ holds in another almost everywhere sense that depends on f.

We could as well have considered the restriction to the subset formed of linear combinations of the p_k with rational coefficients, which is still dense, although not a vector space. This allows for eliminating a negligible set, albeit not the same negligible set, outside of which the above definition works. □

Let us now give a few of the interesting properties of Young measures, and some results that follow from them.

[2] Take two sequences with two different Young measures and mix them.

Corollary 3.1. *If f is convex, there holds $f(u(x)) \leq \bar{f}(x)$ almost everywhere.*

Proof. This is Jensen's inequality, see for example [61]. For almost all x in Ω,

$$f(u(x)) = f\Big(\int_{\mathbb{R}} y\, d\nu_x(y)\Big) \leq \int_{\mathbb{R}} f(y)\, d\nu_x(y) = \bar{f}(x). \qquad \square$$

Thus, for example, $u^2 \leq \lim (u_{n'})^2$ almost everywhere.

An important result in applications of Young measures is that they allows for a characterization of strong convergence.

Proposition 3.2. *Let Ω be a bounded open subset of $\mathbb{R}^d$ and ν_x the Young measures associated with a subsequence $u_{n'}$ bounded in $L^\infty(\Omega)$. If ν_x is a Dirac mass for almost all x, then $u_{n'}$ converges strongly to u in all L^p, $p < +\infty$, and (modulo a subsequence) almost everywhere. Conversely, if u_n tends to u strongly in $L^p(\Omega)$, then the Young measures are Dirac masses for almost all x.*

Proof. We first notice that if ν_x is a Dirac mass, then necessarily $\nu_x = \delta_{u(x)}$ almost everywhere according to formula (3.3). With the choice $f(y) = |y|^p$, we thus obtain

$$|u_{n'}|^p \overset{*}{\rightharpoonup} \int_{\mathbb{R}} |y|^p\, d\nu_x(y) = |u(x)|^p \text{ in } L^\infty(\Omega),$$

hence, integrating over Ω, which is bounded and thus $1 \in L^1(\Omega)$,

$$\|u_{n'}\|^p_{L^p(\Omega)} = \int_\Omega |u_{n'}|^p\, dx \to \int_\Omega |u|^p\, dx = \|u\|^p_{L^p(\Omega)}.$$

Weak convergence plus convergence of the norms imply strong convergence in $L^p(\Omega)$ for all $p > 1$, see Proposition 1.9. Strong convergence in $L^1(\Omega)$ then follows from Hölder's inequality.

Conversely, let us assume that $u_n \to u$ in $L^p(\Omega)$ strong. By hypothesis, u_n is bounded in $L^\infty(\Omega)$, thus the Carathéodory theorem implies that for all continuous f, there holds $f(u_n) \to f(u)$ in $L^p(\Omega)$. This also holds *a fortiori* in $L^\infty(\Omega)$ weak-star. This implies the uniqueness of the Young measures and that

$$f\Big(\int_{\mathbb{R}} y\, d\nu_x(y)\Big) = \int_{\mathbb{R}} f(y)\, d\nu_x(y),$$

for all f.

Let us apply this equality to $f(y) = y^2$. We obtain

$$\Big(\int_{\mathbb{R}} y\, d\nu_x(y)\Big)^2 = \int_{\mathbb{R}} y^2\, d\nu_x(y).$$

Now the Cauchy-Schwarz inequality states that

$$\Big(\int_{\mathbb{R}} y \times 1\, d\nu_x(y)\Big)^2 \leq \int_{\mathbb{R}} y^2\, d\nu_x(y) \int_{\mathbb{R}} d\nu_x(y) = \int_{\mathbb{R}} y^2\, d\nu_x(y).$$

We are thus in a case of equality in the Cauchy-Schwarz inequality. This can only happen if the functions $y \mapsto y$ and $y \mapsto 1$ are collinear as elements of the vector space $L^2(\mathbb{R}, d\nu_x)$. This means that $y = \lambda_x$ for a certain constant λ_x, ν_x-almost everywhere on $\mathbb{R}$, i.e., $\nu_x(\{y \neq \lambda_x\}) = \nu_x(\mathbb{R} \setminus \{\lambda_x\}) = 0$. The measure ν_x is nonnegative, thus this implies that the support ν_x is reduced to a point, the value λ_x in question. Now ν_x is a probability measure, it is thus the Dirac mass $\nu_x = \delta_{\lambda_x}$. Finally, obviously $\lambda_x = u(x)$ almost everywhere in Ω. □

Remark 3.4. By contraposition, if we are faced with a weak-star converging sequence, no subsequence of which converges strongly or almost everywhere, then there is a set of strictly positive measure in Ω where the Young measures are not Dirac masses, and conversely. The Young measures provide a quantitative information on the oscillations of a weakly convergent sequence. The measure ν_x represents in a way the asymptotic repartition of the values taken by the sequence u_n at point x. □

It is possible to gain further information of the structure of a given family of Young measures by using particular choices of f. Thus for exemple, if there exists a closed subset C of R such that $u_n(x) \in C$ for all n and almost all x, then the support of ν_x is included in C for almost all x. This can be seen by considering functions f that vanish on C. It is also possible in certain contexts to deduce restrictions on the form of ν_x, for example that they are Dirac masses, from PDEs satisfied by u_n. This is one of the Yound measures' usefulness in PDEs, see [22, 67]. Finally, let us mention that there are L^p, $W^{1,p}$, etc. versions of Young measures, beyond the simple L^∞ version presented here, see [6, 41, 42].

3.3 Superposition Operators in $W^{1,p}(\Omega)$

The properties of superposition operators in Sobolev spaces are markedly different from their properties in Lebesgue spaces.

In what follows, Ω is an arbitrary open subset of $\mathbb{R}^d$, without any particular regularity property unless otherwise specified. We start by defining the level sets of a locally integrable function. In the case of a continuous function u, the c-level set is simply the preimage $u^{-1}(\{c\})$. In the L^1_{loc} case, the difficulty is that we are dealing not with a single function, but with an equivalence class of functions under almost everywhere equality. We thus cannot use a definition using a preimage, which depends on the choice of a class representative. We have to proceed in a roundabout way.

Definition 3.1. Let $u \in L^1_{\text{loc}}(\Omega)$ and $c \in \mathbb{R}$. We set

$$E_c(u) = \Big\{x \in \Omega;\ \frac{1}{\operatorname{meas} B(x,\rho)} \int_{B(x,\rho)} |u(y) - c|\, dy \to 0 \text{ when } \rho \to 0\Big\}. \tag{3.4}$$

The set $E_c(u)$ is defined based on integral quantities, it thus only depends on the equivalence class of u and not on such or such representative of the class. To be more precise,

Proposition 3.3. *Let u_1 be a measurable function that represents the equivalence class $u \in L^1_{\text{loc}}(\Omega)$. Then $u_1(x) = c$ almost everywhere on $E_c(u)$ and $u_1(x) \neq c$ almost everywhere on $\Omega \setminus E_c(u)$.*

Proof. According to the Lebesgue points theorem, there exists a negligible set $N \subset \Omega$ such that if $x \notin N$, there holds

$$\frac{1}{\operatorname{meas} B(x,\rho)} \int_{B(x,\rho)} |u_1(y) - c|\, dy \to |u_1(x) - c| \text{ when } \rho \to 0.$$

It follows that if $x \in E_c(u) \setminus N$, then $|u_1(x) - c| = 0$, whereas if $x \in \big(\Omega \setminus E_c(u)\big) \setminus N$, then $|u_1(x) - c| \neq 0$. □

Remark 3.5. The set $E_c(u)$ is thus a level set that is defined in a reasonable way for a locally integrable (class of) function(s). When u is continuous, in the sense that there exists a continuous representative, and that this is the representative we choose, then it is easily checked that $E_c(u) = u^{-1}(\{c\})$. □

We are going to prove simultaneously two important results concerning superposition operators in $W^{1,p}(\Omega)$. The first result is actually just a property of $W^{1,p}(\Omega)$ functions, without actual direct reference to superposition operators.

Theorem 3.2. *Let $u \in W^{1,p}(\Omega)$. Then for all $c \in \mathbb{R}$, $\nabla u = 0$ almost everywhere on $E_c(u)$.*

Remark 3.6. i) We note that this result may look contradictory at first glance, since it seems to be implying that ∇u vanishes almost everywhere. This is of course not the case at all, because even if we had $\Omega = \cup_{c \in \mathbb{R}} E_c(u)$, the set $\mathbb{R}$ *is not countable* and a measure is only countably additive. The uncountable union of negligible sets on which ∇u does not vanish has no reason to be negligible, and is indeed of strictly positive measure in general.

Moreover when $\operatorname{meas} E_c(u) = 0$, then the statement is of course correct, but is also empty. Now this the generic situation with respect to c. Let us give an example: let $\Omega =]-2, 2[$ and $u(x) = x + 1$ if $-2 \leq x < -1$, $u(x) = 0$ if $-1 \leq x < 1$ and $u(x) = x - 1$ if $1 \leq x \leq 2$ (here we obviously choose the continuous representative). We easily check that $u \in H^1(\Omega)$ with $u'(x) = 1$ if $-2 < x < -1$ or $1 < x < 2$, and $u'(x) = 0$ if $-1 < x < 1$. Thus, if $c \neq 0$, $E_c(u)$ is either empty or a singleton,

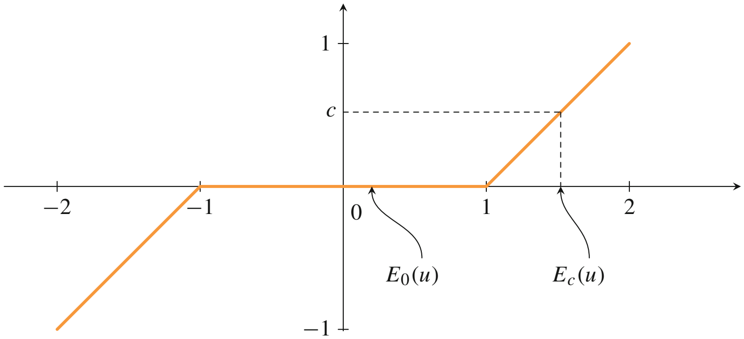


Fig. 3.2 Almost everywhere nullity of the gradient on level sets

hence of zero measure and $u' = 0$ almost everywhere on it of course,[3] and $E_0(u) = [-1, 1]$, a set on which u' effectively vanishes almost everywhere. This is illustrated on Fig. 3.2.

ii) Even a function in $W^{1,p}(\Omega)$ may not have easily defined level sets. In one dimension of space, such functions are continuous, but as soon as the space dimension is higher than 2, there are functions in $W^{1,p}(\Omega)$ that are very discontinuous and for which there is no simple privileged choice of representative. A definition using almost everywhere equality equivalence classes is necessary. □

The second result concerns the superposition operators strictly speaking, but is not independent from the first result. It is a version of a theorem by Stampacchia, see [65].

Theorem 3.3. *Let T be a globally Lipschitz continuous function from $\mathbb{R}$ to $\mathbb{R}$, piecewise C^1 and with only a finite number of points of non differentiability $c_1, c_2, \dots, c_k$. If* meas $\Omega = +\infty$, *we additionally assume that $T(0) = 0$. Then for all $u \in W^{1,p}(\Omega)$,*

i) $T(u) \in W^{1,p}(\Omega)$,

ii) $\nabla(T(u)) = T'(u)\nabla u$ on $\Omega \setminus \cup_{i=1}^k E_{c_i}(u)$ and $\nabla(T(u)) = 0$ almost everywhere on $\cup_{i=1}^k E_{c_i}(u)$,

Remark 3.7. i) In the sequel, instead of using the description of the gradient supplied by Theorem 3.3 ii), we will simply write $\nabla(T(u)) = T'(u)\nabla u$ almost everywhere, with the convention that the product is 0 where $T'(u)$ is not defined, i.e., on $\cup_{i=1}^k E_{c_i}(u)$,[4] since $\nabla u = 0$ almost everywhere on this set anyway.

ii) If meas $\Omega = +\infty$ but $T(0) \neq 0$, then $T(u) \notin L^p(\Omega)$. Nonetheless, it is still true that $T(u) \in L^p_{\text{loc}}(\Omega)$ and $\nabla(T(u)) \in L^p(\Omega; \mathbb{R}^d)$.

[3] As well as taking any other value almost everywhere.

[4] Of course, $T'(u)$ may well be undefined elsewhere, but that would be on a negligible set.

iii) It is not necessary to assume T to be piecewise C^1 with a finite number of points of non differentiability. The result remains true for any globally Lipschitz continuous T, i.e., that $T(u) \in W^{1,p}(\Omega)$ if $u \in W^{1,p}(\Omega)$. This is not too difficult to prove by approximation, as in the following proof. On the other hand, writing a formula for the gradient is more delicate and we will not do it here. The fact that $\nabla u = 0$ almost everywhere on the preimage by u of any negligible set comes into play. □

We decompose the proof of these two theorems into a series of lemmas, proceeding by successive approximations. We first start with the case of a smooth function.

Lemma 3.1. *Let $S\colon \mathbb{R} \to \mathbb{R}$ be a globally Lipschitz function of class C^1 (with $S(0) = 0$ if* $\operatorname{meas}\Omega = +\infty$*). For all $u \in W^{1,p}(\Omega)$, then $S(u) \in W^{1,p}(\Omega)$ and $\nabla(S(u)) = S'(u)\nabla u$.*

Proof. The function S is globally Lipschitz, hence there exists a constant L such that $|S(t) - S(s)| \le L|t - s|$ for all t, s. In particular,

$$|S(t)| \le |S(0)| + L|t|.$$

Moreover, since S is of class C^1, it follows that

$$\left|\frac{1}{t-s}\int_s^t S'(u)\,du\right| \le L.$$

Letting s tend to t, we obtain that $|S'(t)| \le L$ for all $t \in \mathbb{R}$, and $S' \in L^\infty(\mathbb{R})$.

Let us start with the case $p < +\infty$. We consider a function $\phi \in C^\infty(\Omega) \cap W^{1,p}(\Omega)$. By the chain rule, $S(\phi) \in C^1(\Omega)$ and $\nabla(S(\phi)) = S'(\phi)\nabla\phi$. Moreover

$$\int_\Omega |S(\phi)|^p\,dx \le 2^{p-1}\big(|S(0)|^p \operatorname{meas}\Omega + L^p\|\phi\|^p_{L^p(\Omega)}\big) < +\infty,$$

by the first estimate we just noted above on the one hand, and

$$\int_\Omega |\nabla(S(\phi))|^p\,dx \le \|S'\|^p_{L^\infty(\mathbb{R})}\|\nabla\phi\|^p_{L^p(\Omega)} < +\infty,$$

on the other hand. Consequently, $S(\phi) \in W^{1,p}(\Omega)$.

Let now $u \in W^{1,p}(\Omega)$. First of all, it is clear that $S'(u)\nabla u \in L^p(\Omega)$ and that $S(u) \in L^p(\Omega)$. By the Meyers-Serrin theorem, cf. Theorem 1.13 of Chap. 1, there exists a sequence $\phi_n \in C^\infty(\Omega) \cap W^{1,p}(\Omega)$ such that $\phi_n \to u$ in $W^{1,p}(\Omega)$. By extracting a subsequence, we may as well assume that $\phi_n \to u$ almost everywhere in Ω. Consequently, since S is of class C^1, then $S'(\phi_n) \to S'(u)$ almost everywhere.

Using again the global Lipschitz character of S, we see that $|S(\phi_n(x)) - S(u(x))| \le L|\phi_n(x) - u(x)|$. Integrating this inequality to the power p on Ω, then taking the power $\frac{1}{p}$, we obtain,

$$\|S(\phi_n) - S(u)\|_{L^p(\Omega)} \leq L\|\phi_n - u\|_{L^p(\Omega)} \longrightarrow 0 \text{ when } n \to +\infty,$$

(we could also have called on the Carathéodory theorem here).

On the other hand,

$$\begin{aligned}\|\nabla(S(\phi_n)) - S'(u)\nabla u\|_{L^p(\Omega)} \leq \|S'(\phi_n)(\nabla\phi_n - \nabla u)\|_{L^p(\Omega)} \\ + \|(S'(\phi_n) - S'(u))\nabla u\|_{L^p(\Omega)},\end{aligned} \tag{3.5}$$

by the triangle inequality. For the first term in the right-hand side, there clearly holds

$$\|S'(\phi_n)(\nabla\phi_n - \nabla u)\|_{L^p(\Omega)} \leq \|S'\|_{L^\infty(\mathbb{R})}\|\nabla\phi_n - \nabla u\|_{L^p(\Omega)} \longrightarrow 0 \text{ when } n \to +\infty.$$

For the second term in the right-hand side of estimate (3.5), we note that

$$\begin{cases} |S'(\phi_n) - S'(u)|^p|\nabla u|^p \to 0 \text{ almost everywhere,} \\ |S'(\phi_n) - S'(u)|^p|\nabla u|^p \leq 2^p\|S'\|^p_{L^\infty(\mathbb{R})}|\nabla u|^p \in L^1(\Omega). \end{cases}$$

Consequently, the Lebesgue dominated convergence theorem implies that

$$\|(S'(\phi_n) - S'(u))\nabla u\|_{L^p(\Omega;\mathbb{R}^d)} \longrightarrow 0 \text{ when } n \to +\infty.$$

Going back to estimate (3.5), we have just shown that the sequence $\nabla(S(\phi_n))$ converges in $L^p(\Omega; \mathbb{R}^d)$ to $S'(u)\nabla u$.

Since we have already established that $S(\phi_n) \to S(u)$ in $L^p(\Omega)$, it follows that the sequence $S(\phi_n)$ converges in $W^{1,p}(\Omega)$. Consequently, its limit in $L^p(\Omega)$, $S(u)$, belongs to $W^{1,p}(\Omega)$ with $\nabla(S(u)) = S'(u)\nabla u$.

Let us now consider the case $p = +\infty$. Let $u \in W^{1,\infty}(\Omega)$. Clearly, $S(u) \in L^\infty(\Omega)$. Besides, $u \in H^1_{\text{loc}}(\Omega)$, so according to the $p < +\infty$ case, the gradient of $S(u)$ in the sense of distributions is given by $\nabla(S(u)) = S'(u)\nabla u$. Now $S'(u)\nabla u \in L^\infty(\Omega; \mathbb{R}^d)$, hence finally $S(u) \in W^{1,\infty}(\Omega)$. □

We next consider the case of a function T that is piecewise affine with a single point c of non differentiability. Let thus $T \in C^0(\mathbb{R})$ be such that

$$T'(t) = \begin{cases} a & \text{if } t < c, \\ b & \text{if } t > c. \end{cases}$$

with $T(0) = 0$ if $\operatorname{meas}\Omega = +\infty$. We introduce the auxiliary functions

$$\gamma_-(t) = \begin{cases} a & \text{if } t \leq c, \\ b & \text{if } t > c, \end{cases} \qquad \gamma_+(t) = \begin{cases} a & \text{if } t < c, \\ b & \text{if } t \geq c. \end{cases}$$

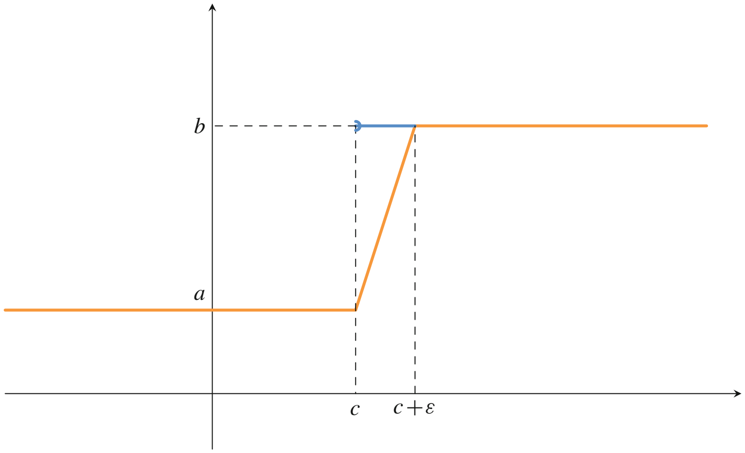


Fig. 3.3 The functions γ_-^ε and γ_-

Lemma 3.2. *For all $u \in W^{1,p}(\Omega)$, there holds $T(u) \in W^{1,p}(\Omega)$ and $\nabla(T(u)) = \gamma_-(u)\nabla u$ almost everywhere.*

Remark 3.8. The formula for the gradient of $T(u)$ is unambiguously defined since γ_- is defined on the whole of $\mathbb{R}$. □

Proof. The idea is of course to regularize T in order to apply Lemma 3.1. With this goal in mind, we set for all $\varepsilon > 0$, see Fig. 3.3.,

$$\gamma_-^\varepsilon(t) = \begin{cases} a & \text{if } t \le c, \\ a + \frac{b-a}{\varepsilon}(t-c) & \text{if } c \le t \le c+\varepsilon, \\ b & \text{if } t \ge c+\varepsilon. \end{cases}$$

The function γ_-^ε is continuous and $\gamma_-^\varepsilon \to \gamma_-$ pointwise when $\varepsilon \to 0$.

We then define

$$S_-^\varepsilon(t) = \int_0^t \gamma_-^\varepsilon(s)\,ds + T(0).$$

It is easy to see that S_-^ε is of class C^1 and globally Lipschitz, that $S(0) = 0$ if $T(0) = 0$ and that

$$\forall t \in \mathbb{R}, \quad |S_-^\varepsilon(t) - T(t)| \le \varepsilon \frac{|b-a|}{2},$$

which shows that S_-^ε tends to T uniformly on $\mathbb{R}$ when $\varepsilon \to 0$. It follows that,

$$\begin{cases} S_-^\varepsilon(u) \to T(u) \text{ everywhere and} \\ |S_-^\varepsilon(u)| \leq \max(|a|, |b|)|u| + |T(0)| \in L^p(\Omega). \end{cases}$$

The Lebesgue dominated convergence theorem then implies that

$$\|S_-^\varepsilon(u) - T(u)\|_{L^p(\Omega)} \longrightarrow 0 \text{ when } \varepsilon \to 0.$$

Likewise,

$$\begin{cases} (\gamma_-^\varepsilon(u) - \gamma_-(u))\nabla u \to 0 \text{ everywhere and} \\ |(\gamma_-^\varepsilon(u) - \gamma_-(u))\nabla u| \leq 2\max(|a|, |b|)|\nabla u| \in L^p(\Omega), \end{cases}$$

and the Lebesgue dominated convergence theorem applies again:

$$\|\nabla(S_-^\varepsilon(u)) - \gamma_-(u)\nabla u\|_{L^p(\Omega)} \longrightarrow 0 \text{ when } \varepsilon \to 0,$$

which proves the Lemma for $p < +\infty$. The $p = +\infty$ case is treated as in the previous lemma. □

We now are in a position to establish Theorem 3.2, as well as point ii) of Theorem 3.3 in the case of Lemma 3.2.

Lemma 3.3. *Let $u \in W^{1,p}(\Omega)$. There holds $\nabla u = \nabla T(u) = 0$ almost everywhere on $E_c(u)$ and $\nabla(T(u)) = T'(u)\nabla u$ almost everywhere on $\Omega \setminus E_c(u)$.*

Proof. By reworking the previous proof with γ_+ instead of γ_-, we likewise obtain $\nabla(T(u)) = \gamma_+(u)\nabla u$. Consequently,

$$[\gamma_+(u) - \gamma_-(u)]\nabla u = 0 \text{ almost everywhere on } \Omega.$$

Since $\gamma_+(u) - \gamma_-(u) = b - a \neq 0$ almost everywhere on $E_c(u)$ by Proposition 3.3, we see that $\nabla u = 0$ almost everywhere on $E_c(u)$. The formula for $\nabla T(u)$ then shows that $\nabla T(u) = 0$ almost everywhere on $E_c(u)$. Finally, since $\gamma_+(u) = \gamma_-(u) = T'(u)$ almost everywhere on $\Omega \setminus E_c(u)$, the proof is complete. □

To conclude the proof of Theorem 3.3, we establish a last lemma.

Lemma 3.4. *Let T be a globally Lipschitz function from $\mathbb{R}$ to $\mathbb{R}$, piecewise C^1 and with only a finite number of points of non differentiability $c_1, c_2, \dots, c_k$. Then, there exists S of class C^1 and functions T_i which are piecewise affine, of class C^1 except at c_i, such that*

$$T = S + \sum_{i=1}^{k} T_i.$$

Moreover, $S(0) = T_i(0) = 0$ if $T(0) = 0$.

Proof. Assume first that $c_1 \geq 0$. Let $\gamma_1 = \lim_{t \to c_1^-} T'(t) - \lim_{t \to c_1^+} T'(t)$ and set $T_1(t) = \gamma_1(t-c_1)_+$. The function $T - T_1$ is of class C^1 at c_1 and has one point of non differentiability less than T. It vanishes at 0 if $T(0) = 0$. We iterate the construction at each c_i until they have all been eliminated, which defines S. If $c_1 < 0$, we set instead $T_1(t) = \gamma_1(t - c_1)_-$ and so on. □

End of the Proof of Theorem 3.3. Let u be an element of $W^{1,p}(\Omega)$. We thus have $T(u) = S(u) + \sum_{i=1}^k T_i(u)$. According to Lemma 3.1, $S(u) \in W^{1,p}(\Omega)$ and according to Lemma 3.2, $T_i(u) \in W^{1,p}(\Omega)$ as well. Therefore, $T(u) \in W^{1,p}(\Omega)$. Finally, Lemmas 3.1–3.3 show that formula ii) for the gradient of $T(u)$ holds. □

Let us now talk about the continuity properties of superposition operators in $W^{1,p}(\Omega)$. The situation differs from that of the case of $L^p(\Omega)$.

Theorem 3.4. *Under the same hypotheses as in Theorem 3.3, the mapping $u \mapsto T(u)$ is*

i) continuous from $W^{1,p}(\Omega)$ strong to $W^{1,p}(\Omega)$ strong for all $p < +\infty$,

ii) sequentially continuous from $W^{1,p}(\Omega)$ weak to $W^{1,p}(\Omega)$ weak (or weak-star for $p = +\infty$).

Proof. i) Let u_n be sequence in $W^{1,p}(\Omega)$ that converges strongly to $u \in W^{1,p}(\Omega)$. Note first that since T is $\|T'\|_{L^\infty(\mathbb{R})}$-Lipschitz continuous,

$$\|T(u_n) - T(u)\|_{L^p(\Omega)} \leq \|T'\|_{L^\infty(\mathbb{R})}\|u_n - u\|_{L^p(\Omega)}.$$

Thus $T(u_n) \to T(u)$ in $L^p(\Omega)$ strong (which also follows from the Carathéodory theorem).

Concerning gradients, we start by extracting an arbitrary subsequence $u_{n'}$. We extract from this subsequence another subsequence $u_{n''}$ that converges almost everywhere. Let $\gamma_{-,i}$ be the function associated with T_i' as in Lemma 3.2. We set $\Gamma_- = S' + \sum_{i=1}^k \gamma_{-,i}$, so that $\nabla(T(u)) = \Gamma_-(u)\nabla u$. This formula is unambiguously defined, because $\Gamma_-(u)$ makes unambiguous sense. Products of the form $\Gamma_-(u)\nabla v$ are also well defined.

Then there holds

$$\|\nabla(T(u_{n''})) - \nabla(T(u))\|_{L^p(\Omega)} \leq \|\Gamma_-(u_{n''})(\nabla u_{n''} - \nabla u)\|_{L^p(\Omega)} + \|(\Gamma_-(u_{n''}) - \Gamma_-(u))\nabla u\|_{L^p(\Omega)}.$$

Obviously

$$\|\Gamma_-(u_{n''})(\nabla u_{n''} - \nabla u)\|_{L^p(\Omega)} \leq \|T'\|_{L^\infty(\mathbb{R})}\|\nabla u_{n''} - \nabla u\|_{L^p(\Omega)} \longrightarrow 0,$$

when $n'' \to +\infty$. Concerning the other term, we remark that

$$\|(\Gamma_-(u_{n''}) - \Gamma_-(u))\nabla u\|^p_{L^p(\Omega)} = \int_{\Omega \setminus \cup_{i=1}^k E_{c_i}(u)} |\Gamma_-(u_{n''}) - \Gamma_-(u)|^p |\nabla u|^p \, dx$$

because $\nabla u = 0$ almost everywhere on $\cup_{i=1}^k E_{c_i}(u)$. Now the function Γ_- is continuous on $\mathbb{R} \setminus \cup_{i=1}^k c_i$, therefore

$$\begin{cases} (\Gamma_-(u_{n''}) - \Gamma_-(u))\nabla u \to 0 \text{ everywhere on } \Omega \setminus \cup_{i=1}^k E_{c_i}(u) \text{ and} \\ |(\Gamma_-(u_{n''}) - \Gamma_-(u))\nabla u| \leq 2\|T'\|_{L^\infty(\mathbb{R})}|\nabla u| \in L^p(\Omega \setminus \cup_{i=1}^k E_{c_i}(u)). \end{cases}$$

We apply the Lebesgue dominated convergence theorem to conclude that

$$\|(\Gamma_-(u_{n''}) - \Gamma_-(u))\nabla u\|_{L^p(\Omega)} \longrightarrow 0 \text{ when } n'' \to +\infty.$$

This shows that, $T(u_{n''}) \to T(u)$ in $W^{1,p}(\Omega)$ strong. The uniqueness of the limit of extracted subsequences then implies that the whole sequence converges.

ii) Let now u_n be a sequence in $W^{1,p}(\Omega)$ that converges weakly to u in $W^{1,p}(\Omega)$ (the stars are understood for $p = +\infty$). By Rellich's theorem, $u_n \to u$ in $L^p_{\text{loc}}(\Omega)$ strong.[5] Thus, as before, $T(u_n) \to T(u)$ in $L^p_{\text{loc}}(\Omega)$ strong, including in the case $p = +\infty$. Furthermore, it follows from Theorem 3.3 that we have an estimate of the form

$$\|T(u_n)\|_{W^{1,p}(\Omega)} \leq C(\|u_n\|_{W^{1,p}(\Omega)} + |T(0)|).$$

Consequently, we can extract from any subsequence $u_{n'}$ a further subsequence $u_{n''}$ such that $T(u_{n''}) \rightharpoonup v$ in $W^{1,p}(\Omega)$ weak for a certain v. Now the convergence also holds true in $L^p_{\text{loc}}(\Omega)$ strong. Thus, $v = T(u)$ and we once more conclude by uniqueness of the limit. □

Remark 3.9. Point i) of the Theorem is false for $p = +\infty$. Consider for this the sequence $u_n(x) = \left(x - \frac{1}{n}\right)$ defined on $]-1, 1[$, and the mapping $T(t) = t_+$. □

Corollary 3.2. *Under the previous hypotheses, and if furthermore* $T(0) = 0$, *then* $u \in W_0^{1,p}(\Omega)$ *implies* $T(u) \in W_0^{1,p}(\Omega)$.

Proof. If $u \in W_0^{1,p}(\Omega)$ then there exists a sequence $\varphi_n \in \mathscr{D}(\Omega)$ such that $\varphi_n \to u$ in $W^{1,p}(\Omega)$ strong. Since $T(0) = 0$ and φ_n is compactly supported in Ω, $T(\varphi_n)$ is also compactly supported. It follows immediately by convolution by a mollifying sequence, that $T(\varphi_n) \in W_0^{1,p}(\Omega)$. Now $T(\varphi_n) \to T(u)$ in $W^{1,p}(\Omega)$ and $W_0^{1,p}(\Omega)$ is closed, thus $T(u) \in W_0^{1,p}(\Omega)$. For the case $p = +\infty$,[6] we can use the weak-star convergence or use the result for $p > d$ in conjunction with the Sobolev embeddings. □

As an example, we see that for all $u \in H_0^1(\Omega)$ and $k \geq 0$, $(u-k)_+ \in H_0^1(\Omega)$.

[5] We need the "loc" here, because we have no regularity hypothesis on Ω.

[6] The space $W_0^{1,\infty}$ can be defined as the space of $W^{1,\infty}$ functions that continuously extend by 0 on $\partial\Omega$.

Let us close this section by a study of a few particular superposition operators. Let us first mention simple, but very useful, relations concerning positive and negative parts. First of all, if $u \in L^p(\Omega)$, then

$$u = u_+ - u_-,\ |u| = u_+ + u_-,\ u_+ = \mathbf{1}_{u>0}u = \mathbf{1}_{u\geq 0}u,\ u_- = -\mathbf{1}_{u<0}u = -\mathbf{1}_{u\leq 0}u,$$

where $\mathbf{1}_{u>0}$ is a quick notation for the characteristic function of the set of points x such that $u(x) > 0$, and so on. If $u \in W^{1,p}(\Omega)$, there furthermore holds

$$\nabla u = \nabla u_+ - \nabla u_-,\ \nabla |u| = \nabla u_+ + \nabla u_-,\ \nabla u_+ = \mathbf{1}_{u>0}\nabla u = \mathbf{1}_{u\geq 0}\nabla u,$$

and similar relations for u_-. Let us note that $|\nabla u_+|.|\nabla u_-| = 0$ almost everywhere.

Another very useful operator is the *truncation at height* k. The truncation is defined for $k > 0$ as the superposition operator associated with the function (Fig. 3.4)

$$T_k(t) = \begin{cases} t & \text{if } |t| \leq k, \\ k\frac{t}{|t|} & \text{if } |t| > k. \end{cases}$$

The truncation at height k provides an approximation of the identity in various spaces.

Theorem 3.5. *If $u \in L^p(\Omega)$, then $T_k(u) \to u$ in $L^p(\Omega)$ strong when $k \to +\infty$. If $u \in W^{1,p}(\Omega)$, then $T_k(u) \to u$ in $W^{1,p}(\Omega)$ strong.*

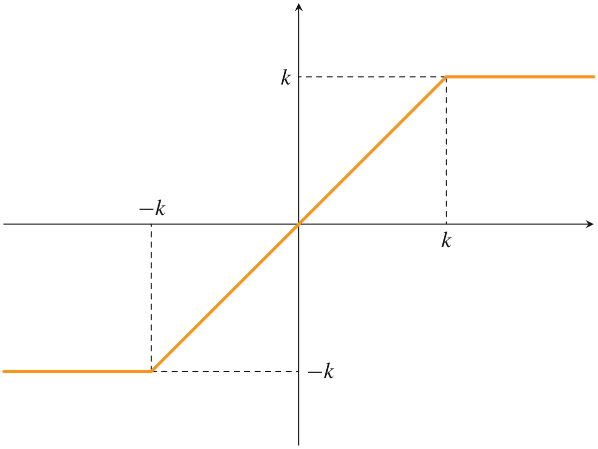


Fig. 3.4 The truncation at height k

Proof. Let us start with the L^p case, $p < +\infty$. There holds

$$\begin{aligned}\|u - T_k(u)\|^p_{L^p(\Omega)} &= \int_{\{u<-k\}} |u+k|^p\,dx + \int_{\{u>k\}} |u-k|^p\,dx\\ &\leq \int_{\{u<-k\}} |u|^p\,dx + \int_{\{u>k\}} |u|^p\,dx\\ &= \int_\Omega |u|^p \mathbf{1}_{|u|>k}\,dx \longrightarrow 0\end{aligned}$$

when $k \to +\infty$ by monotone convergence. Concerning gradients, if u is in $W^{1,p}(\Omega)$, we similarly have

$$\|\nabla u - \nabla(T_k(u))\|^p_{L^p(\Omega)} = \int_\Omega |1 - T'_k(u)|^p |\nabla u|^p\,dx = \int_\Omega |\nabla u|^p \mathbf{1}_{|u|>k}\,dx \longrightarrow 0,$$

when $k \to +\infty$.

Finally, when $p = +\infty$, $T_k(u) = u$ as soon as $k \geq \|u\|_{L^\infty(\Omega)}$. □

Remark 3.10. i) In all cases, there holds $T_k(u) \in L^\infty(\Omega)$ with $\|T_k(u)\|_{L^\infty(\Omega)} \leq k$.

ii) If $u \in W^{1,p}_0(\Omega)$ then $T_k(u) \in W^{1,p}_0(\Omega)$ since $T_k(0) = 0$. □

We conclude this general study with a few remarks.

Remark 3.11. i) The superposition operators do not in general operate on Sobolev spaces of order higher than 1. Thus, for example, if $u \in H^2(\Omega)$, it is not necessarily the case that $u_+ \in H^2(\Omega)$. This is obvious in one dimension of space, since $H^2(\Omega) \hookrightarrow C^1(\bar{\Omega})$ in this case. The problem however is not only connected to a lack of regularity of the function T. Thus, even if T is of class C^∞ with T' and T'' bounded, and $u \in C^\infty(\Omega) \cap H^2(\Omega)$, we do not always have $T(u) \in H^2(\Omega)$.

In effect, by the classical chain rule, $\partial_i(T(u)) = T'(u)\partial_i u$ and $\partial_{ij}(T(u)) = T''(u)\partial_i u \partial_j u + T'(u)\partial_{ij} u$. The second term in the expression of second derivatives actually belongs to $L^2(\Omega)$. However, for the first term, in general $\partial_i u \partial_j u \notin L^2(\Omega)$ (except when $d \leq 4$ by the Sobolev embeddings). Let us mention a more general result in this direction: if T is C^∞, then for real s, if $u \in W^{s,p}(\Omega)$ then $T(u) \in W^{s,p}(\Omega)$ as soon as $s - \frac{d}{p} > 0$, see for example [51].

ii) The vector-valued case is comparable to the scalar case. If $T\colon \mathbb{R}^m \to \mathbb{R}$ is globally Lipschitz, then for all $u \in W^{1,p}(\Omega; \mathbb{R}^m)$, $T(u) \in W^{1,p}(\Omega)$. On the other hand, the chain rule formula is not valid as such, because $DT(u)\nabla u$ makes no sense in general. Consider for example, for $m = 2$, $T(u_1, u_2) = \max(u_1, u_2)$. If $u \in H^1(\Omega; \mathbb{R}^2)$ is of the form $u = (v, v)$ with $v \in H^1(\Omega)$, then $DT(u)$ is nowhere defined on Ω whereas ∇u is not zero almost everywhere on Ω. So we would be hard pressed to give a reasonable definition of such a product as $DT(u)\nabla u$. There are however more complicated formulas to describe this gradient. □

3.4 Superposition Operators and Boundary Trace

In the case of a regular open set, the traces of functions of $W^{1,p}(\Omega)$ are functions of $L^p(\partial\Omega)$. We can thus apply superposition operators to them, and a natural question is to wonder whether these operators commute with the trace mapping.

Theorem 3.6. *Let Ω be a Lipschitz open set and T be as in Theorem 3.3. Then, for all $u \in W^{1,p}(\Omega)$, $\gamma_0(T(u)) = T(\gamma_0(u))$.*

Proof. The case $p = +\infty$ is obvious, since we are dealing with continuous functions. We thus assume that $p < +\infty$. Let us first note that if $u \in W^{1,p}(\Omega)$, then $T(u) \in W^{1,p}(\Omega)$ and therefore $\gamma_0(T(u))$ is well defined. Let us take a sequence $u_n \in C^1(\bar{\Omega})$ such that $u_n \to u$ in $W^{1,p}(\Omega)$ strong. By definition of the trace mapping, $u_{n|\partial\Omega} = \gamma_0(u_n) \to \gamma_0(u)$ in $L^p(\partial\Omega)$ strong. Now the superposition operators are continuous on $L^p(\partial\Omega)$ strong,[7] it follows that

$$T(\gamma_0(u_n)) \longrightarrow T(\gamma_0(u)) \quad \text{in } L^p(\partial\Omega) \text{ strong.}$$

On the other hand, the superposition operators are continuous on $W^{1,p}(\Omega)$ strong, therefore there also holds $T(u_n) \to T(u)$ in $W^{1,p}(\Omega)$ strong. The continuity of the trace mapping thus implies that

$$\gamma_0(T(u_n)) \longrightarrow \gamma_0(T(u)) \quad \text{in } L^p(\partial\Omega) \text{ strong.}$$

At this point, we would like to say that $\gamma_0(T(u_n)) = T(u_n)_{|\partial\Omega} = T(\gamma_0(u_n))$, by definition of the trace, and conclude, but unfortunately, this would be a little premature because $T(u_n)$ has no reason to be C^1 on $\overline{\Omega}$. Nevertheless, $T(u_n) \in C^0(\overline{\Omega})$. Now it also true that for any function $v \in W^{1,p}(\Omega) \cap C^0(\overline{\Omega})$, we have $\gamma_0(v) = v_{|\partial\Omega}$. The Theorem follows. □

Remark 3.12. Let us quickly sketch a proof of the latter trace property. Given $v \in W^{1,p}(\Omega) \cap C^0(\overline{\Omega})$, we can construct a sequence v_n of functions of $C^1(\overline{\Omega})$ by extending v to an open set containing $\overline{\Omega}$, then by convolution by a mollifying sequence. This sequence converges to v simultaneously in $W^{1,p}(\Omega)$ and in $C^0(\overline{\Omega})$, due to the standard properties of such convolutions. Of course then, $\gamma_0(v_n) \to \gamma_0(v)$ in $L^p(\partial\Omega)$ and $v_{n|\partial\Omega} \to v_{|\partial\Omega}$ in $C^0(\partial\Omega)$. □

Remark 3.13. An interesting particular case is the fact that $\gamma_0(u_+) = (\gamma_0(u))_+$. Likewise, we see again that for $k \geq 0$ and $u \in H_0^1(\Omega)$, then $(u-k)_+ \in H_0^1(\Omega)$. Indeed, in this case $\gamma_0((u-k)_+) = (\gamma_0(u-k))_+ = (-k)_+ = 0$. □

[7] Same proof as the Carathéodory theorem, or using the Lipschitz character of T.

3.5 Exercises of Chap. 3

1. Let u_n be the sequence of Proposition 3.1. Show that the Young measures associated with the sequence $v_n = u_n + \sin x$ are given by $\nu_x = \theta\delta_{a+\sin x} + (1-\theta)\delta_{b+\sin x}$.

2. Let $u \in L^\infty(\mathbb{R})$ be a T-periodic function. Show that the sequence $u_n(x) = u(nx)$ converges when $n \to +\infty$ to the average of u, $\frac{1}{T}\int_0^T u(x)\,dx$, in $L^\infty(\mathbb{R})$ weak-star. Assuming in addition that u is of class C^1 and nondecreasing on $[0, T[$, show that the associated Young measures are

$$\nu_x = \frac{1}{T}\mathbf{1}_{[u(0),u(T)^-]}\frac{dy}{u'(u^{-1}(y))}.$$

3. Show that $\sin x \sin nx \overset{*}{\rightharpoonup} 0$ and that $\sin^2 x \sin^2 nx \overset{*}{\rightharpoonup} \frac{1}{2}\sin^2 x$ when $n \to +\infty$. Conclude that the Young measures depend on x in this case.

4. Let f be a function from $\mathbb{R}$ to $\mathbb{R}$ verifying the hypotheses of the Carathéodory Theorem 2.14. The goal of this exercise is to show that the associated superposition operator $\tilde{f}$ from $L^2(\Omega)$ to $L^2(\Omega)$, is Fréchet differentiable at $u = 0$ if and only if the function f is affine. Here, Ω is a bounded open subset of $\mathbb{R}^d$.

4.1. Let $s \in \mathbb{R}$ be fixed. Show that the sequence

$$u_n^s(x) = \begin{cases} s & \text{for } x \in B(0, 1/n), \\ 0 & \text{otherwise,} \end{cases}$$

is such that $\|u_n^s\|_{L^2(\Omega)} = C_d n^{-d/2}|s|$, where C_d is a constant that only depends on the dimension d, and conclude that it tends to 0 in $L^2(\Omega)$ when $n \to +\infty$.

4.2. Assume that $\tilde{f}$ is differentiable at 0 and let $D\tilde{f}(0)$ be its Fréchet derivative. Remember that this means that $D\tilde{f}(0)$ is a continuous linear operator from $L^2(\Omega)$ into $L^2(\Omega)$ such that

$$\tilde{f}(u) = \tilde{f}(0) + D\tilde{f}(0)u + \|u\|_{L^2(\Omega)}\varepsilon(u),$$

where $\|\varepsilon(u)\|_{L^2(\Omega)} \to 0$ when $\|u\|_{L^2(\Omega)} \to 0$.

Show that f is differentiable at 0 and that for all $A \subset \Omega$ measurable, then

$$D\tilde{f}(0)\mathbf{1}_A = f'(0)\mathbf{1}_A.$$

4.3. Use 4.1 to conclude that $f(s) - f(0) - sf'(0) = 0$ for all $s \in \mathbb{R}$.

5. Show that the superposition operator $u \mapsto u^2$ is differentiable—and even of class C^∞—from $L^2(\Omega)$ into $L^1(\Omega)$. Is it possible to generalize this result? How does it compare with the result of Exercise 4?

6. Let $F\colon \Omega \times \mathbb{R}^d \to \mathbb{R}^d$ be a Carathéodory function such that there exists $a \in L^p(\Omega)$ and $C > 0$ such that for almost all $x \in \Omega$ and all $\xi \in \mathbb{R}^d$, there holds

$$|F(x, \xi)| \leq a(x) + C|\xi|.$$

6.1. Show that the operator $\Psi : W_0^{1,p}(\Omega) \to W^{-1,p}(\Omega)$ defined by

$$\Psi(u) = -\operatorname{div} F(x, \nabla u(x)),$$

is differentiable if and only if $F(x, \xi) = b_0(x) + B_1(x)\xi$ with $b_0 \in L^p(\Omega; \mathbb{R}^d)$ and $B_1 \in L^\infty(\Omega; M_d(\mathbb{R}))$.

6.2. Assume that F does not depend on x and is of class C^1. Show that Ψ is differentiable from $W^{2,p}(\Omega) \cap W_0^{1,p}(\Omega)$ to $L^p(\Omega)$ as soon as $p > d$.

Chapter 4
The Galerkin Method

The Galerkin method is a very general framework of methods which is very robust. The idea is as follows. Starting from a variational problem set in an infinite dimensional space, a sequence of finite dimensional approximation spaces is defined. The corresponding finite dimensional approximated problems are then solved, which is usually easier to do than trying it straight in infinite dimension. Finally, the dimension of the approximation spaces is left to tend to infinity, and one has to pass to the limit one way or another in the sequence of approximated solutions in order to construct a solution of the original problem.

It is worth mentioning that, in addition to its theoretical interest, the Galerkin method also provides an effective approximation procedure in some cases. See [48] for many examples of the method at work, mostly for evolution problems.

4.1 Solving the Model Problem by the Galerkin Method

Let us consider again the nonlinear model problem of Chap. 2, which was solved by infinite dimensional fixed point methods, and this time solve it by applying the Galerkin method. Let us write the problem again. Let Ω be a bounded open subset of $\mathbb{R}^d$ and f be a function in $C^0(\mathbb{R}) \cap L^\infty(\mathbb{R})$. We want to find a function $u \in H_0^1(\Omega)$ such that $-\Delta u = f(u)$ in the sense of $\mathscr{D}'(\Omega)$. Equivalently, we would like to solve the variational problem: Find $u \in H_0^1(\Omega)$ such that

$$\forall v \in H_0^1(\Omega), \quad \int_\Omega \nabla u \cdot \nabla v \, dx = \int_\Omega f(u) v \, dx. \tag{4.1}$$

We proceed in little steps, following the principles rapidly sketched above.

H. Le Dret, *Nonlinear Elliptic Partial Differential Equations*, Universitext,
https://doi.org/10.1007/978-3-319-78390-1_4

Lemma 4.1. *Let V be a separable normed vector space of infinite dimension. There exists an increasing sequence of finite dimensional vector subspaces V_m of V such that $\cup_{m=0}^{\infty} V_m$ is dense in V.*

Proof. Let $(w_n)_{n\in\mathbb{N}}$ be a countable dense family of V. We start with $V_0 = \{0\}$ and argue by induction. Assume that we have constructed V_m, $\dim V_m = m$ and $V_m = \text{vect}\{w_k, 0 \leq k \leq n_m\}$ for some n_m (with $n_0 < 0$, i.e., the empty family in this case).

The family $(w_n)_{n\in\mathbb{N}}$ is dense in V, therefore the set of integers $n > n_m$ such that $w_n \not\in V_m$ is not empty. Indeed $V \setminus V_m$ is open, hence contains infinitely many elements of the w_n family. Naturally $\mathbb{N}$ is well-ordered, so we call n_{m+1} the least element of this set. Let $v_{m+1} = w_{n_{m+1}}$ and $V_{m+1} = \text{vect}(V_m \cup \{v_{m+1}\})$. By construction, the family $(v_i)_{1\leq i\leq m+1}$ is a basis of V_{m+1}, and we have $\text{vect}\{w_k, 0 \leq k \leq n_{m+1}\} = V_{m+1}$. Finally, $\cup_{m=0}^{\infty} V_m = \text{vect}\{v_i, i \in \mathbb{N}^*\} = \text{vect}\{w_k, k \in \mathbb{N}\}$ is dense in V, by density of the w_k. □

Remark 4.1. i) Conversely, if there exists such a family v_i, then V is separable. The linear combinations of v_i with rational coefficients form a countable dense family.

ii) The family $\{v_i, i \in \mathbb{N}^*\}$ is by construction a basis of $\cup_{m=0}^{\infty} V_m$. Therefore its linear combinations are dense in V. We also say that it is a *Galerkin basis* of V, even though it does not have much to do with any actual notion of basis for V. The spaces V_m are called *Galerkin spaces.* □

In the sequel, we apply Lemma 4.1 to the space $V = H_0^1(\Omega)$, which is separable. To construct the finite dimensional approximation of the problem, we just restrict the variational formulation (4.1) to the Galerkin space V_m. Let us show that this finite dimensional problem has a solution.

Lemma 4.2. *For all $m \in \mathbb{N}$, the variational problem: Find $u_m \in V_m$ such that*

$$\forall v \in V_m, \quad \int_\Omega \nabla u_m \cdot \nabla v \, dx = \int_\Omega f(u_m) v \, dx, \tag{4.2}$$

admits at least one solution.

Proof. We equip V_m with the inner product inherited from $L^2(\Omega)$, that is to say $(u|v)_m = \int_\Omega uv\,dx$. This makes it a Euclidean space, which we identify with its dual space via this inner product.

The mapping $(u, v) \mapsto a(u, v) = \int_\Omega \nabla u \cdot \nabla v\,dx$ is a bilinear form on V_m. Thus there exists a linear mapping $A_m \in \mathscr{L}(V_m)$ such that $a(u, v) = (A_m(u)|v)_m$ for all u and v in V_m. This mapping is continuous since V_m is finite dimensional.

Likewise, there exists a mapping $F_m : V_m \to V_m$ such that for all pairs (u, v) of V_m, $\int_\Omega f(u)v\,dx = (F_m(u)|v)_m$. We take $F_m = \Pi_m \circ \tilde{f}$, where Π_m is the orthogonal projection of L^2 onto V_m. This nonlinear mapping is also continuous as the composition of two continuous mappings (we use here Carathéodory's theorem).

Problem (4.2) is thus rewritten as: Find $u_m \in V_m$ such that

$$\forall v \in V_m, \quad (A_m(u_m)|v)_m = (F_m(u_m)|v)_m, \tag{4.3}$$

or, introducing the continuous mapping $P_m \colon V_m \to V_m$,

$$P_m(u) = A_m(u) - F_m(u),$$

as

$$P_m(u_m) = 0. \tag{4.4}$$

To solve this problem, we are going to apply Theorem 2.7 of Chap. 2. We must compute for this the product $(P_m(u)|u)_m$ on a sphere, and show that we can choose the sphere in such a way that this inner product is nonnegative.

By definition of the inner product on V_m, we obtain

$$\begin{aligned}(P_m(u)|u)_m &= \int_\Omega P_m(u)u\,dx = a(u,u) - \int_\Omega f(u)u\,dx\\ &\geq \|\nabla u\|^2_{L^2(\Omega)} - \|f\|_{L^\infty(\mathbb{R})}(\text{meas}\,\Omega)^{1/2}\|u\|_{L^2(\Omega)}\\ &\geq \|\nabla u\|^2_{L^2(\Omega)} - C_\Omega\|f\|_{L^\infty(\mathbb{R})}(\text{meas}\,\Omega)^{1/2}\|\nabla u\|_{L^2(\Omega)}\\ &= \|\nabla u\|_{L^2(\Omega)}(\|\nabla u\|_{L^2(\Omega)} - C_\Omega\|f\|_{L^\infty(\mathbb{R})}(\text{meas}\,\Omega)^{1/2}),\end{aligned}$$

where C_Ω is the Poincaré inequality constant. We thus see that if

$$\|\nabla u\|_{L^2(\Omega)} \geq C_\Omega\|f\|_{L^\infty(\mathbb{R})}(\text{meas}\,\Omega)^{1/2},$$

then

$$(P_m(u)|u)_m \geq 0.$$

Now V_m is finite dimensional, so that all norms on V_m are equivalent with each other. Therefore, there exists $\rho_m > 0$ such that $\sqrt{(u|u)_m} \geq \rho_m$ implies that $\|\nabla u\|_{L^2(\Omega)} \geq C_\Omega\|f\|_{L^\infty(\mathbb{R})}(\text{meas}\,\Omega)^{1/2}$. By Theorem 2.7, problem (4.4) admits a solution u_m such that $\sqrt{(u_m|u_m)_m} \leq \rho_m$. □

Remark 4.2. Of course, this finite dimensional variational problem has no interpretation whatsoever in terms of a boundary value problem. On the other hand, it really does not need one. □

We have thus found a sequence u_m of approximated solutions. Remark that we could also have used Brouwer's theorem itself by following the fixed point proof more closely.

We now need to pass to the limit when the dimension $m+1$ goes to infinity. The first step is an estimate that is independent of m.

Lemma 4.3. *The sequence u_m is bounded in $H_0^1(\Omega)$.*

Proof. We go back to the previous computation and observe that

$$a(u_m, u_m) = \int_\Omega f(u_m)u_m\,dx \le C_\Omega \|f\|_{L^\infty(\mathbb{R})}(\text{meas}\,\Omega)^{1/2}\|\nabla u_m\|_{L^2(\Omega)}.$$

It follows that

$$\|\nabla u_m\|_{L^2(\Omega)} \le C_\Omega \|f\|_{L^\infty(\mathbb{R})}(\text{meas}\,\Omega)^{1/2},$$

which proves the Lemma. □

Let us now pass to the limit in the sequence of variational problems.

Lemma 4.4. *Any weakly convergent subsequence of the sequence u_m converges to a solution of problem* (4.1).

Proof. Let us take a subsequence $u_{m'}$ such that $u_{m'} \rightharpoonup u$ in $H_0^1(\Omega)$ (there is at least one according to Lemma 4.3). By Rellich's theorem, it follows that $u_{m'} \to u$ in $L^2(\Omega)$ strong. Consequently, Carathéodory's theorem implies that $f(u_{m'}) \to f(u)$ in $L^2(\Omega)$ strong.

We choose an integer i. The sequence V_m is increasing, thus for all $m \ge i$, $v_i \in V_m$. For these values of m, we can therefore use Eq. (4.2) with the test-function v_i,

$$\int_\Omega \nabla u_m \cdot \nabla v_i\,dx = \int_\Omega f(u_m)v_i\,dx.$$

Now $\nabla u_{m'} \rightharpoonup \nabla u$ in $L^2(\Omega)$ weak, so that on the one hand,

$$\int_\Omega \nabla u_{m'} \cdot \nabla v_i\,dx \to \int_\Omega \nabla u \cdot \nabla v_i\,dx.$$

On the other hand, $f(u_{m'}) \to f(u)$ in $L^2(\Omega)$ strong and thus

$$\int_\Omega f(u_{m'})v_i\,dx \to \int_\Omega f(u)v_i\,dx.$$

Consequently, for all $i \in \mathbb{N}$,

$$\int_\Omega \nabla u \cdot \nabla v_i\,dx = \int_\Omega f(u)v_i\,dx.$$

This relation is linear with respect to v_i, it thus remains true for all linear combinations of the v_i, i.e.,

$$\forall z \in \bigcup_{j=0}^{+\infty} V_j, \quad \int_\Omega \nabla u \cdot \nabla z \, dx = \int_\Omega f(u) z \, dx. \tag{4.5}$$

Finally, $\cup_{j=0}^{+\infty} V_j$ is dense in V. For all $v \in V$, there exists a sequence $z_j \in V_j$ such that $z_j \to v$ in V strong. We apply the previous equality with $z = z_j$ and pass to the limit when $j \to +\infty$ quite effortlessly to conclude that u is actually a solution of problem (4.1). □

Remark 4.3. i) The only place in the argument where strong convergence was actually important as opposed to weak convergence, was for the continuity of the superposition operator allowing us to pass to the limit in the nonlinear term.

ii) The approximated solution sequence $u_{m'}$ converges not only weakly to u, but also strongly. Indeed, $\int_\Omega \nabla u_m \cdot \nabla u_m \, dx = \int_\Omega f(u_m) u_m \, dx$, and the right-hand side tends to $\int_\Omega f(u) u \, dx = \int_\Omega \nabla u \cdot \nabla u \, dx$.

iii) If the model problem has one and only one solution, then the whole sequence u_m converges to this solution.

4.2 A Problem Reminiscent of Fluid Mechanics

There was not much difficulty in solving the model problem, either in terms of existence in finite dimension, estimating the approximated solutions or passing to the limit in infinite dimension. We now give an example of application of the Galerkin method to a problem that presents mathematical similarities with the Navier-Stokes equations of fluid mechanics, although without having anything to do with fluid mechanics from the point of view of modeling.

The Navier-Stokes equations are extremely important equations that describe the velocity of the flow of a viscous fluid, either compressible or incompressible. They have a stationary version and a non-stationary version. In the incompressible stationary case, they assume the following form. We are looking for a pair of unknown functions (u, p) where u is the fluid velocity, which is $\mathbb{R}^3$-valued in the three-dimensional case, and p is the pressure, a scalar function. The pair satisfies

$$\begin{cases} -\nu \Delta u + (u \cdot \nabla) u - \nabla p = f & \text{in } \Omega, \\ \operatorname{div} u = 0 & \text{in } \Omega, \end{cases} \tag{4.6}$$

where Ω is an open subset of $\mathbb{R}^3$ in which the fluid is flowing, with appropriate boundary conditions, for example $u = 0$ on $\partial\Omega$, which means that the fluid sticks to the walls due to viscosity. The constant $\nu > 0$ is the fluid viscosity, the $u \cdot \nabla$ operator is defined by $[(u \cdot \nabla) v]_i = u_j \partial_j v_i$ with summation from 1 to 3 with respect to the repeated index j, and f is an applied force density, for instance gravity. The relation $\operatorname{div} u = 0$ expresses the incompressibility of the fluid. This is the stationary version of the problem and there is no time evolution of the velocity and pressure.[1]

Studying system (4.6) is beyond the scope of this work, and even more so for the non-stationary version. We are nonetheless going to consider a simpler equation, which presents a nonlinear term that is analogous to the nonlinear term in Navier-Stokes, and thus shares some of its mathematical properties. This type of equations was introduced in [48].

We are thus going to look for a *scalar* function u such that

$$\begin{cases} -\Delta u + u\partial_1 u = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{4.7}$$

We will make the functional sense in which this equation must be understood more precise later on. This is a scalar equation, whereas the Navier-Stokes equations are vector-valued. Hence, it contains nothing that can compare with the incompressibility constraint, nor with the presence of the pressure p, which is a Lagrange multiplier for this constraint. On the other hand, the nonlinear term $u\partial_1 u$ shares some common properties with the nonlinear term $(u \cdot \nabla)u$ of the Navier-Stokes equations.

To apply the Galerkin method here, we are going to need a special kind of Galerkin basis. We start with a density result.

Lemma 4.5. *Let Ω be a bounded open subset of $\mathbb{R}^d$ and let $p \in [1, +\infty[$. Then $\mathscr{D}(\Omega)$ is dense in $H_0^1(\Omega) \cap L^p(\Omega)$.*

Remark 4.4. We already know that $\mathscr{D}(\Omega)$ is dense in $H_0^1(\Omega)$ just by definition of $H_0^1(\Omega)$ on the one hand, and also in $L^p(\Omega)$ by convolution by a mollifier sequence on the other hand. Lemma 4.5 asserts that in addition, it is possible to approximate any element in the intersection of the two spaces by a sequence of functions in $\mathscr{D}(\Omega)$ that simultaneously converges for both topologies. Note that this is true in a an arbitrary bounded open set, without any regularity hypothesis. □

Proof. We proceed by successive approximations. Let $u \in H_0^1(\Omega) \cap L^p(\Omega)$. We first truncate u at height k by setting $u_k = T_k(u)$. It follows from the general properties of truncation, viz. Theorem 3.5, that $u_k \in H_0^1(\Omega) \cap L^\infty(\Omega)$ and $u_k \to u$ in $H_0^1(\Omega) \cap L^p(\Omega)$ when $k \to +\infty$.

We now take a sequence $\varphi_{k,m} \in \mathscr{D}(\Omega)$ such that $\varphi_{k,m} \to u_k$ in $H_0^1(\Omega)$ and almost everywhere when $m \to +\infty$. Let $\widetilde{T}_{k+1}$ be a C^∞ function on $\mathbb{R}$ such that for $|t| \leq k$, $\widetilde{T}_{k+1}(t) = t$ and for all s, $|\widetilde{T}_{k+1}(s)| \leq k + 1$ (such a function evidently exists). The associated superposition operator is continuous on $H_0^1(\Omega)$, therefore

[1] The non-stationary Navier-Stokes equations are the real object of study of fluid mechanics, since stationary flows are very rare in nature. An acceleration term $\rho \frac{\partial u}{\partial t}$ is added to the first equation, where ρ is the fluid density, plus initial conditions for u. This is no longer an elliptic problem.

$\widetilde{T}_{k+1}(\varphi_{k,m}) \to \widetilde{T}_{k+1}(u_k) = u_k$ in $H_0^1(\Omega)$ and almost everywhere $m \to +\infty$. The latter equality holds because u_k has already been truncated at height k. By construction $\|\widetilde{T}_{k+1}(\varphi_{k,m})\|_{L^\infty(\Omega)} \leq k+1$ and Ω is bounded. It follows from this that $\widetilde{T}_{k+1}(\varphi_{k,m}) \to u_k$ in $L^p(\Omega)$ by the Lebesgue dominated convergence theorem.

We finally observe that $\widetilde{T}_{k+1}(\varphi_{k,m})$ is compactly supported in Ω and of class C^∞. To conclude, we extract a converging sequence from the two successive approximations $m \to +\infty$ and $k \to +\infty$. □

Remark 4.5. i) If $u \in L^\infty(\Omega)$ then the previous construction provides a sequence of functions $\mathscr{D}(\Omega)$ that converges to u in $H_0^1(\Omega)$ and in $L^\infty(\Omega)$ weak-star. Indeed, all the successive approximations are then bounded in $L^\infty(\Omega)$ and thus weakly-star convergent. The final double limit argument is valid because the weak-star topology is metrizable on bounded sets.

ii) The reason for introducing $\widetilde{T}_{k+1}$ is that even though u_k is bounded in Ω, a sequence of $\mathscr{D}(\Omega)$ functions approximating it in $H_0^1(\Omega)$ is not a priori guaranteed to be likewise bounded. The function $\widetilde{T}_{k+1}$ acts as a smooth truncation at height $k+1$ that preserves the C^∞ character and convergence, while ensuring such a bound. □

Let us now detail the functional setting of the equation.

Lemma 4.6. *Let Ω be a bounded open subset of $\mathbb{R}^d$. If $u \in H_0^1(\Omega)$, then $u\partial_1 u \in L^s(\Omega)$ with*

$$\begin{cases} 1 \leq s \leq 2 & \textit{for } d = 1, \\ 1 \leq s < 2 & \textit{for } d = 2, \\ 1 \leq s \leq \frac{d}{d-1} & \textit{for } d \geq 3. \end{cases}$$

Proof. According to the Sobolev embeddings,

$$\begin{cases} H_0^1(\Omega) \hookrightarrow L^\infty(\Omega), & \text{if } d = 1, \\ H_0^1(\Omega) \hookrightarrow L^q(\Omega) \text{ for all } q < +\infty, & \text{if } d = 2, \\ H_0^1(\Omega) \hookrightarrow L^q(\Omega) \text{ for all } q \leq 2^* = \frac{2d}{d-2}, & \text{if } d \geq 3. \end{cases}$$

By Hölder's inequality, for any pair of positive numbers (θ, θ') such that $1/\theta + 1/\theta' = 1$, there holds

$$\int_\Omega |u\partial_1 u|^s \, dx \leq \Big(\int_\Omega |u|^{s\theta} \, dx\Big)^{\frac{1}{\theta}} \Big(\int_\Omega |\partial_1 u|^{s\theta'} \, dx\Big)^{\frac{1}{\theta'}}.$$

For $d \geq 3$, we will thus be able to conclude under the condition that $1 \leq s\theta \leq 2^*$ and $1 \leq s\theta' \leq 2$, i.e., $s\big(\frac{1}{2^*} + \frac{1}{2}\big) \leq 1$ and $s \geq 1$. Since $\frac{1}{2^*} + \frac{1}{2} = \frac{d-1}{d}$, we obtain the result in this case.

For $d = 2$, the same computation yields $s\big(\frac{1}{q} + \frac{1}{2}\big) \leq 1$ for a certain $q < +\infty$, thus $s < 2$. The case $d = 1$ is trivial. □

Remark 4.6. i) Lemma 4.6 gives a precise meaning to the PDE of problem (4.7). Given $f \in H^{-1}(\Omega)$, we will look for $u \in H_0^1(\Omega)$ such that

$$-\Delta u + u\partial_1 u = f \quad \text{in the sense of } \mathscr{D}'(\Omega). \tag{4.8}$$

This equation makes sense since $\nabla u \in L^2(\Omega; \mathbb{R}^d)$ and $-\Delta u = -\text{div}\,(\nabla u) \in H^{-1}(\Omega)$. Moreover, $u\partial_1 u \in L^s(\Omega)$ for the values of s provided by the lemma. All the terms in the equation therefore are perfectly defined distributions.

ii) If $u \in H_0^1(\Omega)$ happens to be a solution of (4.8), then necessarily $u\partial_1 u \in H^{-1}(\Omega)$. The information that $u\partial_1 u \in H^{-1}(\Omega)$ is an additional information supplied by the equation when $L^s(\Omega) \not\subset H^{-1}(\Omega)$. Since by duality, $L^s(\Omega) \subset H^{-1}(\Omega)$ is equivalent to $H_0^1(\Omega) \subset L^{s'}(\Omega)$ with $s' = 2$ for $d = 1$, $s' > 2$ for $d = 2$ and $s' = d$ for $d \geq 3$, owing to the Sobolev embeddings, we thus see that if $d \geq 5$, then $L^s(\Omega) \not\subset H^{-1}(\Omega)$. In particular, in the "physical" cases, $d = 1, 2, 3$, Eq. (4.8) a priori takes place in the sense of $H^{-1}(\Omega)$. □

After these function space preliminaries, we are now going to prove the following existence result by the Galerkin method.

Theorem 4.1. *Let Ω be a bounded open subset of $\mathbb{R}^d$. For all $f \in H^{-1}(\Omega)$, there exists a solution $u \in H_0^1(\Omega)$ of problem* (4.7).

We start by constructing an appropriate Galerkin basis. In the sequel, s' takes the values indicated in Remark 4.6 ii) following Lemma 4.6.

Lemma 4.7. *There exists a linearly independent, countable family $(w_i)_{i\in\mathbb{N}}$ of elements of $\mathscr{D}(\Omega)$, the linear combinations of which are dense in $H_0^1(\Omega) \cap L^{s'}(\Omega)$.*

Proof. The spaces $H_0^1(\Omega)$ and $L^{s'}(\Omega)$ are both separable for their respective norms. It follows that $V = H_0^1(\Omega) \cap L^{s'}(\Omega)$ is separable for its natural norm $\|v\|_{H_0^1(\Omega)} + \|v\|_{L^{s'}(\Omega)}$. In fact, V est isometric to the subset $\Delta = \{(v, v) \in H_0^1(\Omega) \times L^{s'}(\Omega)\}$ of the Cartesian product $H_0^1(\Omega) \times L^{s'}(\Omega)$ equipped with the norm $\|(v, w)\| = \|v\|_{H_0^1(\Omega)} + \|w\|_{L^{s'}(\Omega)}$, which is clearly separable. Of course, a subset of a separable metric space is also separable for the induced metric topology.

We now use the fact that $\mathscr{D}(\Omega)$ is dense in V, cf. Lemma 4.5, to construct a countable dense family composed of elements of $\mathscr{D}(\Omega)$. We conclude by appealing to Lemma 4.1. □

Let us now consider the approximated finite dimensional variational problem.

Lemma 4.8. *Let $V_m = \text{vect}\{w_0, w_1, w_2, \dots, w_m\}$. The problem: Find $u_m \in V_m$ such that*

$$\forall v \in V_m, \quad \int_\Omega \nabla u_m \cdot \nabla v\,dx + \int_\Omega u_m \partial_1 u_m v\,dx = \langle f, v\rangle, \tag{4.9}$$

admits at least one solution. This solution moreover satisfies

$$\|\nabla u_m\|_{L^2(\Omega)} \leq \|f\|_{H^{-1}(\Omega)}. \tag{4.10}$$

Proof. We first remark that by construction of the w_i, $V_m \subset \mathscr{D}(\Omega)$. We equip V_m with the L^2 inner product (without specific notation for brevity). Just as before, there exist two continuous mappings A_m and B_m from V_m into V_m such that

$$\forall z, v \in V_m, \qquad \begin{cases} \displaystyle\int_\Omega \nabla z \cdot \nabla v \, dx = \int_\Omega A_m(z) v \, dx, \\ \displaystyle\int_\Omega z \partial_1 z v \, dx = \int_\Omega B_m(z) v \, dx. \end{cases}$$

Likewise, there exists $F_m \in V_m$ such that

$$\forall v \in V_m, \quad \langle f, v \rangle = \int_\Omega F_m v \, dx.$$

We set $P_m(z) = A_m(z) + B_m(z) - F_m$. The finite dimensional variational problem then becomes: Find $u_m \in V_m$ such that

$$P_m(u_m) = 0.$$

To solve such an equation with the help of Theorem 2.7, we thus need to compute the following inner products:

$$\int_\Omega P_m(z) z \, dx = \int_\Omega \nabla z \cdot \nabla z \, dx + \int_\Omega z^2 \partial_1 z \, dx - \langle f, z \rangle, \tag{4.11}$$

for $z \in V_m$. Now $V_m \subset \mathscr{D}(\Omega)$, therefore $z^3 \in \mathscr{D}(\Omega)$ and $\partial_1(z^3) = 3z^2\partial_1 z$. Consequently,

$$\int_\Omega z^2 \partial_1 z \, dx = \frac{1}{3} \int_\Omega \partial_1(z^3) \, dx = 0,$$

and the inner product (4.11) thus reduces to

$$\int_\Omega P_m(z) z \, dx = \int_\Omega \nabla z \cdot \nabla z \, dx - \langle f, z \rangle. \tag{4.12}$$

We observe that the nonlinear term has disappeared. It is then elementary to find a sphere on which $\int_\Omega P_m(z) z \, dx \geq 0$, which implies the existence of u_m.

We now write Eq. (4.12) for $z = u_m$, and obtain

$$\|\nabla u_m\|^2_{L^2(\Omega)} = \int_\Omega \nabla u_m \cdot \nabla u_m \, dx = \langle f, u_m \rangle \leq \|\nabla u_m\|_{L^2(\Omega)} \|f\|_{H^{-1}(\Omega)},$$

hence estimate (4.10). □

Remark 4.7. The cancellation of the nonlinear term in the inner product computation is what makes things work for this particular form of nonlinearity. Of course, the nonlinear term will remain in the general variational formulation below. □

According to estimate (4.10), we can extract from the sequence u_m a subsequence, still denoted u_m for brevity, that weakly converges in $H^1_0(\Omega)$ toward a limit u. We now are in a position to complete the proof of Theorem 4.1.

Lemma 4.9. *The weak limit $u \in H^1_0(\Omega)$ is a solution of the variational problem:*

$$\forall v \in V, \quad \int_\Omega \nabla u \cdot \nabla v \, dx + \int_\Omega u \partial_1 u v \, dx = \langle f, v \rangle. \tag{4.13}$$

In particular, u solves problem (4.7)*.*

Proof. Let us take an index i. For all $m \geq i$, $w_i \in V_m$ and there holds

$$\int_\Omega \nabla u_m \cdot \nabla w_i \, dx + \int_\Omega u_m \partial_1 u_m w_i \, dx = \langle f, w_i \rangle. \tag{4.14}$$

Now $\nabla u_m \rightharpoonup \nabla u$ in $L^2(\Omega)$ when $m \to +\infty$, therefore

$$\int_\Omega \nabla u_m \cdot \nabla w_i \, dx \longrightarrow \int_\Omega \nabla u \cdot \nabla w_i \, dx.$$

On the other hand, by Rellich's theorem, $u_m \to u$ in $L^2(\Omega)$ strong. As $w_i \in \mathscr{D}(\Omega)$, it follows at once that $u_m w_i \to u w_i$ in $L^2(\Omega)$ strong. Indeed,

$$\int_\Omega (u_m w_i - u w_i)^2 \, dx \leq \max_{\overline{\Omega}}(w_i^2) \|u_m - u\|^2_{L^2(\Omega)}.$$

This strong convergence combined with the weak convergence of $\partial_1 u_m$ imply that the nonlinear term passes to the limit,

$$\int_\Omega u_m w_i \partial_1 u_m \, dx \longrightarrow \int_\Omega u w_i \partial_1 u \, dx.$$

The right-hand side of Eq. (4.14) does not depend on m, we have therefore obtained

$$\int_\Omega \nabla u \cdot \nabla w_i \, dx + \int_\Omega u \partial_1 u w_i \, dx = \langle f, w_i \rangle.$$

This equality holds true for all $i \in \mathbb{N}$. It follows by linear combinations that

$$\forall z \in \bigcup_{j=0}^{+\infty} V_j, \quad \int_\Omega \nabla u \cdot \nabla z \, dx + \int_\Omega u \partial_1 u z \, dx = \langle f, z \rangle.$$

Here again, $\cup_{j=0}^{+\infty} V_j$ is dense in $V = H_0^1(\Omega) \cap L^{s'}(\Omega)$. For all $v \in V$, there thus exists a sequence $z_k \in \cup_{j=0}^{+\infty} V_j$ such that $z_k \to v$ in $H_0^1(\Omega)$ strong and simultaneously in $L^{s'}(\Omega)$ strong. It follows from the first convergence that

$$\int_\Omega \nabla u \cdot \nabla z_k \, dx \longrightarrow \int_\Omega \nabla u \cdot \nabla v \, dx \text{ and } \langle f, z_k \rangle \longrightarrow \langle f, v \rangle$$

on the one hand, and on the other hand by the second convergence, that

$$\int_\Omega u \partial_1 u z_k \, dx \longrightarrow \int_\Omega u \partial_1 u v \, dx.$$

Indeed, we have seen that $u \partial_1 u \in L^s(\Omega)$. We thus obtain that u solves de variational problem (4.13).

To conclude, we remark that $\mathscr{D}(\Omega) \subset V$, which implies that u solves problem (4.7) in the sense of distributions. □

Remark 4.8. The variational problem is a little unusual inasmuch as the space where the solution is found, $H_0^1(\Omega)$, is in general, i.e., for $d \geq 5$, different from the test-function space V, which is a dense subspace of the solution space. In particular, we cannot take $v = u$ in (4.13), as is commonly done. In fact, there is no reason for u to belong to $L^{s'}(\Omega)$, because the equation offers no control on the $L^{s'}(\Omega)$ norm of potential solutions. □

To circumvent this difficulty, we make the following remark.

Proposition 4.1. *Every solution u of problem* (4.7) *satisfies*

$$\forall v \in H_0^1(\Omega), \quad \int_\Omega \nabla u \cdot \nabla v \, dx + \langle u \partial_1 u, v \rangle_{H^{-1}(\Omega), H_0^1(\Omega)} = \langle f, v \rangle_{H^{-1}(\Omega), H_0^1(\Omega)}. \tag{4.15}$$

Proof. If $u \in H_0^1(\Omega)$ is a solution of problem (4.7), then $u \partial_1 u = f + \Delta u$ belongs to $H^{-1}(\Omega)$ as was already mentioned, and for all $v \in H_0^1(\Omega)$,

$$\begin{aligned}\langle u \partial_1 u, v \rangle_{H^{-1}(\Omega), H_0^1(\Omega)} &= \langle f + \Delta u, v \rangle_{H^{-1}(\Omega), H_0^1(\Omega)} \\ &= \langle f, v \rangle_{H^{-1}(\Omega), H_0^1(\Omega)} - \int_\Omega \nabla u \cdot \nabla v \, dx,\end{aligned}$$

which is exactly Eq. (4.15). □

We need a technical result in order to make use of Eq. (4.15).

Lemma 4.10. *Let Ω be a bounded open subset of $\mathbb{R}^d$ and let $1 \le p \le +\infty$. Let us be given a distribution T such that $T \in H^{-1}(\Omega) \cap L^{p'}(\Omega)$. Then for all $v \in H_0^1(\Omega) \cap L^p(\Omega)$,*

$$\langle T, v\rangle_{H^{-1}(\Omega),H_0^1(\Omega)} = \int_\Omega T(x)v(x)\,dx.$$

Proof. According to Lemma 4.5, for all $v \in H_0^1(\Omega)\cap L^p(\Omega)$, there exists a sequence $\varphi_n \in \mathscr{D}(\Omega)$ such that $\varphi_n \to v$ in $H_0^1(\Omega)$ strong and $\varphi_n \to v$ in $L^p(\Omega)$ strong if $p < +\infty$ and weak-star if $p = +\infty$. Since $T \in L^{p'}(\Omega) \subset L^1_{\text{loc}}(\Omega)$, the canonical identification of locally integrable functions with distributions states that

$$\langle T, \varphi_n\rangle_{\mathscr{D}'(\Omega),\mathscr{D}(\Omega)} = \int_\Omega T(x)\varphi_n(x)\,dx.$$

Using both convergences of the sequence φ_n, we can clearly pass to the limit in each side of this equality when $n \to +\infty$ and thus obtain the Lemma. □

Corollary 4.1 (Energy Equality). *Every solution of problem (4.7) satisfies*

$$\int_\Omega \nabla u \cdot \nabla u\,dx = \langle f, u\rangle_{H^{-1}(\Omega),H_0^1(\Omega)}. \tag{4.16}$$

Proof. It is enough to show that $\langle u\partial_1 u, u\rangle_{H^{-1}(\Omega),H_0^1(\Omega)} = 0$. We proceed by using a truncation that is slightly different from the one used up to now. Let S_n be the continuous piecewise affine function

$$S_n(t) = \begin{cases} 0 & \text{if } |t| \ge 2n, \\ t & \text{if } |t| \le n, \\ -t-2n & \text{if } -2n \le t \le -n, \\ -t+2n & \text{if } n \le t \le 2n. \end{cases}$$

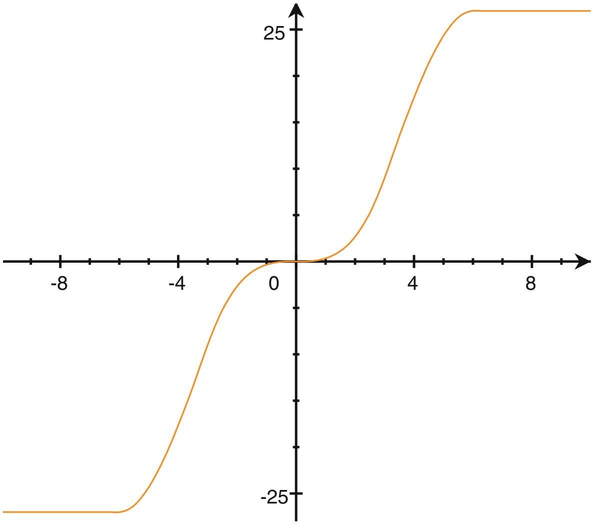
We know that $S_n(u) \in H_0^1(\Omega) \cap L^{s'}(\Omega)$. Therefore, Lemma 4.10 implies that

$$\langle u\partial_1 u, S_n(u)\rangle_{H^{-1}(\Omega),H_0^1(\Omega)} = \int_\Omega u\partial_1 u S_n(u)\,dx. \tag{4.17}$$

We now introduce the function

$$G_n(t) = \int_0^t sS_n(s)\,ds.$$

Fig. 4.1 The function G_3



Since $|sS_n(s)| \leq n^2$, it follows that G_n belongs to $C^1(\mathbb{R})$ and is globally Lipschitz. Consequently, $G_n(u) \in H_0^1(\Omega)$ with $\nabla G_n(u) = uS_n(u)\nabla u$. It thus follows from Eq. (4.17) that

$$\langle u\partial_1 u, S_n(u)\rangle_{H^{-1}(\Omega),H_0^1(\Omega)} = \int_\Omega \partial_1 G_n(u)\,dx = 0. \tag{4.18}$$

It is now easy to establish with the same arguments as those used in the study of the standard truncation at height k, that $S_n(u) \to u$ in $H_0^1(\Omega)$ strong when $n \to +\infty$. We complete the proof by passing to the limit in relation (4.18). □

Remark 4.9. The sequence of functions G_n defines a globally Lipschitz approximation of the function $t \mapsto t^3/3$, see Fig. 4.1. The nullity of the nonlinear term $\langle u\partial_1 u, u\rangle$ thus comes from a slightly refined version of the argument that was already used in the finite dimensional Galerkin approximation. We could not use the standard truncation T_n here, because the primitives of $s \mapsto sT_n(s)$ are not globally Lipschitz on $\mathbb{R}$. Note that the result holds for any solution, not necessarily the one constructed via the Galerkin method.[2] □

Corollary 4.2 (Energy Estimate). *Every solution of problem (4.7) satisfies the estimate*

$$\|\nabla u\|_{L^2(\Omega)} \leq \|f\|_{H^{-1}(\Omega)}.$$

Proof. Immediate application of the energy equality. □

[2] This remark would be more striking if Theorem 4.2, to be seen shortly, did not hold true.

We conclude this chapter by a uniqueness result that shows that there is no other solution than the one already constructed. The result rests on the use of nonlinear test-functions that approximate the sign of the difference of two solutions. More precisely, for all $\delta > 0$, we introduce the continuous piecewise affine function

$$\Sigma_\delta(t) = \begin{cases} -1 & \text{if } t \leq -\delta, \\ \frac{t}{\delta} & \text{if } |t| \leq \delta, \\ +1 & \text{if } t \geq \delta, \end{cases}$$

so that

$$\Sigma'_\delta(t) = \begin{cases} 0 & \text{if } |t| > \delta, \\ \frac{1}{\delta} & \text{if } |t| < \delta. \end{cases}$$

Observe the useful identity $\Sigma'_\delta(t)^2 = \frac{1}{\delta}\Sigma'_\delta(t)$. We start with a technical lemma.

Lemma 4.11. *For all $u_1, u_2 \in H^1_0(\Omega)$, let $w = u_1 - u_2$. There holds*

$$\int_\Omega (u_1\partial_1 u_1 - u_2\partial_1 u_2)\Sigma_\delta(w)\,dx = -\frac{1}{2}\int_\Omega (u_1 + u_2)w\Sigma'_\delta(w)\partial_1 w\,dx. \tag{4.19}$$

Remark 4.10. The integral in the right-hand side makes sense as $|t\Sigma'_\delta(t)| \leq 1$. □

Proof. We consider two sequences $\varphi_1^n, \varphi_2^n \in \mathscr{D}(\Omega)$ that respectively converge to u_1 and u_2 in $H^1_0(\Omega)$ strong. For any n, $(\varphi_1^n)^2 - (\varphi_2^n)^2$ and $\Sigma_\delta(\varphi_1^n - \varphi_2^n)$ both belong to $H^1_0(\Omega)$. We set $\psi^n = \varphi_1^n - \varphi_2^n$ and integrate the left-hand side of (4.19) by parts,

$$\begin{aligned}\int_\Omega (\varphi_1^n\partial_1\varphi_1^n - \varphi_2^n\partial_1\varphi_2^n)\Sigma_\delta(\psi^n)\,dx &= \frac{1}{2}\int_\Omega \partial_1\big((\varphi_1^n)^2 - (\varphi_2^n)^2\big)\Sigma_\delta(\psi^n)\,dx \\ &= -\frac{1}{2}\int_\Omega \big((\varphi_1^n)^2 - (\varphi_2^n)^2\big)\partial_1\Sigma_\delta(\psi^n)\,dx \\ &= -\frac{1}{2}\int_\Omega (\varphi_1^n + \varphi_2^n)\psi^n\Sigma'_\delta(\psi^n)\partial_1\psi^n\,dx.\end{aligned} \tag{4.20}$$

Let us introduce the function $\Gamma_\delta(t) = \int_0^t s\Sigma'_\delta(s)\,ds$. This function is globally Lipschitz with two points of non differentiability. Consequently, $\Gamma_\delta(\psi^n) \in H^1_0(\Omega)$ and $\partial_1\Gamma_\delta(\psi^n) = \psi^n\Sigma'_\delta(\psi^n)\partial_1\psi^n$. It follows that (4.20) may be rewritten in the form

$$\int_\Omega (\varphi_1^n\partial_1\varphi_1^n - \varphi_2^n\partial_1\varphi_2^n)\Sigma_\delta(\psi^n)\,dx = -\frac{1}{2}\int_\Omega (\varphi_1^n + \varphi_2^n)\partial_1\Gamma_\delta(\psi^n)\,dx. \tag{4.21}$$

We now let n tend to infinity. There holds $\varphi_1^n + \varphi_2^n \to u_1 + u_2$ in $L^2(\Omega)$ strong and $\partial_1 \Gamma_\delta(\psi^n) \to \partial_1 \Gamma_\delta(w)$ also in $L^2(\Omega)$ strong, so that

$$\int_\Omega (\varphi_1^n + \varphi_2^n)\partial_1 \Gamma_\delta(\psi^n)\,dx \longrightarrow \int_\Omega (u_1 + u_2)\partial_1 \Gamma_\delta(w)\,dx = \int_\Omega (u_1 + u_2) w \Sigma'_\delta(w)\partial_1 w\,dx.$$

Let us now look at the left-hand side of (4.21). We note that $\Sigma_\delta(\psi^n) \to \Sigma_\delta(w)$ in $H_0^1(\Omega)$ strong and is bounded in $L^\infty(\Omega)$. Consequently, $\Sigma_\delta(\psi^n) \rightharpoonup \Sigma_\delta(w)$ in $L^\infty(\Omega)$ weak-star by uniqueness of the limit. In addition, $\varphi_1^n \partial_1 \varphi_1^n - \varphi_2^n \partial_1 \varphi_2^n \to u_1 \partial_1 u_1 - u_2 \partial_1 u_2$ in $L^1(\Omega)$ strong by the Cauchy-Schwarz inequality. Therefore

$$\int_\Omega (\varphi_1^n \partial_1 \varphi_1^n - \varphi_2^n \partial_1 \varphi_2^n)\Sigma_\delta(\psi^n)\,dx \longrightarrow \int_\Omega (u_1 \partial_1 u_1 - u_2 \partial_1 u_2)\Sigma_\delta(w)\,dx,$$

which proves the Lemma. □

Theorem 4.2. *The solution of problem* (4.15) *is unique.*

Proof. Let u_1 and u_2 be two solutions. Setting $w = u_1 - u_2$ and subtracting the two equations, we see that for all $v \in H_0^1(\Omega)$,

$$\int_\Omega \nabla w \cdot \nabla v\,dx + \langle u_1 \partial_1 u_1 - u_2 \partial_1 u_2, v\rangle_{H^{-1}(\Omega), H_0^1(\Omega)} = 0.$$

In particular, we can take $v = \Sigma_\delta(w)$. Since $\Sigma_\delta(w) \in H_0^1(\Omega) \cap L^\infty(\Omega)$, it follows from Lemma 4.10 that

$$\int_\Omega \nabla w \cdot \nabla \Sigma_\delta(w)\,dx + \int_\Omega (u_1 \partial_1 u_1 - u_2 \partial_1 u_2)\Sigma_\delta(w)\,dx = 0.$$

We now apply Lemma 4.11. This yields

$$\int_\Omega \nabla w \cdot \nabla \Sigma_\delta(w)\,dx = \frac{1}{2}\int_\Omega (u_1 + u_2) w \Sigma'_\delta(w)\partial_1 w\,dx. \tag{4.22}$$

Now $\nabla \Sigma_\delta(w) = \Sigma'_\delta(w)\nabla w = \delta \Sigma'_\delta(w)^2 \nabla w = \delta \Sigma'_\delta(w)\nabla \Sigma_\delta(w)$, so that $\nabla w \cdot \nabla \Sigma_\delta(w) = \delta \nabla \Sigma_\delta(w) \cdot \nabla \Sigma_\delta(w)$, and Eq. (4.22) becomes

$$\int_\Omega \nabla \Sigma_\delta(w) \cdot \nabla \Sigma_\delta(w)\,dx = \frac{1}{2\delta}\int_\Omega (u_1 + u_2) w \Sigma'_\delta(w)\partial_1 w\,dx. \tag{4.23}$$

Let us set

$$E_{\delta,w} = \{x \in \Omega;\ w(x) \neq 0 \text{ and } |w(x)| < \delta\},$$

up to some negligible set. The integrand in the right-hand side of (4.23) vanishes almost everywhere outside of $E_{\delta,w}$. We can thus rewrite it as

$$\int_\Omega \nabla \Sigma_\delta(w)\cdot\nabla\Sigma_\delta(w)\,dx = \frac{1}{2}\int_\Omega \mathbf{1}_{E_{\delta,w}}(u_1+u_2)\Big(\frac{1}{\delta}w\mathbf{1}_{E_{\delta,w}}\Big)\partial_1\Sigma_\delta(w)\,dx. \tag{4.24}$$

At this point, we remark that $|\frac{1}{\delta}w\mathbf{1}_{E_{\delta,w}}| \leq 1$ almost everywhere. Thus by the Cauchy-Schwarz inequality,

$$\|\nabla\Sigma_\delta(w)\|^2_{L^2(\Omega)} \leq \frac{1}{2}\Big(\int_\Omega \mathbf{1}_{E_{\delta,w}}(u_1+u_2)^2\,dx\Big)^{1/2}\|\nabla\Sigma_\delta(w)\|_{L^2(\Omega)},$$

or equivalently

$$\|\nabla\Sigma_\delta(w)\|_{L^2(\Omega)} \leq \frac{1}{2}\Big(\int_{E_{\delta,w}}(u_1+u_2)^2\,dx\Big)^{1/2}.$$

Let us now note that $\cap_{\delta>0}\{x\in\Omega;\,|w(x)|<\delta\} = \{x\in\Omega;\,w(x)=0\}$. Therefore, meas $(E_{\delta,w}) \to 0$ when $\delta\to 0$.[3] Now $(u_1+u_2)^2\in L^1(\Omega)$ and thus by dominated convergence,

$$\int_{E_{\delta,w}}(u_1+u_2)^2\,dx \to 0,$$

when $\delta\to 0$. It follows that $\|\nabla\Sigma_\delta(w)\|_{L^2(\Omega)}\to 0$ when $\delta\to 0$. By the Poincaré inequality, this implies that $\|\Sigma_\delta(w)\|_{H^1(\Omega)}\to 0$, then that $\|\Sigma_\delta(w)\|_{L^1(\Omega)}\to 0$. Since $|\Sigma_\delta(w)| = 1$ almost everywhere on the set $\{x\in\Omega;\,|w|\geq\delta\}$, it follows that

$$\text{meas}\,(\{x\in\Omega;\,|w|\geq\delta\}) \longrightarrow 0 \quad\text{when}\quad \delta\to 0, \delta>0. \tag{4.25}$$

To conclude, we remark that the function $\delta\mapsto$ meas $(\{x\in\Omega;\,|w|\geq\delta\})$ is nonnegative, decreasing on $\mathbb{R}^*_+$. By (4.25), it is thus identically zero, which is equivalent to $w=0$ almost everywhere, or again $u_1=u_2$. □

Remark 4.11. The argument developed here is very specific to this particular equation and the form of its nonlinearity. We can however take note of general purpose techniques such as the use of nonlinear test-functions, and the way we made the nonlinear test-function appear in the bilinear form in passing from Eq. (4.22) to Eq. (4.23).

Now of course, for the actual Navier-Stokes equation, uniqueness is still a major unresolved problem. □

[3] We need to consider every sequence $\delta_n\to 0$ to deduce this from the general properties of measures, for countability reasons.

4.3 Exercises of Chap. 4

1. Let Ω be a bounded open subset of $\mathbb{R}^d$, A a $d \times d$ symmetric matrix with coefficients a_{ij} in $L^\infty(\Omega)$ and such that there exists $\alpha > 0$ with $\sum_{ij} a_{ij}(x)\xi_i\xi_j \geq \alpha\|\xi\|^2$ for all $\xi \in \mathbb{R}^d$ and almost all $x \in \Omega$, and $f\colon \Omega \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ a continuous, bounded mapping. Show that the problem: $u \in H_0^1(\Omega)$, $-\mathrm{div}\,(A(x)\nabla u) = f(x, u, \nabla u)$ admits at least one solution. (*Hint: after having defined a Galerkin approximation u_m of a potential solution, prove that $f(x, u_m, \nabla u_m) \rightharpoonup \bar{f}$ for a certain $\bar{f}$ in $L^2(\Omega)$, then that the sequence u_m actually converge strongly in $H^1(\Omega)$.*)

2. About the necessity of a convoluted proof for Lemma 4.5, let Ω be the unit ball in $\mathbb{R}^d$. Find a real number p and a sequence $\varphi_n \in \mathscr{D}(\Omega)$ such that $\varphi_n \to 0$ in $H_0^1(\Omega)$ but $\varphi_n \not\to 0$ in $L^p(\Omega)$. (*Hint: take $d \geq 3$ and consider a sequence of the form $\varphi_n(x) = n^s\varphi(nx)$ with s well chosen.*) Additional benefit: use this to retrieve the Sobolev exponent in case of blank memory.

3. Let B be the open unit ball of $\mathbb{R}^3$ and let $u \in H_0^1(\Omega; \mathbb{R}^3)$ be such that $\mathrm{div}\, u = 0$. We extend u by 0 outside of the ball and for $0 < \varepsilon < 1$, we let

$$u_\varepsilon = \rho_\varepsilon \star (u((1-\varepsilon)^{-1}x)),$$

where ρ_ε is a mollifier sequence with support in the ball of radius ε. Show that the restriction of u_ε to B belongs to $\mathscr{D}(B; \mathbb{R}^3)$ and that $u_\varepsilon \to u$ in H^1 when $\varepsilon \to 0$. Deduce from this that the space $\mathscr{V} = \{\varphi \in \mathscr{D}(B; \mathbb{R}^3); \mathrm{div}\,\varphi = 0\}$ is dense in $V = \{u \in H_0^1(B; \mathbb{R}^3); \mathrm{div}\, u = 0\}$.

4. With the previous notation, given $f \in L^2(B; \mathbb{R}^3)$, show that problem: Find $u \in V$ such that

$$\forall v \in V, \quad \int_B \big(\nabla u : \nabla v + [(\nabla u)u] \cdot v\big)\, dx = \int_B f \cdot v\, dx,$$

admits at least one solution (if u is a mapping from B into $\mathbb{R}^3$, its gradient ∇u is the 3×3 matrix with coefficients $(\nabla u)_{ij} = \partial_j u_i$ and the inner product of two matrices A_1 and A_2 is defined by $A_1 : A_2 = \mathrm{tr}\,(A_1^T A_2)$. The vector $(\nabla u)u$ thus has components $u_j\partial_j u_i$, $i = 1, 2, 3$, with the Einstein repeated index summation convention. It is usually written $(u\nabla)u$ in the Navier-Stokes literature, $u\nabla$ being the differential operator $u_j\partial_j$, still with summation on j). Remark: we can deduce from this the existence of a solution to the stationary Navier-Stokes problem.

Chapter 5
The Maximum Principle, Elliptic Regularity, and Applications

The expression “maximum principle” is a generic term that covers a set of results of two kinds. One kind concerns the points of maximum or minimum of solutions of certain boundary value problems, elliptic in our case. The other kind is about monotone dependence of the solutions with respect to the data. The two aspects are naturally related to one another. There are furthermore two general contexts, the so-called “strong” context in which classical solutions are considered, and the “weak” context in which variational solutions are considered. This terminology is not universal however.

The connection between classical and variational solutions is one of the aspects of elliptic regularity. Elliptic equations have the property that, as a general rule and loosely speaking, the solution gains some derivatives compared to the data. This is extremely useful in a variety of situations. We give an example of application of both maximum principle and elliptic regularity to the method of super- and sub-solutions to solve certain nonlinear equations.

5.1 The Strong Maximum Principle

Let us start with a first version of the strong maximum principle. We will systematically use the Einstein repeated index summation convention, thus $a_{ij}\xi_i\xi_j = \sum_{i=1}^d \sum_{j=1}^d a_{ij}\xi_i\xi_j$, and so on.

Theorem 5.1. *Let Ω be a bounded open subset of $\mathbb{R}^d$. We are given a $d \times d$ symmetric matrix-valued function A with coefficients a_{ij} belonging to $C^0(\overline{\Omega})$ and such that there exists $\lambda > 0$ with $a_{ij}(x)\xi_i\xi_j \geq \lambda|\xi|^2$ for all $x \in \overline{\Omega}$ and all $\xi \in \mathbb{R}^d$. We are also given a vector-valued function $b \in C^0(\overline{\Omega}; \mathbb{R}^d)$ and a scalar function*

H. Le Dret, *Nonlinear Elliptic Partial Differential Equations*, Universitext,
https://doi.org/10.1007/978-3-319-78390-1_5

$c \in C^0(\overline{\Omega})$ such that $c(x) \geq 0$ in $\overline{\Omega}$. Then any function $u \in C^0(\overline{\Omega}) \cap C^2(\Omega)$ satisfying

$$\begin{cases} -a_{ij}(x)\partial_{ij}u(x) + b_i(x)\partial_i u(x) + c(x)u(x) \geq 0 & \text{in } \Omega, \\ u(x) \geq 0 \quad \text{on } \partial\Omega, \end{cases}$$

is nonnegative in $\overline{\Omega}$.

Remark 5.1. In other words, if we introduce the second order differential operator $L = -a_{ij}\partial_{ij} + b_i\partial_i + c$, when a function u with the indicated regularity is a solution of the Dirichlet boundary value problem $Lu = f$ in Ω, $u = g$ on $\partial\Omega$, with $f \geq 0$ and $g \geq 0$, then $u \geq 0$. This is a monotonicity result: if f somehow represents an "applied force", then the solution u goes in the direction the force is pulling, if the latter is defined.

Condition $A\xi \cdot \xi \geq \lambda|\xi|^2$ is called *uniform coerciveness*, or coercivity, for the matrix-valued function A, which is then said to be *coercive*. It is also this condition that makes the operator *elliptic*. □

Theorem 5.1 is based on the following remark.

Lemma 5.1. *Let $L' = -a_{ij}\partial_{ij}$. If $u \in C^2(\Omega)$ has a local minimum at point x_0 of Ω, then $L'u(x_0) \leq 0$.*

Proof. We first note a linear algebra result that if A and B are two symmetric nonnegative matrices, then $A : B = \operatorname{tr}(A^T B) \geq 0$. Indeed, write $B = Q^T \Lambda Q$ with $Q \in \mathrm{O}(d)$ and Λ a diagonal matrix with nonnegative diagonal entries $\lambda_i \geq 0$. It follows that $A : B = A' : \Lambda$ where $A' = QAQ^T$ is another symmetric nonnegative matrix. In particular, its diagonal entries a'_{ii} (without summation) are also nonnegative. Now, $A' : B = a'_{ij}(\lambda_i\delta_{ij}) \geq 0$ (without summation inside the parentheses), since all nonzero terms in this sum are nonnegative.

Going back to the maximum principle, we see that $L'u(x_0) = -A(x_0) : D^2u(x_0)$. The matrix $A(x_0)$ is symmetric nonnegative by hypothesis. The matrix $D^2u(x_0)$ is symmetric. We pick any vector $\xi \in \mathbb{R}^d$ and consider the function $t \mapsto g(t) = u(x_0 + t\xi)$. This function is well-defined and of class C^2 in an open interval around 0. Moreover, it has a local minimum at $t = 0$, which implies that $g''(0) \geq 0$. By the chain rule, $g''(0) = D^2u(x_0)(\xi, \xi)$, and it follows that the matrix $D^2u(x_0)$ is also nonnegative, hence the Lemma. □

We note that this result does not require $\lambda > 0$ to hold. We use the Lemma to give an important property of the minimum value of u under the hypotheses of Theorem 5.1.

Lemma 5.2. *Let $u \in C^0(\overline{\Omega}) \cap C^2(\Omega)$ be such that $Lu \geq 0$ in Ω. Then*
i) if $c = 0$, there holds $\min_{\overline{\Omega}} u = \min_{\partial\Omega} u$,
ii) if $c \geq 0$, there holds $\min_{\overline{\Omega}} u \geq \min_{\partial\Omega}(-u_-)$.

Proof. First of all Ω is bounded, thus $\overline{\Omega}$ is compact and u attains its minimum somewhere on $\overline{\Omega}$.

Let us prove i). We first assume that $Lu > 0$ in Ω. If u attains its minimum on $\partial\Omega$, then we are done. Assume thus that u attains its minimum at a point x_0 in Ω. This is an interior point, thus $Du(x_0) = 0$, and therefore $Lu(x_0) = L'u(x_0) \leq 0$ according to Lemma 5.1, contradiction.

Let us now assume that $Lu \geq 0$ in Ω. We are going to see that this case reduces to the previous one. For this, we set $u_\varepsilon(x) = u(x) - \varepsilon e^{\gamma x_1}$, where the positive constants ε and γ are to be chosen appropriately. An elementary computation shows that

$$L(e^{\gamma x_1}) = (-a_{11}(x)\gamma^2 + b_1(x)\gamma)e^{\gamma x_1}.$$

Let us pick γ large enough so that $\lambda\gamma^2 - \|b_1\|_{C^0(\overline{\Omega})}\gamma > 0$. This is possible because $\lambda > 0$. The choice $\xi = e_1$ in the coerciveness inequality shows that $a_{11}(x) \geq \lambda$, and of course $|b_1(x)| \leq \|b_1\|_{C^0(\overline{\Omega})}$ for all $x \in \Omega$, therefore

$$-a_{11}(x)\gamma^2 + b_1(x)\gamma \leq -\lambda\gamma^2 + \|b_1\|_{C^0(\overline{\Omega})}\gamma,$$

for all $x \in \Omega$. Consequently,

$$-L(e^{\gamma x_1}) \geq (\lambda\gamma^2 - \|b_1\|_{C^0(\overline{\Omega})}\gamma)e^{\gamma x_1}.$$

If we set $\eta = \varepsilon(\lambda\gamma^2 - \|b_1\|_{C^0(\overline{\Omega})}\gamma)\min_{\overline{\Omega}}(e^{\gamma x_1})$, then $\eta > 0$ since Ω is bounded and

$$Lu_\varepsilon = Lu - \varepsilon L(e^{\gamma x_1}) \geq \eta > 0 \text{ in } \Omega.$$

We are thus back in the previous case, so that u_ε attains its minimum on $\partial\Omega$, i.e.,

$$\min_{x\in\overline{\Omega}}(u(x) - \varepsilon e^{\gamma x_1}) = \min_{x\in\partial\Omega}(u(x) - \varepsilon e^{\gamma x_1}).$$

We now let ε tend to 0. As $\varepsilon e^{\gamma x_1}$ tends uniformly to 0 on $\overline{\Omega}$, both minima converge and case i) follows.

Let us now deal with case ii). If $u \geq 0$ in Ω, then $u \geq 0$ on $\overline{\Omega}$ by continuity, thus $\min_{\overline{\Omega}} u \geq 0$ on the one hand and $u_- = 0$ on the other hand. In such a case, there is nothing to prove.

Let us thus assume that u takes strictly negative values somewhere in Ω and define $\Omega_- = \{x \in \Omega; u(x) < 0\}$. This is a nonempty open set. Moreover, $\bar{L}u = Lu - cu \geq 0$ in Ω_- and $\bar{L}$ has no zero order term. According to case i), there thus holds

$$\min_{x\in\overline{\Omega_-}}(u(x)) = \min_{x\in\partial\Omega_-}(u(x)).$$

Now, it is clear that

$$\min_{x\in\overline{\Omega_-}}(u(x)) = \min_{x\in\overline{\Omega}}(u(x))$$

since u is nonnegative outside of $\overline{\Omega_-}$ and takes strictly negative values in Ω_-. Besides, $u \leq 0$ on $\partial\Omega_-$, so that $u = -u_-$ on $\partial\Omega_-$. Furthermore, $\partial\Omega_- = (\partial\Omega_- \cap \Omega) \cup (\partial\Omega_- \cap \partial\Omega)$. Now if $x \in \partial\Omega_- \cap \Omega$ then $u(x) = 0$, for otherwise $u(x) < 0$ implies $x \in \Omega_-$ by definition. Thus, since $\min_{\overline{\Omega}} u < 0$, we see that

$$\min_{x\in\partial\Omega_-}(u(x)) = \min_{x\in\partial\Omega_-}(-u_-(x)) = \min_{x\in\partial\Omega_-\cap\partial\Omega}(-u_-(x)) = \min_{x\in\partial\Omega}(-u_-(x)),$$

which completes the proof. □

We now are in a position to prove the Theorem.

Proof of Theorem 5.1. We apply Lemma 5.2 ii). If $u \geq 0$ on $\partial\Omega$, then $u_- = 0$ on $\partial\Omega$. Therefore, $\min_{\overline{\Omega}} u \geq 0$. □

Remark 5.2. i) Note that if $Lu > 0$, then the proof of Lemma 5.2 shows that there is no interior local minimum.

ii) There are of course analogous results concerning maxima, by inverting all the signs.

iii) We see that if $c = 0$ or if u takes strictly negative values, then u attains its minimum on the boundary of the open set. A contrario, if u does not take strictly negative values on the boundary and if c is nonzero, then we cannot say anything about where the minimum is attained.

iv) The maximum principle is still true, but with a markedly more delicate proof, under weaker regularity hypotheses, $u \in W^{2,p}(\Omega)$, $p > d$, and $a_{ij}, b_i, c \in L^\infty(\Omega)$, see [36].

v) Let us mention that the maximum principle, strong or otherwise, is specific to second order elliptic equations. In other words, it has no analogue, barring a few exceptions, neither for systems of equations of second order, nor for higher order elliptic equations. There are versions of the maximum principle for parabolic equations, such as the heat equation.

vi) A consequence of the strong maximum principle is the uniqueness of the solution of the Dirichlet problem in the class $C^0(\overline{\Omega}) \cap C^2(\Omega)$. Indeed, $Lu = 0$ and $u = 0$ on $\partial\Omega$ imply that both $u \geq 0$ and $u \leq 0$ in Ω. □

We are going to refine the study of points where u attains its minimum with a result due to Hopf. This result requires some regularity for the boundary of Ω. We will assume the interior sphere condition which states that for each point x of $\partial\Omega$, there exists an open ball $B(y, R)$ included in Ω such that $x \in \bar{B}(y, R)$. Intuitively, this interior ball is "tangent" to $\partial\Omega$, at least when $\partial\Omega$ is smooth. We then define an exterior normal unit vector n to $\partial\Omega$ at x by letting $n = (x - y)/R$. The "exterior"

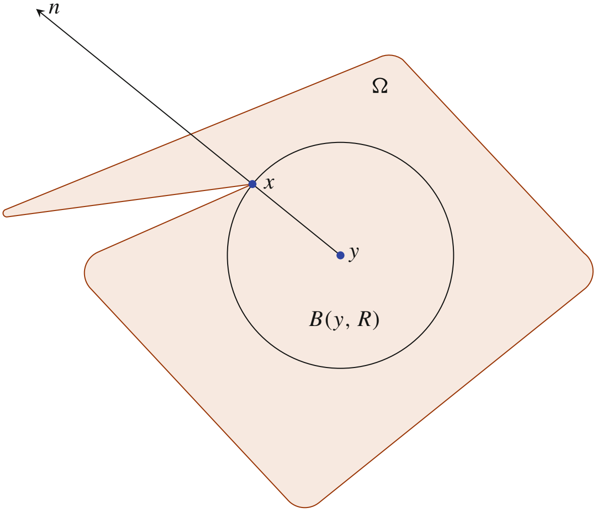


Fig. 5.1 The interior sphere condition

terminology for this vector is a little misleading here, because nothing prevents Ω from being on both sides of $\partial\Omega$ or from being as in Fig. 5.1.

Theorem 5.2 (Hopf). *Let Ω be an open subset of $\mathbb{R}^d$ satisfying the interior sphere condition and let L and $u \in C^1(\overline{\Omega}) \cap C^2(\Omega)$ satisfy the same hypotheses as before. If u attains a strict local minimum at a point x_0 of $\partial\Omega$ in the case $c = 0$, or a strict local nonpositive minimum in the case $c \geq 0$, then*

$$\frac{\partial u}{\partial n}(x_0) = n_i(x_0)\partial_i u(x_0) < 0. \tag{5.1}$$

Remark 5.3. At such a point x_0, the directional derivative in a direction pointing outwards is necessarily nonpositive. Consider for this the function $t \mapsto u(x_0 - tn(x_0))$ for $t > 0$. Hopf's theorem states that it is in fact strictly negative. Heuristically, if this derivative was zero, then $t \mapsto u(x_0 - tn(x_0))$ would tend to be convex in a neighborhood of 0, which is essentially forbidden by $Lu \geq 0$. Of course, this is in no way a proof. □

Proof. Let $B(y_0, R)$ be the ball associated with point x_0 by the interior sphere condition. We can always choose R small enough so that $u(x_0) < u(x)$ for all $x \in B(y_0, R)$, because x_0 is a point of strict local minimum. We set

$$v(x) = e^{-\gamma|x-y_0|^2} - e^{-\gamma R^2},$$

where $\gamma > 0$ is a constant to be appropriately chosen. By construction, $v(x) = 0$ on the sphere $S(y_0, R)$ centered at y_0 and of radius R, thus in particular at x_0, and $v > 0$ in the open ball $B(y_0, R)$.

An elementary computation shows that

$$Lv(x) = \big[-4\gamma^2 a_{ij}(x)(x_i - y_{0i})(x_j - y_{0j}) + 2\gamma\big(a_{ii}(x) - b_i(x)(x_i - y_{0i})\big)\big]e^{-\gamma|x-y_0|^2} + c(x)v(x).$$

Consequently, because A is uniformly coercive and $c(x)e^{-\gamma R^2} \geq 0$, we see that

$$Lv(x) \leq \big[-4\gamma^2\lambda|x - y_0|^2 + 2\gamma\big(a_{ii}(x) + |b_i(x)||x_i - y_{0i}|\big) + c(x)\big]e^{-\gamma|x-y_0|^2}.$$

In particular, if $x \in O = B(y_0, R) \setminus \bar{B}(y_0, R/2)$,

$$Lv(x) \leq \big[-\gamma^2\lambda R^2 + 2\gamma\big(\|a\|_{C^0(\overline{\Omega})} + \|b\|_{C^0(\overline{\Omega})}R\big) + \|c\|_{C^0(\overline{\Omega})}\big]e^{-\gamma|x-y_0|^2}.$$

We pick γ large enough so that

$$-\gamma^2\lambda R^2 + 2\gamma\big(\|a\|_{C^0(\overline{\Omega})} + \|b\|_{C^0(\overline{\Omega})}R\big) + \|c\|_{C^0(\overline{\Omega})} < 0,$$

which implies that $Lv(x) < 0$ for all $x \in \overline{O}$.

Let us now consider the function

$$z(x) = u(x) - u(x_0) - \varepsilon v(x).$$

The following facts hold on the open set O. First of all, $Lz = Lu - cu(x_0) - \varepsilon Lv > 0$. Indeed, either $c = 0$ with no assumption on the sign of $u(x_0)$, or $c \geq 0$ with $c \not\equiv 0$ in which case it is assumed that $u(x_0) \leq 0$.

Secondly, $\partial O = S(y_0, R) \cup S(y_0, R/2)$. On the sphere $S(y_0, R)$, v vanishes and $z = u - u(x_0) \geq 0$. On the sphere $S(y_0, R/2)$, there exists a constant η such that $u - u(x_0) \geq \eta > 0$, because x_0 is a point of strict local minimum. If we choose $\varepsilon \leq \frac{\eta}{e^{-\gamma R^2/4} - e^{-\gamma R^2}}$, it follows that $z = u - u(x_0) - \varepsilon v \geq 0$ on this sphere too.

We then apply the maximum principle of Theorem 5.1 to the function z on the open set O, which shows that

$$\forall x \in O, \quad u(x) \geq u(x_0) + \varepsilon v(x).$$

When we restrict this inequality to the segment, $x_0 - tn(x_0)$, $R/2 \geq t > 0$, we obtain

$$\frac{u(x_0 - tn(x_0)) - u(x_0)}{t} \geq \varepsilon\frac{v(x_0 - tn(x_0))}{t} \longrightarrow 2\varepsilon\gamma R, \quad \text{when} \quad t \to 0^+.$$

It follows that

$$\frac{\partial u}{\partial n}(x_0) \leq -2\varepsilon\gamma R,$$

and the proof is complete. □

Remark 5.4. The trick consists in bounding u from below by a function that agrees with u at x_0, and has a strictly negative normal derivative at x_0, based on the maximum principle. This is an example of the use of so-called *barrier functions.* □

The Hopf theorem implies an even stronger version of the maximum principle: if the minimum is attained at an interior point, then the function u is constant and equal to this minimum value on the connected component of that point.

The idea is very simple. If u attains its minimum at an interior point, then its gradient vanishes there, thus contradicting the Hopf theorem on a small ball with this point on its boundary. Putting this idea in practice is not so easy. The difficulty is that we have no information whatsoever on the set on which u attains its minimum, and we need to be able to construct a ball included in the open set that touches this minimum set at only one single point.

Theorem 5.3. *Let Ω be a bounded, connected open subset of $\mathbb{R}^d$ and $u \in C^0(\overline{\Omega}) \cap C^2(\Omega)$ be such that $Lu \geq 0$ on Ω. Let $m = \min_{\overline{\Omega}} u$. Then if $c = 0$, or if $c \geq 0$ and $m \leq 0$, the following alternative holds: either $u \equiv m$ on $\overline{\Omega}$, or $u > m$ on Ω.*

Proof. Let $M = \{x \in \overline{\Omega};\, u(x) = m\}$. This is a closed subset of $\overline{\Omega}$, hence a compact set. It follows that for all $y \in \mathbb{R}^d$, the distance from y to M is attained, i.e., there exists $p(y) \in M$ such that $\delta(y) = |y - p(y)| = \inf_{z \in M} |y - z|$. Consequently $B(y, \delta(y)) \subset \mathbb{R}^d \setminus M$ and $p(y) \in \bar{B}(y, \delta(y))$.[1]

Let us now set $N = \Omega \setminus M = \{x \in \Omega;\, u(x) > m\}$. This is an open subset of Ω. We are going to show that it is also closed in Ω for the induced topology. We thus consider a sequence $y_k \in N$ such that $y_k \to y_0$ in Ω and we want to show that $y_0 \in N$.

We argue by contradiction and assume that $y_0 \notin N$, that is to say $y_0 \in M \cap \Omega$. In this case, $\delta(y_k) \to 0$ and $p(y_k) \to y_0$, by extracting a subsequence and using the uniqueness of its limit. In particular, for k large enough, we have $B(y_k, \delta(y_k)) \subset \Omega$ and $p(y_k) \in \Omega$. It follows then that $B(y_k, \delta(y_k)) \subset N$. Let us choose one such k and set $B = B\big((p(y_k) + y_k)/2, \delta(y_k)/2\big)$.

By construction, $B \subset N$ and thus, for all x in B, $u(x) > m$. Moreover, the only point in $\bar{B}$ that is not in $B(y_k, \delta(y_k))$ is precisely $p(y_k)$, with $u(p(y_k)) = m$. It follows that $p(y_k)$ is a point of strict minimum of u on the closure of the open set B. This open set obviously satisfies the interior sphere condition since it is a ball. Naturally u is C^1 on $\bar{B}$ as the restriction of a C^2 function on Ω. Finally, we

[1] Caution: p is not a mapping, the distance may be attained at several points of M.

have made the hypothesis that $m \leq 0$ if $c \geq 0$. By the Hopf theorem, it follows that $Du(p(y_k)) \cdot n(p(y_k)) < 0$. Now this is impossible since $p(y_k)$ is an interior minimum point, for which $Du(p(y_k)) = 0$. Contradiction and we see that in fact $y_0 \in N$, that is to say that N is closed.

The set N is an open and closed subset of a connected set Ω, it is thus either empty, or equal to Ω. □

Let us give a first application of the strong maximum principle to an estimate result.

Theorem 5.4. *Let $\eta > 0$ and $u \in C^2(\overline{\Omega})$ be such that $Lu + \eta u = f$ in Ω and $u = g$ on $\partial\Omega$. Then*

$$\|u\|_{C^0(\overline{\Omega})} \leq \max\left\{\|g\|_{C^0(\partial\Omega)}, \frac{\|f\|_{C^0(\overline{\Omega})}}{\eta}\right\}.$$

Proof. First of all, since $u \in C^2(\bar{\Omega})$ and satisfies the equation in Ω, it follows that f has a continuous extension to $\overline{\Omega}$, i.e., $f \in C^0(\overline{\Omega})$. Let us set

$$v = u - \max\left\{\|g\|_{C^0(\partial\Omega)}, \frac{\|f\|_{C^0(\overline{\Omega})}}{\eta}\right\}.$$

There holds

$$v \leq u - \|g\|_{C^0(\partial\Omega)} \leq 0 \text{ on } \partial\Omega,$$

and

$$\begin{aligned} Lv + \eta v &= f - (c+\eta)\max\left\{\|g\|_{C^0(\partial\Omega)}, \frac{\|f\|_{C^0(\overline{\Omega})}}{\eta}\right\} \\ &\leq f - c\frac{\|f\|_{C^0(\overline{\Omega})}}{\eta} - \|f\|_{C^0(\overline{\Omega})} \leq -c\frac{\|f\|_{C^0(\overline{\Omega})}}{\eta} \leq 0, \end{aligned}$$

in Ω. Hence by the strong maximum principle, $v \leq 0$ in $\overline{\Omega}$, or in other words,

$$u \leq \max\left\{\|g\|_{C^0(\partial\Omega)}, \frac{\|f\|_{C^0(\overline{\Omega})}}{\eta}\right\} \text{ in } \overline{\Omega}.$$

We start over the same argument with $v = u + \max\left\{\|g\|_{C^0(\partial\Omega)}, \frac{\|f\|_{C^0(\overline{\Omega})}}{\eta}\right\}$ to conclude the proof. □

Remark 5.5. We could as well have assumed that $u \in C^0(\overline{\Omega}) \cap C^2(\Omega)$. The result still holds if $|f|$ is bounded on Ω by replacing the $C^0(\bar{\Omega})$ norm by its upper bound. If it is not bounded, the upper bound in the right-hand side is $+\infty$, so the estimate also holds true, without saying much. □

Let us mention an estimate result that is rather similar to the previous one.

Theorem 5.5. *Let $u \in C^0(\overline{\Omega}) \cap C^2(\Omega)$ be such that $Lu = f$ on Ω, with f bounded on Ω, and $u = g$ on $\partial\Omega$. Then there exists a constant C that only depends on the diameter of Ω, on $\|b\|_{C^0(\overline{\Omega})}$ and on λ, such that*

$$\|u\|_{C^0(\overline{\Omega})} \leq \|g\|_{C^0(\partial\Omega)} + C \sup_{\Omega} |f|.$$

Proof. The set Ω is bounded, thus there exists a real number δ such that Ω is included in the strip $\{x \in \mathbb{R}^d; -\delta/2 < x_1 < \delta/2\}$. Let $L' = -a_{ij}\partial_{ij} + b_i\partial_i$ and $\beta = \|b\|_{C^0(\overline{\Omega})}/\lambda$. If we set $\alpha = \beta + 1 \geq 1$, then

$$\begin{aligned} L'(e^{\alpha(x_1+\delta/2)}) &= (-\alpha^2 a_{11} + \alpha b_1)e^{\alpha(x_1+\delta/2)} \\ &\leq \lambda(-\alpha^2 + \alpha\beta)e^{\alpha(x_1+\delta/2)} = -\lambda\alpha e^{\alpha(x_1+\delta/2)} \leq -\lambda. \end{aligned}$$

We then define

$$v(x) = \|g\|_{C^0(\partial\Omega)} + (e^{\alpha\delta} - e^{\alpha(x_1+\delta/2)})\frac{\sup_\Omega |f|}{\lambda} \geq 0.$$

Clearly, $Lv = L'v + cv \geq L'v \geq \sup_\Omega |f|$. Consequently,

$$L(v - u) = Lv - Lu \geq \sup_{\Omega} |f| - f \geq 0 \text{ in } \Omega \text{ and } v - u \geq 0 \text{ on } \partial\Omega.$$

By the strong maximum principle, it thus follows that $v - u \geq 0$ in $\overline{\Omega}$, hence

$$u(x) \leq \|g\|_{C^0(\partial\Omega)} + (e^{\alpha\delta} - 1)\frac{\sup_\Omega |f|}{\lambda},$$

for all $x \in \overline{\Omega}$. We conclude by changing u into $-u$. □

Remark 5.6. i) By translating and rotating, we see that we can take δ equal to the diameter of Ω, hence the announced dependance of constant C.

ii) On the other hand, the only boundedness required for Theorem 5.5 to hold, is that Ω be bounded in one direction, thus included in one such strip. So a better constant is actually achieved by taking for δ the infimum of the widths all such strips, which is in general strictly smaller than the diameter.

iii) It is not necessary here to assume that c is bounded below by a strictly positive constant, which is what the constant $\eta > 0$ of Theorem 5.4 was basically here for. □

5.2 The Weak Maximum Principle

In this section, we consider the same kind of questions as before, but for weak solutions and under less restrictive regularity assumptions. Accordingly, the results are less fine than in the previous section. Let us first give a weak analogue of Theorem 5.1.

Theorem 5.6. *Let Ω be a bounded open subset of $\mathbb{R}^d$. We are given a $d \times d$ symmetric matrix-valued function A with coefficients $a_{ij} \in L^\infty(\Omega)$ such that there exists $\lambda > 0$ with $a_{ij}(x)\xi_i\xi_j \geq \lambda|\xi|^2$ for almost all $x \in \Omega$ and all $\xi \in \mathbb{R}^d$, and a scalar function $c \in L^\infty(\Omega)$ such that $c \geq 0$ almost everywhere. Then, any function $u \in H^1(\Omega)$ which satisfies*

$$\begin{cases} -\operatorname{div}(A\nabla u) + cu \geq 0, \\ u_- \in H_0^1(\Omega), \end{cases}$$

is nonnegative almost everywhere in Ω.

Remark 5.7. It is worth mentioning that a distribution $T \in \mathscr{D}'(\Omega)$ is said to be nonnegative if and only if, for all $\varphi \in \mathscr{D}(\Omega)$ such that $\varphi(x) \geq 0$ in Ω, there holds $\langle T, \varphi\rangle \geq 0$. In this case, it is known that T is in fact a nonnegative Radon measure.[2] The first inequality must be understood in the above distributional sense. The second condition is a weak way of expressing that u is nonnegative on the boundary, even if the latter is not regular enough for a trace mapping to exist. Indeed, if Ω is regular, then $\gamma_0(u_-) = (\gamma_0(u))_-$ and since $H_0^1(\Omega) = \ker \gamma_0$, it follows that $\gamma_0(u) \geq 0$ almost everywhere on $\partial\Omega$. □

Let us start with a lemma on nonnegative elements of $H^{-1}(\Omega)$.

Lemma 5.3. *Let $f \in H^{-1}(\Omega)$ be such that $f \geq 0$ in the sense of $\mathscr{D}'(\Omega)$. Then, for all v in $H_0^1(\Omega)$, $\langle f, v_+\rangle_{H^{-1}(\Omega),H_0^1(\Omega)} \geq 0$.*

Proof. Let $v \in H_0^1(\Omega)$. There exists a sequence $\varphi_n \in \mathscr{D}(\Omega)$ such that $\varphi_n \to v$ in $H_0^1(\Omega)$ strong. By the continuity of superposition operators, it follows that $(\varphi_n)_+ \to v_+$ in $H_0^1(\Omega)$ strong. For each n, $(\varphi_n)_+$ is compactly supported. We can thus approximate it in $H_0^1(\Omega)$ strong by convolution by a mollifying sequence, $\rho_\varepsilon \star (\varphi_n)_+$, which is thus well-defined and belongs to $\mathscr{D}(\Omega)$ as soon as ε is smaller than the distance of the support of $(\varphi_n)_+$ to the boundary $\partial\Omega$. Moreover, by the very definition of convolution, $\rho_\varepsilon \star (\varphi_n)_+ \geq 0$. Thus, since $f \geq 0$, we see that $\langle f, \rho_\varepsilon \star (\varphi_n)_+\rangle \geq 0$. We then extract a sequence such that $\rho_{\varepsilon_n} \star (\varphi_n)_+ \to v_+$ in $H_0^1(\Omega)$ strong and conclude by passing to the limit in the inequality, which works because $f \in H^{-1}(\Omega)$. □

[2] When T is moreover L^1_{loc}, it is then nonnegative almost everywhere.

Of course, likewise, $\langle f, v_-\rangle_{H^{-1}(\Omega),H_0^1(\Omega)} \geq 0$ for all v in $H_0^1(\Omega)$.

Proof of Theorem 5.6. We are going to show that $u_- = 0$. Since $u_- \in H_0^1(\Omega)$, it is enough to show $\nabla u_- = 0$ and then apply the Poincaré inequality. Let thus $f = -\operatorname{div}(A\nabla u) + cu$. This distribution belongs to $H^{-1}(\Omega)$ and is nonnegative. Therefore, there holds

$$0 \leq \langle f, u_-\rangle_{H^{-1}(\Omega),H_0^1(\Omega)} = \int_\Omega A\nabla u\nabla(u_-)\,dx + \int_\Omega cuu_-\,dx.$$

Now $\nabla(u_-) = -\mathbf{1}_{u\leq 0}\nabla u = -(\mathbf{1}_{u\leq 0})^2\nabla u$ so that

$$\begin{aligned}A\nabla(u)\nabla(u_-) &= -A\nabla(u)[(\mathbf{1}_{u\leq 0})^2\nabla(u)]\\ &= -[A(\mathbf{1}_{u\leq 0})\nabla(u)][(\mathbf{1}_{u\leq 0})\nabla(u)] = -A\nabla(u_-)\nabla(u_-).\end{aligned}$$

In the same vein, $u_- = -\mathbf{1}_{u\leq 0}u = -(\mathbf{1}_{u\leq 0})^2u$, so that $cuu_- = -c(u_-)^2$. Consequently,

$$\int_\Omega A\nabla(u_-)\nabla(u_-)\,dx + \int_\Omega c(u_-)^2\,dx \leq 0.$$

We use the fact that A is coercive to deduce that

$$\lambda\|\nabla(u_-)\|^2_{L^2(\Omega)} \leq 0,$$

from which the result follows at once. □

Let us note that if $c > 0$ almost everywhere in Ω, then the result still holds when $\lambda = 0$. We now give a weak analogue of Theorem 5.4.

Theorem 5.7. *Let Ω be a bounded open subset of $\mathbb{R}^d$. We are given a $d \times d$ symmetric matrix-valued function A with coefficients $a_{ij} \in L^\infty(\Omega)$ such that $a_{ij}(x)\xi_i\xi_j \geq 0$ for almost all $x \in \Omega$ and all $\xi \in \mathbb{R}^d$, and a scalar function $c \in L^\infty(\Omega)$ such that $c(x) \geq \eta > 0$ almost everywhere. Then any function $u \in H_0^1(\Omega)$ that solves*

$$-\operatorname{div}(A\nabla u) + cu = f \textit{ in the sense of } \mathscr{D}'(\Omega), \tag{5.2}$$

satisfies

$$\|u\|_{L^\infty(\Omega)} \leq \frac{1}{\eta}\|f\|_{L^\infty(\Omega)}. \tag{5.3}$$

Proof. First if all, if $f \notin L^\infty(\Omega)$, there is nothing to prove since the right-hand side is $+\infty$. Let us thus assume that $f \in L^\infty(\Omega)$. Let $k \in \mathbb{R}^+$. As we already know,

$(u-k)_+ \in H_0^1(\Omega)$ and we can take it as a test-function. It follows then from (5.2) that

$$\int_\Omega A\nabla u \cdot \nabla(u-k)_+\,dx + \int_\Omega cu(u-k)_+\,dx = \int_\Omega f(u-k)_+\,dx.$$

Now

$$\begin{aligned}\int_\Omega A\nabla u \cdot \nabla(u-k)_+\,dx &= \int_\Omega A\nabla(u-k)\cdot\nabla(u-k)_+\,dx\\ &= \int_\Omega A\nabla(u-k)_+\cdot\nabla(u-k)_+\,dx \geq 0.\end{aligned}$$

Indeed, $\nabla(u-k)_+ = \mathbf{1}_{u>k}\nabla(u-k)$ and $\mathbf{1}_{u>k} = \mathbf{1}_{u>k}^2$. Replacing this in the previous equality, we obtain

$$\int_\Omega cu(u-k)_+\,dx \leq \int_\Omega f(u-k)_+\,dx.$$

The open set Ω is bounded, so we can subtract $\int_\Omega kc(u-k)_+\,dx$ to both sides of this inequality, which yields

$$\int_\Omega c(u-k)(u-k)_+\,dx \leq \int_\Omega (f-ck)(u-k)_+\,dx.$$

The same kind of positive part trick gives

$$\int_\Omega c(u-k)(u-k)_+\,dx = \int_\Omega c[(u-k)_+]^2\,dx \geq \eta\|(u-k)_+\|^2_{L^2(\Omega)}.$$

Now if we take $k = \|f\|_{L^\infty(\Omega)}/\eta$, then $f - ck \leq 0$ almost everywhere. On the other hand, $(u-k)_+ \geq 0$ almost everywhere, so that $\int_\Omega (f-ck)(u-k)_+\,dx \leq 0$. Therefore $(u-k)_+ = 0$, which means that $u \leq k$ almost everywhere.

We then follow the same argument with $v = (u+k)_-$, and with $k = \|f\|_{L^\infty(\Omega)}/\eta$ again. □

Remark 5.8. There is an analogous result when Ω is regular and $\gamma(u) = g$ for some $g \in H^{1/2}(\partial\Omega)$. In this case, we take $|k| \geq \|g\|_{L^\infty(\partial\Omega)}$. □

5.3 Elliptic Regularity Results

Second order linear elliptic equations have a very important property: roughly speaking, the solutions gain two derivatives compared with the right-hand side of the equation. This is what is called *elliptic regularity*. Elliptic regularity is a wide ranging and very technical theory. We are just going to give the results that are essential for applications, and refer the reader to for instance [36, 38] for proofs.

We can nonetheless try and give an idea of where elliptic regularity comes from by showing two simple examples, involving two different techniques.

Our first example makes use of the Fourier transform. Let $u \in H^1(\mathbb{R}^d)$ be such that $\Delta u \in L^2(\mathbb{R}^d)$. We are going to show that in this case, $u \in H^2(\mathbb{R}^d)$ in fact.[3]

Now, if $u \in H^1(\mathbb{R}^d)$ then $u \in \mathscr{S}'(\mathbb{R}^d)$, which denotes the space of tempered distributions. The space of tempered distributions is a subspace of the space of distributions $\mathscr{D}'(\mathbb{R}^d)$ on $\mathbb{R}^d$, on which the Fourier transform $\mathscr{F}$ is well-defined by transposition of the Fourier transform on the Schwartz space $\mathscr{S}(\mathbb{R}^d)$, and is an isomorphism. Moreover, the following formula holds $\mathscr{F}(\partial_k u) = i\xi_k\hat{u}$ (here $i^2 = -1$), see [15, 39, 68]. Consequently, $\mathscr{F}(\Delta u)(\xi) = -\xi_k\xi_k\hat{u} = -|\xi|^2\hat{u} \in L^2(\mathbb{R}^d)$ since the Fourier transform is also an isomorphism on $L^2(\mathbb{R}^d)$. Likewise, $\mathscr{F}(\partial_{kl}u)(\xi) = -\xi_k\xi_l\hat{u}(\xi)$, in the sense of tempered distributions and therefore[4]

$$\begin{aligned}|\mathscr{F}(\partial_{kl}u)(\xi)|^2 &= \xi_k^2\xi_l^2|\hat{u}(\xi)|^2 \\ &\leq \Bigl(\sum_{m=1}^{d}\xi_m^2\Bigr)\Bigl(\sum_{n=1}^{d}\xi_n^2\Bigr)|\hat{u}(\xi)|^2 = |\xi|^4|\hat{u}(\xi)|^2 \in L^1(\mathbb{R}^d).\end{aligned}$$

It follows that $\mathscr{F}(\partial_{kl}u) \in L^2(\mathbb{R}^d)$ and by the L^2 isomorphism property, that the elliptic regularity $\partial_{kl}u \in L^2(\mathbb{R}^d)$ holds true.

In addition to being an isomorphism, the Fourier transform is also an isometry on $L^2(\mathbb{R}^d)$ (possibly up to a factor $(2\pi)^{-d/2}$ according to the definition used). We thus see that

$$\|\partial_{kl}u\|_{L^2(\mathbb{R}^d)} \leq \|\Delta u\|_{L^2(\mathbb{R}^d)},$$

hence the estimate (there are d^2 second derivatives)

$$\|u\|_{H^2(\mathbb{R}^d)} \leq \bigl(\|u\|^2_{H^1(\mathbb{R}^d)} + d^2\|\Delta u\|^2_{L^2(\mathbb{R}^d)}\bigr)^{1/2}.$$

The second example makes use of what is known as the *Nirenberg translations method*. Let us consider the problem: Find $u \in H^1(\mathbb{R}^d)$ such that $-\Delta u + u = f$ with $f \in H^{-1}(\mathbb{R}^d)$ given. This is a trivially variational problem:

$$\forall v \in H^1(\mathbb{R}^d), \int_{\mathbb{R}^d}(\nabla u \cdot \nabla v + uv)\,dx = \langle f, v\rangle,$$

[3] This is very surprising: we only know that a very specific linear combination of second order derivatives is square integrable, and it turns out that all individual second order derivatives are square integrable.

[4] By hypothesis, $u \in L^2(\mathbb{R}^d)$ so that we already know that $\hat{u}$ is an L^2-function and that writing $\hat{u}(\xi)$ is licit.

with the equally trivial variational estimate

$$\|u\|_{H^1(\mathbb{R}^d)} \leq \|f\|_{H^{-1}(\mathbb{R}^d)}.$$

Let us now assume that furthermore, $f \in L^2(\mathbb{R}^d)$. We are going to show that then $u \in H^2(\mathbb{R}^d)$. Let us choose an index i and for all $h \neq 0$, define the translates $u_{i,h}(x) = u(x + he_i)$ and $f_{i,h}(x) = f(x + he_i)$. Obviously, there holds $-\Delta u_{i,h} + u_{i,h} = f_{i,h}$ with $u_{i,h} \in H^1(\mathbb{R}^d)$, so that, subtracting the two variational forms and dividing by h, we obtain

$$\forall v \in H^1(\mathbb{R}^d), \int_{\mathbb{R}^d} \Big(\nabla\Big(\frac{u_{i,h} - u}{h}\Big) \cdot \nabla v + \frac{u_{i,h} - u}{h} v\Big)\, dx = \Big\langle \frac{f_{i,h} - f}{h}, v\Big\rangle.$$

Of course, $\frac{u_{i,h}-u}{h} \in H^1(\mathbb{R}^d)$, and the variational estimate applies too

$$\Big\|\frac{u_{i,h} - u}{h}\Big\|_{H^1(\mathbb{R}^d)} \leq \Big\|\frac{f_{i,h} - f}{h}\Big\|_{H^{-1}(\mathbb{R}^d)}. \tag{5.4}$$

Let us show that the fact that $f \in L^2(\mathbb{R}^d)$ implies that $\frac{f_{i,h}-f}{h}$ is bounded in $H^{-1}(\mathbb{R}^d)$ uniformly with respect to h. Indeed, for any test function $v \in H_0^1(\mathbb{R}^d)$, there holds

$$\begin{aligned}
\Big\langle \frac{f_{i,h} - f}{h}, v\Big\rangle &= \int_{\mathbb{R}^d} \frac{f_{i,h}(x) - f(x)}{h} v(x)\, dx \\
&= \frac{1}{h}\Big(\int_{\mathbb{R}^d} f_{i,h}(x)v(x)\, dx - \int_{\mathbb{R}^d} f(x)v(x)\, dx\Big) \\
&= \frac{1}{h}\Big(\int_{\mathbb{R}^d} f(x)v_{i,-h}(x)\, dx - \int_{\mathbb{R}^d} f(x)v(x)\, dx\Big) \\
&= \Big\langle f, \frac{v_{i,-h} - v}{h}\Big\rangle,
\end{aligned} \tag{5.5}$$

by the change of variable formula and the fact that all occurring distributions are actually functions.[5] Let us now prove that

$$\Big\|\frac{v_{i,-h} - v}{h}\Big\|_{L^2(\mathbb{R}^d)} \leq \|\partial_i v\|_{L^2(\mathbb{R}^d)}. \tag{5.6}$$

[5] This is a sort of discrete integration by parts.

We start with the case $v = \varphi \in \mathscr{D}(\mathbb{R}^d)$. Let us pick $x \in \mathbb{R}^d$. Setting $g(t) = \varphi(x - the_i)$ for $t \in \mathbb{R}$, we may write

$$\begin{aligned}\varphi_{i,-h}(x) - \varphi(x) &= g(1) - g(0)\\ &= \int_0^1 g'(t)\,dt\\ &= -h\int_0^1 \partial_i\varphi(x - the_i)\,dt.\end{aligned}$$

It follows that

$$\begin{aligned}\Big\|\frac{\varphi_{i,-h} - \varphi}{h}\Big\|^2_{L^2(\mathbb{R}^d)} &= \int_{\mathbb{R}^d}\Big(\int_0^1 \partial_i\varphi(x - the_i)\,dt\Big)^2\,dx\\ &\le \int_{\mathbb{R}^d}\int_0^1 \big(\partial_i\varphi(x - the_i)\big)^2\,dtdx\\ &= \int_0^1\int_{\mathbb{R}^d} \big(\partial_i\varphi(x - the_i)\big)^2\,dxdt\\ &= \|\partial_i\varphi\|^2_{L^2(\mathbb{R}^d)},\end{aligned}$$

by the Cauchy-Schwarz inequality, then by Fubini's theorem. Estimate (5.6) is then a direct consequence of the fact that $\mathscr{D}(\mathbb{R}^d)$ is dense in $H^1(\mathbb{R}^d)$.

Let us now use estimate (5.6) in equality (5.5). We thus obtain

$$\Big\|\frac{f_{i,h} - f}{h}\Big\|_{H^{-1}(\mathbb{R}^d)} \le \|f\|_{L^2(\mathbb{R}^d)},$$

once more by the Cauchy-Schwarz inequality and by the definition of the dual norm.

Due to the variational estimate (5.4), it follows that $\frac{u_{i,h}-u}{h}$ is bounded in $H^1(\mathbb{R}^d)$ independently of h. In particular, for all j, $\frac{\partial_j u_{i,h}-\partial_j u}{h}$ is bounded in $L^2(\mathbb{R}^d)$.

Now, on the other hand, $\frac{\partial_j u_{i,h}-\partial_j u}{h} \to \partial_{ij}u$ in the sense of $\mathscr{D}'(\mathbb{R}^d)$ when $h \to 0$. Indeed, for all $\varphi \in \mathscr{D}(\mathbb{R}^d)$, we see as in the proof of equality (5.5) that

$$\Big\langle\frac{\partial_j u_{i,h} - \partial_j u}{h}, \varphi\Big\rangle = \Big\langle\partial_j u, \frac{\varphi_{i,-h} - \varphi}{h}\Big\rangle.$$

By the fundamental theorem of calculus, we can write for all x and h

$$\begin{aligned}\Big(\frac{\varphi_{i,-h} - \varphi}{h}\Big)(x) &= -\frac{1}{h}\int_0^h \partial_i\varphi(x - se_i)\,ds\\ &= -\partial_i\varphi(x) - \frac{1}{h}\int_0^h \big(\partial_i\varphi(x - se_i) - \partial_i\varphi(x)\big)\,ds\\ &= -\partial_i\varphi(x) - \frac{1}{h}\int_0^h\int_0^s \partial_{ii}\varphi(x - ste_i)\,dtds.\end{aligned}$$

It follows that

$$\left|\Big(\frac{\varphi_{i,-h}-\varphi}{h}\Big)(x)+\partial_i\varphi(x)\right|\le\frac{h}{2}\max_{\mathbb{R}^d}|\partial_{ii}\varphi|,$$

for all x, so that $\frac{\varphi_{i,-h}-\varphi}{h}\to-\partial_i\varphi$ uniformly when $h\to 0$. Additionally, for say $0<|h|\le 1$, the supports of the functions $\frac{\varphi_{i,-h}-\varphi}{h}$ are included in a compact set. Since $\partial_j u\in L^2(\mathbb{R}^d)\subset L^1_{\text{loc}}(\mathbb{R}^d)$, it follows that

$$\Big\langle\partial_j u,\frac{\varphi_{i,-h}-\varphi}{h}\Big\rangle=\int_{\mathbb{R}^d}\partial_j u(x)\Big(\frac{\varphi_{i,-h}-\varphi}{h}\Big)(x)\,dx\to-\langle\partial_j u,\partial_i\varphi\rangle=\langle\partial_{ij}u,\varphi\rangle,$$

when $h\to 0$, which establishes the announced distributional convergence.

Now, a sequence of distributions T_n which is bounded in $L^2(\mathbb{R}^d)$ and converges toward a distribution T in the sense of $\mathscr{D}'(\mathbb{R}^d)$, is such that the limit T is also in $L^2(\mathbb{R}^d)$. We can for instance extract a subsequence that converges weakly in $L^2(\mathbb{R}^d)$, thus also in the sense of $\mathscr{D}'(\mathbb{R}^d)$, and conclude by the uniqueness of distributional limits. It follows from all this that $\partial_{ij}u\in L^2(\mathbb{R}^d)$ with the estimate

$$\|u\|_{H^2(\mathbb{R}^d)}\le\sqrt{1+d}\|f\|_{L^2(\mathbb{R}^d)},$$

due to the fact that $\|f\|_{H^{-1}(\mathbb{R}^d)}\le\|f\|_{L^2(\mathbb{R}^d)}$ and $\|\partial_i f\|_{H^{-1}(\mathbb{R}^d)}\le\|f\|_{L^2(\mathbb{R}^d)}$. We also deduce from the previous arguments that $\partial_i u$ is *in fine* the unique solution belonging to $H^1(\mathbb{R}^d)$ of the equation $-\Delta(\partial_i u)+\partial_i u=\partial_i f$. □

Remark 5.9. The fact that $-\Delta(\partial_i u)+\partial_i u=\partial_i f$ holds true in the sense of distributions is trivial. Even though the right-hand side belongs to $H^{-1}(\mathbb{R}^d)$ when $f\in L^2(\mathbb{R}^d)$, we could not use the variational formulation and estimate for this equation to establish elliptic regularity, because we do not a priori know that $\partial_i u\in H^1(\mathbb{R}^d)$. Indeed, this is actually the conclusion of elliptic regularity in this context. On the other hand, the translations method uses differential quotients in place of derivatives, and differential quotients do not suffer from this drawback, thus making the method ultimately successful. □

As their name makes it clear, boundary value problems tend to involve boundary values and the open set $\mathbb{R}^d$ is singularly devoid of any boundary. It so happens that a large part of the technicality of elliptic regularity theory stems from having to deal with boundaries of open sets and boundary conditions. Let us take a quick glance at this in the context of Nirenberg translations in a half-space with a homogeneous Dirichlet boundary condition.

We thus consider $\mathbb{R}^d_+=\{x\in\mathbb{R}^d; x_d>0\}$ and the problem: Find $u\in H^1_0(\mathbb{R}^d_+)$ such that $-\Delta u+u=f$ with $f\in H^{-1}(\mathbb{R}^d_+)$. This problem is just as variational as the previous one, with the same variational estimate.

It not too hard to see that the translations method still works with all the translations that leave $\mathbb{R}^d_+$ invariant, that is to say for all $i<d$. Indeed, in this

case $\frac{u_{i,h}-u}{h} \in H_0^1(\mathbb{R}^d_+)$. It follows that $\partial_i u \in H_0^1(\mathbb{R}^d_+)$, thus $\partial_{ij} u \in L^2(\mathbb{R}^d_+)$ for all $i < d$ and all j. There is only one second derivative missing, namely $\partial_{dd} u$, which we recover via the equation itself

$$\partial_{dd} u = -\sum_{i<d} \partial_{ii} u + u - f \in L^2(\mathbb{R}^d_+),$$

since we just saw that all second derivatives in the right-hand side belong to $L^2(\mathbb{R}^d_+)$. The upshot of all this is that $u \in H^2(\mathbb{R}^d_+)$ with an estimate of its $H^2(\mathbb{R}^d_+)$ norm in terms of that of f in $L^2(\mathbb{R}^d_+)$. See [11] for more details on the Nirenberg translations method in a more general situation. □

We now start a catalogue of elliptic regularity results in greater generality, without proofs. In the sequel, we are given a second order differential operator $L = -a_{ij}\partial_{ij} + b_i\partial_i + c$ with $c \geq 0$, which is strictly elliptic in the sense that there exists a constant $\lambda > 0$ with $a_{ij}(x)\xi_i\xi_j \geq \lambda|\xi|^2$ for all $x \in \overline{\Omega}$ and all $\xi \in \mathbb{R}^d$.

Let us begin with Hölder regularity, which stems from the Schauder estimates.

Theorem 5.8 (Schauder Estimates). *Let $0 < \alpha < 1$ and Ω be a bounded open subset of $\mathbb{R}^d$ of class $C^{2,\alpha}$. Assume that the coefficients of L, a_{ij}, b_i and c, belong to $C^{0,\alpha}(\overline{\Omega})$ and let Λ be an upper bound for their norms in this space. Given $f \in C^{0,\alpha}(\overline{\Omega})$ and $g \in C^{2,\alpha}(\overline{\Omega})$, let $u \in C^0(\overline{\Omega}) \cap C^2(\Omega)$ be a function such that*

$$Lu = f \text{ in } \Omega, \quad u = g \text{ on } \partial\Omega.$$

Then u belongs to $C^{2,\alpha}(\overline{\Omega})$ with the estimate

$$\|u\|_{C^{2,\alpha}(\overline{\Omega})} \leq C(\|f\|_{C^{0,\alpha}(\overline{\Omega})} + \|g\|_{C^{2,\alpha}(\overline{\Omega})}), \tag{5.7}$$

where C only depends on d, α, λ, Λ and Ω.

Remark 5.10. i) More generally, in the Schauder estimates, there is no sign hypothesis made on the zero order coefficient c. In this case, a term $\|u\|_{C^0(\overline{\Omega})}$ must be added to the right-hand side of (5.7). When $c \geq 0$, this term becomes redundant due to Theorem 5.5.

ii) Let us once more emphasize how a priori surprising elliptic regularity is. Let us take the simple case of $L = -\Delta$ with $g = 0$. The sole information that Δu belongs to a certain $C^{0,\alpha}(\overline{\Omega})$, that is to say that a certain, very specific, linear combination of second derivatives of u is Hölder continuous, is enough to ensure that all individual second derivatives have the same Hölder continuity.[6] This includes not only the derivatives that appear in L, albeit in a sum, but also all the other mixed derivatives that do not appear in L. So in a way, even though it could be thought that passing from u to Lu entails a loss of information on these individual derivatives, somehow,

[6] Under the assumption that u and Ω also have some minimal regularity in this context.

part of this information remains hidden in Lu. This is thus an extremely strong and profound property of elliptic operators.

iii) It should be noted that *elliptic regularity fails for* $\alpha = 0$ *and* $\alpha = 1$. Thus for example, there exists a function u such that $\Delta u \in C^0(\overline{\Omega})$ but $u \notin C^2(\overline{\Omega})$, see the exercises of this chapter. Hölder continuity is very important in this respect.

iv) As a general rule, elliptic regularity results are of a local nature. For instance, if ω is an open set compactly included in Ω, and if the restriction of f to ω is of class $C^{0,\alpha}$, then the restriction of u to ω is of class $C^{2,\alpha}$. □

The estimate result comes along with a companion existence and uniqueness result.

Theorem 5.9. *Under the same hypotheses as before on the open set and on the operator L, for all* $f \in C^{0,\alpha}(\overline{\Omega})$ *and* $g \in C^{2,\alpha}(\overline{\Omega})$*, there exists a unique function* $u \in C^0(\overline{\Omega}) \cap C^2(\Omega)$ *such that*

$$Lu = f \text{ in } \Omega, \quad u = g \text{ on } \partial\Omega.$$

Remark 5.11. Theorem 5.9 shows that the operator L^{-1} defines an isomorphism between $C^{0,\alpha}(\overline{\Omega}) \times (C^{2,\alpha}(\overline{\Omega})/C_0^{2,\alpha}(\overline{\Omega}))$ and $C^{2,\alpha}(\overline{\Omega})$. □

There also are higher order regularity results.

Theorem 5.10. *Let* $0 < \alpha < 1$*, k a natural number, and* Ω *a bounded open set of class* $C^{k+2,\alpha}$*. Assume that the coefficients of L,* a_{ij}*,* b_i *and c, belong to* $C^{k,\alpha}(\overline{\Omega})$ *and let* Λ *be an upper bound for their norms in this space. Given* $f \in C^{k,\alpha}(\overline{\Omega})$ *and* $g \in C^{k+2,\alpha}(\overline{\Omega})$*, let* $u \in C^0(\overline{\Omega}) \cap C^2(\Omega)$ *be a function such that*

$$Lu = f \text{ in } \Omega, \quad u = g \text{ on } \partial\Omega.$$

Then $u \in C^{k+2,\alpha}(\overline{\Omega})$ *with the estimate*

$$\|u\|_{C^{k+2,\alpha}(\overline{\Omega})} \le C(\|f\|_{C^{k,\alpha}(\overline{\Omega})} + \|g\|_{C^{k+2,\alpha}(\overline{\Omega})}), \tag{5.8}$$

where C only depends on d, α*,* λ*,* Λ*, k and* Ω*.*

In other words, the solution of a second order linear elliptic equation gains two derivatives compared to the data f. This implies that if the open set, the coefficients of the differential operator, and the data are of class C^∞, then the solution is C^∞ as well.

There is an analogous theory in Sobolev spaces. We must distinguish between weak solutions, i.e., in the sense of distributions, and strong solutions, meaning that satisfy the equation almost everywhere. The minimal regularity hypotheses that must be made on the coefficients of the operators are not the same in the two cases.

In the case of weak solutions, we consider the differential operator in the so-called divergence form, $Lu = -\text{div}\,(A\nabla u) + cu = -\partial_i(a_{ij}\partial_j u) + cu$, with

$a_{ij}, c \in L^\infty(\Omega)$, A coercive and $c \geq 0$.[7] If the coefficients of A are smooth, then the divergence form reduces to the form used for the Schauder estimates.

Theorem 5.11. *Let Ω be a bounded open set of class C^2 and let us assume that the coefficients of A belong to $C^{0,1}(\overline{\Omega})$. Given $f \in L^2(\Omega)$ and $g \in H^2(\Omega)$, let $u \in H^1(\Omega)$ be a function such that*

$$Lu = f \text{ in the sense of } \mathscr{D}'(\Omega), \quad u - g \in H_0^1(\Omega).$$

Then $u \in H^2(\Omega)$ with the estimate

$$\|u\|_{H^2(\Omega)} \leq C(\|f\|_{L^2(\Omega)} + \|g\|_{H^2(\Omega)}), \tag{5.9}$$

where C does not depend on f and g. Moreover, the equation is also satisfied almost everywhere in the form

$$-a_{ij}\partial_{ij}u - \partial_i a_{ij}\partial_j u + cu = f.$$

Remark 5.12. i) The existence and uniqueness of u follow here directly from the Lax-Milgram theorem (we do not assume A to be symmetric).

ii) If $a_{ij} \in C^{0,1}(\overline{\Omega})$, then $a_{ij} \in W^{1,\infty}(\Omega)$, see [11], and therefore $\partial_i a_{ij}\partial_j u \in L^2(\Omega)$. This term thus makes sense and we can apply the Leibniz formula to differentiate the product $a_{ij}\partial_j u$.

iii) The assumed regularity of the open set is not a necessary condition for the result to hold. Thus, in two dimensions and when Ω is a convex polygon, then $f \in L^2(\Omega)$ and $g = 0$ imply $u = (-\Delta)^{-1}f \in H^2(\Omega)$. On the other hand, if Ω is a polygon with a reentrant angle, then there exists data f in $L^2(\Omega)$ such that $u \notin H^2(\Omega)$, see [38]. □

More generally, higher order regularity also holds true, under appropriate regularity hypotheses.

Theorem 5.12. *Let $k \geq 1$, Ω a bounded open set of class C^{k+2}, $a_{ij} \in C^{k,1}(\overline{\Omega})$ and $c \in C^{k-1,1}(\overline{\Omega})$. Given $f \in H^k(\Omega)$ and $g \in H^{k+2}(\Omega)$, if $u \in H^1(\Omega)$ is a function such that*

$$Lu = f \text{ in the sense of } \mathscr{D}'(\Omega), \quad u - g \in H_0^1(\Omega),$$

then $u \in H^{k+2}(\Omega)$ with the estimate

$$\|u\|_{H^{k+2}(\Omega)} \leq C_k(\|f\|_{H^k(\Omega)} + \|g\|_{H^{k+2}(\Omega)}), \tag{5.10}$$

where C_k does not depend on f and g.

[7] We can also add first order terms.

We recover the fact that if the open set, the coefficients and the data are of class C^∞, then the solution is also of class C^∞.

We now consider strong solutions, i.e., functions $u \in W^{2,p}(\Omega)$ such that $Lu = -a_{ij}\partial_{ij}u + b_i\partial_i u + cu = f$ almost everywhere, for which there is also an existence and regularity theory. Note that if the coefficients are for instance bounded, then all the terms in the differential operator make sense as L^p functions (assuming Ω bounded as always).

Theorem 5.13. *Let Ω be a bounded open set of class $C^{1,1}$, $a_{ij} \in C^0(\overline{\Omega})$ and $b_i, c \in L^\infty(\Omega)$. For all $f \in L^p(\Omega)$ and $g \in W^{2,p}(\Omega)$ with $1 < p < +\infty$, there exists a unique function $u \in W^{2,p}(\Omega)$ such that*

$$Lu = f \text{ almost everywhere in } \Omega, \quad u - g \in W_0^{1,p}(\Omega),$$

with the estimate

$$\|u\|_{W^{2,p}(\Omega)} \le C_p(\|f\|_{L^p(\Omega)} + \|g\|_{W^{2,p}(\Omega)}), \tag{5.11}$$

where C_p does not depend on f and g.

Remark 5.13. The result fails for $p = 1$ and $p = +\infty$. □

The situation is similar for higher order derivatives.

Theorem 5.14. *Let $k \ge 1$, Ω a bounded open set of class $C^{k+1,1}$, and $a_{ij}, b_i, c \in C^{k-1,1}(\overline{\Omega})$. For all $f \in W^{k,p}(\Omega)$ and $g \in W^{k+2,p}(\Omega)$ with $1 < p < +\infty$, there exists a unique function $u \in W^{k+2,p}(\Omega)$ such that*

$$Lu = f \text{ almost everywhere in } \Omega, \quad u - g \in W_0^{1,p}(\Omega),$$

with the estimate

$$\|u\|_{W^{k+2,p}(\Omega)} \le C_{k,p}(\|f\|_{W^{k,p}(\Omega)} + \|g\|_{W^{k+2,p}(\Omega)}), \tag{5.12}$$

where $C_{k,p}$ does not depend on f and g.

This kind of elliptic regularity results admit considerable generalizations to systems, see [3, 4, 33, 49]. They are also not limited to just Dirichlet boundary conditions, but extend to more complicated boundary operators as well. There must be some kind of compatibility between on the one hand, the differential operator inside the open set and the differential operator on the boundary on the other hand.

Let us also note that global elliptic regularity does not hold in the case of mixed conditions, even when everything else is smooth. Such is the case of a Dirichlet boundary condition on one part of the boundary and a Neumann boundary condition on another part of the boundary, except when these parts have disjoint closures, see [38]. Singularities may develop at the interface between the two boundary regions.

To complete this brief catalogue of regularity results, let us mention two theorems that are useful in slightly different contexts. The first one is a theorem of De Giorgi.

Theorem 5.15. *Let Ω be a bounded regular open subset of $\mathbb{R}^d$, A a coercive matrix with $L^\infty(\Omega)$ coefficients, $f_0 \in L^{d/2+\varepsilon}(\Omega)$, $f \in L^{d+\varepsilon}(\Omega;\mathbb{R}^d)$, $\varepsilon > 0$, and let $u \in H_0^1(\Omega)$ be such that $-\mathrm{div}\,(A\nabla u) = f_0 + \mathrm{div}\, f$ in the sense of distributions. Then there exists $0 < \alpha < 1$ and $C > 0$ such that $u \in C^{0,\alpha}(\overline{\Omega})$ and*

$$\|u\|_{C^{0,\alpha}(\overline{\Omega})} \leq C(\|f_0\|_{L^{d/2+\varepsilon}(\Omega)} + \|f\|_{L^{d+\varepsilon}(\Omega;\mathbb{R}^d)}).$$

Remark 5.14. i) De Giorgi's theorem is easily shown in the case when the coefficients are regular, based on the L^p regularity results and the Sobolev embeddings. The difficulty stems from the fact that the coefficients are not assumed to be continuous.

ii) The theorem is not true for systems. In this case, a typical result that can be shown is that $u \in C^{0,\alpha}(\overline{\Omega} \setminus H)$, where H is a "small" set, with smallness here understood in the sense of its Hausdorff dimension, see [26, 35, 37]. □

The second result is a theorem of Meyers, see [52].

Theorem 5.16. *Let Ω be a bounded regular open subset of $\mathbb{R}^d$, A a coercive matrix with $L^\infty(\Omega)$ coefficients. Then, there exists $2 < p_0 < +\infty$ such that the operator $u \mapsto -\mathrm{div}\,(A\nabla u)$ is an isomorphism between $W_0^{1,p}(\Omega)$ and $W^{-1,p}(\Omega)$ for all $p_0' \leq p \leq p_0$.*

Remark 5.15. The Lax-Milgram theorem implies that this operator is an isomorphism between $H_0^1(\Omega)$ and $H^{-1}(\Omega)$. The Meyers theorem thus makes it possible to gain a little bit of integrability, even with discontinuous coefficients. It remains valid for systems. □

5.4 The Method of Super- and Sub-Solutions

The maximum principle and elliptic regularity can be combined to solve certain nonlinear equations. We develop here the so-called method of super- and sub-solutions. The problem is as follows. Let $L = -a_{ij}\partial_{ij} + b_i\partial_i + c$ be an elliptic operator with $C^{0,\alpha}(\overline{\Omega})$ coefficients, where Ω is a bounded open subset of $\mathbb{R}^d$ of class $C^{2,\alpha}$. Let us be given a locally Lipschitz function $f : \overline{\Omega} \times \mathbb{R} \to \mathbb{R}$. We would like to solve the boundary value problem of finding $u \in C^0(\overline{\Omega}) \cap C^2(\Omega)$ such that

$$\begin{cases} Lu(x) = f(x, u(x)) \text{ in } \Omega, \\ u(x) = 0 \text{ on } \partial\Omega. \end{cases} \tag{5.13}$$

Definition 5.1. We say that $\overline{u}$ (resp. $\underline{u}$) is a super-solution (resp. sub-solution) if $\overline{u}$ (resp. $\underline{u}$) belongs to $C^0(\overline{\Omega}) \cap C^2(\Omega)$ and verifies $L\overline{u} \geq f(x, \overline{u})$ (resp. $L\underline{u} \leq f(x, \underline{u})$) in Ω and $\overline{u} \geq 0$ (resp. $\underline{u} \leq 0$) on $\partial\Omega$.

We are going to show the following result.

Theorem 5.17. *Let us assume that there exists a super-solution $\overline{u}$ and a sub-solution $\underline{u}$ that are such that $\overline{u} \geq \underline{u}$. Then problem* (5.13) *admits a maximal solution $\overline{u}^*$ and a minimal solution $\underline{u}_*$ such that $\overline{u} \geq \overline{u}^* \geq \underline{u}_* \geq \underline{u}$ and that there exists no solution u between $\overline{u}$ and $\underline{u}$ such that $u(x) > \overline{u}^*(x)$ or $u(x) < \underline{u}_*(x)$ at one point x of Ω.*

The end of the statement is actually an explanation of the meaning of maximal and minimal in this context. The proof follows the same pattern as a fixed point proof, by iterating an operator. The maximum principle and elliptic regularity are used to show that these iterations are well defined and converge.

Lemma 5.4. *There exists a constant $\mu > 0$ such that the following relations*

$$\begin{cases} \overline{u}^0 = \overline{u}, \\ L\overline{u}^{n+1}(x) + \mu\overline{u}^{n+1}(x) = f(x, \overline{u}^n(x)) + \mu\overline{u}^n(x) \text{ in } \Omega, \\ \overline{u}^{n+1}(x) = 0 \text{ on } \partial\Omega, \end{cases}$$

and

$$\begin{cases} \underline{u}^0 = \underline{u}, \\ L\underline{u}^{n+1}(x) + \mu\underline{u}^{n+1}(x) = f(x, \underline{u}^n(x)) + \mu\underline{u}^n(x) \text{ in } \Omega, \\ \underline{u}^{n+1}(x) = 0 \text{ on } \partial\Omega, \end{cases}$$

define two sequences $\overline{u}^n$ and $\underline{u}^n$ of functions on $\overline{\Omega}$ that converge pointwise.

Proof. Let us first note that if f is a locally Lipschitz mapping from a metric space (X, d) to a metric space (Y, δ), then it is globally Lipschitz on any compact subset K of X. Indeed, the sets $U_t = \{(y, y') \in K \times K, \delta(f(y), f(y')) < td(y, y')\}$, $t \in \mathbb{R}_+^*$, are open and cover the compact $K \times K$. We extract from this family a finite subcover and adopt as a global Lipschitz constant on K the largest of the numbers t thus retained.

Let $M = \max\{\|\overline{u}\|_{C^0(\overline{\Omega})}, \|\underline{u}\|_{C^0(\overline{\Omega})}\}$. We apply the previous remark to the compact set $K = \overline{\Omega} \times [-M, M]$. Therefore, there exists a constant λ such that

$$|f(x, s) - f(x', s')| \leq \lambda(|x - x'| + |s - s'|)$$

for all (x, s) and (x', s') in K. We set $\mu = \lambda + 1$.

Let us then show that the function $f_\mu(x,s) = f(x,s) + \mu s$ is nondecreasing with respect to s for (x,s) in K. Let us thus take $s, s' \in [-M, M]$ with $s \geq s'$. Since $|f(x,s) - f(x,s')| \leq \lambda |s - s'|$, it follows that

$$f_\mu(x,s) - f_\mu(x,s') \geq (\mu - \lambda)(s - s') = s - s' \geq 0.$$

At this stage, it worth noticing that if $v \in C^{0,\alpha}(\overline{\Omega}; [-M, M])$ (meaning that v takes its values in $[-M, M]$), then the function $x \mapsto f_\mu(x, v(x))$ also belongs to $C^{0,\alpha}(\overline{\Omega})$. Indeed, we can write

$$\begin{aligned}|f_\mu(x, v(x)) - f_\mu(y, v(y))| &\leq \lambda \big| |x-y| + |v(x) - v(y)| \big| + \mu |v(x) - v(y)| \\ &\leq C(|x-y| + |x-y|^\alpha),\end{aligned}$$

where C depends on λ, μ and $\|v\|_{C^{0,\alpha}(\overline{\Omega})}$. Consequently, we see that for $x \neq y$,

$$\frac{|f_\mu(x, v(x)) - f_\mu(y, v(y))|}{|x-y|^\alpha} \leq C(|x-y|^{1-\alpha} + 1),$$

and the right-hand side is bounded on $\overline{\Omega}$ since Ω is itself bounded.

We set $L_\mu u = Lu + \mu u$. The boundary value problem

$$\begin{cases} L_\mu(T(v))(x) = f_\mu(x, v(x)) & \text{in } \Omega, \\ \quad T(v)(x) = 0 \quad \text{on } \partial\Omega, \end{cases}$$

then defines a mapping $T\colon C^{0,\alpha}(\overline{\Omega}; [-M, M]) \to C^{0,\alpha}(\overline{\Omega})$ for all $0 < \alpha < 1$, by the previous remark and the Hölder space existence and uniqueness theory.

Moreover, if $\underline{u} \leq v \leq \overline{u}$, then $\underline{u} \leq T(v) \leq \overline{u}$. Indeed,

$$\begin{cases} L_\mu(T(v) - \overline{u})(x) = f_\mu(x, v(x)) - L_\mu \overline{u}(x) \leq f_\mu(x, v(x)) - f_\mu(x, \overline{u}(x)) \leq 0 \text{in } \Omega, \\ \quad T(v)(x) - \overline{u}(x) = -\overline{u}(x) \leq 0 \quad \text{on } \partial\Omega, \end{cases}$$

since f_μ is nondecreasing with respect to its second argument on the compact set K. By the strong maximum principle, it follows that $T(v) \leq \overline{u}$. We likewise show that $\underline{u} \leq T(v)$. Consequently, the set $\{v \in C^{0,\alpha}(\overline{\Omega}; [-M, M]); \underline{u} \leq v \leq \overline{u}\}$ is stable under T, by definition of M. In the sequel, we pick a value for $\alpha \in]0, 1[$.

It follows from the previous remarks that there exists a well defined sequence $\overline{u}^n$ satisfying the recursion formula

$$\begin{cases} \quad \overline{u}^0 = \overline{u}, \\ \overline{u}^{n+1} = T(\overline{u}^n). \end{cases}$$

The only possible difficulty is that we do not assume $\overline{u}$ to belong to $C^{0,\alpha}(\overline{\Omega})$, but only to $C^0(\overline{\Omega})$. Therefore, $f_\mu(\cdot, \overline{u}(\cdot)) \in C^0(\overline{\Omega}) \subset L^p(\Omega)$ for all p. Choosing p such that $d < p < +\infty$, we obtain that $\overline{u}^1 \in W^{2,p}(\Omega) \subset C^1(\overline{\Omega}) \subset C^{0,\alpha}(\overline{\Omega})$. From then on, $\overline{u}^n \in C^{0,\alpha}(\overline{\Omega})$ by the Schauder estimates. We have moreover established that $\underline{u} \leq \overline{u}^n \leq \overline{u}$ for all n.

Let us show that the sequence $\overline{u}^n$ is nonincreasing. We argue by induction. First of all, $\overline{u}^1 \leq \overline{u}^0 = \overline{u}$ according to what was said above. Let us now assume that $\overline{u}^n \leq \overline{u}^{n-1}$. By the same computations as before, there holds

$$\begin{cases} L_\mu(\overline{u}^{n+1} - \overline{u}^n)(x) = f_\mu(x, \overline{u}^n(x)) - f_\mu(x, \overline{u}^{n-1}(x)) \leq 0 & \text{in } \Omega, \\ (\overline{u}^{n+1} - \overline{u}^n)(x) = 0 \quad \text{on } \partial\Omega, \end{cases}$$

since $f_\mu(x, \cdot)$ is nondecreasing, thus by the strong maximum principle, $\overline{u}^{n+1} \leq \overline{u}^n$.

We have shown that for each $x \in \Omega$, the real-valued sequence $\overline{u}^n(x)$ is nonincreasing and is bounded below by $\underline{u}(x)$. This sequence is therefore convergent.

We likewise show that the sequence $\underline{u}^n(x)$ is well defined, nondecreasing and bounded above, thus convergent for all x. □

Let us call $\overline{u}^*$ and $\underline{u}_*$ the respective pointwise limits of the sequences $\overline{u}^n$ and $\underline{u}^n$. At this stage, it is worth noticing that we have no information whatsoever on the regularity of these limits. In particular, we have so far no way of passing to the limit in the differential operator.

Lemma 5.5. *There holds $\underline{u}_* \leq \overline{u}^*$.*

Proof. Pick any natural number n. We show that $\underline{u}^l \leq \overline{u}^n$ for all l by induction, exactly as in the previous proof using the maximum principle. We then pass to the pointwise limit when $l \to +\infty$, which yields $\underline{u}_* \leq \overline{u}^n$, then pass again to the pointwise limit when $n \to +\infty$ to obtain the lemma. □

Let us now obtain some regularity.

Lemma 5.6. *The limits $\underline{u}_*$ and $\overline{u}^*$ belong to $C^2(\overline{\Omega})$, and the sequences $\underline{u}_n$ and $\overline{u}^n$ converge in $C^2(\overline{\Omega})$.*

Proof. We use elliptic regularity estimates. Since $\underline{u} \leq \overline{u}^n \leq \overline{u}$ and f_μ is nondecreasing with respect to its second argument, we see that $f_\mu(x, \underline{u}(x)) \leq f_\mu(x, \overline{u}^n(x)) \leq f_\mu(x, \overline{u}(x))$ for all $x \in \overline{\Omega}$. Consequently,

$$\|f_\mu(\cdot, \overline{u}^n(\cdot))\|_{C^0(\overline{\Omega})} \leq C = \max\{\|f_\mu(\cdot, \underline{u}(\cdot))\|_{C^0(\overline{\Omega})}, \|f_\mu(\cdot, \overline{u}(\cdot))\|_{C^0(\overline{\Omega})}\}.$$

Here and in the sequel, the actual value of the generic constant C may change from line to line, but never depends on n.

Since Ω is bounded, it follows that the right-hand side of the equation that defines $\overline{u}^n$ is bounded in $L^p(\Omega)$ for any p. According to the $W^{2,p}$-estimate of Theorem 5.13, this implies that $\|\overline{u}^n\|_{W^{2,p}(\Omega)} \leq C_p$ for any $p \in]1, +\infty[$. Let us pick $p > d$.

By the Sobolev embeddings, $W^{2,p}(\Omega) \hookrightarrow C^1(\overline{\Omega})$, and therefore, $\|\overline{u}^n\|_{C^1(\overline{\Omega})} \leq C$. Now, the same computation as in the proof of Lemma 5.4, using this time the mean value inequality, shows that $\|f_\mu(\cdot, \overline{u}^n(\cdot))\|_{C^{0,\beta}(\overline{\Omega})} \leq C$ for any $\beta \in [0, 1]$. Let us choose one value of $0 < \beta < 1$. By the Schauder estimates (5.7), we finally obtain

$$\|\overline{u}^n\|_{C^{2,\beta}(\overline{\Omega})} \leq C,$$

for some constant C independent of n.

Now as $\beta > 0$, the embedding $C^{2,\beta}(\overline{\Omega}) \hookrightarrow C^2(\overline{\Omega})$ is compact, due to Ascoli's theorem. The family $\{\overline{u}^n\}_{n\in\mathbb{N}}$ is thus relatively compact in $C^2(\overline{\Omega})$. We already know that it converges pointwise to $\overline{u}^*$. Extracting a C^2 convergent subsequence, we thus see that $\overline{u}^* \in C^2(\overline{\Omega})$ and by uniqueness of the limit that $\overline{u}^n \to \overline{u}^*$ strongly in $C^2(\overline{\Omega})$.

We proceed in the exact same manner for $\underline{u}_*$. □

Lemma 5.7. *The limits $\underline{u}_*$ and $\overline{u}^*$ are solutions of Problem* (5.13).

Proof. The strong C^2 convergence of Lemma 5.6 implies right away that $L_\mu \overline{u}^{n+1} \to L_\mu \overline{u}^*$ and that $f_\mu(\cdot, \overline{u}^n(\cdot)) \to f_\mu(\cdot, \overline{u}^*(\cdot))$, both uniformly in Ω. We thus see that $L_\mu \overline{u}^* = f_\mu(\cdot, \overline{u}^*(\cdot))$, which is obviously equivalent to $L\overline{u}^* = f(\cdot, \overline{u}^*(\cdot))$. Moreover, since $\overline{u}^n = 0$ on $\partial\Omega$ for $n \geq 1$, and since the sequence converges pointwise on $\overline{\Omega}$, there also holds $\overline{u}^* = 0$ on $\partial\Omega$. We do the same for $\underline{u}_*$. □

It remains to show that the two solutions thus constructed are respectively minimal and maximal.

Lemma 5.8. *Let u be a solution of Problem* (5.13) *such that $\underline{u} \leq u \leq \overline{u}$. Then $\underline{u}_* \leq u \leq \overline{u}^*$.*

Proof. Using the maximum principle, we show exactly as before by induction that $\underline{u}^n \leq u \leq \overline{u}^n$ for all n, and then pass to the limit. □

Remark 5.16. By the Schauder estimates, we see that $\underline{u}_*$ and $\overline{u}^*$ belong in fact to $C^{2,\alpha}(\overline{\Omega})$ for all $0 < \alpha < 1$. We can of course gain additional regularity if the operator coefficients, the open set and the function f are themselves more regular. Now it could very well happen that $\underline{u}_* = \overline{u}^*$, the result says nothing about uniqueness.

The kind of argument used in the proof of Lemma 5.6 in order to nibble little by little the regularity that is necessary to conclude by switching from the left-hand side to the right-hand side of the equation and vice-versa, is an exemple of what is called a *bootstrap argument*. It is in fact strongly reminiscent of the method advocated by Cyrano de Bergerac to fly to the Moon. □

Clearly, the use of the method rests on the possibility of constructing super- and sub-solutions. This possibility in turns depends on the specific form of f, and there is no general recipe.

Examples. i) Assume that there are two constants $m_- < 0$ and $m_+ > 0$ such that, for all x, $f(x, m_-) \geq 0$ and $f(x, m_+) \leq 0$. Then there exists a solution u such that $m_- \leq u(x) \leq m_+$. The inequalities are actually strict, as can be seen by using Theorem 5.3.

ii) Let $f \in C^1(\mathbb{R})$ be such that $f'(0) > 0$ and that there exists $\beta > 0$ with $f(0) = f(\beta) = 0$. Let $\lambda_1 > 0$ be the first eigenvalue of $-\Delta$ in Ω and $\phi_1 > 0$ the first eigenfunction normalized in $C^0(\overline{\Omega})$. Then, for all $\lambda > \lambda_1/f'(0)$, the problem

$$\begin{cases} -\Delta u = \lambda f(u) & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases}$$

admits a nontrivial solution u, which is strictly positive in Ω. Indeed, $\overline{u} = \beta$ is a super-solution and $\underline{u} = \varepsilon\phi_1$ is a sub-solution for $\varepsilon > 0$ small enough.

5.5 Exercises of Chap. 5

1. Let $\Omega =]-1, 1[$ and $c(x) = 2/(1+x^2)$. Find a function u such that $-u'' + cu \geq 0$ in Ω, $u(\pm 1) \geq 0$ and u attains its minimum in Ω.

2. Let $\Omega = B(0, 1)$ the unit ball of $\mathbb{R}^2$. Let us take a function ζ of class C^∞ on $\mathbb{R}_+$ and such that $\zeta(t) = 1$ for $0 \leq t \leq 1/2$ and $\zeta(t) = 0$ for $t \geq 3/4$. We set $\varphi(x) = x_1 x_2 \zeta(|x|)$.

2.1. Show that $\varphi \in \mathscr{D}(\Omega)$.

2.2. Let

$$u(x) = \sum_{k=1}^{\infty} \frac{2^{-2k}}{k} \varphi(2^k x).$$

Show that $u \in C^1(\overline{\Omega})$ and that u is C^∞ outside of $\{0\}$.

2.3. Show that u admits second order partial derivatives on Ω in the classical sense $\frac{\partial^2 u}{\partial x_1^2}$ and $\frac{\partial^2 u}{\partial x_2^2}$ that are continuous on $\overline{\Omega}$ (the only difficulty is the continuity at 0).

2.4. Show that if $|x| = 2^{-n}$, then $u(x) = \sum_{k=0}^{n-1} \frac{2^{-2k}}{k} \varphi(2^k x)$. Conclude in this case that $\frac{\partial^2 u}{\partial x_1 \partial x_2}(x) = \sum_{k=1}^{n-1} \frac{1}{k}$.

2.5. Conclude that, even though $\Delta u \in C^0(\overline{\Omega})$, nonetheless $u \notin C^2(\overline{\Omega})$.

3. Let $\Omega = B(0, 1)$ the unit ball of $\mathbb{R}^d$ with $d \geq 2$. Let $F \in L^1(\Omega)$ be a radial function, i.e., such that there exists f defined on $[0, 1]$ with $F(x) = f(|x|)$ almost everywhere on both sides and $\int_0^1 r^{d-1}|f(r)|\,dr < +\infty$. We define another radial function $U(x) = u(r)$ with $r = |x|$ by

$$u(r) = \int_1^r s^{1-d}\Big(\int_0^s t^{d-1} f(t)\,dt\Big)ds.$$

3.1. Show that $U \in W_0^{1,1}(\Omega)$.

3.2. Show that $\Delta U = F$ in the sense of $\mathscr{D}'(\Omega)$.

3.3. Show that $U \in W^{2,1}(\Omega)$ if and only if

$$\int_0^1 r^{-1}\Big(\int_0^r s^{d-1}|f(s)|\,ds\Big)dr < +\infty.$$

3.4. Find an example of function U in $W_0^{1,1}(\Omega)$ such that $\Delta U \in L^1(\Omega)$ but $U \notin W^{2,1}(\Omega)$.

3.5. Retrace the same computations with $F \in L^p(\Omega)$, $p \in]1, +\infty]$. What can we say, in particular for $p = +\infty$ in the light of Exercise 2? (We remind the reader of the Hardy inequality: let $p \in]1, +\infty[$, $g \in L^p(\mathbb{R}_+)$ and $G(x) = \frac{1}{x}\int_0^x g(t)\,dt$, then $\|G\|_{L^p(\mathbb{R}_+)} \le \frac{p}{p-1}\|g\|_{L^p(\mathbb{R}_+)}$.)

4. Let Ω be a bounded open subset of $\mathbb{R}^d$, a_0 a function in $L^\infty(\Omega)$ such that there exists $\alpha_0 > 0$ with $a_0(x) \ge \alpha_0$ almost everywhere in Ω, A a $d \times d$ matrix with $L^\infty(\Omega)$ coefficients such that there exists $\alpha > 0$ with $A(x)\xi \cdot \xi \ge \alpha|\xi|^2$ almost everywhere in Ω and for all $\xi \in \mathbb{R}^d$, $f\colon \Omega \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ a Carathéodory function such that there exists $C_0 \ge 0$ and $C_1 \ge 0$ with $|f(x, s, \xi)| \le C_0 + C_1|\xi|^2$ almost everywhere in Ω, and for all $s \in \mathbb{R}$ and $\xi \in \mathbb{R}^d$.

We are interested in the following problem:

$$\begin{cases} u \in L^\infty(\Omega) \cap H_0^1(\Omega), \\ -\mathrm{div}\,(A\nabla u) + a_0 u + f(x, u, \nabla u) = 0 \text{ in the sense of } \mathscr{D}'(\Omega). \end{cases} \tag{5.14}$$

4.1. Show that Eq. (5.14) makes sense.

4.2. For all $\varepsilon > 0$, we set

$$f^\varepsilon(x, s, \xi) = \frac{f(x, s, \xi)}{1 + \varepsilon|f(x, s, \xi)|}.$$

Show that there exists u^ε solution of problem:

$$\begin{cases} u^\varepsilon \in H_0^1(\Omega), \\ -\mathrm{div}\,(A\nabla u^\varepsilon) + a_0 u^\varepsilon + f^\varepsilon(x, u^\varepsilon, \nabla u^\varepsilon) = 0 \text{ in the sense of } \mathscr{D}'(\Omega). \end{cases} \tag{5.15}$$

(Either show or admit that if $g\colon \Omega \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is a bounded Carathéodory function, then the mapping $v \mapsto g(x, v, \nabla v)$ is continuous from $H_0^1(\Omega)$ strong to $L^2(\Omega)$ strong.)

4.3. Show that any solution de (5.15) is in $L^\infty(\Omega)$ and satisfies

$$\|u^\varepsilon\|_{L^\infty(\Omega)} \le \frac{1}{\alpha_0 \varepsilon}.$$

4.4. Assume for this question that Ω, A, a_0 are f of class C^∞. Show that

$$\|u^\varepsilon\|_{L^\infty(\Omega)} \le \frac{C_0}{\alpha_0}. \tag{5.16}$$

In the sequel, we will assume that estimate (5.16) holds true, even if the regularity hypotheses are not satisfied.

4.5. Consider the function $\phi\colon \mathbb{R} \to \mathbb{R}$, $\phi(t) = te^{\lambda t^2}$ where λ is a parameter. Show that if $\lambda \ge \frac{C_1^2}{4\alpha^2}$ then $\alpha\phi'(t) - C_1|\phi(t)| \ge \frac{\alpha}{2}$ for all $t \in \mathbb{R}$.

4.6. Let $v^\varepsilon = \phi(u^\varepsilon)$. Show that $v^\varepsilon \in L^\infty(\Omega) \cap H_0^1(\Omega)$. Using this test-function with a well chosen λ, show that u^ε is bounded in $H_0^1(\Omega)$ independently of ε.

4.7. We extract a subsequence (still denoted u^ε) such that $u^\varepsilon \rightharpoonup u$ in $H_0^1(\Omega)$ weak when $\varepsilon \to 0$. Using the test-function $w^\varepsilon = \phi(u^\varepsilon - u)$ with λ well chosen, show that u^ε tends to u in $H_0^1(\Omega)$ strong.

4.8. Conclude that problem (5.14) admits at least one solution.

4.9. Prove estimate (5.16) without extra regularity hypotheses.

5. Let Ω be a bounded open subset of $\mathbb{R}^d$ of class C^3. We are going to show the existence of a nonzero solution u of problem

$$\begin{cases} -\Delta u = \sqrt{u} & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{5.17}$$

5.1. For all $\varepsilon > 0$, let

$$f_\varepsilon(t) = \begin{cases} 0 & \text{if } t \le 0, \\ \frac{t}{\sqrt{\varepsilon}} & \text{if } 0 \le t \le \varepsilon, \\ \sqrt{t} & \text{if } \varepsilon \le t. \end{cases}$$

Picking $\alpha > 0$, show that problem

$$\begin{cases} -\Delta u_{\alpha,\varepsilon} = f_\varepsilon(u_{\alpha,\varepsilon}) - \alpha u_{\alpha,\varepsilon} & \text{in } \Omega, \\ u_{\alpha,\varepsilon} = 0 & \text{on } \partial\Omega. \end{cases} \tag{5.18}$$

has a nontrivial solution $u_{\alpha,\varepsilon} \in C^{2,\beta}(\overline{\Omega})$ (for all $\beta \in]0,1[$) as soon as ε is small enough. If ϕ_1 is the first positive eigenfunction of $(-\Delta)$ in Ω normalized in $C^0(\bar{\Omega})$, show that there exists $\eta > 0$ independent of ε and α such that

$$\forall x \in \Omega, \quad \eta\phi_1(x) \leq u_{\alpha,\varepsilon}(x) \leq \frac{1}{\alpha^2}.$$

5.2. Show that for all $p \in]1, +\infty[$, $\|u_{\alpha,\varepsilon}\|_{W^{2,p}(\Omega)} \leq C_{\alpha,p}$, where the constant $C_{\alpha,p}$ does not depend on ε. Deduce from this that there exists a solution $u_\alpha \in W^{2,p}(\Omega) \cap W_0^{1,p}(\Omega)$ of problem

$$-\Delta u_\alpha = \sqrt{u_\alpha} - \alpha u_\alpha \quad \text{in } \Omega \tag{5.19}$$

that also satisfies

$$\forall x \in \Omega, \quad \eta\phi_1(x) \leq u_\alpha(x) \leq \frac{1}{\alpha^2}.$$

5.3. Show that $u_\alpha \in C^{2,1/2}(\bar{\Omega})$.

5.4. Show that

$$\|\nabla u_\alpha\|_{L^2(\Omega)} \leq C_\Omega^3 (\operatorname{mes}\Omega)^{1/2},$$

where C_Ω is the Poincaré inequality constant. Deduce from this that there exists $u \in H_0^1(\Omega)$ such that

$$\forall v \in H_0^1(\Omega), \quad \int_\Omega \nabla u \cdot \nabla v \, dx = \int_\Omega v\sqrt{u}\, dx,$$

with

$$\eta\phi_1 \leq u \quad \text{a.e. in } \Omega.$$

(*Hint:* show that there exists a sequence $\alpha_n \to 0$ such that $\sqrt{u_{\alpha_n}}$ strongly converges to a limit in $L^4(\Omega)$.)

5.5. Show that $u \in W^{2,p}(\Omega)$ for all $1 < p < +\infty$, then that $u \in C^{2,1/2}(\bar{\Omega})$ and that u is a solution of (5.18). Can we expect more regularity?

6. Let Ω be a regular bounded open subset of $\mathbb{R}^d$ and $f : \overline{\Omega} \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ a function of class C^1 such that there exists C and $0 < a < 1$ with

$$\forall x \in \overline{\Omega}, \forall s \in \mathbb{R}, \forall \xi \in \mathbb{R}^d, \qquad |f(x,s,\xi)| \leq C(1 + |s|^a + |\xi|^a).$$

We consider the problem: Find u such that

$$\begin{cases} -\Delta u = f(x, u, \nabla u) \text{ in } \Omega, \\ \quad u = 0 \text{ on } \partial\Omega. \end{cases} \tag{5.20}$$

Let $0 < \alpha < 1$ and $E = \{v \in C^{1,\alpha}(\overline{\Omega});\ v = 0 \text{ on } \partial\Omega\}$. For all $v \in E$, we define $T(v)$ by means of the boundary value problem

$$\begin{cases} -\Delta T(v) = f(x, v, \nabla v) \text{ in } \Omega, \\ \quad T(v) = 0 \text{ on } \partial\Omega. \end{cases}$$

6.1. Show that the mapping T is well defined from E into E, continuous and compact.

6.2. Show that there exists R such that $T(\bar{B}_R) \subset \bar{B}_R$ where $\bar{B}_R$ denotes the closed ball of center 0 and radius R in E.

6.3. Show that there exists a solution u of problem (5.20).

Chapter 6
Calculus of Variations and Quasilinear Problems

The nonlinear elliptic problems studied up to now were what is called *semilinear* problems. Semilinear problems are nonlinear problems in which the nonlinearity only concerns the terms that involve derivatives of order strictly less than the maximum differentiation order appearing in the operator. Such is the case for instance of the term $f(u)$ of the first model problem, with derivatives of order 0, compared to the term $-\Delta u$ which contains all the derivatives of the highest order, namely here 2, involved in the problem. In a semilinear problem, the principal part of the operator remains a linear operator.

Semilinear problems are in a sense the simplest nonlinear problems. At the other end of the spectrum of possible second order equations, we find equations in which the second order derivatives appear in a *completely nonlinear* fashion. Such equations assume the very general form $F(x, u, \nabla u, \nabla^2 u) = 0$ with some structure hypotheses on the function F in order for the equation to qualify as elliptic. We will not consider completely nonlinear equations here, which require another range of techniques.

An interesting in-between case is when the second derivatives appear linearly, with coefficients that however depend nonlinearly on lower order derivatives. These operators thus have a principal part that looks like $A(x, u, \nabla u) : \nabla^2 u$, where A is a $d \times d$ matrix-valued function. In this case, we talk about *quasilinear* problems.

Many quasilinear problems are naturally associated with functional minimization problems, which lead to problems in the calculus of variations. Powerful techniques are available for these problems, which are all based on weak lower semicontinuity. This chapter is mostly devoted to a description of some of these techniques in two different cases, the scalar case and the vectorial case. The scalar case refers to scalar-valued unknown functions. In this case, the keyword is convexity, and we provide a very concise introduction to convex analysis. In the vector-valued case, more involved convexity conditions come into play, quasiconvexity, rank-1-convexity, polyconvexity, that we also describe.

H. Le Dret, *Nonlinear Elliptic Partial Differential Equations*, Universitext,
https://doi.org/10.1007/978-3-319-78390-1_6

6.1 Lower Semicontinuity and Convexity

To begin with, let us go back to a few basic topological concepts, independently of the convexity context. First of all, a function from a topological space X with values in $[-\infty, +\infty]$ is said to be *lower semicontinuous* (abbreviated as lsc) if, for all $a \in [-\infty, +\infty]$, the preimage of $[-\infty, a]$ by this function is closed for the topology of X, see Fig. 6.1. Equivalently, it is lsc if $J^{-1}(]a, +\infty])$ is open for all a. The importance of lsc functions for the calculus of variations stems from the following result.

Theorem 6.1. *Let X be a compact topological space and $J\colon X \to [-\infty, +\infty]$ a lsc function. Then J attains its greatest lower bound on X.*

Proof. Let A be the set of those $a \in [-\infty, +\infty]$ such that $C_a = J^{-1}([-\infty, a]) \neq \emptyset$. For all $a \in A$, C_a is a nonempty closed subset of X. Let us take a finite family $(a_i)_{1 \leq i \leq p}$ of elements of A. There clearly holds $\bigcap_{i=1}^{p} C_{a_i} = C_{\min a_i} \neq \emptyset$ by definition of the C_a. Now X is compact, therefore $\bigcap_{a \in A} C_a \neq \emptyset$. By construction, for all $x \in \bigcap_{a \in A} C_a$, there holds $J(x) = \inf_X J$. □

An immediate consequence of this is:

Corollary 6.1. *Let X be a compact and $J\colon X \to]-\infty, +\infty]$ a lsc function. Then J is bounded below and attains its greatest lower bound on X.*

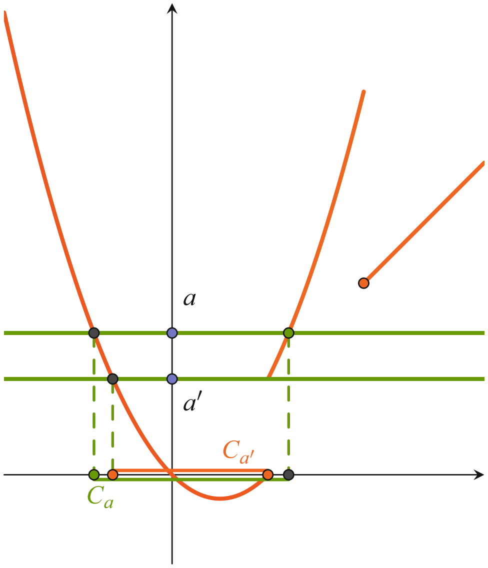


Fig. 6.1 A lsc function

We are now going to be interested in convex functions defined on a convex subset C of a Banach space E. In the context of convex analysis, we use the notation $]-\infty, +\infty] = \bar{\mathbb{R}}$. Convex functions are $\bar{\mathbb{R}}$-valued functions such that,

$$J(\lambda x + (1-\lambda)y) \le \lambda J(x) + (1-\lambda)J(y)$$

for all $x, y \in C$ and all $\lambda \in [0,1]$. The reason why the value $-\infty$ is excluded is that a convex function that takes this value is either identically equal to $-\infty$, or the very notion of convexity is ill-defined with indeterminate expressions of the form $+\infty - \infty$ in case it also takes the value $+\infty$. This case thus presents no interest whatsoever. On the other hand, allowing the $+\infty$ value for convex functions turns out to be extremely useful.

Convexity and weak lower semicontinuity go hand in hand.

Theorem 6.2. *Let C be a closed convex subset of E and $J : C \to \bar{\mathbb{R}}$ be convex and strongly lsc. Then J is weakly lsc.*

Proof. For all $a \in [-\infty, +\infty]$, there holds $C_a = \{x \in C; J(x) \le a\}$ by definition. The function J is convex, thus C_a is a convex subset of E. Indeed, if $x, y \in C_a$ and $\lambda \in [0,1]$, then

$$J(\lambda x + (1-\lambda)y) \le \lambda J(x) + (1-\lambda)J(y) \le a,$$

so that $\lambda x + (1-\lambda)y \in C_a$. Now J is assumed to be strongly lsc, so this set is also closed for the strong topology of E. According to Theorem 1.20, C_a is thus a weakly closed set, which means that J is weakly lsc. □

Remark 6.1. Because the weak topology has less closed sets than the strong topology, there are fewer weakly lsc functions than strongly lsc functions. In particular, it is relatively easy to define strongly continuous, hence strongly lsc functions. It is remarkable that convexity is then enough to ensure weak lower semicontinuity. □

The above result has a sequential version.

Corollary 6.2. *Let C be a closed convex subset of E and $J : C \to \bar{\mathbb{R}}$ be a convex, strongly lsc function. For any sequence $x_n \rightharpoonup x$, there holds* $\liminf J(x_n) \ge J(x)$.

Proof. Let us take a sequence x_n such that $x_n \rightharpoonup x$ for some $x \in E$. Since C is weakly closed, it follows that $x \in C$. We set $a = \liminf J(x_n) \in [-\infty, +\infty]$. We can extract a subsequence, $x_{n'}$, such that $J(x_{n'}) \to a$ (this is just a property of $[-\infty, +\infty]$).

There are three a priori possible cases. The first case is when $a = +\infty$ and there is nothing to prove.

The second case is when $+\infty > a > -\infty$. In this case, for all $\varepsilon > 0$, there exists n_0 such that for all $n' \ge n_0$, $J(x_{n'}) \le a + \varepsilon$. In other words, $x_{n'} \in C_{a+\varepsilon}$ for all $n' \ge n_0$. Now $C_{a+\varepsilon}$ is a weakly closed subset, it thus contains the weak limit of the sequence $(x_{n'})_{n' \ge n_0}$, that is to say x. We have thus shown that for all $\varepsilon > 0$, $J(x) \le a + \varepsilon$, hence the result in this case.

The third and last case is when $a = -\infty$. In this case, for all $M < 0$, there exists n_0 such that for all $n' \geq n_0$, $J(x_{n'}) \leq M$. This means that $x_{n'} \in C_M$ for all $n' \geq n_0$ and since C_M is a weakly closed subset, it follows that $x \in C_M$. We have thus shown that $x \in \cap_{M<0} C_M$. Now $\cap_{M<0} C_M = \emptyset$ because J does not take the value $-\infty$. This is a contradiction, and the third case therefore cannot happen. □

Remark 6.2. i) We could also have gotten summarily rid of the third case by noticing that the set $\{x\} \bigcup \big(\bigcup_{n\in\mathbb{N}}\{x_n\}\big)$ is weakly compact, hence J is bounded below on it.

ii) Another way of proving Corollary 6.2 is to appeal to Mazur's lemma, cf. Corollary 1.1. □

Corollary 6.3. *Let E be a reflexive Banach space, C a nonempty, closed convex subset of E and $J : C \to \bar{\mathbb{R}}$ a convex lsc function such that*

$$\lim_{\substack{x\in C\\ \|x\|\to+\infty}} J(x) = +\infty. \tag{6.1}$$

Then J attains its greatest lower bound on C, that is to say that there exists $x_0 \in C$ such that $J(x_0) = \inf_{y\in C} J(y)$.

Proof. If J is identically equal to $+\infty$, then there is nothing to prove. Assume thus that J is not identically $+\infty$. There exists $\bar{x} \in C$ such that $J(\bar{x}) < +\infty$. We set $\widetilde{C} = \{x \in C; J(x) \leq J(\bar{x})\} = C_{J(\bar{x})}$. This is a closed convex subset that is nonempty since it contains $\bar{x}$. It is moreover bounded by condition (6.1). It is consequently weakly compact since E est reflexive, see Corollary 1.3. The function J is weakly lsc and thus attains its infimum on this compact set at a point x_0. On the other hand, if $x \in C \setminus \widetilde{C}$, then $J(x) > J(\bar{x}) \geq J(x_0)$, so the minimum on $\widetilde{C}$ is also the minimum on the whole of C. □

Remark 6.3. i) Condition (6.1) is called a *coercivity condition*, and the function J is said to be *coercive*. If C is bounded, this condition is empty. Note that by convexity, there is no need to specify whether we are talking about strong lower semicontinuity or weak lower semicontinuity.

ii) We can also establish Corollary 6.3 by using sequences. Actually, this is the most common way used in practice when applying the so-called *direct method of the calculus of variations*. Let us quickly outline the method in question.

There always exists a *minimizing sequence* $x_n \in C$ for J, i.e., a sequence in C such that $J(x_n) \to \inf_{y\in C} J(y)$. This is just an elementary property of $\mathbb{R}$. Since $\inf_{y\in C} J(y) \leq J(\bar{x})$, it follows from the coercivity (6.1) of J that this minimizing sequence is bounded. Now E is reflexive, we can thus extract a subsequence $x_{n'}$ that converges weakly toward a point x_0 of E. The set C is weakly closed, therefore $x_0 \in C$. This implies on the one hand that $J(x_0)$ makes sense, and that $J(x_0) \geq \inf_{y\in C} J(y)$, by definition of a lower bound. On the other hand, J is weakly lsc., so that $\inf_{y\in C} J(y) = \lim J(x_{n'}) = \liminf J(x_{n'}) \geq J(x_0)$. We thus see that J attains its infimum at x_0.

iii) A function J such that $\liminf J(x_n) \geq J(x)$ whenever x_n is a sequence that tends to x for some topology, is said to be *sequentially lower semicontinuous* (slsc) for this topology. We have seen that a lsc function is slcs, cf. the proof of Corollary 6.2. The two properties are equivalent in a metric space, but not necessarily equivalent in a general topological space. The direct method of the calculus of variations obviously only requires slsc functions to work. □

The set of minimum points of a convex function is always a convex set, but there is no reason in general for it to be reduced to a single point. We nonetheless have the following almost obvious uniqueness result.

Proposition 6.1. *Under the previous hypotheses, if J is in addition strictly convex, then its minimum point is unique.*

Proof. Let x_1 and x_2 be two minimum points of J. We see that $\frac{x_1+x_2}{2} \in C$ is also a minimum point. If $x_1 \neq x_2$, strict convexity implies that $J\big(\frac{x_1+x_2}{2}\big) < \frac{1}{2}J(x_1) + \frac{1}{2}J(x_2) = \min_C J$, which is a contradiction. Consequently, $x_1 = x_2$. □

Of course, this is just a sufficient condition for uniqueness, which is in no way necessary.

The reader may consult [11, 25, 60] for more details on convex analysis.

6.2 Application to Scalar Quasilinear Elliptic Boundary Value Problems

In order to apply the previous abstract results to problems in the calculus of variations that are associated with quasilinear elliptic boundary value problems, we consider the following situation. Let Ω be a bounded open subset of $\mathbb{R}^d$ and F a convex function from $\mathbb{R}^d$ to $\mathbb{R}$, such that there exists $p \in]1, +\infty[$, C, β and $\alpha > 0$ with

$$|F(\xi)| \leq C(1 + |\xi|^p), \tag{6.2}$$

$$F(\xi) \geq \alpha|\xi|^p - \beta, \tag{6.3}$$

for all $\xi \in \mathbb{R}^d$. Condition (6.2) is called a growth condition and condition (6.3) a coercivity condition. Given $f \in L^{p'}(\Omega)$ with $\frac{1}{p} + \frac{1}{p'} = 1$, we introduce the functional $J\colon W_0^{1,p}(\Omega) \to \mathbb{R}$ by letting

$$J(v) = \int_\Omega F(\nabla v)\,dx - \int_\Omega fv\,dx. \tag{6.4}$$

We then prove a first minimization result.

Theorem 6.3. *There exists $u \in W_0^{1,p}(\Omega)$ that minimizes J on $W_0^{1,p}(\Omega)$.*

Proof. The function F is convex and the integral involving f is linear with respect to v, hence it is clear that J itself is convex. Moreover, the growth condition (6.2) and the fact that p' is the Hölder conjugate exponent of p, ensure that J is $\mathbb{R}$-valued.

We notice that the space $W_0^{1,p}(\Omega)$ is reflexive since $1 < p < +\infty$.

Let us show that J is strongly continuous on $W_0^{1,p}(\Omega)$. Let v_n be a sequence such that $v_n \to v$ strongly in $W_0^{1,p}(\Omega)$. First of all, it is clear that $\int_\Omega f v_n\,dx \to \int_\Omega f v\,dx$.

Next, we extract a subsequence $v_{n'}$ such that $\nabla v_{n'}$ converges almost everywhere to ∇v and such that there exists $g \in L^p(\Omega)$ with $|\nabla v_{n'}| \le g$ almost everywhere by Theorem 1.6. The function F is convex and with values in $\mathbb{R}$, it is thus continuous. It follows that

$$\begin{cases} F(\nabla v_{n'}) \to F(\nabla v) \quad \text{almost everywhere,} \\ |F(\nabla v_{n'})| \le C(1 + |\nabla v_{n'}|^p) \le C(1 + g^p) \in L^1(\Omega). \end{cases}$$

The Lebesgue dominated convergence theorem then allows us to conclude that

$$\int_\Omega F(\nabla v_{n'})\,dx \longrightarrow \int_\Omega F(\nabla v)\,dx \quad \text{when } n' \to +\infty.$$

Consequently, J is strongly continuous by uniqueness of the limit.

It follows from this that J is weakly lsc on $W_0^{1,p}(\Omega)$, due to Corollary 6.2. In order to apply Corollary 6.3, we need to check that hypothesis (6.1) is satisfied. We use the coercivity hypothesis (6.3) for this purpose. Indeed, the latter implies that

$$J(v) \ge \alpha \|\nabla v\|^p_{L^p(\Omega)} - \beta \operatorname{meas}\Omega - \|v\|_{L^p(\Omega)} \|f\|_{L^q(\Omega)}$$

and we conclude by the Poincaré inequality and the fact that $p > 1$. □

Remark 6.4. We could have just used Carathéodory's theorem for strong continuity, instead of essentially proving it again. It is in fact not necessary to invoke strong continuity at all. The following more general result is available. Let $f : \mathbb{R}^m \to \bar{\mathbb{R}}$ be lsc and bounded below. Then the mapping $I : L^1_{\text{loc}}(\Omega; \mathbb{R}^m) \to \bar{\mathbb{R}}$ defined by $I(z) = \int_\Omega f(z)\,dx$ is strongly slsc. Indeed, let $z_n \to z$ in $L^1_{\text{loc}}(\Omega; \mathbb{R}^n)$. It is sufficient to extract a subsequence that recovers the inferior limit of the sequence $I(z_n)$, then to extract from this subsequence a further subsequence that converges almost everywhere. Fatou's lemma then applies and yields the result. The fact that J is strongly lsc on $W_0^{1,p}(\Omega)$ follows immediately. □

We have the following result concerning uniqueness.

Theorem 6.4. *If F is in addition strictly convex, then the minimum point is unique.*

Proof. Let $u_1 \ne u_2$. By the Poincaré inequality, there holds $\nabla u_1 \ne \nabla u_2$ on a set of strictly positive measure. By strict convexity of F, we thus see that for all $\lambda \in]0,1[$, $F(\lambda \nabla u_1 + (1-\lambda)\nabla u_2) < \lambda F(\nabla u_1) + (1-\lambda)F(\nabla u_2)$ on a set of strictly positive measure. Integrating over Ω, we deduce that J is strictly convex and we apply Proposition 6.1. □

A problem in the calculus of variations consists in minimizing a functional of the kind (6.4) over a function space. Passing from such a problem to a boundary value problem is called finding the *Euler-Lagrange equation* of the minimization problem. More precisely, let us assume now that F is of class C^1. Its gradient is thus a mapping from $\mathbb{R}^d$ to $\mathbb{R}^d$.[1] We suppose that this mapping satisfies a growth condition that is compatible with that satisfied by F, namely

$$|\nabla F(\xi)| \leq C(1 + |\xi|^{p-1}), \tag{6.5}$$

for a certain constant C and for all $\xi \in \mathbb{R}^d$.

Theorem 6.5. *Any minimum point $u \in W_0^{1,p}(\Omega)$ for J is a solution of the variational problem*

$$\forall v \in W_0^{1,p}(\Omega), \qquad \int_\Omega \nabla F(\nabla u) \cdot \nabla v \, dx = \int_\Omega f v \, dx. \tag{6.6}$$

Proof. Let v be an arbitrary function in $W_0^{1,p}(\Omega)$. For all $t \in \mathbb{R}$, there holds $u + tv \in W_0^{1,p}(\Omega)$, so that $J(u+tv)$ is well defined. Moreover, the function $j\colon [-1,1] \to \mathbb{R}$, $j(t) = J(u+tv)$, admits a minimum at $t = 0$. Let us show that this function is of class C^1. By definition of J, there holds

$$j(t) = \int_\Omega F(\nabla u + t\nabla v)\, dx - \int_\Omega f u \, dx - t \int_\Omega f v \, dx.$$

We set $G(x,t) = F(\nabla u(x) + t\nabla v(x))$. This function is of class C^1 with respect to t for almost all x, with

$$\frac{\partial G}{\partial t}(x,t) = \nabla F(\nabla u(x) + t\nabla v(x)) \cdot \nabla v(x).$$

Due to the Cauchy-Schwarz inequality and the growth hypothesis (6.5), we see that

$$\begin{aligned}\left|\frac{\partial G}{\partial t}(x,t)\right| &\leq |\nabla F(\nabla u(x) + t\nabla v(x))|\, |\nabla v(x)| \\ &\leq C(1 + |\nabla u(x) + t\nabla v(x)|^{p-1})|\nabla v(x)| \\ &\leq C'(1 + |\nabla u(x)|^{p-1} + |\nabla v(x)|^{p-1})|\nabla v(x)|.\end{aligned}$$

[1] We identify $\mathbb{R}^d$ with its dual via the usual Euclidean inner product, in order to represent the differential of F, which is a linear form DF, by a vector ∇F.

Indeed, there exists a constant $C'' > 0$ such that for all pairs $a, b \in \mathbb{R}^+$, there holds $(a+b)^{p-1} \leq C''(a^{p-1} + b^{p-1})$. To see this, consider the function $\mathbb{R}^+ \to \mathbb{R}^+$, $z \mapsto (1+z)^{p-1}/(1+z^{p-1})$, show that it is bounded and then set $z = b/a$.[2]

Since $\nabla u \in L^p(\Omega; \mathbb{R}^d)$, it follows that $|\nabla u|^{p-1} \in L^{p/(p-1)}(\Omega)$. Now it turns out that $q = p/(p-1)$ is none other than the Hölder conjugate exponent of p. We can thus apply the Hölder inequality to conclude that $|\nabla u|^{p-1}|\nabla v| \in L^1(\Omega)$. Since it is otherwise clear that $|\nabla v|^p \in L^1(\Omega)$ and that $|\nabla v| \in L^1(\Omega)$ because Ω is bounded, we see that $\partial G/\partial t$ is dominated by a function of $L^1(\Omega)$, uniformly with respect to $t \in [-1, 1]$.

By the Lebesgue version of the Leibniz integral rule, it follows that the function $t \mapsto \int_\Omega G(x,t)\,dx$ is of class C^1 and that its derivative is given by $t \mapsto \int_\Omega \frac{\partial G}{\partial t}(x,t)\,dx$. This obviously implies that j is of class C^1 with

$$j'(t) = \int_\Omega \nabla F\big(\nabla u(x) + t\nabla v(x)\big) \cdot \nabla v(x)\,dx - \int_\Omega f v\,dx.$$

Since j has a minimum at $t = 0$, it follows that $j'(0) = 0$, which is exactly the Euler-Lagrange equation (6.6). □

The Euler-Lagrange equation is the variational form of a boundary value problem.

Corollary 6.4. *Any minimum point of J is a solution of the quasilinear boundary value problem*

$$\begin{cases} u \in W_0^{1,p}(\Omega), \\ -\operatorname{div} \nabla F(\nabla u) = f \quad \textit{in the sense of } \mathscr{D}'(\Omega). \end{cases}$$

Proof. Just take $v \in \mathscr{D}(\Omega)$. □

Remark 6.5. i) To check that this is really a quasilinear problem, let assume that u and F are of class C^2. The gradient of F is $\nabla F(\xi) = (\partial_1 F(\xi), \dots, \partial_d F(\xi))^T$ and we see that the distributional equation above reads

$$-\partial_i[\partial_i F(\nabla u)] = -\partial_{ik} F(\nabla u)\partial_{ki} u = f.$$

Letting $a_{ik}(\xi) = \partial_{ik} F(\xi)$, we actually obtain an operator of the form $-a_{ik}(\nabla u)\partial_{ki} u$, as expected. The convexity of F implies that its Hessian matrix $(a_{ij}(\xi))$ is positive, hence the elliptic character of the equation.

ii) Conversely, it may happen that a given quasilinear boundary value problem is the Euler-Lagrange equation of a minimization problem, at least formally. In this case, it may be advantageous to solve the minimization problem and then possibly derive a solution of the initial boundary value problem.

[2] In the case $p \geq 2$, we can simply use the convexity of the function $a \mapsto a^{p-1}$ and find $C'' = 2^{p-2}$.

iii) The proof of Theorem 6.5 consists in fact in showing that the functional J is Gateaux-differentiable, and that its differential vanishes at any minimum point.

iv) More general calculus of variations problems may be considered, with integrands of the form $F(x, u, \nabla u)$ satisfying appropriate measurability, continuity and convexity hypotheses with respect to their various variables.

v) The solutions of the minimization problem satisfy an elliptic equation. It is thus to be expected that they possess additional regularity due to elliptic regularity, see Sect. 5.3 of Chap. 5. See [13] for a general discussion of calculus of variations methods when Ω is one-dimensional, including regularity. □

In the case of a convex integrand $F(\nabla u)$, we have thus found all the solutions of the associated boundary value problem.

Theorem 6.6. *Any solution $u \in W_0^{1,p}(\Omega)$ of the variational problem (6.6) is a minimum point of J.*

Proof. Since F is convex of class C^1, it follows that for all pairs of vectors $\xi, \zeta \in \mathbb{R}^d$, there holds

$$F(\xi) - F(\zeta) \geq \nabla F(\zeta) \cdot (\xi - \zeta).$$

Replacing ξ by $\nabla v(x)$ and ζ by $\nabla u(x)$, and integrating over Ω, we obtain that for all $v \in W_0^{1,p}(\Omega)$,

$$\int_\Omega F(\nabla v)\,dx - \int_\Omega F(\nabla u)\,dx \geq \int_\Omega \nabla F(\nabla u) \cdot (\nabla v - \nabla u)\,dx = \int_\Omega f(v-u)\,dx$$

due to the Euler-Lagrange equation (6.6) with the test-function $v - u$. This shows that $J(v) \geq J(u)$. □

Example. The function $F(\xi) = (1/p)|\xi|^p$ is strictly convex, of class C^1 for $p > 1$ with $\nabla F(\xi) = |\xi|^{p-2}\xi$. We have thus found that the equation

$$-\mathrm{div}\,(|\nabla u|^{p-2}\nabla u) = f,$$

has a unique solution in $W_0^{1,p}(\Omega)$, which is a minimizer of the associated functional. We could as well have taken f in $W^{-1,p'}(\Omega) = (W_0^{1,p}(\Omega))'$. Expanding formally, if u is smooth,[3] the equation reads

$$-|\nabla u|^{p-2}\Delta u - (p-2)|\nabla u|^{p-4}(\partial_i u \partial_j u \partial_{ij} u) = f,$$

in other words, the operator has quasilinear matrix coefficients $-a_{ij}(\nabla u)\partial_{ij} u$ with

$$a_{ij}(\xi) = |\xi|^{p-2}\delta_{ij} + (p-2)|\xi|^{p-4}\xi_i\xi_j.$$

[3] and ∇u does not vanish when $p < 2$.

The above quasilinear operator is called the p-Laplacian. It reduces to the usual, linear, Laplacian when $p = 2$.

6.3 Calculus of Variations in the Vectorial Case, Quasiconvexity

The results of the previous section naturally admit many generalisations. We are going to consider the case of functions u with values in $\mathbb{R}^m$, $m \geq 1$, which in terms of Euler-Lagrange equations, correspond to systems of scalar quasilinear equations. In the scalar case $m = 1$, the basic hypothesis on F that ensures existence in all cases, modulo a few technical hypotheses, is convexity. The same hypothesis can be made when $m > 1$ and the theory is essentially unchanged.

It turns out however that the convexity hypothesis is too restrictive in a number of applications in the vectorial case. It is even sometimes totally unrealistic, as in nonlinear elasticity for example, see [16]. It must thus be replaced by a more general condition that still implies the weak lower semicontinuity of functionals of the type (6.4), while being acceptable for such applications.

In what follows, Ω is still a bounded open subset of $\mathbb{R}^d$ but the mappings u considered on Ω take their values in $\mathbb{R}^m$. We thus have m scalar components u_i in the canonical basis. The differential of u is then represented by the Jacobian matrix, a $m \times d$ matrix given by $(\nabla u)_{ij} = \partial_j u_i$. In the context of the vectorial calculus of variations, it is customary to call this matrix the gradient of u, even though this is not exactly correct vocabulary.[4] Let M_{md} be the space of $m \times d$ matrices. It is equipped with the usual inner product $A : B = \operatorname{tr}(A^T B)$ (note: $A^T B$ is a $d \times d$ matrix).

Given a continuous function $F \colon M_{md} \to \mathbb{R}$, we consider the functional

$$I(u) = \int_\Omega F(\nabla u)\,dx,$$

without talking precisely about function spaces for the time being, but still in the general context of $W^{1,p}$ spaces. We are going to study the weak lower semicontinuity properties of this class of functionals.

Definition 6.1. We say that F is quasiconvex if there exists a bounded open subset D of $\mathbb{R}^d$ such that, for any matrix $A \in M_{md}$ and any function $\varphi \in \mathscr{D}(D; \mathbb{R}^m)$, there

[4] In the scalar case, the word gradient actually refers to the identification of the differential, a linear form, with a vector via an inner product, that is to say that the gradient is the adjoint of the differential. The same word is nonetheless often used for the Jacobian matrix in the present vectorial context, even though there is no adjunction here. Indeed, in the scalar case $m = 1$, the Jacobian matrix is a row matrix and the gradient a column matrix.

holds

$$\int_D F(A+\nabla\varphi(x))\,dx \geq (\text{meas}\,D)F(A). \tag{6.7}$$

Here, the notation $\mathscr{D}(D;\mathbb{R}^m)$ stands for the space of C^∞, compactly supported functions with values in $\mathbb{R}^m$. See [55, 56] for the original introduction of quasiconvexity. Let us first check that this definition is reasonable. Indeed, it seems to depend on an arbitrary open set D with no relationship with F, which looks sort of strange.

Lemma 6.1. *Definition 6.1 does not depend on the open set D.*

Proof. Let F be a quasiconvex function and D be the corresponding open set in Definition 6.1. Now let D_1 be another bounded open subset of $\mathbb{R}^d$. Clearly, there exists a point $x_0 \in \mathbb{R}^d$ and a number $\eta > 0$ such that $x_0 + \eta D_1 \subset D$. For all $\varphi \in \mathscr{D}(D_1;\mathbb{R}^m)$, we define $\varphi_* \in \mathscr{D}(D;\mathbb{R}^m)$ by

$$\varphi_*(x) = \begin{cases} \eta\varphi(\frac{x-x_0}{\eta}) & \text{if } x \in x_0+\eta D_1,\\ 0 & \text{otherwise.}\end{cases}$$

Since F is quasiconvex, it follows in particular that

$$\int_D F(A+\nabla\varphi_*(x))\,dx \geq (\text{meas}\,D)F(A).$$

By definition of φ_*, there holds

$$\nabla\varphi_*(x) = \begin{cases} \nabla\varphi(\frac{x-x_0}{\eta}) & \text{if } x \in x_0+\eta D_1,\\ 0 & \text{otherwise.}\end{cases}$$

Replacing these expressions in the quasiconvexity inequality, we obtain

$$\int_{D\setminus(x_0+\eta D_1)} F(A)\,dx + \int_{x_0+\eta D_1} F(A+\nabla\varphi_*(x))\,dx \geq (\text{meas}\,D)F(A),$$

that is to say

$$\int_{x_0+\eta D_1} F\Big(A+\nabla\varphi\Big(\frac{x-x_0}{\eta}\Big)\Big)\,dx \geq (\text{meas}(x_0+\eta D_1))F(A) = \eta^d(\text{meas}\,D_1)F(A),$$

since $F(A) < +\infty$. We now perform the change of variable $y = (x-x_0)/\eta$ in the integral, and deduce that F satisfies the quasiconvexity inequality on D_1 for all $\varphi \in \mathscr{D}(D_1;\mathbb{R}^m)$. □

Remark 6.6. We can replace $\mathscr{D}(D;\mathbb{R}^m)$ by $W_0^{1,\infty}(D;\mathbb{R}^m)$ without changing anything in Definition 6.7. □

We now prove that there are quasiconvex functions.

Proposition 6.2. *Any convex real-valued function is quasiconvex.*

Proof. Let F be a convex real-valued function on M_{md}, D an open bounded subset of measure 1 and $\varphi \in \mathscr{D}(D;\mathbb{R}^m)$. We apply Jensen's inequality with the restriction of the Lebesgue measure to D, which is a probability measure. There thus holds

$$\int_D F(A+\nabla\varphi(x))\,dx \geq F\Big(\int_D (A+\nabla\varphi(x))\,dx\Big).$$

Now D has measure 1 so that

$$\int_D (A+\nabla\varphi(x))\,dx = A + \int_D \nabla\varphi(x)\,dx = A,$$

since the second integral vanishes after integration by parts. We thus obtain that

$$\int_D F(A+\nabla\varphi(x))\,dx \geq F(A),$$

which means that F is quasiconvex. □

Quasiconvexity would not be too thrilling if it coincided with convexity. Fortunately, this is not the case as soon as $m > 1$ and $d > 1$. We will go back to this later on, when introducing a sufficient condition for quasiconvexity. The two however coincide when $m = 1$ or $d = 1$.

Quasiconvexity intervenes in problems in the calculus of variations in the vectorial case due to the following two theorems.

Theorem 6.7. *If the functional I is weakly-$*$ slsc on $W^{1,\infty}(\Omega;\mathbb{R}^m)$, then F is quasiconvex.*

Proof. Let $A \in M_{md}$. Without loss of generality, we can assume that $0 \in \Omega$. Let then $Q \subset \Omega$ be a hypercube centered at 0 and of width L. Consider $\varphi \in \mathscr{D}(Q;\mathbb{R}^m)$. Let $k \neq 0$ be a natural number. We subdivide Q into k^d hypercubes $(Q_l)_{l=1,\dots,k^d}$ of disjoint interiors with edges parallel to the edges of Q and of length L/k. Let x_l denote the centers of these hypercubes. We then set

$$u_k(x) = \begin{cases} Ax + \frac{1}{k}\varphi(k(x-x_l)) & \text{if } x \in Q_l \text{ for some } l, \\ Ax & \text{otherwise,} \end{cases}$$

see Fig. 6.2.

Since φ vanishes in a neighborhood of the boundary of Q, it is clear that $u_k \in \mathscr{D}(\Omega;\mathbb{R}^m)$. Moreover,

$$\nabla u_k(x) = \begin{cases} A + \nabla\varphi(k(x-x_l)) & \text{if } x \in Q_l, \\ A & \text{otherwise,} \end{cases}$$

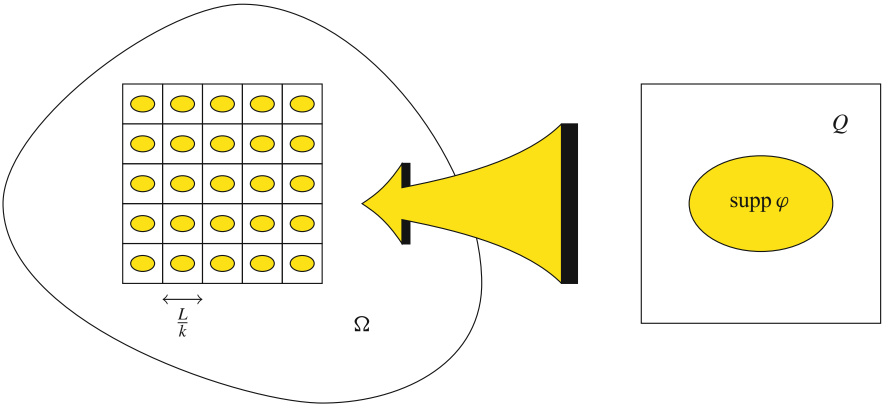


Fig. 6.2 Constructing the sequence u_k

so that there exists a constant C independent of k such that

$$\|u_k\|_{L^\infty(\Omega;\mathbb{R}^m)} \leq C, \qquad \|\nabla u_k\|_{L^\infty(\Omega;M_{md})} \leq C.$$

The sequence u_k is bounded in $W^{1,\infty}(\Omega;\mathbb{R}^m)$, it thus admits a weakly-$*$ convergent subsequence. Now on the other hand, the whole sequence converges uniformly to $u(x) = Ax$. Therefore, $u_k \rightharpoonup u$ in $W^{1,\infty}(\Omega;\mathbb{R}^m)$ weak-$*$. The weak-$*$ slsc of I implies then that

$$\liminf I(u_k) \geq I(u) = \int_\Omega F(A)\,dx = (\text{meas}\,\Omega)F(A). \tag{6.8}$$

Let us compute $I(u_k)$.

$$\begin{aligned} I(u_k) &= \int_{\Omega\setminus Q} F(A)\,dx + \int_Q F(A+\nabla u_k(x))\,dx \\ &= (\text{meas}\,\Omega - \text{meas}\,Q)F(A) + \sum_{l=1}^{k^d}\int_{Q_l} F(A+\nabla\varphi(k(x-x_l)))\,dx, \end{aligned}$$

so that, by the change of variable $y_l = k(x - x_l)$ in each small cube,

$$\begin{aligned} I(u_k) &= (\text{meas}\,\Omega - \text{meas}\,Q)F(A) + \sum_{l=1}^{k^d}\frac{1}{k^d}\int_Q F(A+\nabla\varphi(y_l))\,dy_l \\ &= (\text{meas}\,\Omega - \text{meas}\,Q)F(A) + \int_Q F(A+\nabla\varphi(y))\,dy. \end{aligned}$$

Combining this expression (which actually does not depend on k) with inequality (6.8), we obtain that F is quasiconvex. □

Theorem 6.7 has a markedly more difficult converse, a proof of which can be found in appendix to this chapter, see also [55].

Theorem 6.8. *If $F: M_{md} \to \mathbb{R}$ is quasiconvex, then the functional I is weakly-$*$ slsc on $W^{1,\infty}(\Omega; \mathbb{R}^m)$.*

Quasiconvexity thus appears as a necessary and sufficient condition for weak lower semicontinuity of functionals of the calculus of variations in the vectorial case. The result is also true in $W^{1,p}(\Omega; \mathbb{R}^m)$, on the condition of adding appropriate growth and bound below conditions. Here is an example due to [1].

Theorem 6.9. *Let $F: M_{md} \to \mathbb{R}$ be quasiconvex and such that*

$$\begin{cases} |F(A)| \leq C(1 + |A|^p), \\ F(A) \geq 0, \end{cases}$$

for a certain $p \in]1, +\infty[$. Then the functional I is weakly slsc on $W^{1,p}(\Omega; \mathbb{R}^m)$.

We will also show this difficult result in appendix.

A number of existence results of minimum points for such functionals follow from this theorem. Let us give an example.

Corollary 6.5. *Let F satisfy the hypotheses of Theorem 6.9 and be such that there exists $\alpha > 0$ with*

$$F(A) \geq \alpha |A|^p. \tag{6.9}$$

Let us be given $f \in L^{p'}(\Omega; \mathbb{R}^m)$ and set

$$J(u) = I(u) - \int_\Omega f \cdot u \, dx.$$

Then the functional J attains its minimum on $W_0^{1,p}(\Omega; \mathbb{R}^m)$.

Proof. We apply the direct method of the calculus of variations. Consider a minimizing sequence, that is to say a sequence $u_k \in W_0^{1,p}(\Omega; \mathbb{R}^m)$ such that $J(u_k) \to \inf J$. Due to the coercivity (6.9) of F and the Poincaré inequality in $W_0^{1,p}(\Omega; \mathbb{R}^m)$, it follows that u_k is bounded in $W_0^{1,p}(\Omega; \mathbb{R}^m)$. We extract a weakly convergent subsequence, that converges to a certain u, still denoted u_k. Now J is weakly slsc, because the added linear term is weakly continuous, so that $\liminf J(u_k) \geq J(u)$. Hence u is a minimum point of J. □

Remark 6.7. We establish the Euler-Lagrange equation just as before, except that it is here a system of m scalar equations:

$$-\partial_j\Big[\frac{\partial F}{\partial A_{ij}}(\nabla u)\Big] = f_i \text{ for } i = 1, \dots, m, \tag{6.10}$$

in the sense $\mathscr{D}'(\Omega)$, in the m unknown scalar functions u_i.

This is a quasilinear system, since it can be written, at least formally, in the form

$$C_{ijkl}(\nabla u)\partial_{jl}u_k = f_i, i = 1, \dots, m,$$

where the fourth order tensor C is given componentwise by

$$C_{ijkl}(A) = \frac{\partial^2 F}{\partial A_{ij}\partial A_{kl}}(A).$$

Regularity issues are more complicated than in the scalar case, typically with the possibility of singular sets of small Hausdorff dimension, see for example [26, 34, 35, 37]. □

6.4 Quasiconvexity: A Necessary Condition and a Sufficient Condition

Even though quasiconvexity is a necessary and sufficient condition of weak lower semicontinuity, it suffers from an important drawback in practice. In fact, it is in a way almost as difficult to show that a given function is quasiconvex, as it is to show that the associated functional is weakly slsc. It can be shown that quasiconvexity is a nonlocal condition, that must in addition be checked for an infinite set of test-functions. We thus need more workable conditions, either necessary, or sufficient, that can be actually used in practice. We first give such a necessary condition.

Definition 6.2. We say that $F\colon M_{md} \to \bar{\mathbb{R}}$ is *rank-1-convex* if for all pairs of matrices $A, B \in M_{md}$ such that rank $(B - A) = 1$, there holds

$$\forall\lambda \in [0, 1], \quad F(\lambda A + (1-\lambda)B) \le \lambda F(A) + (1-\lambda)F(B). \tag{6.11}$$

In other words, a function is rank-1-convex if it is convex on all the segments whose extremities differ by a rank one matrix. Note that we allow here the value $+\infty$ for F. We moreover assume it to bounded below on bounded sets in order to be able to define the associated functional I on $W^{1,\infty}(\Omega; \mathbb{R}^m)$ with values in $\bar{\mathbb{R}}$.

Theorem 6.10. *If the functional I is weakly-$*$ slsc on $W^{1,\infty}(\Omega; \mathbb{R}^m)$, then F is rank-1-convex.*

Proof. If $F(A) = +\infty$ or $F(B) = +\infty$, there is nothing to prove. Let us thus assume that $F(A) < +\infty$ and $F(B) < +\infty$, with rank $(B - A) = 1$. The latter condition means that there exists two nonzero vectors $a \in \mathbb{R}^d$ and $b \in \mathbb{R}^m$ such that $B - A = b \otimes a$.[5] We take $\lambda \in]0, 1[$ and define a function $h \colon \mathbb{R} \to \mathbb{R}$, 1-periodic and such that

$$h(t) = \begin{cases} 0 & \text{if } 0 \leq t \leq \lambda, \\ 1 & \text{if } \lambda < t \leq 1. \end{cases}$$

We then set

$$g_1(t) = \int_0^t h(s)\,ds, \text{ and } \forall k \in \mathbb{N}^*, g_k(t) = \frac{1}{k} g_1(kt).$$

There holds $g_k'(t) = h(kt)$ almost everywhere in $\mathbb{R}$. If we now define

$$u_k(x) = Ax + g_k(a \cdot x)b$$

then

$$\nabla u_k(x) = A + h(ka \cdot x) b \otimes a = \begin{cases} A & \text{if } x \in \Omega_k^\lambda, \\ B & \text{if } x \in \Omega_k^{1-\lambda}, \end{cases}$$

where

$$\Omega_k^\lambda = \Omega \cap \{x \in \mathbb{R}^d; ka \cdot x - \lfloor ka \cdot x \rfloor < \lambda\} \text{ and } \Omega_k^{1-\lambda} = \Omega \cap \{x \in \mathbb{R}^d; \lambda \leq ka \cdot x - \lfloor ka \cdot x \rfloor\},$$

see Figs. 6.3 and 6.4.

In other words,

$$\nabla u_k(x) = A \mathbf{1}_{\Omega_k^\lambda}(x) + B \mathbf{1}_{\Omega_k^{1-\lambda}}(x),$$

and likewise

$$F(\nabla u_k(x)) = F(A) \mathbf{1}_{\Omega_k^\lambda}(x) + F(B) \mathbf{1}_{\Omega_k^{1-\lambda}}(x).$$

The sequence u_k is clearly bounded $W^{1,\infty}(\Omega; \mathbb{R}^m)$. It is also easy to show that $\mathbf{1}_{\Omega_k^\lambda} \overset{*}{\rightharpoonup} \lambda$ and that $\mathbf{1}_{\Omega_k^{1-\lambda}} \overset{*}{\rightharpoonup} 1 - \lambda$ in $L^\infty(\Omega)$ weak-$*$, cf. Chap. 3, Sect. 3.1. It follows immediately that the sequence u_k converges toward the function $x \mapsto \lambda Ax + (1 - \lambda)$

[5] The tensor product notation means here the rank one matrix $(b \otimes a)_{ij} = b_i a_j$, so that for all $c \in \mathbb{R}^d$, $(b \otimes a)c = (a \cdot c)b$.

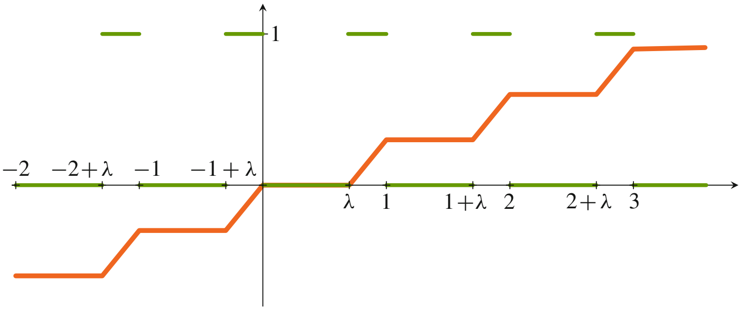


Fig. 6.3 Graphs of g_1 in orange and g_1' in green

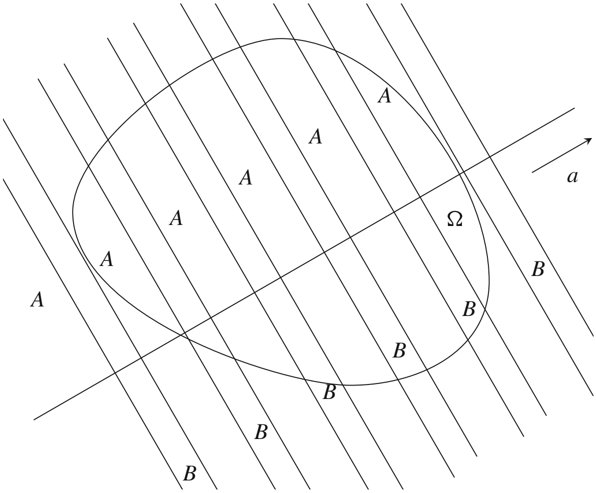


Fig. 6.4 The values taken by the gradient of u_k

Bx in $W^{1,\infty}(\Omega;\mathbb{R}^m)$ weak-$*$. Likewise, the sequence $F(\nabla u_k)$ converges toward the function $x \mapsto \lambda F(A) + (1-\lambda)F(B)$ in $L^\infty(\Omega)$ weak-$*$.

According to the above remarks, there holds on the one hand

$$\int_\Omega F(\nabla u_k(x))\,dx \longrightarrow \operatorname{meas}\Omega[\lambda F(A) + (1-\lambda)F(B)],$$

and on the other hand,

$$\liminf \int_\Omega F(\nabla u_k(x))\,dx \geq \operatorname{meas}\Omega\, F(\lambda A + (1-\lambda)B),$$

since I is weakly-$*$ slsc. The last two relations show that F is rank-1-convex. □

The following necessary condition is now immediate.

Corollary 6.6. *If F is real-valued, then F quasiconvex implies F rank-1-convex.*

Remark 6.8. i) For some insights on what happens when F can take the value $+\infty$, see [29].

ii) If F is of class C^2, the rank-1-convexity of F is equivalent to the *Legendre-Hadamard condition*,

$$\forall A \in M_{md}, \forall \xi \in \mathbb{R}^d, \forall \eta \in \mathbb{R}^m, \quad \frac{\partial^2 F}{\partial A_{ij} \partial A_{kl}}(A)\xi_j \xi_l \eta_i \eta_k \geq 0. \tag{6.12}$$

When it is slightly reinforced in the form

$$\exists c > 0, \forall A \in M_{md}, \forall \xi \in \mathbb{R}^d, \forall \eta \in \mathbb{R}^m, \quad \frac{\partial^2 F}{\partial A_{ij} \partial A_{kl}}(A)\xi_j \xi_l \eta_i \eta_k \geq c|\xi|^2|\eta|^2, \tag{6.13}$$

this condition takes the name of *strong ellipticity* for system (6.10).

iii) When $d = 1$ or $m = 1$, Theorem 6.10 shows that a necessary condition for sequential weak lower semicontinuity is that F be convex.

iv) One of the uses of rank-1-convexity is a negative use: if a finite valued F is not rank-1-convex, then it is certainly not quasiconvex. To show that a given function is not rank-1-convex, it is enough to find two matrices differing by a rank one matrix such that the function is not convex on the segment they define. On the other hand, it may not be so easy to find a test-function φ such that the definition of quasiconvexity is violated. □

Let us now switch to sufficient conditions. Let $T(m, d)$ be the total number of minors of all orders that can be extracted from a $m \times d$ matrix.[6] We denote by $M(A) \in \mathbb{R}^{T(m,d)}$ the family of all minors of A, ordered one way or another.

Definition 6.3. We say that $F\colon M_{md} \to \bar{\mathbb{R}}$ is polyconvex if there exists a convex function $G\colon \mathbb{R}^{T(m,d)} \to \bar{\mathbb{R}}$, such that

$$\forall A \in M_{md}, \quad F(A) = G(M(A)). \tag{6.14}$$

Theorem 6.11. *If F is polyconvex and real-valued, then F is quasiconvex.*

Proof. For simplicity, we just show the theorem in the simplest nontrivial case, $m = d = 2$. The proof in the general case is analogous, with more complicated algebra involved. Since $T(2, 2) = 5$, there thus exists $G\colon \mathbb{R}^5 \to \mathbb{R}$ which is convex (thus continuous) such that $F(A) = G(A, \det A)$. Without loss of generality, we can

[6] This number turns out to be $T(m, d) = \sum_{i=1}^{d} \frac{m!d!}{(i!)^2(m-i)!(d-i)!}$.

assume that D is the unit square. By Jensen's inequality, there holds

$$\begin{aligned}\int_D F(A+\nabla\varphi)\,dx &= \int_D G(A+\nabla\varphi, \det(A+\nabla\varphi))\,dx\\ &\geq G\Big(\int_D (A+\nabla\varphi)\,dx, \int_D \det(A+\nabla\varphi)\,dx\Big).\end{aligned} \tag{6.15}$$

Now φ has compact support and an integration by parts shows that

$$\int_D (A+\nabla\varphi)\,dx = A,$$

on the one hand. On the other hand, if we let $\psi(x) = Ax + \varphi(x)$, then

$$\det(A+\nabla\varphi) = \det(\nabla\psi) = \partial_1\psi_1\partial_2\psi_2 - \partial_1\psi_2\partial_2\psi_1 = \partial_1(\psi_1\partial_2\psi_2) - \partial_2(\psi_1\partial_1\psi_2). \tag{6.16}$$

Consequently, by another integration by parts,

$$\int_D \det(A+\nabla\varphi)\,dx = \int_{\partial D} (\psi_1\partial_2\psi_2 n_1 - \psi_1\partial_1\psi_2 n_2)\,d\sigma.$$

Again, φ has compact support, so that $\psi(x) = Ax$ and $\nabla\psi(x) = A$ on ∂D. Thus, letting $\psi_0(x) = Ax$, the same computation conducted backwards shows that

$$\begin{aligned}\int_D \det(A+\nabla\varphi)\,dx &= \int_{\partial D} ((\psi_0)_1\partial_2(\psi_0)_2 n_1 - (\psi_0)_1\partial_1(\psi_0)_2 n_2)\,d\sigma\\ &= \int_D \det(\nabla\psi_0)\,dx = \det A.\end{aligned}$$

Replacing in inequality (6.15), we obtain

$$\int_\Omega F(A+\nabla\varphi)\,dx \geq G(A, \det A) = F(A),$$

that is to say that F is quasiconvex. □

Remark 6.9. i) The proof rests crucially on the fact that the determinant of a gradient of a function, or more generally, any minor of a gradient, can actually be written as the divergence of a certain vector field. Therefore, its integral on an open set only depends on the values taken by the function in a neighborhood of the boundary of the open set. We say that the determinant is a *null Lagrangian*, see [7].

ii) Polyconvex functions provide examples of quasiconvex functions that are not convex. Thus, for $m = d = 2$, $A \mapsto \det A$ is polyconvex, with $G(A, d) = d$, and thus quasiconvex, but certainly not convex.

iii) If it is relatively easy in principle to construct polyconvex functions—it is enough to take any convex G for that—it is in general a delicate question to determine whether a given function F is polyconvex or not. First of all, there is no uniqueness of the function G. Secondly, the way F is written does not necessarily indicate a clear candidate for G. For instance, the clearly polyconvex function $F(A) = |A|^2 + 2\det A$ could also be written in the form

$$F(A) = (A_{11} + A_{22})^2 + (A_{12} + A_{21})^2 - 4A_{12}A_{21},$$

under which an adequate G is not exactly obviously apparent. There are however some general criteria for polyconvexity.

iv) Polyconvex functions have the added interest of making it possible to bypass the growth conditions needed for the existence of minimum points in the quasiconvex case. Problems in the vectorial calculus of variations in which the functional can take the value $+\infty$, as is the case in nonlinear elasticity, become amenable. This rests on a quite technical study of the weak continuity properties of the minors of a gradient, which are due to their divergence form, see [5].

v) In the case when F only takes finite values, we have shown the following implications: F convex $\Rightarrow$ F polyconvex $\Rightarrow$ F quasiconvex $\Rightarrow$ F rank-1-convex. It is remarkable that none of the converse implications is true, unless $m = 1$ or $d = 1$ in which case the four properties coincide. The latter converse implication is the most difficult of all and was shown not to hold, at least for $d \geq 2$ and $m \geq 3$, see [66].

There is thus no easily checkable criterion of sequential weak lower semicontinuity for functionals in the vectorial calculus of variations. The very concept of quasiconvexity is not local in the space of matrices, see [44], and it is thus very likely that no such criterion exists.

vi) Let us note than when F is quadratic, that is to say when the Euler-Lagrange system (6.10) is linear, then F rank-1-convex is equivalent to F quasiconvex, which is easily shown using the Fourier transform. Incidentally, this is how quasiconvex but non polyconvex functions were exhibited, see [18]. □

6.5 Exercises of Chap. 6

1. Let $F\colon \mathbb{R} \to \mathbb{R}$. Show that if the functional $u \mapsto \int_0^1 F(u'(x))\,dx$ is weakly-$*$ sequentially lower semicontinuous on $W^{1,\infty}(]0,1[)$, then F is convex. (*Hint:* draw inspiration from the proof with rank-one convexity.)

2. Let $\Omega =]0,1[$ and consider the functional

$$I(v) = \int_0^1 \big(v^2 + ((v')^2 - 1)^2\big)\,dx.$$

We are interested in the minimization problem: Find $u \in V = W_0^{1,4}(\Omega)$ such that

$$I(u) = \inf_{v \in V} I(v). \tag{6.17}$$

2.1. We assume that problem (6.17) has a solution $u \in V$. Write its Euler-Lagrange equation first in variational form, then in the sense of distributions.

2.2. Assume that this solution is of class C^2. Show that this is a quasilinear problem. What can you say about the sign of the coefficient of the principal part of the differential operator? Does it sound good?

2.3. Using the sequence

$$v_n(x) = \frac{1}{n}\Big(\frac{1}{2} - \Big|nx - \lfloor nx \rfloor - \frac{1}{2}\Big|\Big),$$

where $\lfloor t \rfloor$ denotes the integer part of t, show that

$$\inf_{v \in V} I(v) = 0.$$

2.4. Deduce from this that in fact, problem (6.17) has no solution (thus it did not sound so good in retrospect).

2.5. Show that $v_n \rightharpoonup 0$ in V and that the functional I is not weakly slsc on V.

2.6. Show that the seminorm $\|v''\|_{L^2(\Omega)}$ is a norm on $H^2(\Omega) \cap H_0^1(\Omega)$, equivalent to the $H^2(\Omega)$ norm.

2.7. For all $k \in \mathbb{N}^*$, we set

$$I^k(v) = I(v) + \frac{1}{k}\int_0^1 (v'')^2\,dx.$$

Show that the problem: Find $u^k \in W = H^2(\Omega) \cap H_0^1(\Omega)$ such that

$$I^k(u^k) = \inf_{v \in W} I^k(v), \tag{6.18}$$

admits at least on solution. Write its Euler-Lagrange equation.

2.8. Show that the sequence u^k is bounded in $W_0^{1,4}(\Omega)$.

2.9. Show that $I^k(u^k) \to 0$ when $k \to +\infty$. (*Hint:* You can start by showing that $\limsup I^k(u^k) \leq I(w)$ for all $w \in W$.) Deduce from this that $u^k \rightharpoonup 0$ in $W_0^{1,4}(\Omega)$.

2.10. We introduce a new functional

$$\bar{I}(v) = \int_0^1 \big(v^2 + [((v')^2 - 1)_+]^2\big)\,dx.$$

Show that the problem: Find $u \in V = W_0^{1,4}(\Omega)$ such that

$$\bar{I}(u) = \inf_{v \in V} \bar{I}(v),$$

admits one (and only one) solution.

2.11. Show that $\bar{I}$ is of class C^1 on V. Write its Euler-Lagrange equation. What is the relation between the solution of this problem and the weak limit of solutions of problem (6.18)?

2.12. Show that $\bar{I}$ is the largest weakly slsc functional on V that is smaller than I.

3. Let Ω be a bounded open subset of $\mathbb{R}^d$, $p \geq 1$ and $f \in L^q(\Omega)$. Show that if $q \geq \frac{2d}{d+2}$ then $L^q(\Omega) \subset H^{-1}(\Omega)$. In this case, we consider the functional defined on $H_0^1(\Omega)$ by

$$I(v) = \int_\Omega \Big(\frac{1}{2}|\nabla v|^2 + \frac{1}{p+1}|v|^{p+1}\Big)dx - \int_\Omega f v\,dx$$

if $v \in L^{p+1}(\Omega)$, $I(v) = +\infty$ otherwise. Show that the minimization problem

$$I(u) = \inf_{v \in H_0^1(\Omega)} I(v)$$

admits a solution $u \in H_0^1(\Omega) \cap L^{p+1}(\Omega)$, that satisfies the Euler-Lagrange equation

$$-\Delta u + |u|^{p-1}u = f \text{ in the sense of } \mathscr{D}'(\Omega).$$

4. Let V be a uniformly convex Banach space (for example, a Hilbert space, a Sobolev space based on L^p with $1 < p < +\infty$, etc.) and I a functional on V convex, lsc., differentiable in the sense of Gateaux and such that there exists $\delta > 0$ with

$$I(v) \geq I(u) + \langle DJ(u), v - u\rangle + \delta\|u - v\|_V^2,$$

for all $u, v \in V$ (the bracket denotes V'-V duality). Show that I attains its minimum on V and furthermore that any minimizing sequence converges strongly in V to a minimizer.

5. Show that the functional I of Exercise 3 is twice differentiable in the sense of Gateaux and that its second differential satisfies an inequality of the form $D^2I(u)(v, v) \geq 2\delta\|\nabla v\|^2_{L^2(\Omega;\mathbb{R}^d)}$ (take $p \leq \frac{d+2}{d-2}$). Use then Exercise 4 to show that the minimizing sequences of Exercise 3 converge strongly in $H_0^1(\Omega)$.

6. We are going to give a second proof of the existence of a nontrivial solution to the boundary value problem

$$\begin{cases} -\Delta u = \sqrt{u} & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases}$$

where Ω is a bounded open subset of $\mathbb{R}^d$.

6.1. Let $I\colon H_0^1(\Omega) \to \mathbb{R}$ be the functional defined by

$$I(v) = \frac{1}{2}\int_\Omega |\nabla v|^2\,dx - \frac{2}{3}\int_\Omega v|v|^{1/2}\,dx.$$

Show that it is well defined and that $\inf_{H_0^1(\Omega)} I(v) > -\infty$.

6.2. Let ϕ_1 be the first positive eigenfunction of $(-\Delta)$ in Ω normalized in $L^2(\Omega)$. Show that there exists $\sigma \in \mathbb{R}$ such that $I(\sigma\phi_1) < 0$.

6.3. Show that the mapping $v \mapsto \int_\Omega v|v|^{1/2}\,dx$ is sequentially continuous from $H_0^1(\Omega)$ weak into $\mathbb{R}$. Deduce from this that I is weakly sequentially lower semicontinuous on $H_0^1(\Omega)$.

6.4. Show that any minimizing sequence of I is bounded in $H_0^1(\Omega)$. Conclude that I attains its infimum at a certain $u \in H_0^1(\Omega)$ with $u \neq 0$.

6.5. Show that u satisfies the Euler-Lagrange equation

$$-\Delta u = |u|^{1/2} \quad \text{in the sense of } \mathscr{D}'(\Omega).$$

6.6. Deduce from this that

$$-\Delta u = \sqrt{u} \quad \text{in the sense of } \mathscr{D}'(\Omega).$$

6.7. What does elliptic regularity say starting from there?

7. Let Ω be a bounded open subset of $\mathbb{R}^2$, and $F\colon M_{2,2} \to \mathbb{R}$ be defined by

$$F(A) = \frac{1}{2}|A|^2 + \sqrt{(\det A)^2 + 1},$$

and $\Phi\colon \mathbb{R}^2 \to \mathbb{R}$ be continuous and such that

$$|\Phi(\xi)| \le C(1 + |\xi|^q)$$

for a certain $1 \le q < 2$. We set

$$J(u) = \int_\Omega F(\nabla u)\,dx - \int_\Omega \Phi(u)\,dx.$$

Show that J is well defined on $H_0^1(\Omega; \mathbb{R}^2)$, that it attains its infimum there, and write the system of quasilinear PDEs thus solved.

8. Let Ω be a bounded open subset of $\mathbb{R}^2$.

8.1. Prove that the relation

$$\det(\nabla u) = \partial_1 u_1 \partial_2 u_2 - \partial_1 u_2 \partial_2 u_1 = \partial_1(u_1 \partial_2 u_2) - \partial_2(u_1 \partial_1 u_2).$$

still holds in the sense of distributions when $\psi \in H^1(\Omega; \mathbb{R}^2)$.

8.2. Let $u(x) = x/|x|$. Show that $u \in W^{1,p}(\Omega; \mathbb{R}^2)$ for all $p < 2$ but not for $p \geq 2$. We let

$$\det \nabla u = \partial_1 u_1 \partial_2 u_2 - \partial_2 u_1 \partial_1 u_2, \quad \operatorname{Det} \nabla u = \partial_1(u_1 \partial_2 u_2) - \partial_2(u_1 \partial_1 u_2).$$

Show that $\det \nabla u = 0$ whereas $\operatorname{Det} \nabla u = -\frac{1}{\pi}\delta_0$, where δ_0 is the Dirac mass at 0.

9. Let Ω be a bounded, regular open subset of $\mathbb{R}^2$, F a function from $M_{2,2}$ into $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ such that $F(A) \geq \alpha|A|^p - \beta$ with $\alpha > 0$ and $p > 2$. Let us be given $u_0 \in W^{1,p}(\Omega; \mathbb{R}^2)$ and set $\mathscr{A} = \{u \in W^{1,p}(\Omega; \mathbb{R}^2); \gamma u = \gamma u_0\}$ where γ denotes the trace mapping on $\partial\Omega$.

Consider the functional

$$J(v) = \int_\Omega F(\nabla v)\, dx,$$

(note that J can take the value $+\infty$). We want to solve the minimization problem: Find $u \in \mathscr{A}$ such that

$$J(u) = \inf_{v \in \mathscr{A}} J(v). \tag{6.19}$$

9.1. We assume that there exists $v_0 \in \mathscr{A}$ such that $J(v_0) < +\infty$. Show that any minimizing sequence u_n is bounded in $W^{1,p}(\Omega; \mathbb{R}^2)$.

9.2. Let $u_n \in W^{1,p}(\Omega; \mathbb{R}^2)$ be a sequence that weakly converges to some u in this space. Show that

$$\det \nabla u_n \rightharpoonup \det \nabla u \text{ in } L^{p/2}(\Omega) \text{ weak}.$$

(*Hint:* use Exercise 8.)

9.3. We assume that F is polyconvex. Show that any limit point in $W^{1,p}(\Omega; \mathbb{R}^2)$ weak of a minimizing sequence u_n is a solution of the minimization problem. (Careful not to use a quasiconvexity argument! Setting $\nabla u_n = z_n$ and $\det \nabla u_n = w_n$ and working with this quintuple of scalar functions is a good idea however.) Deduce that there exists a solution.

9.4. Let

$$G(A, \delta) = \begin{cases} |A|^p - \ln \delta & \text{if } \delta > 0, \\ +\infty & \text{otherwise.} \end{cases}$$

and set $F(A) = G(A, \det A)$. Show that the previous results apply. Show that any solution u satisfies $\det \nabla u > 0$ almost everywhere in Ω.

9.5. We take $\Omega =]-1, 1[^2$. Assume that the function

$$u(x) = \begin{pmatrix} x_1^3 \\ x_2 \end{pmatrix}$$

is a minimizer of (6.19), with the function F given in 9.4. Show that the usual method used to obtain the Euler-Lagrange equations associated with the minimization problem fails. (*Hint:* consider a test-function of the form $(x_1, 0)^T$ in a neighborhood of $x = 0$.)

10. Let $\Omega =]0, 1[^2$. We consider the sequence

$$u_n(x) = n^{-1/2}(1 - x_2)^n \begin{pmatrix} \sin nx_1 \\ \cos nx_1 \end{pmatrix}.$$

Show that $u_n \rightharpoonup 0$ in $H^1(\Omega)$ when $n \to +\infty$, that $\det \nabla u_n \to 0$ in the sense of $\mathscr{D}'(\Omega)$, but that

$$\int_\Omega \det \nabla u_n \, dx = -1.$$

Conclude that the functional $J(u) = \int_\Omega \det \nabla u \, dx$ is not weakly slsc on $H^1(\Omega)$. Show however that the functional $I(u) = \int_\Omega |\det \nabla u| \, dx$ is weakly slsc. Meditate on this pretty mysterious example.

Appendix: Weak Lower Semicontinuity Proofs

In this appendix, in addition to proving some of the previously admitted results, we introduce a certain number of techniques that are useful in the calculus of variations: *blow-up*, *De Giorgi slicing*, among others. These proofs may be skipped at first reading. The reader may also consult more specialized works such as [18, 56].

We begin with a $W^{1,\infty}$-extension lemma. We let B be the unit ball of $\mathbb{R}^d$ and S the unit sphere.

Lemma 6.2. *Let $\zeta^* \in W^{1,\infty}(B; \mathbb{R}^m)$ be such that $\|\zeta^*\|_{L^\infty(S;\mathbb{R}^m)} \le k < 1$. There exists a function $\zeta \in W^{1,\infty}(B; \mathbb{R}^m)$ such that $\zeta = \zeta^*$ on S, $\|\zeta\|_{L^\infty(B;\mathbb{R}^m)} \le k$, $\zeta = 0$ on $(1-k)B$ and $\|\nabla\zeta\|_{L^\infty(B;M_{md})} \le 2M + 1$ where $M = \|\nabla\zeta^*\|_{L^\infty(B;M_{md})}$.*

Proof. For $x \in \bar{B}$, we set

$$\zeta(x) = \begin{cases} 0 & \text{if } |x| \le 1 - k, \\ \left(\frac{|x|+k-1}{k}\right)\zeta^*\left(\frac{x}{|x|}\right) & \text{otherwise,} \end{cases}$$

or equivalently

$$\zeta(x) = \Big(\frac{|x| + k - 1}{k}\Big)_+ \zeta^*\Big(\frac{x}{|x|}\Big),$$

so that $\zeta \in W^{1,\infty}(B; \mathbb{R}^m)$, $\zeta = \zeta^*$ on S, $\|\zeta\|_{L^\infty(B;\mathbb{R}^m)} \le k$ and $\zeta = 0$ on $(1-k)B$. Moreover,

$$\nabla\zeta(x) = \begin{cases} 0 & \text{if } |x| \le 1-k, \\ \frac{x}{k|x|} \otimes \zeta^*\big(\frac{x}{|x|}\big) + \big(\frac{|x|+k-1}{k|x|}\big)\nabla\zeta^*\big(\frac{x}{|x|}\big)\big(I - \frac{x\otimes x}{|x|^2}\big) & \text{otherwise,} \end{cases}$$

hence the upper bound for the norm of the gradient of ζ. □

Remark 6.10. This extension is not especially clever. In fact, we could have used the McShane extension: if u is a real-valued Lipschitz continuous function on a compact set $K \subset \mathbb{R}^d$, with Lipschitz constant M, then it is easy to check that $\tilde{u}(y) = \min_{x\in K}(u(x) + M|x-y|)$ is a Lipschitz extension of u to the whole of $\mathbb{R}^d$, whose Lipschitz constant is still M. To use it here, we need the fact that $W^{1,\infty}(B) = C^{0,1}(\bar{B})$, and then we need to work componentwise. □

Let us introduce the *blow-up* idea for scalar functions, which extends immediately to vector-valued functions, see for example [31] in a more general setting. Let $v \in W^{1,\infty}(\Omega)$. For $x_0 \in \Omega$ and $0 < \rho < d(x_0, \partial\Omega)$, we let $y = (x - x_0)/\rho$ and

$$v_{x_0,\rho}(y) = \frac{v(x_0 + \rho y) - v(x_0)}{\rho}. \tag{6.20}$$

It is clear that $v_{x_0,\rho} \in W^{1,\infty}(B)$, with $\|v_{x_0,\rho}\|_{L^\infty(B)} \le \|\nabla v\|_{L^\infty(\Omega;\mathbb{R}^d)}$ and $\nabla v_{x_0,\rho}(y) = \nabla v(x_0 + \rho y)$ so that

$$\|v_{x_0,\rho}\|_{W^{1,\infty}(B)} \le \|v\|_{W^{1,\infty}(\Omega)}. \tag{6.21}$$

The interest of blow-up stems from the next lemma.

Lemma 6.3. *Let ρ_l be a sequence that tends to 0. For almost all $x_0 \in \Omega$, the sequence v_{x_0,ρ_l} defined by (6.20) tends uniformly to the linear function $z_{x_0}\colon y \mapsto \nabla v(x_0)\cdot y$ on $\bar{B}$.*

This means that blow-up makes it possible to "see" the gradient of $v \in W^{1,\infty}(\Omega)$ at almost all point. The lemma also shows that $W^{1,\infty}(\Omega)$ functions are, in a sense, almost everywhere very close to their tangential affine mapping.

Proof. To alleviate the notation, we write the sequence ρ_l simply as ρ, but it has to be kept in mind that this actually is a sequence. Let A a representative of ∇v, that is to say a measurable $\mathbb{R}^d$-valued function belonging to the equivalence class ∇v. Let $\Omega' \subset \Omega$ be an open set such that $\bar{\Omega}' \subset \Omega$. For $\rho \le d(\bar{\Omega}', \partial\Omega)$, we introduce a function in two variables $h_\rho(x, y)\colon \Omega' \times B \to \mathbb{R}$ defined by

$$h_\rho(x, y) = v_{x,\rho}(y) - A(x) \cdot y = \frac{v(x + \rho y) - v(x)}{\rho} - A(x) \cdot y.$$

We are first going to show that this sequence of functions tends to 0 strongly in $L^1(\Omega' \times B)$ when $\rho \to 0$.

For this, we fix y and consider the sequence of functions $h_{y,\rho}(x) = h_\rho(x, y)$. For all v in $C^\infty(\bar{\Omega}')$, there is no reason to distinguish between ∇v and A, and there holds

$$\begin{aligned}\int_{\Omega'} |h_{y,\rho}(x)|\,dx &= \int_{\Omega'} \left| \int_0^1 [\nabla v(x + t\rho y) - \nabla v(x)] \cdot y\,dt \right| dx \\ &\leq \int_0^1 \int_{\Omega'} |[\nabla v(x + t\rho y) - \nabla v(x)] \cdot y|\,dxdt.\end{aligned}$$

By the Meyers-Serrin theorem, we can approximate any function of $W^{1,1}(\Omega)$ by a sequence of functions of $C^\infty(\Omega)$, thus the inequality still holds for $v \in W^{1,1}(\Omega)$. Now the translation in direction y is continuous in $L^1(\Omega')$, see [11, 61], therefore

$$g_\rho(t) = \int_{\Omega'} |[\nabla v(x + t\rho y) - \nabla v(x)] \cdot y|\,dx \to 0 \text{ when } \rho \to 0,$$
$$|g_\rho(t)| \leq 2 \operatorname{meas} \Omega' \, \|v\|_{W^{1,\infty}(\Omega)}.$$

By the dominated convergence theorem, it follows that

$$\int_{\Omega'} |h_{y,\rho}(x)|\,dx \longrightarrow 0 \text{ quand } \rho \to 0$$

for all $y \in B$. Besides and as above, $\|h_\rho\|_{L^\infty(\Omega' \times B)} \leq \|v\|_{W^{1,\infty}(\Omega)}$, thus

$$\int_{\Omega'} |h_{y,\rho}(x)|\,dx \leq \operatorname{meas} \Omega' \|v\|_{W^{1,\infty}(\Omega)},$$

hence, applying the dominated convergence theorem once more,

$$\int_{\Omega' \times B} |h_\rho(x, y)|\,dxdy = \int_B \Big(\int_{\Omega'} |h_{y,\rho}(x)|\,dx \Big) dy \longrightarrow 0 \text{ when } \rho \to 0.$$

We have thus shown that $h_\rho \to 0$ strongly in $L^1(\Omega' \times B)$ when $\rho \to 0$. By Fubini's theorem, we deduce that for almost all $x \in \Omega'$, $v_{x,\rho}(y) - A(x) \cdot y$ tends to 0 in $L^1(B)$ when $\rho \to 0$.

On the other hand, we also have

$$\|v_{x,\rho} - z_x\|_{W^{1,\infty}(B)} \leq 2\|v\|_{W^{1,\infty}(\Omega)}.$$

Consequently, if we take x_0 in the set where the above convergence holds true, we can extract from $v_{x_0,\rho} - z_{x_0}$ a subsequence that converges in $W^{1,\infty}(B)$ weak-$*$ toward a certain h_{x_0}. By the Sobolev embeddings and the Rellich-Kondrašov theorem, the convergence is thus uniform. Since the sequence also converges in $L^1(B)$ to 0, we see that $h_{x_0} = 0$, which completes the proof of the lemma. □

Remark 6.11. Lemma 6.3 has generalizations to $W^{1,p}(\Omega)$ spaces, see [70]. □

Proof of Theorem 6.8. Let u_n be a sequence that converges to u in the sense of $W^{1,\infty}(\Omega;\mathbb{R}^m)$ weak-$*$ and let $J = \liminf \int_\Omega F(\nabla u_n)\,dx$. We can extract a subsequence (still denoted u_n) such that $\int_\Omega F(\nabla u_n)\,dx \to J$. Since the sequence ∇u_n is bounded in $L^\infty(\Omega; M_{md})$ and since the function F is continuous, $F(\nabla u_n)$ is bounded in $L^\infty(\Omega)$. We can thus extract another subsequence such that $F(\nabla u_n) \overset{*}{\rightharpoonup} g$ in $L^\infty(\Omega)$ weak-$*$. Consequently, for all measurable $A \subset \Omega$, there holds

$$\int_A F(\nabla u_n)\,dx \to \int_A g\,dx. \tag{6.22}$$

It is enough to show that $g \geq F(\nabla u)$ almost everywhere in order to conclude. For this, we consider a Lebesgue point x_0 of g.[7] By definition of the Lebesgue points, there holds

$$g(x_0) = \lim_{\rho\to 0}\Big(\frac{1}{\rho^d \text{meas } B}\int_{B(x_0,\rho)} g(x)\,dx\Big).$$

Consequently, due to (6.22), we obtain,[8]

$$g(x_0) = \lim_{\rho\to 0}\Big\{\lim_{n\to+\infty}\Big(\frac{1}{\rho^d \text{meas } B}\int_{B(x_0,\rho)} F(\nabla u_n(x))\,dx\Big)\Big\}.$$

We now perform a blow-up at x_0, which yields

$$g(x_0) = \lim_{\rho\to 0}\Big\{\lim_{n\to+\infty}\Big(\frac{1}{\text{meas } B}\int_B F(\nabla u_{n,x_0,\rho}(y))\,dy\Big)\Big\}.$$

[7] The Lebesgue points set is of full measure.

[8] Pay close attention to the order of the limits.

By Lemma 6.3, we can assume that x_0 is a convergence point for the blow-up of u as well. Let us take $\rho = 1/k$, $k \in \mathbb{N}^*$, and set

$$I_{n,k} = \frac{1}{\text{meas } B} \int_B F(\nabla u_{n,x_0,1/k}(y))\, dy.$$

We are going to extract a diagonal sequence that simultaneously achieves the double limit and the uniform convergence. For this, we notice that for all $l \in \mathbb{N}^*$, there exists k_l such that

$$\Big| \lim_{k\to+\infty} \lim_{n\to+\infty} I_{n,k} - \lim_{n\to+\infty} I_{n,k_l} \Big| \leq \frac{1}{2l} \quad \text{and} \quad \|u_{x_0,1/k_l} - \nabla u(x_0)y\|_{L^\infty(B;\mathbb{R}^m)} \leq \frac{1}{2l},$$

by Lemma 6.3 for the second estimate. Let us set k to this value k_l. Then, there exists n_l such that

$$\Big| \lim_{n\to+\infty} I_{n,k_l} - I_{n_l,k_l} \Big| \leq \frac{1}{2l} \quad \text{and} \quad \|u_{x_0,1/k_l} - u_{n_l,x_0,1/k_l}\|_{L^\infty(B;\mathbb{R}^m)} \leq \frac{1}{2l}.$$

Indeed, for the second estimate, there is no gradient involved and since $u_n \to u$ strongly in $L^\infty(B;\mathbb{R}^m)$ when $n \to +\infty$ by the Rellich-Kondrašov theorem, for fixed k it follows that $u_{n,x_0,1/k} \to u_{x_0,1/k}$ strongly in $L^\infty(B;\mathbb{R}^m)$ when $n \to +\infty$. We have thus constructed a diagonal sequence (k_l, n_l) such that

$$g(x_0) = \lim_{l\to+\infty} \Big(\frac{1}{\text{meas } B} \int_B F(\nabla u_{n_l,x_0,1/k_l}(y))\, dy \Big) \quad \text{and}$$

$$\|u_{n_l,x_0,1/k_l} - \nabla u(x_0)y\|_{L^\infty(B;\mathbb{R}^m)} \leq \frac{1}{l}. \tag{6.23}$$

At this point, we set

$$w_l(y) = u_{n_l,x_0,1/k_l}(y) - \nabla u(x_0)y.$$

By Lemma 6.2, there exists $\zeta_l \in W^{1,\infty}(B;\mathbb{R}^m)$ such that $\zeta_l = w_l$ on S, $\zeta_l = 0$ on $(1-\frac{1}{l})B$, and $\|\zeta_l\|_{W^{1,\infty}(B;\mathbb{R}^m)} \leq C$. Indeed, since $u_n \overset{*}{\rightharpoonup} u$ in $W^{1,\infty}(\Omega, \mathbb{R}^m)$ weak-$*$, $u_{n,x_0,\rho}$ is bounded in $W^{1,\infty}(B, \mathbb{R}^m)$ independently of n, x_0 and ρ, cf. (6.21). Letting $A = \nabla u(x_0)$, there holds

$$\nabla u_{n_l,x_0,1/k_l}(y) = A + \nabla w_l(y) = A + \nabla(w_l - \zeta_l)(y) + \nabla \zeta_l(y), \tag{6.24}$$

with $w_l - \zeta_l \in W_0^{1,\infty}(B;\mathbb{R}^m)$, $\|\nabla \zeta_l\|_{L^\infty(B;M_{md})} \leq C$ and $\nabla \zeta_l = 0$ on $(1-\frac{1}{l})B$.

Now the function F is continuous, thus uniformly continuous on the compact subsets of M_{md}. Let K be the closed ball of radius C and M_{md}, and let ω_K be the uniform modulus of continuity of F on this compact. This is a nondecreasing function from $[0, 2C]$ into $\mathbb{R}_+$ such that $\omega_K(s) \to 0$ when $s \to 0$ and such that for all pairs of matrices $A, B \in K$,

$$|F(A) - F(B)| \leq \omega_K(|A - B|).$$

Using Eq. (6.24), we obtain

$$F(\nabla u_{n_l,x_0,1/k_l}(y)) = F(A + \nabla(w_l - \zeta_l)(y)) + r_l(y),$$

with

$$|r_l(y)| \leq \omega_K(|\nabla\zeta_l(y)|).$$

Consequently, if we integrate the above equality on B, we obtain

$$\frac{1}{\text{meas}\, B}\int_B F(\nabla u_{n_l,x_0,1/k_l}(y))\,dy = \frac{1}{\text{meas}\, B}\int_B F(A + \nabla(w_l - \zeta_l)(y))\,dy + \frac{1}{\text{meas}\, B}\int_B r_l(y)\,dy. \tag{6.25}$$

For the first term of the right-hand side of (6.25), we use (at last!) the quasiconvexity of F to obtain

$$\frac{1}{\text{meas}\, B}\int_B F(A + \nabla(w_l - \zeta_l)(y))\,dy \geq F(A) = F(\nabla u(x_0)). \tag{6.26}$$

For the remainder, we notice that

$$\begin{aligned}\left|\int_B r_l(y)\,dy\right| &\leq \int_B \omega_K(|\nabla\zeta_l(y)|)\,dy = \int_{B\setminus(1-1/l)B} \omega_K(|\nabla\zeta_l(y)|)\,dy\\ &\leq (1 - (1 - 1/l)^d)(\text{meas}\, B)\,\omega_K(C) \longrightarrow 0 \text{ when } l \to +\infty.\end{aligned} \tag{6.27}$$

since $\nabla\zeta_l = 0$ on $(1 - \frac{1}{l})B$. Putting (6.23), (6.25), (6.26) and (6.27) together, we finally see that

$$g(x_0) \geq F(\nabla u(x_0)),$$

which shows that the functional I is sequentially weakly-$*$ lower semicontinuous on $W^{1,\infty}(\Omega; \mathbb{R}^m)$. □

Remark 6.12. i) The difficulty stems from the fact that, even though weak-star convergence in $W^{1,\infty}$ implies the uniform convergence of the functions, *it in no way implies almost everywhere convergence of the gradients*, quite to the contrary.

ii) We have used weak-$*$ limits of functions of the form $F(\nabla u_n)$, which brings Young measures to mind. Effectively, there exists proofs of this kind of results that are based on Young measures, see [42].

iii) Let us point out that there is no need for growth or bound below hypotheses on F for weak-$*$ lower semicontinuity on $W^{1,\infty}$. □

Let us now turn to a proof of Theorem 6.9. This particular proof is due to [50]. We proceed in four steps. The first step consists in extending the quasiconvexity inequality to $W_0^{1,p}$ functions.

Lemma 6.4. *Let F be a quasiconvex function satisfying the growth and bound below hypotheses of Theorem 6.9. Then, for all $A \in M_{md}$ and $v \in W_0^{1,p}(\Omega;\mathbb{R}^m)$, there holds*

$$\int_\Omega F(A+\nabla v)\,dx \geq \text{meas}\,\Omega\,F(A). \tag{6.28}$$

Proof. By definition, $W_0^{1,p}(\Omega;\mathbb{R}^m)$ is the closure of $\mathscr{D}(\Omega;\mathbb{R}^m)$ in $W^{1,p}(\Omega;\mathbb{R}^m)$. For all $v \in W_0^{1,p}(\Omega;\mathbb{R}^m)$, we can thus find a sequence $\varphi_n \in \mathscr{D}(\Omega;\mathbb{R}^m)$ such that $\nabla\varphi_n \to \nabla v$ almost everywhere and $|\nabla\varphi_n| \leq g$ with $g \in L^p(\Omega)$ by the partial converse of the dominated convergence theorem. Due to the growth and bound below hypotheses, we can thus apply the Lebesgue dominated convergence theorem to deduce that $\int_\Omega F(A+\nabla\varphi_n)\,dx \to \int_\Omega F(A+\nabla v)\,dx$, hence the result. □

The second step consists in dealing with the case when the weak limit is an affine function.

Lemma 6.5. *Let $A \in M_{md}$ be a given matrix, $b \in \mathbb{R}^m$ a given vector and define $u(x) = Ax + b$. Then, for any sequence u_n of functions of $W^{1,p}(\Omega;\mathbb{R}^m)$ such that $u_n \rightharpoonup u$, there holds* $\liminf I(u_n) \geq I(u)$.

Proof. If the boundary values of u_n were the same as those of u, then the result would be an immediate consequence of the $W^{1,p}$ version of quasiconvexity, (6.28), by setting $v = u_n - u$. The idea is to reduce the general case to this particular situation. For this, we use the so-called *slicing* technique invented by De Giorgi.

Let $\Omega_0 \subset \Omega$ be an open set, compactly included in Ω. We set $R = \frac{1}{2}d(\overline{\Omega}_0, \partial\Omega) > 0$, pick an integer k and for $i = 1,\ldots,k$, set

$$\Omega_i = \Big\{x \in \Omega;\, d(x,\Omega_0) < \frac{i}{k}R\Big\},$$

see Fig. 6.5, in which the distances are not drawn exactly for simplicity.

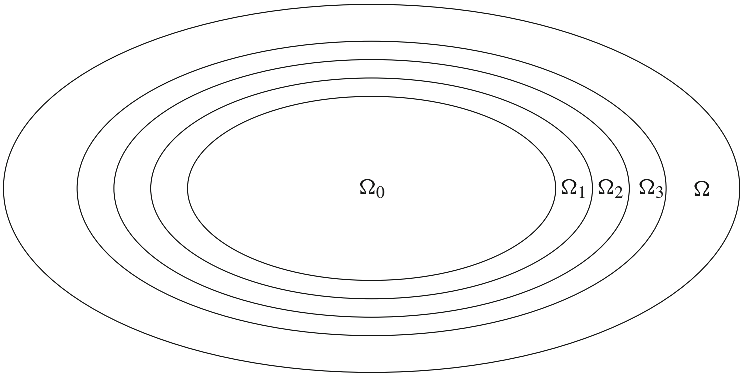


Fig. 6.5 A three-slice slicing

By construction, Ω_i is an open set compactly included in Ω and that contains $\overline{\Omega}_0$. For $i = 1, \dots, k$, there exist C^∞-functions ϕ_i such that

$$0 \leq \phi_i \leq 1, \phi_i = 1 \text{ on } \Omega_{i-1}, \phi_i = 0 \text{ on } \Omega \setminus \Omega_i \text{ and } |\nabla \phi_i| \leq \frac{k+1}{R}.$$

To construct them, we can for example regularize the Lipschitz continuous functions $x \mapsto \frac{d(x,\Omega\setminus\Omega_i)}{d(x,\Omega\setminus\Omega_i)+d(x,\Omega_{i-1})}$.[9] Such functions are called cut-off functions.

We then set

$$v_{n,i} = u + \phi_i(u_n - u).$$

This construction has two crucial properties. One is that $v_{n,i} = u_n$ on Ω_{i-1}, that is to say on most of Ω when Ω_0 is "large", and the second is that it agrees with u in a neighborhood of $\partial\Omega$. Moreover, since $\phi_i(u_n - u) \in W_0^{1,p}(\Omega; \mathbb{R}^m)$, it follows by quasiconvexity that

$$\operatorname{meas}\Omega\, F(A) = \int_\Omega F(\nabla u)\,dx \leq \int_\Omega F(\nabla v_{n,i})\,dx. \tag{6.29}$$

It follows from both crucial properties that

$$\int_\Omega F(\nabla v_{n,i})\,dx = \int_{\Omega\setminus\Omega_i} F(\nabla u)\,dx + \int_{\Omega_i\setminus\Omega_{i-1}} F(\nabla v_{n,i})\,dx + \int_{\Omega_{i-1}} F(\nabla u_n)\,dx. \tag{6.30}$$

[9] Notice that regularization by convolution of a Lipschitz continuous function lowers its Lipschitz constant, hence the estimate on the gradients.

The middle term is an integral over one slice. We note that since $F \geq 0$, the first and last terms satisfy

$$\int_{\Omega\setminus\Omega_i} F(\nabla u)\,dx \leq \int_{\Omega\setminus\Omega_0} F(\nabla u)\,dx \quad \text{and} \quad \int_{\Omega_{i-1}} F(\nabla u_n)\,dx \leq \int_{\Omega} F(\nabla u_n)\,dx. \tag{6.31}$$

Consequently, adding equalities (6.30) from $i = 1$ to $i = k$ after having taken estimates (6.29) and (6.31) into account, and dividing the result by k, we obtain the estimate

$$\int_{\Omega} F(\nabla u)\,dx \leq \int_{\Omega\setminus\Omega_0} F(\nabla u)\,dx + \frac{1}{k}\int_{\Omega_k\setminus\Omega_0}\Big(\sum_i \mathbf{1}_{\Omega_i\setminus\Omega_{i-1}} F(\nabla v_{n,i})\Big)dx + \int_{\Omega} F(\nabla u_n)\,dx \tag{6.32}$$

The first and last terms in the right-hand side are ideally suited to passing to the inferior limit when $n \to +\infty$. We need to control the middle term by showing that it can be made as small as we want in the $n \to +\infty$ limit. For this, we go back to the definition of $v_{n,i}$, which implies that

$$\nabla v_{n,i} = (1-\phi_i)\nabla u + \phi_i \nabla u_n + (u_n - u)\otimes \nabla\phi_i.$$

Consequently,

$$|\nabla v_{n,i}|^p \leq C\bigg(1 + |\nabla u|^p + |\nabla u_n|^p + \Big(\frac{k+1}{R}\Big)^p |u_n - u|^p\bigg).$$

We now use the growth hypothesis on F to deduce that

$$\int_{\Omega_k\setminus\Omega_0}\Big(\sum_i \mathbf{1}_{\Omega_i\setminus\Omega_{i-1}} F(\nabla v_{n,i})\Big)dx \leq C\bigg(1+\|\nabla u\|^p_{L^p(\Omega;M_{md})}+\|\nabla u_n\|^p_{L^p(\Omega;M_{md})} + \Big(\frac{k+1}{R}\Big)^p \|u_n - u\|^p_{L^p(\Omega_k\setminus\Omega_0;M_{md})}\bigg), \tag{6.33}$$

where the constant C does not depend on k.

Now $u_n \rightharpoonup u$, so on the one hand, ∇u_n is bounded in L^p. On the other hand, and this is where the subtlety of the slicing argument lies, $u_n \to u$ in $L^p_{\text{loc}}(\Omega;\mathbb{R}^m)$ strong, by Rellich's theorem. In particular, we see that $\|u_n - u\|_{L^p(\Omega_k\setminus\Omega_0;M_{md})} \to 0$ when $n \to +\infty$. Letting thus n tend to $+\infty$ in estimates (6.32) and (6.33), we obtain,

$$\int_{\Omega} F(\nabla u)\,dx \leq \int_{\Omega\setminus\Omega_0} F(\nabla u)\,dx + \frac{C}{k} + \liminf_{n\to+\infty}\int_{\Omega} F(\nabla u_n)\,dx,$$

with a constant C that still does not depend on k. We conclude by letting first k tend to $+\infty$, then by taking a sequence of open sets Ω_0 such that meas $(\Omega \setminus \Omega_0) \to 0$. □

Remark 6.13. A simpler gluing between u_n and u consisting in using only one cut-off function ϕ would not permit the appearance of the $1/k$ crucial factor of slicing. It is however possible to carry out an argument with one or a small number of cut-off functions, see [45, 46].

Note that the different slices are offset relative to each other, so that the addition of their contributions amounts to an integral over the union of all slices. In this contribution, the terms that are a priori bad are those that contain the gradients of the cut-off functions ϕ_i. These terms are controlled by strong convergence in L^p. In this respect, it is important to work in a compact subset of Ω in order to be able to apply Rellich's theorem in the absence of regularity hypotheses on Ω. □

In a third step, we show that a quasiconvex function with p-growth is actually locally Lipschitz, with an estimate of its Lipschitz constant.

Lemma 6.6. *Let F be a quasiconvex function satisfying the growth estimate of Theorem 6.9. Then there exists a constant C such that for all $A, B \in M_{md}$, there holds*

$$|F(B) - F(A)| \leq C(1 + |A|^{p-1} + |B|^{p-1})|B - A|. \tag{6.34}$$

Proof. We write the proof in the case when F is differentiable. The general case is similar by making use of the property of almost everywhere differentiability of convex functions, i.e., Rademacher's theorem.

Let E_{ij} be the elementary matrix with 0 entries everywhere except at row i, column j, where there is a 1. This is a rank one matrix, thus the function $t \mapsto F(A + tE_{ij})$ is convex. It is therefore larger than its tangent affine mapping,

$$F(A + tE_{ij}) \geq F(A) + t\frac{\partial F}{\partial A_{ij}}(A)$$

for all $t \in \mathbb{R}$. In particular, taking $t = \pm(1 + |A|)$, we deduce that

$$\begin{aligned}\Big|\frac{\partial F}{\partial A_{ij}}(A)\Big| &\leq \frac{\max(|F(A + (1 + |A|)E_{ij}) - F(A)|, |F(A - (1 + |A|)E_{ij}) - F(A)|)}{1 + |A|}\\ &\leq \frac{C(1 + |A|^p)}{1 + |A|} \leq C(1 + |A|^{p-1}).\end{aligned} \tag{6.35}$$

We then write

$$F(B) - F(A) = \int_0^1 DF(tB + (1 - t)A)(B - A)\,dt,$$

and conclude by estimate (6.35) of the derivatives of F. □

In the last step, we treat the case of an arbitrary weak limit by approximating it locally with appropriately matched affine functions.

Lemma 6.7. *Let* $u_n \rightharpoonup u$ *in* $W^{1,p}(\Omega;\mathbb{R}^m)$*. Then* $\liminf I(u_n) \geq I(u)$.

Proof. We first use the fact that if $v \in L^p(\Omega)$, then for all $\varepsilon > 0$, there exists a countable family of disjoint open hypercubes Q_i such that $\overline{\Omega} = \cup_{i\in\mathbb{N}}\bar{Q}_i$ and if we denote by $\bar{v}_i = \frac{1}{\operatorname{meas} Q_i}\int_{Q_i} v(y)\,dy$ the average of v over Q_i, and let $\bar{v}(x) = \sum_{i\in\mathbb{N}} \bar{v}_i \mathbf{1}_{Q_i}(x)$ be the piecewise constant function constructed from these averages, there holds

$$\int_\Omega |v(x) - \bar{v}(x)|^p\,dx = \sum_{i\in\mathbb{N}} \int_{Q_i} |v(x) - \bar{v}_i|^p\,dx \leq \varepsilon^p.$$

Let us sketch the proof of this. We use the density of compactly supported continuous functions in L^p, that the above estimate is easy to obtain for a simple choice of Q_i for a compactly supported continuous function, and finally, that given a family Q_i, the mapping $v \mapsto \bar{v}$ is continuous from $L^p(\Omega)$ into $L^p(\Omega)$. Indeed, it is immediate that

$$\int_{Q_i} |v(x) - \bar{v}_i|^p\,dx \leq \frac{1}{\operatorname{meas} Q_i}\int_{Q_i\times Q_i} |v(x) - v(y)|^p\,dxdy.$$

Therefore, if v is uniformly continuous, it is enough to choose the diameter of the cubes small enough so that $|v(x) - v(y)| \leq \varepsilon \operatorname{meas} \Omega^{-1/p}$ to conclude in this case.

Let us thus pick $\varepsilon > 0$ and $\{Q_i\}_{i\in\mathbb{N}}$ the family of hypercubes associated with ∇u by the previous remark. For each i, we define the sequence of functions on Q_i,

$$u_{n,i}(x) = u_n(x) - u(x) + \overline{(\nabla u)}_i x.$$

By construction, $u_{n,i} \rightharpoonup \overline{(\nabla u)}_i x$ in $W^{1,p}(Q_i;\mathbb{R}^m)$. By Lemma 6.5, there thus holds

$$\liminf_{n\to+\infty} \int_{Q_i} F(\nabla u_{n,i})\,dx \geq \int_{Q_i} F(\overline{(\nabla u)}_i)\,dx,$$

so that, summing over all the hypercubes

$$\liminf_{n\to+\infty} \sum_{i\in\mathbb{N}} \int_{Q_i} F(\nabla u_{n,i})\,dx \geq \sum_{i\in\mathbb{N}} \liminf_{n\to+\infty} \int_{Q_i} F(\nabla u_{n,i})\,dx \geq \sum_{i\in\mathbb{N}} \int_{Q_i} F(\overline{(\nabla u)}_i)\,dx. \tag{6.36}$$

Indeed, the inferior limit of a sum is clearly larger than the sum of inferior limits.

Let us take a look at the left-hand side of this inequality.

$$\begin{aligned}\Big|\int_\Omega F(\nabla u_n)\,dx-\sum_{i\in\mathbb{N}}\int_{Q_i}F(\nabla u_{n,i})\,dx\Big| &= \Big|\int_\Omega\Big[F(\nabla u_n)-\sum_{i\in\mathbb{N}}\mathbf{1}_{Q_i}F(\nabla u_{n,i})\Big]dx\Big|\\ &\le \int_\Omega\sum_{i\in\mathbb{N}}\mathbf{1}_{Q_i}|F(\nabla u_n)-F(\nabla u_{n,i})|\,dx.\end{aligned}$$

Now by Lemma 6.6, we know that

$$\begin{aligned}\sum_{i\in\mathbb{N}}\mathbf{1}_{Q_i}|F(\nabla u_n)-F(\nabla u_{n,i})| &\le C\sum_{i\in\mathbb{N}}\mathbf{1}_{Q_i}(1+|\nabla u_n|^{p-1}+|\nabla u_{n,i}|^{p-1})|\nabla u_n-\nabla u_{n,i}|\\ &= C\sum_{i\in\mathbb{N}}\mathbf{1}_{Q_i}(1+|\nabla u_n|^{p-1}+|\nabla u_{n,i}|^{p-1})|\nabla u-\overline{(\nabla u)}_i|\\ &\le C\Big(\sum_{i\in\mathbb{N}}\mathbf{1}_{Q_i}(1+|\nabla u_n|^{p-1}+|\nabla u_{n,i}|^{p-1})^{\frac{p}{p-1}}\Big)^{\frac{p-1}{p}}\\ &\qquad\times\Big(\sum_{i\in\mathbb{N}}\mathbf{1}_{Q_i}|\nabla u-\overline{(\nabla u)}_i|^p\Big)^{\frac{1}{p}}\\ &\le C\Big(\sum_{i\in\mathbb{N}}\mathbf{1}_{Q_i}(1+|\nabla u_n|^{p}+|\nabla u_{n,i}|^{p})\Big)^{\frac{p-1}{p}}\\ &\qquad\times\Big(\sum_{i\in\mathbb{N}}\mathbf{1}_{Q_i}|\nabla u-\overline{(\nabla u)}_i|^p\Big)^{\frac{1}{p}}\\ &= C(1+|\nabla u_n|^p+|\overline{(\nabla u_n)}|^p)^{\frac{p-1}{p}}\\ &\qquad\times\Big(\sum_{i\in\mathbb{N}}\mathbf{1}_{Q_i}|\nabla u-\overline{(\nabla u)}_i|^p\Big)^{\frac{1}{p}}\end{aligned}$$

using the notation $\overline{(\nabla u_n)}=\sum_{i\in\mathbb{N}}\mathbf{1}_{Q_i}\nabla u_{n,i}$. We have used Hölder's inequality in the meantime. Integrating this over Ω and re-using Hölder's inequality, we obtain

$$\begin{aligned}\Big|\int_\Omega F(\nabla u_n)\,dx-\sum_{i\in\mathbb{N}}\int_{Q_i}F(\nabla u_{n,i})\,dx\Big| &\le C(1+\|\nabla u_n\|_{L^p}^{p-1}+\|\overline{(\nabla u_n)}\|_{L^p}^{p-1})\\ &\qquad\times\Big(\sum_{i\in\mathbb{N}}\int_{Q_i}|\nabla u-\overline{(\nabla u)}_i|^p\,dx\Big)^{\frac{1}{p}}\end{aligned}$$

Now ∇u_n and $\overline{(\nabla u_n)}$ are clearly bounded in $L^p(\Omega;M_{md})$ independently of n and ε, therefore

$$\Big|\int_\Omega F(\nabla u_n)\,dx-\sum_{i\in\mathbb{N}}\int_{Q_i}F(\nabla u_{n,i})\,dx\Big|\le C\varepsilon,$$

which implies that

$$\liminf_{n\to+\infty}\int_\Omega F(\nabla u_n)\,dx \geq \liminf_{n\to+\infty}\sum_{i\in\mathbb{N}}\int_{Q_i} F(\nabla u_{n,i})\,dx - C\varepsilon. \tag{6.37}$$

For the right-hand side of inequality (6.36), we establish in the same way that

$$\sum_{i\in\mathbb{N}}\int_{Q_i} F(\overline{(\nabla u)}_i)\,dx \geq \int_\Omega F(\nabla u)\,dx - C\varepsilon. \tag{6.38}$$

The weak lower semicontinuity then immediately follows from inequalities (6.36)–(6.38), and by letting $\varepsilon \to 0$. □

Remark 6.14. It is possible to give a simpler proof of the last step in the case when Ω is smooth and if we admit the density of piecewise affine functions in $W^{1,p}(\Omega)$. This is a classical result in finite element theory, a theory that is however only interested in low dimensional open sets. In general, it is necessary to use triangulations of open sets of arbitrary dimension.

Assuming this density, we begin by showing the result when u is piecewise affine. let Ω_i, $i = 1, \ldots, k$, be a subdivision of Ω such that u is affine on each Ω_i. There holds

$$\begin{aligned}\liminf_{n\to+\infty}\int_\Omega F(\nabla u_n)\,dx &= \liminf_{n\to+\infty}\sum_{i=1}^k\int_{\Omega_i} F(\nabla u_n)\,dx\\ &\geq \sum_{i=1}^k\liminf_{n\to+\infty}\int_{\Omega_i} F(\nabla u_n)\,dx\\ &\geq \sum_{i=1}^k\int_{\Omega_i} F(\nabla u)\,dx\\ &= \int_\Omega F(\nabla u)\,dx,\end{aligned}$$

due to Lemma 6.5.

Let us now take an arbitrary $u \in W^{1,p}(\Omega;\mathbb{R}^m)$ and v_q a sequence of piecewise affine functions that tends strongly to u in $W^{1,p}(\Omega;\mathbb{R}^m)$. We see that $u_n - u + v_q \rightharpoonup v_q$ in $W^{1,p}(\Omega;\mathbb{R}^m)$ when $n \to +\infty$, so that

$$\liminf_{n\to+\infty}\int_\Omega F(\nabla u_n - \nabla u + \nabla v_q)\,dx \geq \int_\Omega F(\nabla v_q)\,dx,$$

from the above. Now $F(\nabla u_n) = F(\nabla u_n - \nabla u + \nabla v_q) + F(\nabla u_n) - F(\nabla u_n - \nabla u + \nabla v_q)$. Therefore, by Lemma 6.6, there holds

$$\begin{aligned}\liminf_{n\to+\infty} \int_\Omega F(\nabla u_n)\,dx &\geq \int_\Omega F(\nabla v_q)\,dx \\ &- C \limsup_{n\to+\infty} (1 + \|\nabla u_n\|_{L^p} + \|\nabla u\|_{L^p} + \|\nabla v_q\|_{L^p})^{p-1} \|\nabla u - \nabla v_q\|_{L^p},\end{aligned} \tag{6.39}$$

by Hölder's inequality, hence the result by letting q tend to infinity (we use the partial converse of the dominated convergence theorem, then the growth of F and then the dominated convergence theorem itself to pass to the limit in the right-hand side integral, as usual). □

Chapter 7
Calculus of Variations and Critical Points

We now return to semilinear problems from the point of view of the calculus of variations, not only by minimizing a functional as in the previous chapter, but also by looking more generally for critical points of this functional.

If V is a Banach space and $J : V \to \mathbb{R}$ is a functional on V that is differentiable in the sense of Fréchet, a *critical point* of J is an element u of V at which the differential DF of F vanishes and a *regular point* of J is a point u such that $DJ(u) \neq 0$. A *critical value* of J is a real number c such that there exists a critical point $u \in V$ of J with $J(u) = c$. A value that is not critical is called a *regular value* of J. Each point in the preimage of a regular value is regular. Naturally, a minimum point for such a functional is a critical point, but there may exist more critical points that are not minimizers. Finally, to show the existence of a critical point, it is clearly enough to exhibit a critical value.

7.1 Why Look for Critical Points?

Let us go back to the semilinear model problem of Chap. 2. Given Ω an open bounded subset of $\mathbb{R}^d$ and f a function of $C^0(\mathbb{R}) \cap L^\infty(\mathbb{R})$, we are looking for a function $u \in H_0^1(\Omega)$ such that $-\Delta u = f(u)$ in the sense of $\mathscr{D}'(\Omega)$. Let F be the primitive of f on $\mathbb{R}$ that vanishes at 0. It is clear that $|F(t)| \leq \|f\|_{L^\infty(\mathbb{R})}|t|$. We associate with this boundary value problem the functional

$$J(u) = \frac{1}{2}\int_\Omega \|\nabla u\|^2\,dx - \int_\Omega F(u)\,dx. \tag{7.1}$$

H. Le Dret, *Nonlinear Elliptic Partial Differential Equations*, Universitext,
https://doi.org/10.1007/978-3-319-78390-1_7

Proposition 7.1. *The functional J is well defined and of class C^1 on $H_0^1(\Omega)$. Its differential at point u is given by*

$$DJ(u)v = \int_\Omega \nabla u \cdot \nabla v \, dx - \int_\Omega f(u)v \, dx, \tag{7.2}$$

for all $v \in H_0^1(\Omega)$.

Proof. First of all, it is clear that the two integrals in (7.1) are meaningful for any $u \in H_0^1(\Omega)$. For the second one, it is enough to use the upper bound on $|F|$ and the fact that $H_0^1(\Omega) \hookrightarrow L^1(\Omega)$ since Ω is bounded.

Let us now show that J is differentiable in the sense of Fréchet and that its differential is given by (7.2). Since the quadratic part is trivially C^1, it is enough to show that the mapping $u \mapsto I(u) = \int_\Omega F(u)\,dx$ is differentiable. For all $u, v \in H_0^1(\Omega)$, there holds

$$\begin{aligned} I(u+v) - I(u) - \int_\Omega f(u)v\,dx &= \int_\Omega \Bigl(\int_0^1 f(u+tv)v\,dt - f(u)v\Bigr)\,dx \\ &= \int_\Omega \Bigl(\int_0^1 \bigl(f(u+tv) - f(u)\bigr)\,dt\Bigr) v\,dx. \end{aligned}$$

Indeed, $\frac{d}{dt}\bigl(F(u+tv)\bigr) = f(u+tv)v$ almost everywhere in Ω. By the Cauchy-Schwarz inequality, it follows that

$$\Bigl| I(u+v) - I(u) - \int_\Omega f(u)v\,dx \Bigr| \le \Bigl(\int_\Omega \Bigl(\int_0^1 \bigl(f(u+tv) - f(u)\bigr)\,dt\Bigr)^2 dx\Bigr)^{1/2} \|v\|_{L^2(\Omega)}.$$

Now, still by the Cauchy-Schwarz inequality, there holds

$$\Bigl(\int_0^1 \bigl(f(u+tv) - f(u)\bigr)\,dt\Bigr)^2 \le \int_0^1 \bigl(f(u+tv) - f(u)\bigr)^2\,dt.$$

Thus we have obtained the estimate

$$\begin{aligned} \Bigl| I(u+v) - I(u) - \int_\Omega f(u)v\,dx \Bigr| &\le \|f(u+tv) - f(u)\|_{L^2(\Omega\times[0,1])} \|v\|_{L^2(\Omega)} \\ &\le \|f(u+tv) - f(u)\|_{L^2(\Omega\times[0,1])} \|v\|_{H^1(\Omega)}. \end{aligned}$$

To conclude this part of the proof, we now need to show that for any sequence v_n that tends to 0 in $H_0^1(\Omega)$, we have $\|f(u+tv_n) - f(u)\|_{L^2(\Omega\times[0,1])} \to 0$. We start by extracting a subsequence, still denoted v_n, that recovers the upper limit of the latter real-valued sequence and that converges almost everywhere. For this sequence, there holds $|f(u+tv_n) - f(u)|^2 \to 0$ almost everywhere in $\Omega \times [0,1]$ and $|f(u+tv_n) - f(u)|^2 \le 4\|f\|^2_{L^\infty(\mathbb{R})} \in L^1(\Omega \times [0,1])$. The Lebesgue dominated convergence theorem then implies the result.

We now show that the mapping $u \mapsto DJ(u)$ is continuous from $H_0^1(\Omega)$ into its dual $H^{-1}(\Omega)$. The linear part, that is to say the differential of the quadratic part of the functional, does not pose any problem. We just need to deal with $u \mapsto DI(u)$. For all u, h and v in $H_0^1(\Omega)$, there holds

$$\big(DI(u+h) - DI(u)\big)v = \int_\Omega (f(u+h) - f(u))v\,dx.$$

Consequently, the Cauchy-Schwarz inequality implies that

$$\begin{aligned}\big|\big(DI(u+h) - DI(u)\big)v\big| &\leq \|f(u+h) - f(u)\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)}\\ &\leq C\|f(u+h) - f(u)\|_{L^2(\Omega)}\|v\|_{H_0^1(\Omega)}\end{aligned}$$

by also using the Poincaré inequality. Therefore, taking the supremum for v in the unit ball of $H_0^1(\Omega)$, we obtain

$$\|DI(u+h) - DI(u)\|_{H^{-1}(\Omega)} \leq C\|f(u+h) - f(u)\|_{L^2(\Omega)}.$$

Now, Carathéodory's theorem 2.14 asserts that $\|f(u+h) - f(u)\|_{L^2(\Omega)} \to 0$ when $\|h\|_{L^2(\Omega)} \to 0$, which completes the proof. □

Remark 7.1. i) We have in fact shown that I is of class C^1 on $L^2(\Omega)$.

ii) We could also have proved that J is differentiable in the sense of Gateaux (that is to say on straight lines of the form $u + tv$), as in the proof of Theorem 6.5, and then showed that its Gateaux differential is continuous, which implies Fréchet C^1, see for instance [40] □

Corollary 7.1. *Any critical point of J is a solution of the model problem and conversely.*

Proof. Let u be a critical point of J. If we take $v \in \mathscr{D}(\Omega)$ in the equation $DJ(u)v = 0$, we see that u is a solution of the model problem in the sense of distributions. Conversely, if u is such a solution, that is to say if $\langle -\Delta u, \varphi\rangle = \langle f(u), \varphi\rangle$ for all $\varphi \in \mathscr{D}(\Omega)$, we obtain

$$\int_\Omega \nabla u \cdot \nabla\varphi\,dx = \int_\Omega f(u)\varphi\,dx,$$

and we conclude by the density of $\mathscr{D}(\Omega)$ in $H_0^1(\Omega)$. □

Remark 7.2. It is thus equivalent either to solve the model problem or to find critical points of J, which are in fact solutions of the Euler-Lagrange equation associated with the functional J. Let us notice that J is not necessarily convex due to the term $u \mapsto I(u)$ which is not concave, since f has no property of this kind. There thus may exist other critical points than just minimum points, as well as other critical values than just the minimum. □

We are thus going to focus on finding critical points, or equivalently critical values, for rather general functionals J. The applications to partial differential equations are typically semilinear boundary value problems.

7.2 Ekeland's Variational Principle

Let us begin with an abstract result that plays an important role in a certain number of situations in the calculus of variations, called the *Ekeland variational principle*, see [24].

This result only concerns functional minimization, in a very general context and not specifically the search for critical points. We will however see later on how it can also be used in the context of this chapter to actually find critical points.

Theorem 7.1 (Ekeland's Variational Principle). *Let (X, d) be a complete metric space and $J: X \to \mathbb{R}$ a lsc functional, bounded below on X. Let $c = \inf_X J$. Then for all $\varepsilon > 0$, there exists $x_\varepsilon \in X$ such that*

$$\begin{cases} c \leq J(x_\varepsilon) \leq c + \varepsilon, \\ \forall x \in X, x \neq x_\varepsilon,\ J(x) - J(x_\varepsilon) + \varepsilon d(x, x_\varepsilon) > 0. \end{cases} \tag{7.3}$$

Remark 7.3. Let us note right away that Ekeland's principle in this form only has an interest if we do not know yet whether or not the infimum is attained, or if the infimum is not attained. Indeed, if we know that there is a minimum point, it suffices to take such a point for x_ε and the relations (7.3) are then trivially satisfied for all $\varepsilon > 0$.

The set of points $(x, a) \in X \times \mathbb{R}$ such that $a > J(x_\varepsilon) - \varepsilon d(x, x_\varepsilon)$ can be visualized as the complement of a sort of "cone" of slope $-\varepsilon$ in the product space $X \times \mathbb{R}$, as in Fig. 7.1.

The theorem says that the epigraph of J is entirely contained in this set, to the exception of point $(x_\varepsilon, J(x_\varepsilon))$. Careful with the geometrical intuition though: the picture of Fig. 7.1 corresponds to $X = \mathbb{R}$ equipped with its usual distance. An arbitrary metric space has no reason to look even remotely like $\mathbb{R}$ with the usual distance. □

Proof. Let

$$\text{Epi}\, J = \{(x, a) \in X \times \mathbb{R};\ J(x) \leq a\}$$

be the epigraph of J, that is to say the set of points in $X \times \mathbb{R}$ situated "above" the graph of J. This is a closed subset of $X \times \mathbb{R}$ for the product topology, because J is lsc. Let us be given $\varepsilon > 0$. We introduce an order relation on $X \times \mathbb{R}$ in the following way: We say that $(x, a) \preccurlyeq_\varepsilon (y, b)$ if and only if, see Fig. 7.2.,

$$\varepsilon d(x, y) \leq b - a. \tag{7.4}$$

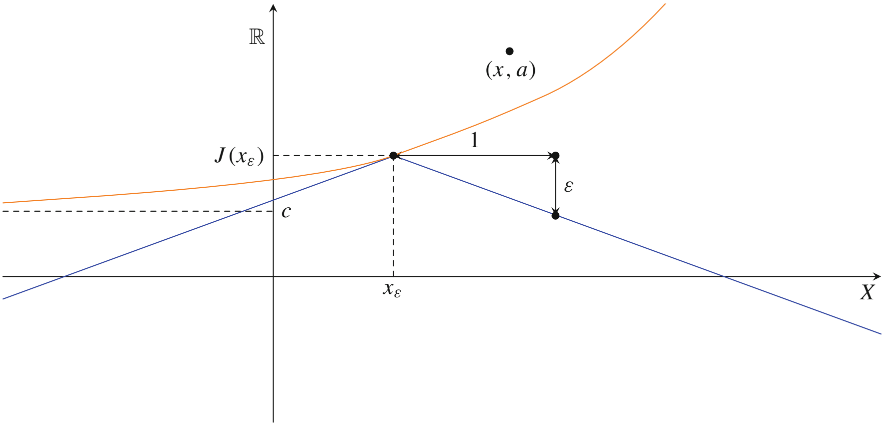


Fig. 7.1 A visualization of Ekeland's variational principle

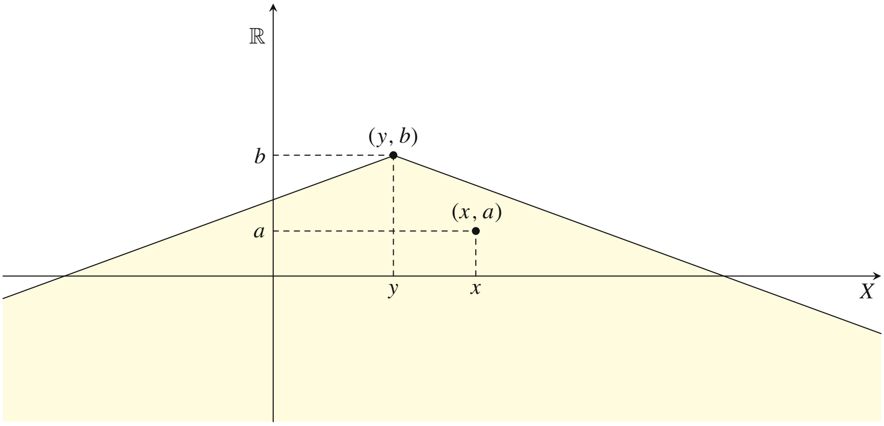


Fig. 7.2 The set of points (x, a) such that $(x, a) \preccurlyeq_\varepsilon (y, b)$, a closed subset of $X \times \mathbb{R}$

This is obviously an order relation, i.e., it is reflexive, transitive because of the triangle inequality, and antisymmetric because $(x, a) \preccurlyeq_\varepsilon (y, b)$ and $(y, b) \preccurlyeq_\varepsilon (x, a)$ imply $b - a \geq 0$ and $a - b \geq 0$, thus $a = b$, thus $d(x, y) = 0$, thus *in fine* $(x, a) = (y, b)$. This relation makes $X \times \mathbb{R}$ into a poset. We are going to construct a totally ordered subset of $X \times \mathbb{R}$.

Let us pick a point $x_1 \in X$ such that

$$c \leq J(x_1) \leq c + \varepsilon.$$

Such a point exists by the very definition of the infimum of a subset of $\mathbb{R}$. We set $a_1 = J(x_1)$ and

$$A_1 = \{(x, a) \in \operatorname{Epi} J; \ (x, a) \preccurlyeq_\varepsilon (x_1, a_1)\}.$$

This is a closed set, as an intersection of closed sets, which is nonempty since $(x_1, a_1) \in A_1$. Let P be the canonical projection of $X \times \mathbb{R}$ onto X. We remark that if $x \in P(A_1)$, then $c \leq J(x) \leq c + \varepsilon$. Indeed, let $a \in \mathbb{R}$ be such that $(x, a) \in A_1$. Since $A_1 \subset \operatorname{Epi} J$, there holds

$$J(x) \leq a \leq a_1 - \varepsilon d(x, x_1) \leq J(x_1) \leq c + \varepsilon.$$

We now argue by induction. Let us assume that we have constructed a sequence (x_i, a_i) for $i = 1, \ldots, n$ such that the nonempty closed sets

$$A_i = \{(x, a) \in \operatorname{Epi} J;\, (x, a) \preccurlyeq_\varepsilon (x_i, a_i)\}$$

are nested, that is to say $A_{i+1} \subset A_i$ for $i \leq n - 1$ and such that setting $c_i = \inf_{x \in P(A_i)} J(x)$, there holds

$$0 \leq a_i - c_i \leq 2^{1-i}(a_1 - c_1).$$

We have just seen that this is doable for $n = 1$, the nesting condition being empty in this case. Now we go on with the induction.

First Case $c_n < a_n$. In this case, we pick $x_{n+1} \in P(A_n)$ such that

$$0 \leq J(x_{n+1}) - c_n \leq \frac{1}{2}(a_n - c_n),$$

a point that exists by the definition of an infimum, and set

$$a_{n+1} = J(x_{n+1}).$$

Let us check that this works. The closed set A_{n+1} is nonempty because it contains (x_{n+1}, a_{n+1}). Let us show that $A_{n+1} \subset A_n$. For this, we take $b_{n+1} \in \mathbb{R}$ such that $(x_{n+1}, b_{n+1}) \in A_n$, which exists from our choice of x_{n+1} as an element of $P(A_n)$, and notice that

$$a_{n+1} \leq b_{n+1} \leq a_n - \varepsilon d(x_n, x_{n+1}),$$

which implies that $(x_{n+1}, a_{n+1}) \preccurlyeq_\varepsilon (x_n, a_n)$, hence the desired inclusion by transitivity of the order relation. Finally, there clearly holds $c \leq c_n \leq c_{n+1}$ since $P(A_{n+1}) \subset P(A_n)$. It follows that

$$0 \leq a_{n+1} - c_{n+1} \leq a_{n+1} - c_n \leq \frac{1}{2}(a_n - c_n) \leq \frac{1}{2^n}(a_1 - c_1),$$

which completes the induction in the first case.

Second Case $c_n = a_n$. In this case, we just take $x_{n+1} = x_n$ and $a_{n+1} = a_n$, and thus trivially satisfy the induction conditions.

We now proceed to show that the diameters of the sets A_n tend to 0 when $n \to +\infty$. Let us thus take $(x, a), (y, b) \in A_n$. Since $a \geq c_n$ and by definition of the order relation, it follows that

$$\varepsilon d(x, x_n) \leq a_n - a \leq a_n - c_n \leq 2^{1-n}(a_1 - c_1),$$

and the same for y, hence by the triangle inequality

$$d(x, y) \leq \frac{1}{\varepsilon} 2^{2-n}(a_1 - c_1) \to 0 \text{ when } n \to +\infty.$$

Besides, since $c_n \leq a, b \leq a_n$, there also holds

$$|a - b| \leq 2^{2-n}(a_1 - c_1) \to 0 \text{ when } n \to +\infty,$$

hence the claim on diameters.

To sum things up so far, we have constructed a countable family of nonempty closed subsets of the complete metric space $X \times \mathbb{R}$ that are nested and whose diameters tend to 0. It follows that the intersection of this family is a singleton, a classical property of complete metric spaces,

$$\bigcap_{n \in \mathbb{N}^*} A_n = \{(x_\varepsilon, a_\varepsilon)\}.$$

By construction, there holds $(x_\varepsilon, a_\varepsilon) \in A_1$, so that

$$c \leq J(x_\varepsilon) \leq a_\varepsilon \leq c + \varepsilon.$$

Let us now show that point $(x_\varepsilon, a_\varepsilon)$ is minimal in Epi J for the $\preccurlyeq_\varepsilon$ order relation. This means that any lesser point in Epi J must be $(x_\varepsilon, a_\varepsilon)$ itself. Let thus $(y, b) \in$ Epi J be such that $(y, b) \preccurlyeq_\varepsilon (x_\varepsilon, a_\varepsilon)$. By construction, the family (x_n, a_n) is totally ordered. Moreover, since $(x_\varepsilon, a_\varepsilon) \in A_n$, it follows that $(x_\varepsilon, a_\varepsilon) \preccurlyeq_\varepsilon (x_n, a_n)$ for all n. By transitivity, we thus see that $(y, b) \preccurlyeq_\varepsilon (x_n, a_n)$ for all n. Now since $(y, b) \in$ Epi J, this implies that $(y, b) \in A_n$ for all n. We have just seen that the intersection of the A_n reduces to $\{(x_\varepsilon, a_\varepsilon)\}$. Consequently, $(y, b) = (x_\varepsilon, a_\varepsilon)$ and the announced minimality is proved, see Fig. 7.3.

We deduce from this that no point in Epi J distinct from $(x_\varepsilon, a_\varepsilon)$ is lesser than $(x_\varepsilon, a_\varepsilon)$ for the order relation. This means in particular that if $x \neq x_\varepsilon$, then

$$\varepsilon d(x_\varepsilon, x) > a_\varepsilon - J(x) \geq J(x_\varepsilon) - J(x),$$

thus completing the proof of the Theorem. □

Remark 7.4. Since $x_n \to x_\varepsilon$, $a_n \to a_\varepsilon$ and $a_n = J(x_n)$, it follows that $a_\varepsilon \geq J(x_\varepsilon)$ by lower semicontinuity of J. But if $a_\varepsilon > J(x_\varepsilon)$, minimality is contradicted by the

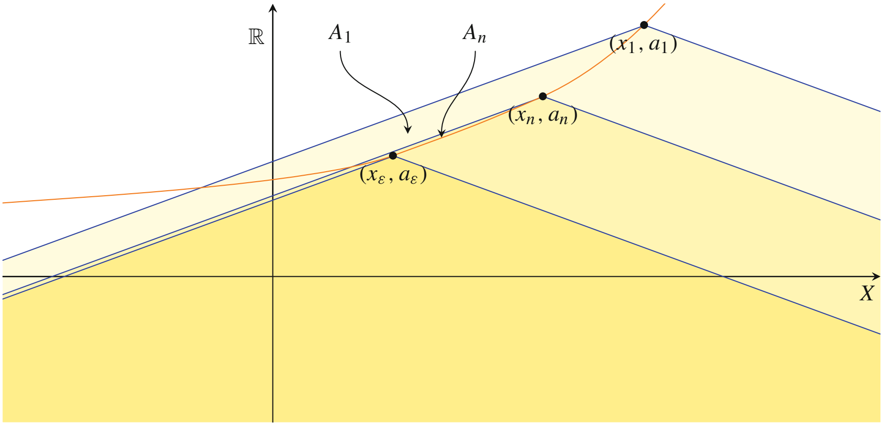


Fig. 7.3 A visualization of how the proof works

fact the $(x_\varepsilon, J(x_\varepsilon))$ is strictly lesser. There thus holds $a_\varepsilon = J(x_\varepsilon)$, that is to say that the minimal point is precisely on the graph of J. □

In the case of a functional of class C^1 on a Banach space, Ekeland's variational principle takes a more striking form, that makes it easier to appreciate its power.

Corollary 7.2. *Let J be a functional of class C^1 on a Banach space V that is bounded below and let $c = \inf_V J$. Then, for all $\varepsilon > 0$, there exists $u_\varepsilon \in V$ such that*

$$\begin{cases} c \le J(u_\varepsilon) \le c + \varepsilon, \\ \|DJ(u_\varepsilon)\|_{V'} \le \varepsilon. \end{cases} \tag{7.5}$$

Proof. In this case, the second relation of (7.3) can be rewritten as

$$J(u) - J(u_\varepsilon) + \varepsilon\|u - u_\varepsilon\|_V > 0$$

for all $u \neq u_\varepsilon$. We take $u = u_\varepsilon + tv$ with $\|v\|_V = 1$ and $t > 0$. There holds

$$J(u_\varepsilon + tv) - J(u_\varepsilon) > -\varepsilon t,$$

so that dividing by $-t$,

$$-\frac{J(u_\varepsilon + tv) - J(u_\varepsilon)}{t} < \varepsilon.$$

Since J is differentiable, if we let t go to 0, we obtain that

$$-DJ(u_\varepsilon)v \le \varepsilon,$$

then by changing v into $-v$ that

$$|DJ(u_\varepsilon)v| \le \varepsilon,$$

for all $v \in V$ such that $\|v\|_V = 1$. This implies the result by the definition of the dual norm. □

Remark 7.5. i) The assumption that J is of class C^1 is not necessary. Clearly, it is enough for J to be Gateaux-differentiable.

ii) It is instructive take a peek at example ii) of Remark 7.7 in Sect. 7.3, in the light of this version of Ekeland's variational principle. We can even spice it up a little bit by looking at $J(u) = \sin u^2 + \frac{1}{1+u^2}$. □

There is also a local version of the previous results which is equally striking.

Corollary 7.3. *Let J be a lsc, bounded below functional on a complete metric space X and let $c = \inf_X J$. Consider $x_\varepsilon \in X$ such that $c \le J(x_\varepsilon) \le c + \varepsilon$. Then there exists $\bar{x}_\varepsilon \in X$ such that*

$$\begin{cases} c \le J(\bar{x}_\varepsilon) \le c + \varepsilon, \\ d(\bar{x}_\varepsilon, x_\varepsilon) \le 2\sqrt{\varepsilon}, \\ \forall x \in X, x \neq \bar{x}_\varepsilon,\ J(x) - J(\bar{x}_\varepsilon) + \sqrt{\varepsilon} d(x, \bar{x}_\varepsilon) > 0. \end{cases} \tag{7.6}$$

Proof. We follow exactly the same proof as in the first version, with the following slight modification of the order relation:

$$(x, a) \preccurlyeq_\varepsilon (y, b) \text{ if and only if } \sqrt{\varepsilon} d(x, y) \le b - a. \tag{7.7}$$

We then perform the same construction starting from $x_1 = x_\varepsilon$, and the estimates $\sqrt{\varepsilon} d(x, x_n) \le 2^{1-n}(a_1 - c_1) \le 2^{1-n}\varepsilon$ for all $x \in A_n$ make it possible to deduce that $d(x_n, x_\varepsilon) = d(x_n, x_1) \le 2(1 - 2^{-n})\sqrt{\varepsilon}$, hence the result by passing to the limit when $n \to +\infty$. □

This result is again better appreciated in its differential version.

Corollary 7.4. *Let V be a Banach space, J a C^1 functional bounded below on a closed subset F of V and set $c = \inf_F J$. Let $u_\varepsilon \in F$ be such that $c \le J(u_\varepsilon) \le c+\varepsilon$. Then there exists $\bar{u}_\varepsilon \in F$ such that*

$$\begin{cases} c \le J(\bar{u}_\varepsilon) \le c + \varepsilon, \\ \|\bar{u}_\varepsilon - u_\varepsilon\|_V \le 2\sqrt{\varepsilon}, \\ \forall u \in F, u \neq \bar{u}_\varepsilon,\ J(u) - J(\bar{u}_\varepsilon) + \sqrt{\varepsilon}\|u - \bar{u}_\varepsilon\|_V > 0. \end{cases} \tag{7.8}$$

Moreover if $\bar{u}_\varepsilon$ belongs to the interior of F, then

$$\|DJ(\bar{u}_\varepsilon)\|_{V'} \le \sqrt{\varepsilon}. \tag{7.9}$$

Proof. The first part (7.8) is but an immediate rewriting of Corollary 7.3 in the complete metric space F. The second part (7.9) follows the same argument as in the proof of Corollary 7.2. □

Remark 7.6. Corollary 7.4 shows that if we are given a point u_ε where J is almost minimized by at most $\varepsilon > 0$, then there exists very close to this point, at a distance at most of the order of $\sqrt{\varepsilon}$, another point $\bar{u}_\varepsilon$ where J is also almost minimized—in fact taking a lower value—and where the differential of J almost vanishes! Which is quite surprising when you think about it. □

7.3 The Palais-Smale Condition

A crucial ingredient in minimizing a functional of the calculus of variations is the compactness of minimizing sequences, for a certain topology. The Palais-Smale condition plays quite a similar role for sequences on which the functional takes values that tend to a potential critical value, which is not necessarily the infimum of the functional. This is an a priori condition, to be checked on a case by case basis for each functional, independently of the existence or nonexistence of critical values. The Palais-Smale condition will be an essential tool in showing such existence in certain situations.

Definition 7.1. Let V be a Banach space and $J\colon V \to \mathbb{R}$ be a functional of class C^1. We say that J verifies the *Palais-Smale condition* (at level c) if given any sequence u_n of V such that

$$J(u_n) \to c \text{ in } \mathbb{R} \quad \text{and} \quad DJ(u_n) \to 0 \text{ in } V',$$

we can extract from it a convergent subsequence.

Remark 7.7. i) The Palais-Smale condition does not prejudge the existence of a critical value or the existence of such a sequence, also called a Palais-Smale sequence. It just says that if we happen to have one such Palais-Smale sequence at hand, then the image of this sequence is necessarily relatively compact.

ii) The two Palais-Smale hypotheses are independent from one another. Indeed, even if $c = \inf_V J$, we can easily have a minimizing sequence u_n such that $DJ(u_n) \not\to 0$. Just take $V = \mathbb{R}$, $J(u) = \sin u^2$, $c = -1$ and $u_n = \big(\frac{3\pi}{2} + n2\pi + \frac{1}{\sqrt{n2\pi}}\big)^{1/2}$. There holds $J(u_n) \to -1$ and $J'(u_n) \to 2$. Likewise, it is easy to construct an example in which $DJ(u_n) = 0$ and $J(u_n)$ does not converge.

iii) An important difference with the previous chapter is that we are working here with the strong topology of V, instead of with the weak topology.

iv) There are several variants of the Palais-Smale condition to be found in the literature, see [40]. □

Corollary 7.2 of Sect. 7.2 immediately suggests to make joint use of the Palais-Smale condition and of Ekeland's variational principle.

Theorem 7.2. *Let J be a functional of class C^1 on a Banach space V that is bounded below and satisfies the Palais-Smale condition. Then J attains its minimum.*

Proof. This is almost obvious. We take $\varepsilon = \frac{1}{n}$ and Ekeland's variational principle ensures the existence of a minimizing sequence u_n such that $DJ(u_n) \to 0$ in V'. Due to the Palais-Smale condition, this sequence has a convergent subsequence, that thus converges to a minimum point. □

This is only a sufficient condition. The example of Fig. 7.5 does not satisfy the Palais-Smale condition, which does not prevent it from attaining its minimum.

Let us now give a few examples of functions that satisfy or do not satisfy the Palais-Smale condition. First of all, it is clear that the function $J(u) = e^u$ defined on $V = \mathbb{R}$ does not satisfy the Palais-Smale condition at level $c = 0$. It does however verify it for all other real values, simply because there are no Palais-Smale sequences at these levels. It is instructive to look at other finite dimensional examples, such as those pictured in Figs. 7.4 and 7.5.

Here is a more interesting example in the context of partial differential equations. Let Ω be a bounded, regular open subset of $\mathbb{R}^d$. By Theorems 1.23 and 1.24, if

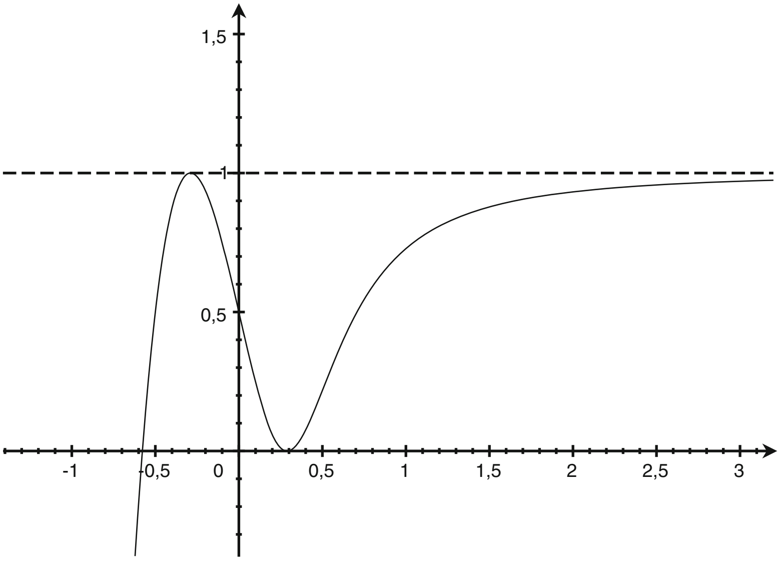


Fig. 7.4 The Palais-Smale condition is satisfied at level $c = 0$ but not at level $c = 1$

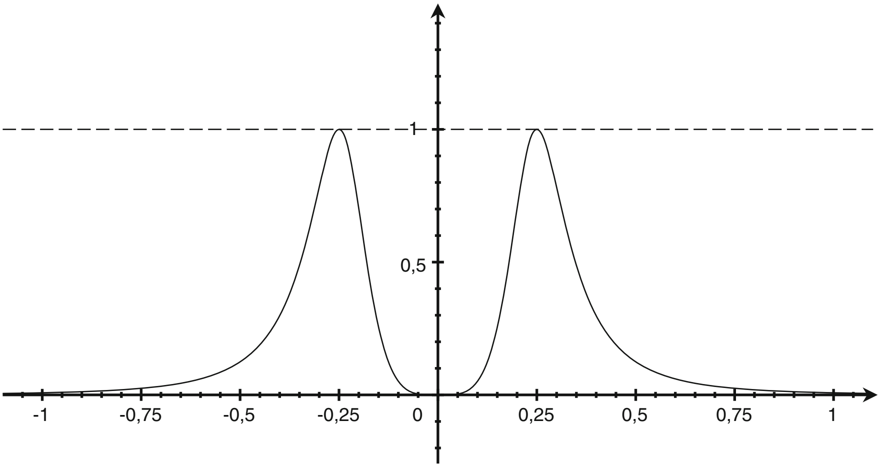


Fig. 7.5 The Palais-Smale condition is satisfied at level $c = 1$ but not at level $c = 0$

$T\colon L^2(\Omega) \to L^2(\Omega)$ denotes the operator that associates with $f \in L^2(\Omega)$, the solution $u \in H_0^1(\Omega)$ of $-\Delta u = f$, i.e., $T = (-\Delta)^{-1}$, there exists a sequence $\lambda_k > 0$ that tends to $+\infty$ such that for $\lambda \neq \lambda_k$, $\lambda \neq 0$, the operator $T_\lambda = (-\Delta)^{-1} - \frac{1}{\lambda}\mathrm{Id}$ is an isomorphism of $L^2(\Omega)$.

It follows from this that the restriction of T_λ to $H_0^1(\Omega)$ is injective. It is moreover surjective on $H_0^1(\Omega)$. Indeed, for all $u \in H_0^1(\Omega)$, there exists a unique $f \in L^2$ such that $T_\lambda f = u$, so that $f = \lambda(-\Delta)^{-1} f - \lambda u$, which implies that $f \in H_0^1(\Omega)$. Finally, since $(-\Delta)^{-1}$ is continuous from $L^2(\Omega)$ into $H_0^1(\Omega)$ by the Lax-Milgram theorem, it is fortiori continuous from $H_0^1(\Omega)$ into $H_0^1(\Omega)$. Consequently, the restriction of T_λ to $H_0^1(\Omega)$ is an isomorphism.

Proposition 7.2. *For all $f \in L^2(\Omega)$, the functional on $H_0^1(\Omega)$*

$$J(u) = \frac{1}{2}\int_\Omega \big(\|\nabla u\|^2 - \lambda u^2\big)\,dx - \int_\Omega f u\,dx$$

satisfies the Palais-Smale condition at all levels c if $\lambda \neq \lambda_k$ for all k, and does not satisfy it for $c = 0$ and $f = 0$, if $\lambda = \lambda_k$ for a given k.

Proof. As a rare exception to the usual rule, we identify here the dual of $H_0^1(\Omega)$ with $H_0^1(\Omega)$ itself via its inner product,[1] and not with $H^{-1}(\Omega)$. For all $v \in H_0^1(\Omega)$, there holds

$$(DJ(u)|v) = \int_\Omega \big(\nabla u \cdot \nabla v - \lambda uv\big)\,dx - \int_\Omega f v\,dx,$$

[1] The differential DJ is thus identified here with the gradient.

which means that $DJ(u) \in H_0^1(\Omega)$ is the unique solution of the variational problem

$$\int_\Omega \nabla(DJ(u)) \cdot \nabla v \, dx = \int_\Omega \big(\nabla u \cdot \nabla v - \lambda uv\big) \, dx - \int_\Omega fv \, dx,$$

for all $v \in H_0^1(\Omega)$. In terms of partial differential equations, this entails that

$$-\Delta(DJ(u)) = -\Delta u - \lambda u - f,$$

which can be rewritten as

$$DJ(u) = u - \lambda(-\Delta)^{-1}u - (-\Delta)^{-1}f \in H_0^1(\Omega).$$

Let us consider a sequence u_n in $H_0^1(\Omega)$ such that $DJ(u_n) \to 0$ in $H_0^1(\Omega)$ when $n \to +\infty$ (the other condition actually plays no role whatsoever). For $\lambda \neq 0$, $\lambda \neq \lambda_k$ for all k, there thus holds

$$DJ(u_n) = -\lambda\Big((-\Delta)^{-1}u_n - \frac{1}{\lambda}u_n\Big) - (-\Delta)^{-1}f = -\lambda T_\lambda u_n - (-\Delta)^{-1}f.$$

Now we have seen that T_λ is an isomorphism, therefore

$$u_n = -\frac{1}{\lambda}T_\lambda^{-1}(DJ(u_n) + (-\Delta)^{-1}f) \to -\frac{1}{\lambda}T_\lambda^{-1}((-\Delta)^{-1}f) \text{ when } n \to +\infty,$$

and the Palais-Smale condition is satisfied. For $\lambda = 0$, we have directly $u_n = DJ(u_n) + (-\Delta)^{-1}f \to (-\Delta)^{-1}f$, hence the Palais-Smale condition is also satisfied.

Let us now assume that $\lambda = \lambda_k$ for a certain k and let us just for simplicity consider the case $f = 0$. There exists an associated nonzero eigenfunction $\varphi_k \in H_0^1(\Omega)$ such that $-\Delta\varphi_k = \lambda_k\varphi_k$. The sequence $u_n = n\varphi_k$ is such that $J(u_n) = 0$, $DJ(u_n) = 0$, but certainly does not contain any convergent subsequence. □

Remark 7.8. Let us notice that as soon as $\lambda > \lambda_1$, then the functional J is not bounded below on $H_0^1(\Omega)$ (consider the sequence $n\varphi_1$). This does not prevent it from satisfying the Palais-Smale condition. □

Let us now turn to an even more interesting example, markedly less linear than the previous example in terms of the underlying PDE.

Proposition 7.3. *Let Ω be a bounded open subset of $\mathbb{R}^d$ and $1 < p < \frac{d+2}{d-2}$ for $d \geq 3$, $1 < p < +\infty$ for $d \leq 2$. Then the functional on $H_0^1(\Omega)$*

$$J(u) = \frac{1}{2}\int_\Omega \|\nabla u\|^2 \, dx + \frac{1}{p+1}\int_\Omega |u|^{p+1} \, dx$$

satisfies the Palais-Smale condition at all levels c.

Proof. Under the hypotheses made on p, there holds $p+1 < 2^*$, where 2^* is the Sobolev critical exponent ($2^* = 2d/(d-2)$ for $d \geq 3$), thus $H_0^1(\Omega) \hookrightarrow L^{p+1}(\Omega)$ and the functional is well defined on $H_0^1(\Omega)$.

We show with similar arguments as those used before, that this functional is of class C^1 with

$$DJ(u)v = \int_\Omega \nabla u \cdot \nabla v \, dx + \int_\Omega |u|^{p-1} uv \, dx,$$

so that identifying the dual of $H_0^1(\Omega)$ with $H^{-1}(\Omega)$ as usual this time, we obtain

$$DJ(u) = -\Delta u + |u|^{p-1}u.$$

Let us be given a sequence u_n such that $J(u_n) \to c$ and $DJ(u_n) \to 0$ in $H^{-1}(\Omega)$ when $n \to +\infty$. Since u_n belongs to $H_0^1(\Omega)$, there holds

$$\begin{aligned} DJ(u_n)u_n &= \int_\Omega \|\nabla u_n\|^2 \, dx + \int_\Omega |u_n|^{p+1} \, dx \\ &= (p+1)J(u_n) - \frac{p-1}{2} \int_\Omega \|\nabla u_n\|^2 \, dx. \end{aligned}$$

Now, by definition of the dual norm, it follows that

$$|DJ(u_n)u_n| \leq \|DJ(u_n)\|_{H^{-1}(\Omega)} \|\nabla u_n\|_{L^2(\Omega)}.$$

We consequently derive the estimate

$$\frac{p-1}{2} \|\nabla u_n\|^2_{L^2(\Omega)} \leq (p+1)J(u_n) + \|DJ(u_n)\|_{H^{-1}(\Omega)} \|\nabla u_n\|_{L^2(\Omega)}.$$

We thus see that for n large enough so that $J(u_n) \leq c+1$ and $\|DJ(u_n)\|_{H^{-1}(\Omega)} \leq 1$, the quantity $X_n = \|\nabla u_n\|_{L^2(\Omega)}$ satisfies the inequality

$$\frac{p-1}{2} X_n^2 - X_n - (p+1)(c+1) \leq 0.$$

Since $p-1 > 0$, it follows that X_n is bounded above independently of n, which implies that u_n is bounded in $H_0^1(\Omega)$. The assumption $p+1 < 2^*$ implies that the Sobolev embedding $H_0^1(\Omega) \hookrightarrow L^{p+1}(\Omega)$ is compact. We can thus extract a subsequence, still denoted u_n, and find a $u \in H_0^1(\Omega)$ such that

$$u_n \rightharpoonup u \text{ in } H_0^1(\Omega) \quad \text{and} \quad u_n \to u \text{ in } L^{p+1}(\Omega).$$

We remark that if $v \in L^{p+1}(\Omega)$, then trivially $|v|^p \in L^{\frac{p+1}{p}}(\Omega)$, and arguing as in the proof of Carathéodory's theorem, we realize that

$$|u_n|^{p-1}u_n \to |u|^{p-1}u \text{ in } L^{\frac{p+1}{p}}(\Omega).$$

Now the Hölder conjugate exponent of $p+1$ is nothing but $\frac{p+1}{p}$ and since $H_0^1(\Omega) \hookrightarrow L^{p+1}(\Omega)$, it follows by duality that $L^{\frac{p+1}{p}}(\Omega) \hookrightarrow H^{-1}(\Omega)$. Finally, we have obtained that

$$-\Delta u_n = DJ(u_n) - |u_n|^{p-1}u_n \to -|u|^{p-1}u \text{ in } H^{-1}(\Omega).$$

We have already noted that $(-\Delta)^{-1}$ is an isomorphism from $H^{-1}(\Omega)$ into $H_0^1(\Omega)$, which shows that

$$u_n \to (-\Delta)^{-1}(-|u|^{p-1}u) \text{ in } H_0^1(\Omega),$$

and the Palais-Smale condition is satisfied. □

Remark 7.9. We have in fact obtained an additional information, which is that $u = (-\Delta)^{-1}(-|u|^{p-1}u)$, or again that $-\Delta u = -|u|^{p-1}u$ with $u \in H_0^1(\Omega)$. We could then think that we had succeeded in solving this particular semilinear boundary value problem. Unfortunately, this is not at all the case, because we have not shown that such a Palais-Smale sequence u_n exists! Actually, there always exists one such sequence, but it is not very thrilling, since it is $u_n = 0$ for all n, and we have thus only succeeded in recovering the fact that $-\Delta 0 = -|0|^{p-1}0$.

In order to make this proposition say something interesting as a direct consequence, we would have for example to show the existence of a Palais-Smale sequence at a level $c \neq 0$, which would then preclude $u = 0$, and imply the existence of a nontrivial solution to the semilinear problem.

Now of course, this could prove to be a little difficult here, since this particular functional is strictly convex, and thus has just one critical value 0 and only one critical point $u = 0$. So the only interest of the proposition is actually to show an example of how one can prove that the Palais-Smale condition is satisfied in the case of such functionals, typically associated with semilinear elliptic boundary value problems.

In particular, note that one of the crucial ingredients in the above proof is how the quantities $DJ(u_n)u_n$ and $J(u_n)$ relate to one another and how this can imply the first estimates for a Palais-Smale sequence. □

7.4 The Deformation Lemma

In order to proceed further, we need a way of constructing critical values that are not just infima. We are going to make extensive use of the following notation:

$$\{J \leq c\} = \{u \in V;\, J(u) \leq c\}, \{J > c\} = \{u \in V;\, J(u) > c\}, \text{ etc.}$$

The crucial remark concerning critical values is that something changes in the topological nature of the sets $\{J \leq c\}$ when c crosses a critical value. Thus for example, when $c = \min_V J$, then for all $\varepsilon > 0$, $\{J \leq c-\varepsilon\} = \emptyset$ whereas $\{J \leq c+\varepsilon\}$ is not empty. For an exemple that is a little subtler, take $V = \mathbb{R}^2$ and $J(x) = x_1x_2$. The critical value $c = 0$, which is not a mimimum, is such that $\{J \leq -\varepsilon\}$ has two connected components, whereas $\{J \leq \varepsilon\}$ is connected, see Fig. 7.6.

We actually realize that the sets $\{J \leq -\varepsilon\}$ and $\{J \leq -\varepsilon'\}$ with $\varepsilon > 0$, $\varepsilon' > 0$ are homeomorphic to each other, whereas the sets $\{J \leq -\varepsilon\}$ and $\{J \leq \varepsilon'\}$ are not. The reason is that there is a critical value, $c = 0$, between $-\varepsilon$ and ε'. The following deformation lemma gives a precise content to this observation.

Theorem 7.3. *Let V be a Banach space and $J \colon V \to \mathbb{R}$ a functional of class C^1 satisfying the Palais-Smale condition. Let $c \in \mathbb{R}$ be a regular value of J. Then there*

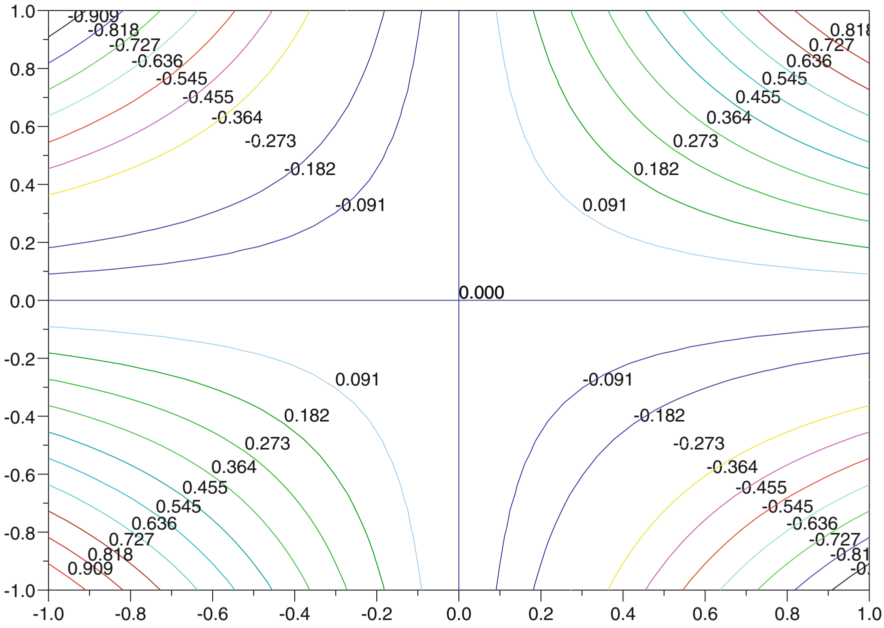


Fig. 7.6 The level lines of $J(x) = x_1x_2$

exists $\varepsilon_0 > 0$ such that for all ε with $0 < \varepsilon < \varepsilon_0$, there exists a homeomorphism $\eta\colon V \to V$ satisfying

i) For all $u \in \{J \le c - \varepsilon_0\} \cup \{J \ge c + \varepsilon_0\}$, there holds $\eta(u) = u$,

ii) There holds $\eta(\{J \le c + \varepsilon\}) \subset \{J \le c - \varepsilon\}$.

See Figs. 7.7 and 7.8 for an illustration of the meaning of Theorem 7.3. We first give a technical lemma.

Lemma 7.1. *Let V, W be Banach spaces, $f\colon V \to \mathbb{R}$ and $g\colon V \to W$ two locally Lipschitz continuous mappings. Then the mapping fg is locally Lipschitz continuous. If f is locally bounded away from 0, then $1/f$ is locally Lipschitz continuous.*

Proof. Let $u \in V$ and $B(u,r)$ be an open ball on which f and g are Lipschitz continuous with respective Lipschitz constants λ_f and λ_g. Both f and g are bounded on $B(u,r)$ by Lipschitz continuity. For instance, $|f(v)| \le |f(u)| + \lambda_f r$ for all $v \in B(u,r)$. For all v, w in U, we write $(fg)(v) - (fg)(w) = f(v)(g(v) - g(w)) + (f(v) - f(w))g(w)$ and estimate

$$\begin{aligned}\|(fg)(v) - (fg)(w)\|_W &\le |f(v)|\|g(v) - g(w)\|_W + |f(v) - f(w)|\|g(w)\|_W \\ &\le \Big(\lambda_g \sup_{B(u,r)} |f| + \lambda_f \sup_{B(u,r)} \|g\|_W\Big)\|v - w\|_V,\end{aligned}$$

hence the result in the case of the product.

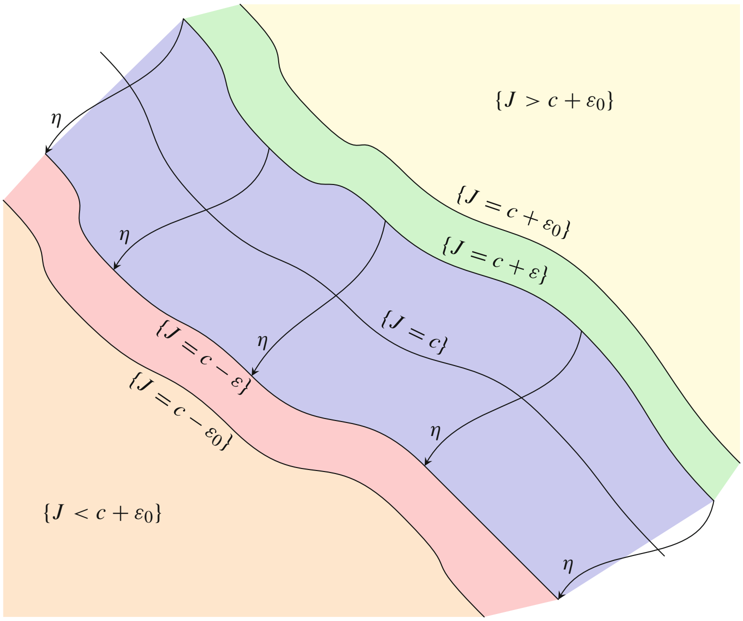

$\{J > c + \varepsilon_0\}$
$\{J = c + \varepsilon_0\}$
$\{J = c + \varepsilon\}$
$\{J = c\}$
$\{J = c - \varepsilon\}$
$\{J = c - \varepsilon_0\}$
$\{J < c + \varepsilon_0\}$
η
η
η
η
η

Fig. 7.7 Illustrating the deformation lemma, before…

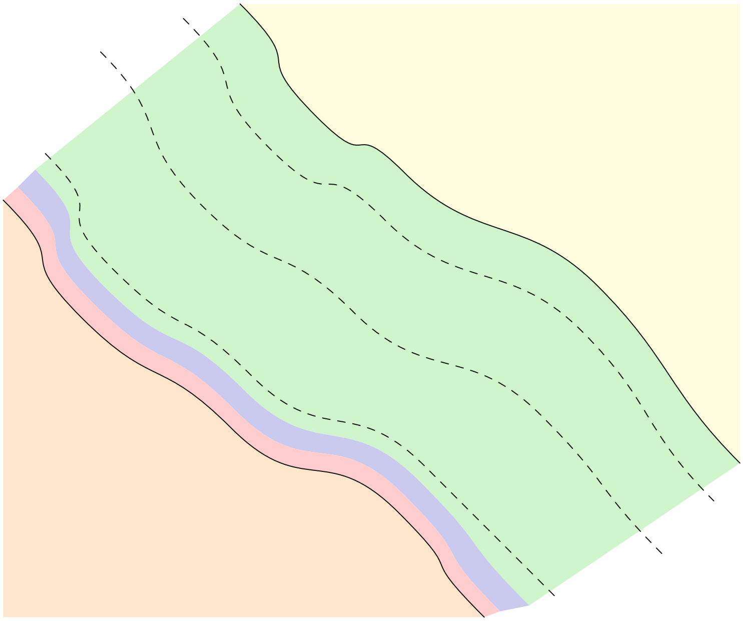


Fig. 7.8 . . . and after

Assuming now that $|f(v)| \geq \alpha > 0$ on $B(u, r)$. Then

$$\left|\frac{1}{f(v)} - \frac{1}{f(w)}\right| \leq \frac{\lambda_f}{\alpha^2}\|v - w\|_V,$$

hence the result in the case of the inverse. □

We now return to the proof of the deformation lemma, Theorem 7.3.

Proof. We are only going to write the complete proof in the case when V is a Hilbert space and when J is of class $C^{1,1}_{\text{loc}}$. We will afterwards give a few indications on how to treat the general case, see also [40]. Since V is a Hilbert space, we identify V and V' via the inner product and thus identify differentials and gradients.

Let c be regular value of J. We now show that there exists $\varepsilon_0 > 0$ and $\delta > 0$ such that $\|DJ(u)\|_V \geq \delta$ for all $u \in \{c - \varepsilon_0 \leq J \leq c + \varepsilon_0\}$. We argue by contradiction, thus assuming there exists a sequence $u_n \in \{c - 1/n \leq J \leq c + 1/n\}$ such that $\|DJ(u_n)\|_V \to 0$. Now J satisfies the Palais-Smale condition, therefore we can extract a subsequence that converges to a certain u. By continuity of J, there holds $J(u) = c$ and by continuity of DJ, that $DJ(u) = 0$. In other words, c is a critical value of J, which contradicts the initial hypothesis. This is the sole spot in the proof where the Palais-Smale condition intervenes, albeit crucially.

Let us now pick $0 < \varepsilon < \varepsilon_0$. All the objects introduced from then on should be indexed by ε, but we do not do it in order to keep a simpler notation. The sets

$$A = \{J \leq c - \varepsilon_0\} \cup \{J \geq c + \varepsilon_0\} \text{ and } B = \{c - \varepsilon \leq J \leq c + \varepsilon\}$$

are closed and disjoint. We denote by $d(u, A) = \inf_{v \in A} \|u - v\|_V$ the distance from u to the set A. It follows that the function $\gamma \colon V \to \mathbb{R}_+$,

$$\gamma(u) = \frac{d(u, A)}{d(u, A) + d(u, B)}$$

is such that $0 \leq \gamma(u) \leq 1$, $\gamma(u) = 0$ if and only if $u \in A$ and $\gamma(u) = 1$ if and only if $u \in B$.

Let us now show that this function is locally Lipschitz on V. By Lemma 7.1, it is enough to show that the function $u \mapsto d(u, A) + d(u, B)$ is locally bounded below by a strictly positive constant, since the mappings $v \mapsto d(v, A)$ and $v \mapsto d(v, B)$ are 1-Lipschitz. To see this, we note that if $u \in V \setminus B$, then there exists $r > 0$ such that $B(u, r) \subset V \setminus B$ since B closed, thus $d(v, A) + d(v, B) \geq d(u, B) - r/2 \geq r/2$ for all $v \in B(u, r/2)$. If $u \in B$, then $u \in V \setminus A$ and the same argument applies replacing B with A.

At this point, we set

$$\Phi(u) = -\gamma(u) \frac{DJ(u)}{\max(\|DJ(u)\|_V, \delta)}.$$

This mapping is well defined from V into V. By Lemma 7.1, it is locally Lipschitz, since DJ is assumed to be locally Lipschitz. Moreover, there clearly holds $\|\Phi(u)\|_V \leq 1$ for all u. Additionally, $\Phi(u) = 0$ on A and $\Phi(u) = -DJ(u)/\|DJ(u)\|_V$ on B.

By the Picard-Lindelöf theorem, the Cauchy problem

$$\begin{cases} \dfrac{dz}{dt} = \Phi(z), \\ z(0) = u, \end{cases}$$

has a unique solution $t \mapsto z(t)$ for all $u \in V$, which is defined on an interval of the form $]t_{\min}, t_{\max}[$. Now the right-hand side of the ODE is bounded which implies that $t_{\min} = -\infty$ and $t_{\max} = +\infty$. We let $\eta_t(u) = z(t)$.

The solution of an ordinary differential equation depends continuously on its initial value, therefore η_t is continuous for all t. Furthermore, still by Picard-Lindelöf, we have the one parameter group property

$$\eta_t(\eta_s(u)) = \eta_{t+s}(u).$$

It follows that η_t is invertible and that its inverse is η_{-t}, which is also continuous. We have thus shown that η_t is a homeomorphism on V for all t.

We now remark that if $u \in A$, then $\eta_t(u) = u$ for all t since $\Phi(u) = 0$, by Picard-Lindelöf uniqueness.

Let us now choose an appropriate value of t. First of all, we remark that J is nonincreasing along the trajectories $\eta_t(u)$. Indeed,

$$\frac{d}{dt}J(\eta_t(u)) = \Big(DJ(\eta_t(u))\Big|\frac{d}{dt}\eta_t(u)\Big) = (DJ(\eta_t(u))|\Phi(\eta_t(u)))$$
$$= -\gamma(\eta_t(u))\frac{\|DJ(\eta_t(u))\|_V^2}{\max(\|DJ(\eta_t(u))\|_V, \delta)} \leq 0.$$

In particular, if $u \in \{J \leq c - \varepsilon\}$, then $J(\eta_t(u)) \leq c - \varepsilon$ for all $t \geq 0$. The set we are interested in may be decomposed as $\{J \leq c + \varepsilon\} = \{J \leq c - \varepsilon\} \cup B$ with $B = \{c - \varepsilon < J \leq c + \varepsilon\}$. It is therefore sufficient to consider the set $\eta_t(B)$. Let us thus take $u \in B$. Let

$$t_0(u) = \sup\{t \geq 0; \forall s \leq t, J(\eta_s(u)) \geq c - \varepsilon\}.$$

Since the function $t \mapsto J(\eta_t(u))$ is continuous nonincreasing, we have $t_0(u) \in [0, +\infty]$. Thus, for all $0 \leq t \leq t_1 \leq t_0(u)$, there holds $\eta_t(u) \in B$, in which case

$$\frac{d}{dt}J(\eta_t(u)) = -\|DJ(\eta_t(u))\|_V \leq -\delta.$$

Integrating this inequality between 0 and t_1, we obtain

$$J(\eta_{t_1}(u)) - J(u) \leq -\delta t_1,$$

so that

$$t_1 \leq \frac{1}{\delta}\big(J(u) - J(\eta_{t_1}(u))\big) \leq \frac{1}{\delta}(c + \varepsilon - (c - \varepsilon)) = \frac{2\varepsilon}{\delta}.$$

It follows on the one hand that $t_0(u) \leq 2\varepsilon/\delta$, with $J(\eta_{t_0(u)}(u)) = c - \varepsilon$ by continuity, and on the other hand, since the upper bound on $t_0(u)$ is uniform with respect to u, that $\eta_{2\varepsilon/\delta}(B) \subset \{J \leq c - \varepsilon\}$. We then let $\eta = \eta_{2\varepsilon/\delta}$, which completes the proof of the deformation lemma. □

Remark 7.10. i) We have proved a little bit more than what is asserted in the deformation lemma: We end up not only with a homeomorphism, but with a homotopy since η_t depends continuously on t. Of course, this is just the flow of the ODE we started from.

ii) There are numerous variants of the deformation lemma. For instance, we can impose that $\eta(\{J \leq c + \varepsilon\}) = \{J \leq c - \varepsilon\}$.

iii) Even though the Palais-Smale condition only made a very discreet appearance in the proof of the deformation lemma, the latter nonetheless depends crucially on the former. We take the version of remark ii) and consider the function $J(x) = \frac{x}{1+x^2}$ on $V = \mathbb{R}$. It does not satisfy the Palais-Smale condition at level $c = 0$. In addition, $c = 0$ is a regular value of J. Nonetheless, $\{J \leq \varepsilon\}$ is certainly not homeomorphic to $\{J \leq -\varepsilon\}$ because the latter is compact, whereas the former is not. The same counterexample works with the version of Theorem 7.3, but it is slightly harder to see. □

Let us now give a few indications for the general case. When V is not a Hilbert space, there is no way of identifying V and V', and the proof, which requires a V-valued vector field as a right-hand side for the ODE, does not work as is. Even if V is a Hilbert space, but J is only C^1 instead of $C^{1,1}_{\text{loc}}$, then the right-hand side is not locally Lipschitz and the Picard-Lindelöf theorem does not apply.

In both cases, we replace the gradient of J in the definition of the function Φ by what is called a *pseudo-gradient*. Let us say a few words about pseudo-gradients.

Definition 7.2. Let V be a Banach space and $J \in C^1(V;\mathbb{R})$. We say that $v \in V$ is a *pseudo-gradient*of J at u if

$$\|v\|_V \leq 2\|DJ(u)\|_{V'} \text{ and } \langle DJ(u), v\rangle \geq \|DJ(u)\|^2_{V'}. \tag{7.10}$$

Let V_r be the set of regular points of J. A mapping $v\colon V_r \to V$ is a *pseudo-gradient field* for J if it is locally Lipschitz and for all $u \in V_r$, $v(u)$ is a pseudo-gradient of J at u.

Let us remark that if V is a Hilbert space and J is of class $C^{1,1}_{\text{loc}}$, then DJ is visibly a pseudo-gradient field, which is in addition defined on the whole of V.

Lemma 7.2. *Let V be a Banach space and $J \in C^1(V;\mathbb{R})$. There exists a pseudo-gradient field for J.*

Proof. Let $u \in V_r$. We will first show the existence of a pseudo-gradient v_u at u. For this, we note that by definition of the dual norm, $\|DJ(u)\|_{V'} = \sup_{\|v\|_V=1}\langle DJ(u), v\rangle \neq 0$. There thus exists $w_u \in V$ such that $\|w_u\|_V = 1$ and $\langle DJ(u), v\rangle > \frac{2}{3}\|DJ(u)\|_{V'}$. Setting then $v_u = \frac{3}{2}\|DJ(u)\|_{V'}w_u$, it follows that

$$\|v_u\|_V = \frac{3}{2}\|DJ(u)\|_{V'} < 2\|DJ(u)\|_{V'} \text{ and } \langle DJ(u), v_u\rangle > \|DJ(u)\|^2_{V'}.$$

Since both inequalities above are strict, by continuity of DJ, there exists an open subset U_u of V such that for all $z \in U_u$,

$$\|v_u\|_V \leq 2\|DJ(z)\|_{V'} \text{ and } \langle DJ(z), v_u\rangle \geq \|DJ(z)\|^2_{V'}.$$

In other words, v_u is also a pseudo-gradient at any point z of U_u. Let us remark that since $v_u \neq 0$, $U_u \subset V_r$ as follows from the first inequality above. We have thus constructed an open covering of $V_r = \cup_{u\in V_r}U_u$ with for each $u \in V_r$, a pseudo-gradient v_u constant in each U_u. The question now is to glue this constant vectors together to construct a locally Lipschitz field.

We use for this the fact that every metric space is *paracompact*, see [21]. This means that every open covering admits a locally finite open refinement. In our context, this translates as the existence of another open covering of V_r, $\{\omega_\lambda\}_{\lambda\in\Lambda}$ such that for all $\lambda \in \Lambda$, there exists $u_\lambda \in V_r$ such that $\omega_\lambda \subset U_{u_\lambda}$, and that for any point u of V_r, there exists an open neighborhood O of u such that $O \cap \omega_\lambda = \emptyset$ except for a finite number of indices $\lambda \in \Lambda$, which is the locally finite character of the new covering.

We then set $\psi_\lambda(u) = d(u, V_r \setminus \omega_\lambda)$. The function ψ_λ is 1-Lipschitz, its support is exactly $\bar{\omega}_\lambda$ and ψ_λ is locally bounded below by a strictly positive number in ω_λ because if $B(u,r) \subset \omega_\lambda$, then $\psi_\lambda \geq r$ on $B(u,r)$. Moreover, the sum $\sum_{\mu\in\Lambda} \psi_\mu$ is locally finite, hence locally Lipschitz and locally bounded below by a strictly positive number because of the fact that $\{\omega_\lambda\}_{\lambda\in\Lambda}$ is a covering of V_r. We now define on V_r,

$$\theta_\lambda = \frac{\psi_\lambda}{\sum_{\mu\in\Lambda} \psi_\mu},$$

so that $0 \leq \theta_\lambda \leq 1$, $\sum_{\mu\in\Lambda} \theta_\mu = 1$, $\operatorname{supp}\theta_\lambda = \bar{\omega}_\lambda$ and θ_λ is locally Lipschitz.[2]

We can now set

$$v(u) = \sum_{\lambda\in\Lambda} \theta_\lambda(u) v_{u_\lambda}.$$

This is a locally finite sum, hence it is well defined and locally Lipschitz. At each point u, its value is a convex combination of a finite number of pseudo-gradients at u. Indeed, $\theta_\lambda(u) > 0$ if and only if $u \in \omega_\lambda$, with $\omega_\lambda \subset U_{u_\lambda}$, a set on which v_{u_λ} is a (constant) pseudo-gradient. Now it is clear that any convex combination of pseudo-gradients is a pseudo-gradient, and the proof is complete. □

The rest of the proof of the deformation lemma follows along the same lines as in the Hilbert and $C^{1,1}_{\text{loc}}$ case, using a pseudo-gradient in place of the gradient, see Exercise 3.

7.5 The Min-Max Principle and the Mountain Pass Lemma

The deformation lemma thus allows for a topological characterization of regular values that is only based on the values taken by the functional itself and not on its differential. It is used in the context of finding critical points via the

[2] See the proof of Theorem 7.3 for the details in the case of two sets.

following min-max principle. For all $c \in \mathbb{R}$ and $\varepsilon_0 > 0$, we introduce a set of homeomorphisms of V

$$D_c^{\varepsilon_0} = \{\eta; \forall u \in \{J \leq c - \varepsilon_0\} \cup \{J \geq c + \varepsilon_0\}, \eta(u) = u\}.$$

This is of course condition i) of Theorem 7.3.

Theorem 7.4. *Let $\mathscr{A}$ be a nonempty set of subsets of V and let*

$$c = \inf_{A \in \mathscr{A}} \sup_{u \in A} J(u).$$

We assume that $c \in \mathbb{R}$ and that J satisfies the Palais-Smale condition at level c. We assume in addition that there exists α such that $\mathscr{A}$ is stable under $D_c^{\varepsilon_0}$ for all $\alpha \geq \varepsilon_0 > 0$, that is to say that if $A \in \mathscr{A}$, then $\eta(A) \in \mathscr{A}$ for all $\eta \in D_c^{\varepsilon_0}$. Then c is a critical value of J.

Proof. We argue by contradiction. We thus assume that c is a regular value of J. Let ε_0 be given by the deformation lemma, which we can always assume to be less than α, and pick $\varepsilon \in]0, \varepsilon_0[$. By definition of an infimum, there exists $A \in \mathscr{A}$ such that

$$\sup_{u \in A} J(u) \leq c + \varepsilon,$$

which can be otherwise formulated as $A \subset \{J \leq c + \varepsilon\}$. By the deformation lemma, there exists $\eta \in D_c^{\varepsilon_0}$ such that $\eta(\{J \leq c + \varepsilon\}) \subset \{J \leq c - \varepsilon\}$, hence a fortiori $A' = \eta(A) \subset \{J \leq c - \varepsilon\}$, which implies that $\sup_{u \in A'} J(u) \leq c - \varepsilon$. Now the hypothesis is that $A' \in \mathscr{A}$, which contradicts the definition of c as the infimum of such suprema. □

Remark 7.11. i) If J satisfies the Palais-Smale condition, so does $-J$. We thus have a similar result with the quantities

$$d = \sup_{A \in \mathscr{A}} \inf_{u \in A} J(u).$$

ii) If we take $\mathscr{A} = \{\{u\}, u \in V\}$ the set of singletons, then we recover that fact that a functional satisfying the Palais-Smale condition and bounded below attains its infimum. By the previous remark, if it is bounded above, then it attains its supremum.

iii) The use of the min-max principle rests on specific choices for $\mathscr{A}$. The latter is generally taken as a class of subsets that share a common topological invariant (genus, category, homotopy class, homology class, and so on) that may be conserved by the flow η_t, see [58]. □

We now give a first application of the min-max principle, which is known under the colorful name of mountain pass lemma.

Theorem 7.5. *Let $J \in C^1(V;\mathbb{R})$ satisfying the Palais-Smale condition and such that*

i) $J(0) = 0$,

ii) There exists $R > 0$ and $a > 0$ such that if $\|u\|_V = R$ then $J(u) \geq a$,

iii) There exists $v \in V$, $\|v\|_V > R$, such that $J(v) < a$.

Then J admits a critical value $c \geq a$.

Proof. We apply the min-max principle with, as a choice for $\mathscr{A}$, the set of images of all continuous paths joining 0 to v, i.e.,

$$\mathscr{A} = \{A = \gamma([0,1]); \gamma \in C^0([0,1]; V), \gamma(0) = 0, \gamma(1) = v\},$$

and we take $\alpha = \frac{1}{2}\min\{a, a - J(v)\} > 0$.

Let γ be a continuous path joining 0 to v and $A = \gamma([0,1])$ the associated element of $\mathscr{A}$. Since the function $t \mapsto \|\gamma(t)\|_V$ is continuous from $[0,1]$ into $\mathbb{R}$, is 0 at $t = 0$ and $\|v\|_V$ at $t = 1$, the intermediate value theorem implies that there exists $s \in [0,1]$ such that $\|\gamma(s)\|_V = R$. Consequently, $\sup_A J(u) \geq a$, and therefore

$$c = \inf_{A \in \mathscr{A}} \sup_{u \in A} J(u) \geq a.$$

In addition, it is clear that $c < +\infty$ since the image of a path is compact in V, so that the supremum of J on each path is attained.[3]

Let us pick ε_0 such that $0 < \varepsilon_0 \leq \alpha$. There holds $D_c^{\varepsilon_0} \subset D_c^\alpha$. Let us consider a homeomorphism η in D_c^α. By construction, we have $J(0) = 0 \leq a - \alpha \leq c - \alpha$ and $J(v) = a + J(v) - a \leq a - \alpha \leq c - \alpha$, so that $\eta(0) = 0$ and $\eta(v) = v$ by definition of D_c^α. Consequently, $\eta \circ \gamma(0) = 0$ and $\eta \circ \gamma(1) = v$ and since $\eta \circ \gamma$ is a continuous path, this is equivalent to saying that $\eta(A) \in \mathscr{A}$. We thus see that $\mathscr{A}$ is stable under D^{ε_0} for all $0 < \varepsilon_0 \leq \alpha$ and the min-max principle implies that c is a critical value of J. □

Remark 7.12. i) The expression "mountain pass" lemma is better understood if we interpret conditions i) to iii) geometrically, or rather geographically, in the case when $V = \mathbb{R}^2$ and $J(u)$ represents the altitude of a point on the surface of such a flat earth that projects vertically on u. Conditions i) and ii) mean that the origin is located in a bowl surrounded by mountains which are all at least as high as a. Condition iii) means that beyond these mountains, there exists a lower point v, let's say in a valley.

It then seems intuitively clear that if we want to go continuously from 0 to v, the best way to do it is to go through a mountain pass, which is bound to exist. In fact, the min-max construction tells us what to do: look at the maximum altitude reached on each path and find a path that minimizes this maximum altitude among all paths.

If we take a path that culminates at an altitude that is a regular value, then the deformation lemma constructs another better path for us, that culminates strictly

[3] Or since $J \circ \gamma$ is continuous on $[0,1]$.

COPYRIGHTED BY E. A. HEGG
CORDOVA ALASKA
983
"Over Chilkoot Pass during The Gold Rush in Alaska.
Thousands of Gold Seekers used This Trail."

Fig. 7.9 The mountain pass lemma rush

lower. If a path culminates at the infimum of the altitudes of the culminating points of all paths, then its altitude is a critical value and we are at a mountain pass (Fig. 7.9).

ii) The mountaineering intuition must however be taken with a grain of salt. Thus the mountain pass lemma is true even without the Palais-Smale condition when $V = \mathbb{R}$, by the intermediate value and the Rolle theorems. It fails without the Palais-Smale condition in dimensions higher than 2. There may exist no mountain pass because the infimum of the path maximum altitude is not attained.

Here is an example of this counter-intuitive situation. Let us consider the function in two variables

$$J(x_1, x_2) = x_1^2(1 + x_2)^3 + x_2^4$$

which is such that

$$DJ(x_1, x_2) = \begin{pmatrix} 2x_1(1 + x_2)^3 \\ 3x_1^2(1 + x_2)^2 + 4x_2^3 \end{pmatrix}.$$

By inspection of this gradient, we see that there is only one critical point on $\mathbb{R}^2$, namely the origin, where $J = 0$. This critical point is a strict local minimum because $J(x_1, x_2) \sim x_1^2 + x_2^4$ in a neighborhood of 0. The origin is thus actually situated in a bowl surrounded by mountains. It is possible to go even lower outside of the bowl, since $\inf_{\mathbb{R}^2} J = -\infty$. This is thus an example of a function with just one critical point which is a local, but not global, minimum, see Fig. 7.10.

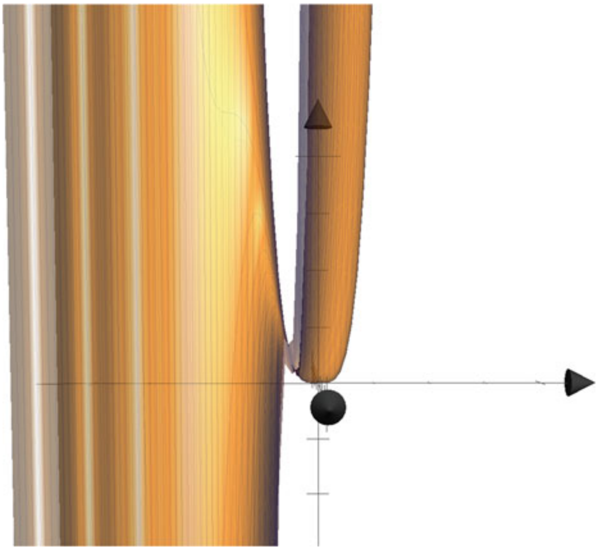


Fig. 7.10 The graph of J, seen from afar

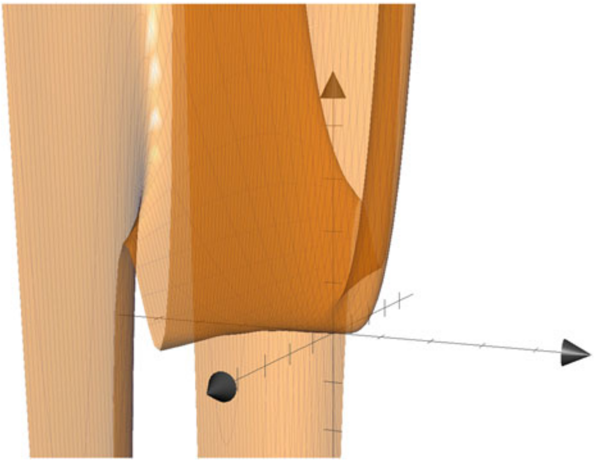


Fig. 7.11 The bowl of J, seen from a little closer

Since there is no other critical point than the local minimum, this means that there exists no mountain pass to exit the bowl and go down into the valley beyond, see Fig. 7.11. This can only happen if the minimizing paths go to infinity. Such a loss of compactness is of course linked to the fact that J does not satisfy the Palais-Smale condition at the level of the inf-max.

If we look at what is going on on the curves $x_1 = \pm\frac{2|x_2|^{3/2}}{\sqrt{3}(1+x_2)}$, parametrized by $-1 < x_2 \leq 0$, we see that $x_1 \to \pm\infty$, $DJ \to 0$ and $J \to 1$ when $x_2 \to -1$. By

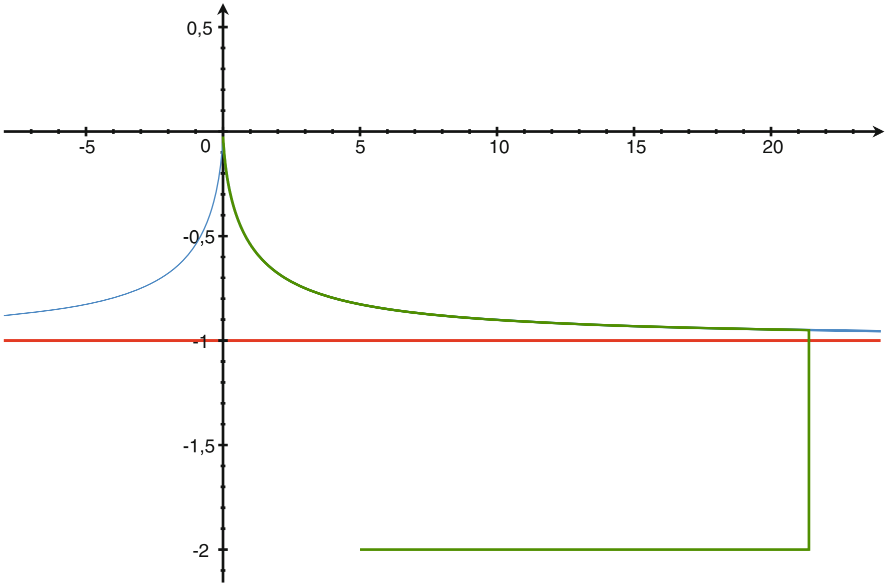


Fig. 7.12 In blue, the curves on which we can pick a non relatively compact Palais-Smale sequence, in green, a path going from $(0, 0)$ a $(5, -2)$ almost reaching the inf-max $(n = 20)$

moving along these curves, we can thus construct Palais-Smale sequences at level $c = 1$ that are not relatively compact, for example by setting $(x_2)_n = -1 + \frac{1}{n}$. Moreover, it can be checked that the level $c = 1$ is precisely the inf-max of J on paths exiting the bowl.

To see this, we can for example join $(0, 0)$ to $(5, -2)$ (just to pick a point where $J < 0$) by following the above curve up to $x_2 = -1 + \frac{1}{n}$, go down $x_2 = -2$ while keeping x_1 constant, then join $x_1 = 5$ with $x_2 = -2$ constant. On the curve itself $J < 1$, on the x_1 constant segment, its maximum tends to 1 when $n \to +\infty$, and on the $x_2 = -2$ segment, $J < 0$, see Fig. 7.12.

We have given above the traditional form of the mountain pass lemma. Actually, hypothesis ii) of the mountain pass lemma turns out to sometimes be automatically satisfied, due to Ekeland's variational principle.

Corollary 7.5. *Let J be as in Theorem 7.5, except hypothesis ii) and such that* 0 *is a strict local minimum of J. Then, hypothesis ii) is satisfied.*

Proof. Let $r > 0$ be such that for all $u \in B(0, r)\setminus\{0\}$, there holds $J(u) > 0$. Let $S_{r/2}$ be the sphere of half radius. We argue by contradiction. Assume that $\inf_{S_{r/2}} J = 0$. There thus exists a sequence v_n such that

$$v_n \in S_{r/2} \quad \text{and} \quad J(v_n) \leq \frac{r^2}{16n^2}.$$

Let us apply Corollary 7.4 of the Ekeland variational principle on the closed set $F = \bar{B}(0, r)$. We thus obtain a sequence $\bar{v}_n \in \bar{B}(0, r)$ such that

$$\|\bar{v}_n - v_n\|_V \leq \frac{r}{2n} \quad \text{and} \quad J(\bar{v}_n) \leq \frac{r^2}{16n^2},$$

plus the other condition of the Corollary. Now, as soon as $n > 1$, we see that $\|\bar{v}_n\|_V \leq \frac{r}{2}\bigl(1 + \frac{1}{n}\bigr) < r$ so that $\bar{v}_n$ belongs to the interior of F. It follows that

$$\|DJ(\bar{v}_n)\|_{V'} \leq \frac{r}{4n}.$$

The sequence $\bar{v}_n$ is thus a Palais-Smale sequence. It admits a convergent subsequence that tends to a certain $\bar{v}$, which is obviously such that $\bar{v} \in S_{r/2}$ and $J(\bar{v}) = 0$. This contradicts the fact that 0 is a point of strict minimum in this ball.

We thus have $\inf_{S_{r/2}} J = a > 0$, which is hypothesis ii) of the mountain pass lemma. □

An immediate corollary of this is that, if J has two strict local minima and satisfies the Palais-Smale condition, then J admits a third critical point.

We now apply the mountain pass lemma to a concrete example of semilinear boundary value problem. Let Ω be a bounded open subset of $\mathbb{R}^d$, $d \geq 3$, and $\lambda_1 > 0$ be the first eigenvalue of $-\Delta$ in $H_0^1(\Omega)$. We are given a function $g \in C^0(\mathbb{R}; \mathbb{R})$ and call G its primitive vanishing at 0. This function is such that

i) $g(0) = 0$,

ii) $\limsup_{s\to 0} \frac{g(s)}{s} < \lambda_1$,

iii) there exist $\theta > 2$, $R > 0$ such that $0 < \theta G(s) \leq sg(s)$ for $|s| \geq R$,

iv) $\frac{g(s)}{|s|^{\frac{d+2}{d-2}}} \to 0$ when $s \to \pm\infty$.

We are going to try to solve the problem: Find $u \in H_0^1(\Omega)$ such that

$$-\Delta u = g(u) \text{ in } \Omega. \tag{7.11}$$

Of course, this problem always has the trivial solution $u = 0$ and we are going to look for another, nontrivial solution!

We proceed in a sequence of lemmas. The letters C, C', etc., denote strictly positive generic constants whose value may change from line to line. First of all, let us give a few elementary properties of the function g stemming from hypotheses i) to iv).

Lemma 7.3. *Let g be a function satisfying hypotheses* i) *to* iv). *Then*

a) For all $\varepsilon > 0$, there exists a constant $C(\varepsilon) \geq 0$ such that

$$\forall s \in \mathbb{R}, \quad |g(s)| \leq \varepsilon |s|^{\frac{d+2}{d-2}} + C(\varepsilon).$$

b) For all $|s| \geq R$, there holds

$$C|s|^\theta \leq |G(s)| \leq C'|s|^{2^*}.$$

c) $\theta \leq 2^* = \frac{2d}{d-2}$.
d) For all $|s| \geq R$, there holds $|g(s)| \geq C|s|^{\theta-1}$.

Proof. For a), we use hypothesis iv). Indeed, for all $\varepsilon > 0$, there exists $M > 0$ such that for $|s| \geq M$, $\frac{|g(s)|}{|s|^{\frac{d+2}{d-2}}} \leq \varepsilon$. We then set $C(\varepsilon) = \big(\max_{[-M,M]}(|g(s)| - \varepsilon|s|^{\frac{d+2}{d-2}})\big)_+$.

For the rightmost inequality of b), we take $\varepsilon = 1$ in the estimate of a). For $s \geq R$, we see that

$$|G(s)| \leq \int_0^s \big(|t|^{\frac{d+2}{d-2}} + C(1)\big)\,dt = \frac{1}{2^*}s^{2^*} + C(1)s \leq \Big(\frac{1}{2^*} + \frac{C(1)}{R^{2^*-1}}\Big)s^{2^*}.$$

For $s \leq -R$, we change s into $-s$. For the leftmost inequality, we use hypothesis iii). For $s \geq R$, there holds

$$\big(s^{-\theta}G(s)\big)' = s^{-\theta}g(s) - \theta s^{-\theta-1}G(s) = s^{-\theta-1}(sg(s) - \theta G(s)) \geq 0.$$

Consequently,

$$s^{-\theta}G(s) \geq R^{-\theta}G(R),$$

hence

$$G(s) \geq R^{-\theta}G(R)s^\theta,$$

with $G(R) > 0$. Likewise for $s \leq -R$.

Inequality c) follows immediately from the fact that $Cs^\theta \leq C's^{2^*}$ for s large, and from the fact that $C > 0$.

Inequality d) follows from hypothesis iii) and from b), since

$$g(s) \geq \frac{\theta G(s)}{s} \geq Cs^{\theta-1}$$

for $s \geq R$ and likewise for $s \leq -R$. □

Since $\theta > 2$, we say that G has super-quadratic growth at infinity and that g is super-linear (Fig. 7.13).

Using these properties and the same kind of arguments as before, it is easy to see that finding solutions of problem (7.11) is equivalent to finding the critical points of the functional

$$J(u) = \frac{1}{2}\int_\Omega \|\nabla u\|^2\,dx - \int_\Omega G(u)\,dx,$$

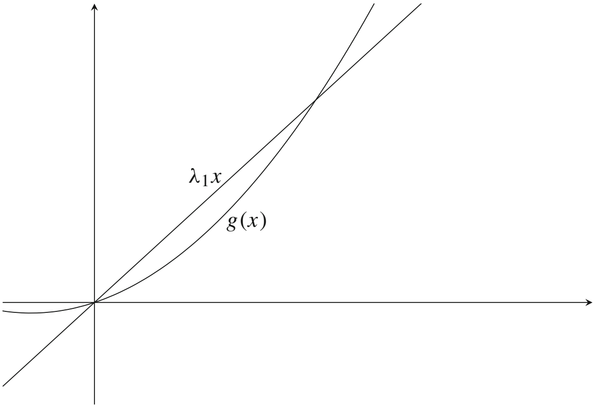


Fig. 7.13 The graph of a typical g

which is well defined and of class C^1 on $H_0^1(\Omega)$ and whose differential is given by

$$DJ(u) = -\Delta u - g(u) \text{ in the sense of } H^{-1}(\Omega).$$

We first give a compactness result.

Lemma 7.4. *Let Ω be a bounded open subset of $\mathbb{R}^d$, $g \in C^0(\mathbb{R};\mathbb{R})$, $Q \in C^0(\mathbb{R};\mathbb{R}_+)$ such that $\frac{|g(s)|}{Q(s)} \to 0$ when $s \to \pm\infty$. We are given a sequence of measurable functions u_n that tends almost everywhere to u and is such that $\int_\Omega Q(u_n)^p\,dx \leq C$ for a certain $p \geq 1$. Then $g(u_n) \to g(u)$ in $L^p(\Omega)$ strong.*

Proof. By Egorov's theorem, for all $\varepsilon > 0$, there exists a measurable set $\Omega_\varepsilon \subset \Omega$ such that meas $(\Omega \setminus \Omega_\varepsilon) \leq \varepsilon$ and u_n converges uniformly to u on Ω_ε.

Since $|g(s)|^p \leq \varepsilon'^p Q(s) + C(\varepsilon')$, we can write on the complement of Ω_ε,

$$\begin{aligned}\int_{\Omega\setminus\Omega_\varepsilon} |g(u_n) - g(u)|^p\,dx &\leq 2^{p-1}\int_{\Omega\setminus\Omega_\varepsilon} (|g(u_n)|^p + |g(u)|^p)\,dx\\ &\leq \varepsilon'^d \int_{\Omega\setminus\Omega_\varepsilon} (Q(u_n)^p + Q(u)^p)\,dx + \text{meas}\,(\Omega\setminus\Omega_\varepsilon)C(\varepsilon')^p\\ &\leq C\varepsilon'^p + \varepsilon C(\varepsilon')^p,\end{aligned}$$

for all n. Indeed, $\int_\Omega Q(u)^p\,dx \leq C$ by Fatou's lemma. We first choose ε' to make the first term small, then we choose ε to make the second term small. Once ε is set, there holds

$$\int_{\Omega_\varepsilon} |g(u_n) - g(u)|^p\,dx \to 0 \text{ quand } n \to +\infty.$$

by uniform convergence, hence the result. □

Lemma 7.5. *The functional J satisfies the Palais-Smale condition at all levels.*

Proof. Let u_n be a sequence such that $J(u_n) \to c$ and $DJ(u_n) \to 0$ in $H^{-1}(\Omega)$. Noticing that $\theta G(s) \leq sg(s) - C$ for all s for some C, we obtain

$$\begin{aligned}\langle DJ(u_n), u_n\rangle &= \int_\Omega \|\nabla u_n\|^2\,dx - \int_\Omega g(u_n)u_n\,dx \\ &\leq \int_\Omega \|\nabla u_n\|^2\,dx - \theta \int_\Omega G(u_n)\,dx + C \operatorname{meas}\Omega \\ &= \theta J(u_n) - \Big(\frac{\theta}{2} - 1\Big)\int_\Omega \|\nabla u_n\|^2\,dx + C \operatorname{meas}\Omega.\end{aligned}$$

Since $\theta > 2$, it follows as in the proof of Proposition 7.3 that u_n is uniformly bounded in $H_0^1(\Omega)$.

We can thus extract a subsequence still denoted u_n and find a $u \in H_0^1(\Omega)$ such that

$$u_n \rightharpoonup u \text{ in } H_0^1(\Omega) \quad \text{and} \quad u_n \to u \text{ a.e. in } \Omega.$$

We then set $Q(s) = |s|^{\frac{d+2}{d-2}}$ and $p = \frac{2d}{d+2} = 1 + \frac{d-2}{d+2} \geq 1$. There holds

$$\int_\Omega Q(u_n)^p\,dx = \int_\Omega |u_n|^{2^*}\,dx \leq C$$

by the Sobolev embeddings. The compactness Lemma 7.4 thus implies that

$$g(u_n) \to g(u) \text{ in } L^{\frac{2d}{d+2}}(\Omega) \text{ strong .}$$

Now the Hölder conjugate exponent of 2^* is nothing else than $\frac{2d}{d+2}$ and since $H_0^1(\Omega) \hookrightarrow L^{2^*}(\Omega)$, by duality, there holds $L^{\frac{2d}{d+2}}(\Omega) \hookrightarrow H^{-1}(\Omega)$. Finally, we have obtained that

$$-\Delta u_n = DJ(u_n) + g(u_n) \to g(u) \text{ in } H^{-1}(\Omega) \text{ strong.}$$

Of course, $(-\Delta)^{-1}$ is an isomorphism from $H^{-1}(\Omega)$ to $H_0^1(\Omega)$, which shows that

$$u_n \to (-\Delta)^{-1}(g(u)) \text{ in } H_0^1(\Omega) \text{ strong,}$$

and the Palais-Smale condition is satisfied. □

Lemma 7.6. *The functional J satisfies the hypotheses of the mountain pass lemma.*

Proof. First of all, $J(0) = 0$. Since λ_1 is the first eigenvalue of $-\Delta$, there holds $\int_\Omega \|\nabla u\|^2\,dx \geq \lambda_1 \|u\|^2_{L^2(\Omega)}$ for all $u \in H_0^1(\Omega)$.

By hypothesis ii), there exists $\mu < \lambda_1$ such that $g(s)/s \leq \mu$ in a neighborhood of 0. Besides, in a neighborhood of infinity, there holds $|g(s)| \leq C|s|^{\frac{d+2}{d-2}}$. We can thus write globally $g(s) \leq \mu s + Cs^{\frac{d+2}{d-2}}$ for $s \geq 0$ and $g(s) \geq \mu s - C|s|^{\frac{d+2}{d-2}}$ for $s \leq 0$. In both cases, integrating from 0, we obtain

$$G(s) \leq \frac{\mu}{2}s^2 + C|s|^{2^*}.$$

In order to bound J from below, we pick $\varepsilon > 0$ such that $(1-\varepsilon)\lambda_1 - \mu \geq 0$ and notice that

$$\begin{aligned}
J(u) &\geq \frac{\varepsilon}{2}\int_\Omega \|\nabla u\|^2\,dx + \Big(\frac{1-\varepsilon}{2}\lambda_1 - \frac{\mu}{2}\Big)\int_\Omega u^2\,dx - C\int_\Omega |u|^{2^*}\,dx \\
&\geq \frac{\varepsilon}{2}\|u\|^2_{H^1_0(\Omega)} - C\|u\|^{2^*}_{L^{2^*}(\Omega)} \\
&\geq \frac{\varepsilon}{2}\|u\|^2_{H^1_0(\Omega)} - C\|u\|^{2^*}_{H^1_0(\Omega)} \geq C\|u\|^2_{H^1_0(\Omega)}
\end{aligned}$$

in a neighborhood of 0 in $H^1_0(\Omega)$ by the Sobolev embedding and because $2^* > 2$. This implies condition ii) of the mountain pass lemma.

Finally, let $\phi_1 \in H^1_0(\Omega)$ be an eigenfunction of $-\Delta$ associated with the eigenvalue λ_1, which we assume to be normalized in $L^2(\Omega)$ and nonnegative without loss of generality. For all $\sigma \geq 0$, there holds

$$J(\sigma\phi_1) = \frac{\sigma^2\lambda_1}{2} - \int_\Omega G(\sigma\phi_1)\,dx.$$

Now g grows super-linearly at infinity, thus there exists $\alpha > \lambda_1$ such that $g(s) \geq \alpha s$ for s large enough, hence $g(s) \geq \alpha s - C$ for some C for all $s \geq 0$. Consequently, there holds $G(s) \geq \frac{\alpha}{2}s^2 - C$ for all $s \geq 0$. It follows that

$$J(\sigma\phi_1) \leq \frac{\sigma^2(\lambda_1 - \alpha)}{2} - C.$$

Therefore, there exists $\sigma \geq 0$ such that $J(\sigma\varphi_1) < 0$, that is to say condition iii) of the mountain pass lemma. □

Theorem 7.6. *Problem* (7.11) *admits a nontrivial solution.*

Proof. We apply the mountain pass lemma which ensures the existence of a strictly positive critical value c. There thus exists at least one corresponding critical point u and this point is nonzero since $J(0) = 0 < c = J(u)$. □

A much more comprehensive overview of critical point techniques applied to the solution of semilinear partial differential equations can be found in [40].

7.6 Exercises of Chap. 7

1. What does Ekeland's variational principle say when the distance d is a multiple of the discrete distance? Draw pictures that illustrate this case.

2. Give an example of a function that is bounded below, does not satisfy the Palais-Smale condition at the level of its infimum and nonetheless attains its minimum.

3. Completely prove Theorem 7.3 in the general case: V a Banach space, J of class C^1.

4. We are interested in a slightly more sophisticated application of the min-max principle.

4.1. Let ω be a bounded, nonempty open subset of $\mathbb{R}^k$ and $F \in C^0(\bar{\omega}; \mathbb{R}^k)$ such that $F(x) = x$ for all $x \in \partial\omega$. Show that for all $y \in \omega$, there exists $x \in \omega$ such that $F(x) = y$. (*Hint:* extend F to a Euclidean ball containing ω and consider $G(x) = F(x) - y$.)

4.2. Let V be a Banach space decomposed as a direct sum $V = V_0 \oplus V_1$ where V_0 is a closed vector subspace and V_1 a finite dimensional vector subspace. Let us be given $u_0 \in V_0$, $\|u_0\|_V = 1$, $R_0, R_1 > 0$ and set

$$\omega = \{u = su_0 + u_1; 0 < s < R_0, u_1 \in V_1, \|u_1\|_V < R_1\} \subset \mathbb{R}u_0 \oplus V_1.$$

Let $\varphi \in C^0(\bar{\omega}; V)$ be such that $\varphi(u) = u$ for all $u \in \partial\omega$ and $0 < R < \inf(R_0, R_1)$. Consider the mapping $F : \mathbb{R}u_0 \oplus V_1 \to \mathbb{R}u_0 \oplus V_1$ defined by

$$F(u) = \|\varphi(u) - \pi(\varphi(u))\|_V u_0 + \pi(\varphi(u))$$

where π is the projection of V on V_1 along V_0. Show that there exists $u \in \omega$ such that $F(u) = Ru_0$.

4.3. We set $A = \varphi(\bar{\omega})$. Show that there exists $v \in A$ such that $\|v\|_V = R$.

4.4. Let $J \in C^1(V; \mathbb{R})$ satisfying the Palais-Smale condition and such that

i) $J(0) = 0$,

ii) there exist $R > 0$, $a > 0$ such that if $u \in V_0$ and $\|u\|_V = R$ then $J(u) \geq a$,

iii) there exist $u_0 \in V_0$, $\|u_0\|_V = 1$, $R_0, R_1 > R$ such that $J(u) \leq 0$ for all $u \in \partial\omega$.

Setting $\mathscr{A} = \{\varphi(\bar{\omega}); \varphi \in C^0(\bar{\omega}; V), \varphi = \mathrm{Id} \text{ on } \partial\omega\}$, show that

$$c = \inf_{A \in \mathscr{A}} \max_{v \in A} J(v)$$

is a critical value of J and that $c \geq a$. What about the case $V_1 = \{0\}$?

5. Let Ω be a regular bounded open subset of $\mathbb{R}^d$ with $d \geq 3$ and let $V = H_0^1(\Omega)$. We are given two real numbers $1 < p < \frac{d+2}{d-2}$ and $2 < \theta \leq p + 1$.

5.1. Show that for all $\varepsilon > 0$, there exists a constant $C(\varepsilon) > 0$ such that for all $v \in V$, $\|v\|_{L^2(\Omega)} = 1$, there holds

$$1 \leq \varepsilon \int_\Omega \|\nabla v\|^2\, dx + C(\varepsilon)\Bigl(\int_\Omega |v|^\theta\, dx\Bigr)^{\frac{2}{\theta}}.$$

(*Hint:* argue by contradiction.) Deduce that for all $v \in V$,

$$\int_\Omega v^2\, dx \leq \varepsilon \int_\Omega \|\nabla v\|^2\, dx + C(\varepsilon)\Bigl(\int_\Omega |v|^\theta\, dx\Bigr)^{\frac{2}{\theta}}.$$

5.2. Let $g \in C^0(\mathbb{R}; \mathbb{R})$ such that $g(0) = 0$ and G be its primitive vanishing at 0. We assume that $|g(s)| \leq C(1 + |s|^p)$ for all s and that $0 \leq \theta G(s) \leq sg(s)$ for $|s|$ large enough. Let $\lambda > 0$ be a given number. We consider the functional on V

$$J(v) = \frac{1}{2}\int_\Omega (\|\nabla v\|^2 - \lambda v^2)\, dx - \int_\Omega G(v)\, dx.$$

Show that this functional is well defined, of class C^1, and that

$$DJ(v) = -\Delta v - \lambda v - g(v) \in H^{-1}(\Omega).$$

5.3. Let u_n be a Palais-Smale sequence for J. Show that

$$\int_\Omega G(u_n)\, dx \leq C(1 + \|\nabla u_n\|_{L^2}),$$

where C does not depend on n. (*Hint:* Remark that $G(s) = \frac{1}{\theta-2}(\theta G(s) - 2G(s))$ and that $\int_\Omega u_n g(u_n)\, dx = 2J(u_n) + 2\int_\Omega G(u_n)\, dx - \langle DJ(u_n), u_n\rangle$.) Deduce that u_n is bounded in V.

5.4. Show that the functional J satisfies the Palais-Smale condition. (Careful with compactness! Prove that $g(u_n) \to g(u)$ in $L^{\frac{p+1}{p}}(\Omega)$ strong, up to a subsequence.)

5.5. Let λ_i, $i \in N^*$, be the nondecreasing sequence of eigenvalues of $-\Delta$ on $H_0^1(\Omega)$ and ϕ_i a corresponding sequence of eigenfunctions. Assume that there exists j such that $\lambda_j \leq \lambda < \lambda_{j+1}$. Define $V_1 = \operatorname{vect}\{\phi_1, \dots, \phi_j\}$ and $V_0 = V_1^\perp$. Assume also that $\limsup_{s\to 0}(g(s)/s) \leq 0$. Show that J satisfies conditions ii) and iii) of Exercise 4. (*Hint:* Show that $G(s) \leq \varepsilon s^2 + C(\varepsilon)|s|^{p+1}$ and remember that $\int_\Omega \|\nabla v\|^2\, dx \geq \lambda_{j+1}\int_\Omega v^2\, dx$ for all $v \in V_0$.)

5.6. Setting $\tilde{g}(s) = \lambda s + g(s)$, what can we conclude? Why was it not possible to use the mountain pass lemma?

6. Let Ω a bounded open subset of $\mathbb{R}^2$ of class C^∞. For all $\lambda \geq 0$, we consider the functional

$$I_\lambda(v) = \frac{1}{2}\int_\Omega |\nabla v|^2\,dx + \frac{\lambda}{8}\int_\Omega (v^4-1)^2\,dx.$$

6.1. Show that I_λ is well defined on $H_0^1(\Omega)$ with values in $\mathbb{R}$, and is of class C^1.

6.2. Show that any critical point $u \in H_0^1(\Omega)$ of I_λ is a solution of the equation

$$-\Delta u = \lambda u^3(1-u^4) \quad \text{in the sense of } \mathscr{D}'(\Omega). \tag{7.12}$$

6.3. Show that any solution of (7.12) belonging to $H_0^1(\Omega)$ is such that

$$|u| \leq 1 \quad \text{a.e. in } \Omega.$$

(*Hint:* Use the test-function $v = (u-1)_+$.)

6.4. Show that any solution of (7.12) belonging to $H_0^1(\Omega)$ belongs in fact to $C^\infty(\overline{\Omega})$.

6.5. Show that I_λ is of class C^2 on $H_0^1(\Omega)$ and deduce from this that the zero solution of (7.12) is a local minimum of I_λ.

6.6. Show that there exists $\lambda^* > 0$ such that for all $0 \leq \lambda < \lambda^*$, 0 is the unique solution of (7.12).

6.7. Let $\varphi \in \mathscr{D}(\Omega)$ be such that $0 \leq \varphi \leq 1$ and $\varphi = 1$ on a closed ball $\bar{B} \subset \Omega$. Show that there exists $\lambda_1 > \lambda^*$ such that $I_\lambda(\varphi) < I_\lambda(0)$ for all $\lambda \geq \lambda_1$.

6.8. Show that if $\lambda \geq \lambda_1$, then there exists a nonzero solution $u_\lambda \in H_0^1(\Omega)$ of Eq. (7.12).

6.9. Using the mountain pass lemma, show that there then exists a third solution v_λ.

Chapter 8
Monotone Operators and Variational Inequalities

A quasilinear operator is not always the differential of a functional of the calculus of variations. The abstract concept of monotone operator, and more generally of quasi-monotone operator, makes it possible to go further than the calculus of variations in the convex case. They are associated with variational inequalities, which appear in many situations, such as obstacle problems for example.

8.1 Monotone Operators, Definitions and First Properties

In what follows, V is a reflexive and separable Banach space and A is a mapping from V into V' (which is nonlinear in general[1]).

Definition 8.1. We say that

i) A is monotone if

$$\forall u, v \in V, \quad \langle A(u) - A(v), u - v\rangle \geq 0. \tag{8.1}$$

ii) A is strictly monotone if in addition $\langle A(u) - A(v), u - v\rangle = 0$ implies $u = v$.

iii) A is hemicontinuous if for all $u, v \in V$, the mapping $t \mapsto \langle A(u + tv), v\rangle$ is continuous from $\mathbb{R}$ into $\mathbb{R}$.

Remark 8.1. i) If $V = \mathbb{R}$, monotonicity in this sense is just saying that A is nondecreasing.[2] We generalize the idea to Banach spaces.

[1] But not multi-valued. Monotone operators can be generalized to the multi-valued case, with a rich theory of so called maximal monotone operators. We will not consider them here, but the reader can consult [10].

[2] But not nonincreasing, so the vocabulary is maybe a little misleading in this case.

H. Le Dret, *Nonlinear Elliptic Partial Differential Equations*, Universitext,
https://doi.org/10.1007/978-3-319-78390-1_8

ii) Let $J\colon V \to \mathbb{R}$ be convex and differentiable in the sense of Gateaux. Then its differential $DJ\colon V \to V'$ is monotone. Indeed, let $w = u - v$. Then $j(t) = J(v + tw)$ is convex from $\mathbb{R}$ into $\mathbb{R}$, $j(0) = J(v)$, $j(1) = J(u)$, j is differentiable with $j'(t) = \langle DJ(v + tw), w\rangle$. Since j is convex, j' is nondecreasing, and writing that $j'(0) \leq j'(1)$ is nothing but writing the monotonicity of DJ. If J is strictly convex, then DJ is strictly monotone. Finally, as a convex function from $\mathbb{R}$ to $\mathbb{R}$ that is differentiable is in fact of class C^1, we see that DJ is also hemicontinuous.

iii) We can always assume that $A(0) = 0$. We otherwise replace A with $A - A(0)$.

iv) If A is continuous from V strong into V' weak, then A is hemicontinuous.

v) If V is a Hilbert space and A is the linear operator from V to V' associated with a continuous bilinear form $a(\cdot, \cdot)$, then A is hemicontinuous. It is monotone if and only if a is positive and strictly monotone if and only if a is positive definite.

□

Remark 8.1 iv) admits a rather surprising converse, inasmuch as hemicontinuity looks a priori like a very weak condition.

Lemma 8.1. *Let A be a bounded, hemicontinuous and monotone operator. Then A is continuous from V strong into V' weak.*

Proof. Let u_n be a sequence such that $u_n \to u$ in V strong. This sequence is thus bounded, and since A maps bounded sets to bounded sets, $A(u_n)$ is bounded in V'. We extract a subsequence n' such that $A(u_{n'}) \rightharpoonup \psi$ for some $\psi \in V'$. This is possible because V' is reflexive. Due to the monotonicity, for all $v \in V$, there holds

$$0 \leq \langle A(u_{n'}) - A(v), u_{n'} - v\rangle = \langle A(u_{n'}), u_{n'} - v\rangle - \langle A(v), u_{n'} - v\rangle.$$

It is quite clear that $\langle A(v), u_{n'} - v\rangle \to \langle A(v), u - v\rangle$ by strong convergence of $u_{n'}$. Moreover, since $A(u_{n'})$ converges weakly in V' and $u_{n'} - v$ converges strongly in V, their duality bracket converge

$$\langle A(u_{n'}), u_{n'} - v\rangle \longrightarrow \langle \psi, u - v\rangle.$$

Consequently, we obtain in the limit

$$\forall v \in V, \quad 0 \leq \langle \psi - A(v), u - v\rangle. \tag{8.2}$$

We are going to show that this inequality actually determines ψ by using a technique that is characteristic of monotone operators, called *Minty's trick*. Let $w \in V$ be arbitrary and $t \in \mathbb{R}_+^*$. By applying (8.2) with $v = u + tw$ and dividing the resulting inequality by $t > 0$, we obtain

$$\langle \psi - A(u + tw), w\rangle \leq 0.$$

We let now t go to 0. Since A is hemicontinuous, it follows that

$$\forall w \in V, \quad \langle \psi - A(u), w\rangle \leq 0.$$

This inequality is also true for $-w$, thus there in fact holds,

$$\forall w \in V, \quad \langle \psi - A(u), w\rangle = 0,$$

therefore

$$\psi = A(u).$$

We conclude by uniqueness of the weak limit of convergent subsequences extracted from $A(u_n)$. □

See [9, 48] for more details.

8.2 Examples of Monotone Operators

In this section, we give a few examples of monotone operators, in the context of quasilinear boundary value problems.

We begin with an example of a (linear) monotone operator that is not the differential of a functional of the calculus of variations. Let Ω be a bounded open subset of $\mathbb{R}^d$, $V = H_0^1(\Omega)$ and $b \in \mathbb{R}^d \setminus \{0\}$. We set $A(u) = -\Delta u + b \cdot \nabla u$. The operator A maps $H_0^1(\Omega)$ into its dual $H^{-1}(\Omega)$. It is monotone. Indeed, for all $u \in V$

$$\langle A(u), u\rangle = \int_\Omega |\nabla u|^2\,dx + \int_\Omega (b \cdot \nabla u)u\,dx.$$

Now, if $u \in H_0^1(\Omega)$, it is clear that $\int_\Omega u\partial_i u\,dx = 0$ for all $i = 1, \dots, d$, which implies that

$$\langle A(u), u\rangle = \int_\Omega |\nabla u|^2\,dx \geq 0,$$

since $b \cdot \nabla u = b_i \partial_i u$, and the monotonicity is established, by linearity of A.

The operator $B(u) = -\Delta u$ is the differential of a well-known functional. It is thus enough to prove that $C(u) = b \cdot \nabla u$ is not a differential. Let us thus assume for contradiction that there exists $J : V \to \mathbb{R}$ such that $DJ(u)v = \int_\Omega (b \cdot \nabla u)v\,dx$. If we set $j(t) = J(tu)$, then we have $j'(t) = DJ(tu)u = t\int_\Omega (b \cdot \nabla u)u\,dx = 0$. Thus $J(u) = j(1) = j(0) = 0$ for all u in V, so that $DJ(u) = 0$ for all u in V, which is not the case.

Indeed, by a change of coordinates, we may assume that $b = e_d$ so that $b \cdot \nabla u = \partial_d u$. We take $u \in \mathscr{D}(\Omega)$ of the form $u(x) = \psi(x')\varphi(x_d) \neq 0$, with $\varphi \in \mathscr{D}(\mathbb{R})$ and $\psi \in \mathscr{D}(\mathbb{R}^{d-1})$, using the notation $x' = (x_1, x_2, \dots, x_{d-1})$. Now, $b \cdot \nabla u = \psi(x')\varphi'(x_d)$. We take $v(x) = \psi(x')\varphi'(x_d)$, which is also in $\mathscr{D}(\Omega)$. So with this choice of u and v, we obtain

$$DJ(u)v = \int_\Omega (b \cdot \nabla u)v\,dx = \int_\Omega \big(\psi(x')\varphi'(x_d)\big)^2\,dx \neq 0,$$

hence the result. □

Let now Ω be a bounded open subset of $\mathbb{R}^d$, $p \in \,]1, +\infty[$ and $V = W_0^{1,p}(\Omega)$. Let us be given a mapping $F\colon \mathbb{R}^d \to \mathbb{R}^d$ continuous and monotone, i.e., for all pairs $\xi, \zeta \in \mathbb{R}^d$,

$$(F(\xi) - F(\zeta)) \cdot (\xi - \zeta) \geq 0,$$

where $\cdot$ denotes the usual Euclidean inner product in $\mathbb{R}^d$. We assume that F satisfies the growth condition

$$\forall \xi \in \mathbb{R}^d, \quad |F(\xi)| \leq C(1 + |\xi|^{p-1})$$

for a certain constant C independent of ξ.

Proposition 8.1. *The operator $A(u) = -\mathrm{div}\,(F(\nabla u))$ is well defined from $W_0^{1,p}(\Omega)$ into $W^{-1,p'}(\Omega)$. It is bounded, hemicontinuous and monotone.*

Proof. Let p' be the Hölder conjugate exponent of p. If $u \in W_0^{1,p}(\Omega)$, then $\nabla u \in L^p(\Omega; \mathbb{R}^d)$ and $F(\nabla u) \in L^{p'}(\Omega; \mathbb{R}^d)$. Indeed,

$$|F(\nabla u)|^{p'} = |F(\nabla u)|^{\frac{p}{p-1}} \leq C(1 + |\nabla u|^{p-1})^{\frac{p}{p-1}} \leq C'(1 + |\nabla u|^p) \in L^1(\Omega).$$

Consequently, $-\mathrm{div}\,(F(\nabla u)) \in W^{-1,p'}(\Omega)$ with the duality bracket

$$\langle -\mathrm{div}\,(F(\nabla u)), v\rangle = \int_\Omega F(\nabla u) \cdot \nabla v\,dx,$$

for all v in $W_0^{1,p}(\Omega)$. This duality is established by density of $\mathscr{D}(\Omega)$ in $W_0^{1,p}(\Omega)$.

Moreover, if u belongs to a bounded subset of $W_0^{1,p}(\Omega)$, then ∇u belongs to a bounded subset of $L^p(\Omega; \mathbb{R}^d)$ and by the previous estimate, $F(\nabla u)$ belongs to a bounded subset of $L^{p'}(\Omega; \mathbb{R}^d)$. Therefore, $A(u)$ belongs to a bounded subset of $W^{-1,p'}(\Omega)$.

By Carathéodory's theorem, the mapping $z \mapsto F(z)$ is continuous from $L^p(\Omega; \mathbb{R}^d)$ strong into $L^{p'}(\Omega; \mathbb{R}^d)$ strong. Since we then compose it on both sides with continuous linear mappings, it follows that A is continuous from $W_0^{1,p}(\Omega)$ strong into $W^{-1,p'}(\Omega)$ strong, hence *a fortiori* hemicontinuous.

Finally, for all pairs $u, v \in W_0^{1,p}(\Omega)$, there holds

$$\langle A(u) - A(v), u - v\rangle = \int_\Omega (F(\nabla u) - F(\nabla v)) \cdot (\nabla u - \nabla v)\, dx \geq 0$$

by the monotonicity of F on $\mathbb{R}^d$. The operator A is thus monotone. □

Remark 8.2. i) An example of such a function F is given by $F(\xi) = |\xi|^{p-2}\xi$. It is the gradient of the convex function $\xi \mapsto \frac{1}{p}|\xi|^p$ and gives rise as an operator to the p-Laplacian that we have already encountered in Chap. 6. See [9, 12, 54] for more.

ii) Assuming that all functions are regular, the operator reads

$$-\mathrm{div}\,(F(\nabla u)) = -\frac{\partial F_i}{\partial \xi_k}(\nabla u)\partial_{ki} u,$$

which shows that it is quasilinear. Using the monotonicity of F, we also see that for all $\xi, \zeta \in \mathbb{R}^d$ and $t > 0$

$$\big(F(\xi + t\zeta) - F(\xi)\big) \cdot (t\zeta) \geq 0,$$

so that dividing by t and then letting $t \to 0$, we obtain

$$DF(\xi)\zeta \cdot \zeta \geq 0.$$

In other words, the matrix $\frac{\partial F_i}{\partial \xi_k}(\xi)$ is nonnegative for all $\xi \in \mathbb{R}^d$ and the operator is elliptic. □

8.3 Variational Inequalities

Monotone operators are well suited to the resolution of abstract variational inequalities. To motivate their study, we will give a few concrete examples related to nonlinear boundary value problems later on in Sect. 8.4, see also [43, 48].

Theorem 8.1. *Let $A : V \to V'$ be a bounded, hemicontinuous and monotone operator and let C be a nonempty, closed and bounded convex subset of V. Then, for all $f \in V'$, the variational inequality: Find $u \in C$ such that*

$$\forall v \in C, \quad \langle A(u) - f, v - u\rangle \geq 0 \tag{8.3}$$

admits at least one solution.

Notice that if u belongs to the interior of C, then $A(u) = f$. Interesting things happen when u is not an interior point of C.

To prove Theorem 8.1, we use the Galerkin method through a series of lemmas. We first adapt the basic idea of the Galerkin method to the case of convex sets.

Lemma 8.2. *There exists a countable family of closed convex subsets $(C_m)_{m\in\mathbb{N}}$, which is increasing, such that each C_m is finite dimensional, included in C and $\bigcup_{m=0}^{+\infty} C_m$ is dense in C.*

Proof. Since V is a separable metric space, C is also separable and there thus exists a countable family $(w_m)_{m\in\mathbb{N}}$ of points of C that it dense in C. We set for $m \in \mathbb{N}$,

$$C_m = \overline{\text{conv}}\{w_0, w_1, \dots, w_m\}.$$

By construction, C_m is a closed convex subset of dimension less than m for each m. Clearly $C_m \subset C_{m+1} \subset C$. Finally, the set $\bigcup_{m=0}^{+\infty} C_m$ is dense in C, since it contains each w_m. □

Once the C_m are constructed, we solve the corresponding finite dimensional problems.

Lemma 8.3. *The variational inequality: Find $u_m \in C_m$ such that*

$$\forall v \in C_m, \quad \langle A(u_m) - f, v - u_m\rangle \geq 0 \tag{8.4}$$

admits at least one solution.

Proof. Let $V_m = \text{vect}\, C_m$ be the vector space spanned by C_m. This space is finite dimensional, we equip it with a Euclidean structure induced by some inner product $(\cdot|\cdot)_m$. By the Riesz theorem in V_m, there exists a continuous linear mapping J_m from V' weak into V_m such that

$$\forall g \in V', \forall v \in V_m, \quad \langle g, v\rangle = (J_m g|v)_m.$$

By definition, C_m is a nonempty, bounded, closed convex subset of V_m for each m. The orthogonal projection of V_m on C_m, denoted by Π_m, is characterized by $\Pi_m(v) \in C_m$ and

$$\forall w \in C_m, \quad (v - \Pi_m(v)|\Pi_m(v) - w)_m \geq 0. \tag{8.5}$$

This is a continuous mapping from V_m into C_m, which is nonlinear since C_m is bounded, hence not a vector subspace of V. We now define a mapping $T_m\colon C_m \to C_m$ by

$$\forall v \in C_m, \quad T_m(v) = \Pi_m(v - J_m A(v) + J_m f). \tag{8.6}$$

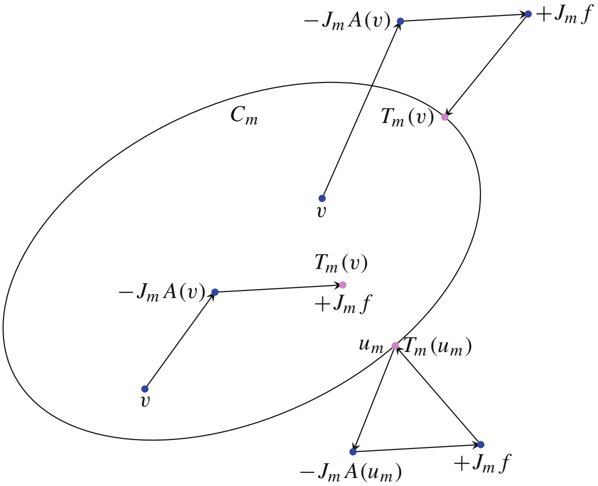


Fig. 8.1 The mapping T_m and its fixed point

By Lemma 8.1, A is continuous from V strong into V' weak. We thus see that T_m is continuous as a composition of continuous mappings. Due to Brouwer's fixed point theorem, T_m admits at least one fixed point $u_m \in C_m$. In particular, according to (8.5) and (8.6), there holds for all $v \in C_m$,

$$(u_m - J_m A(u_m) + J_m f - T_m(u_m)|T_m(u_m) - v)_m \geq 0,$$

which boils down to

$$(J_m f - J_m A(u_m)|u_m - v)_m = \langle f - A(u_m), u_m - v\rangle \geq 0,$$

since u_m is a fixed point of T_m and by definition of the mapping J_m. □

Remark 8.3. It is noteworthy that monotonicity plays practically no role in the existence of a solution of a finite dimensional variational inequality. Monotonicity only intervenes through the continuity of A. Lemma 8.3 actually gives the existence of such a solution in the finite dimensional case under the sole hypothesis that A is continuous, see Fig. 8.1. □

Of course, if V is finite dimensional, the proof can stop here. Let us now start moving to the infinite dimensional case.[3]

Lemma 8.4. *There exists a subsequence (still denoted m), $u \in C$ and $\psi \in V'$ such that $u_m \rightharpoonup u$ in V weak and $A(u_m) \rightharpoonup \psi$ in V' weak.*

[3] The rest of the proof still works if V is finite dimensional, it just serves no purpose in this case.

Proof. Since $C_m \subset C$, which is bounded, the sequence u_m is bounded in V. By hypothesis, A is a bounded operator, hence the sequence $A(u_m)$ is bounded in V'. We thus extract a subsequence such that $u_m \rightharpoonup u$ in V weak and $A(u_m) \rightharpoonup \psi$ in V' weak. Moreover, C is a strongly closed convex, it is thus weakly closed, which implies that $u \in C$. □

The point is now to identify ψ, and this is where monotonicity plays a crucial role. The problem is that we only have weak convergences, so that we cannot say anything a priori about $\langle A(u_m), u_m\rangle$ in the $m \to +\infty$ limit.

Lemma 8.5. *The following inequality holds* $\liminf_{m\to+\infty}\langle A(u_m), u_m\rangle \geq \langle\psi, u\rangle$.

Proof. We use the monotonicity of A. There holds

$$\langle A(u_m) - A(u), u_m - u\rangle \geq 0,$$

so that expanding the duality bracket, we obtain

$$\langle A(u_m), u_m\rangle - \langle A(u_m), u\rangle - \langle A(u), u_m\rangle + \langle A(u), u\rangle \geq 0. \tag{8.7}$$

Due to the respective weak convergences of $A(u_m)$ to ψ and u_m to u, we see that $\langle A(u_m), u\rangle \to \langle\psi, u\rangle$ and $\langle A(u), u_m\rangle \to \langle A(u), u\rangle$. Consequently, the inferior limit of the left-hand side of (8.7) is nonnegative, which says that

$$\liminf_{m\to+\infty}\langle A(u_m), u_m\rangle - \langle\psi, u\rangle - \langle A(u), u\rangle + \langle A(u), u\rangle \geq 0,$$

hence the result. □

Lemma 8.6. *The following inequality also holds* $\limsup_{m\to+\infty}\langle A(u_m), u_m\rangle \leq \langle\psi, u\rangle$.

Proof. This time, we use the finite dimensional variational inequality (8.4).[4] Let us take an arbitrary natural number i and some $v \in C_i$. For all $m \geq i$, $C_i \subset C_m$, so that

$$\langle A(u_m) - f, v - u_m\rangle \geq 0,$$

and expanding the duality bracket, we obtain

$$-\langle A(u_m), u_m\rangle - \langle f, v\rangle + \langle f, u_m\rangle + \langle A(u_m), v\rangle \geq 0. \tag{8.8}$$

[4] Which has little to do with monotonicity, as we recall.

We again take the inferior limit of this inequality, taking into account all known weak convergences. This yields

$$-\limsup_{m\to+\infty}\langle A(u_m), u_m\rangle - \langle f, v\rangle + \langle f, u\rangle + \langle \psi, v\rangle \geq 0,$$

an inequality which holds true for all $v \in C_i$. Now the union of C_i for $i \geq 0$ is dense in C, we can thus choose a sequence $v_i \in C_i$ that tends strongly to u when $i \to +\infty$. Passing to the limit when $i \to +\infty$ in the above inequality with this sequence v_i, we obtain the Lemma. □

We can now complete the proof of Theorem 8.1.

Lemma 8.7. *There holds $\langle A(u_m), u_m\rangle \to \langle \psi, u\rangle$, $A(u) = \psi$ and u is a solution of the variational inequality* (8.3)*.*

Proof. The first convergence follows immediately from Lemmas 8.5 and 8.6. We go back to the monotonicity inequality

$$\langle A(u_m), u_m\rangle - \langle A(u_m), v\rangle - \langle A(v), u_m\rangle + \langle A(v), v\rangle \geq 0$$

for all v in V. Passing to the limit when $m \to +\infty$, we obtain

$$\langle \psi, u\rangle - \langle \psi, v\rangle - \langle A(v), u\rangle + \langle A(v), v\rangle \geq 0,$$

or again

$$\langle \psi - A(v), u - v\rangle \geq 0,$$

for all $v \in V$. At this point, we use Minty's trick to deduce that $\psi = A(u)$.

We now go back to the finite dimensional variational inequality (8.8). Passing to the limit when $m \to +\infty$, we see that for all v in $\bigcup_{i=0}^{+\infty} C_i$,

$$-\langle A(u), u\rangle - \langle f, v\rangle + \langle f, u\rangle + \langle A(u), v\rangle \geq 0,$$

or again

$$\langle A(u) - f, v - u\rangle \geq 0,$$

hence the final result by density of $\bigcup_{i=0}^{+\infty} C_i$ in C. □

Remark 8.4. Let us pinpoint the main difficulty. The sequence u_m converges weakly to u, but A is a nonlinear operator. There is thus no reason in general for $A(u_m)$ to converge in any sense, reasonable or not, toward $A(u)$. It is quite remarkable that monotonicity and Minty's trick make this unlikely convergence nonetheless true, when u_m solves a finite dimensional version of the variational inequality. □

Theorem 8.2. *If A is strictly monotone, then the solution of the variational inequality* (8.3) *is unique.*

Proof. Let $u_1, u_2 \in C$ be two solutions. For all $v_1, v_2 \in C$,

$$\langle A(u_1) - f, v_1 - u_1\rangle \geq 0 \quad \text{and} \quad \langle A(u_2) - f, v_2 - u_2\rangle \geq 0.$$

Let us take $v_1 = u_2$ and $v_2 = u_1$, so that

$$\langle A(u_1) - A(u_2), u_1 - u_2\rangle \leq 0.$$

Consequently, by monotonicity of A,

$$\langle A(u_1) - A(u_2), u_1 - u_2\rangle = 0,$$

from which $u_1 = u_2$ follows by strict monotonicity. □

To treat the case of an unbounded convex set C, we need an additional hypothesis.

Definition 8.2. We say that A is coercive is there exists $v_0 \in C$ ($v_0 = 0$ if $C = V$) such that

$$\lim_{\|v\|_V \to +\infty} \frac{\langle A(v), v - v_0\rangle}{\|v\|_V} = +\infty. \tag{8.9}$$

Theorem 8.3. *Let $A\colon V \to V'$ be an operator that is bounded, hemicontinuous, monotone and coercive and let C be a nonempty closed convex subset of V. Then, for all $f \in V'$, the variational inequality* (8.3) *admits at least one solution.*

Proof. For $R > 0$, we set $C_R = \{v \in C; \|v\|_V \leq R\}$. This is a bounded, closed convex subset of V, that is nonempty if R is large enough. Due to Theorem 8.1, there then exists at least one solution u_R to the variational inequality on C_R. We can always start with R large enough so that $v_0 \in C_R$ and thus

$$\langle A(u_R) - f, v_0 - u_R\rangle \geq 0,$$

or

$$\langle A(u_R), u_R - v_0\rangle \leq \langle f, u_R - v_0\rangle \leq \|f\|_{V'}(\|u_R\|_V + \|v_0\|_V).$$

Let us show that u_R (when defined) is bounded in V independently of R. Dividing the last inequality by $\|u_R\|_V$ (which we assume to be nonzero, otherwise there is nothing to prove), we obtain

$$\frac{\langle A(u_R), u_R - v_0\rangle}{\|u_R\|_V} \leq \|f\|_{V'}\Big(1 + \frac{\|v_0\|_V}{\|u_R\|_V}\Big).$$

Let us assume for contradiction that there exists a sequence $R_n \to +\infty$ such that $\|u_{R_n}\|_V \to +\infty$. Then $\|v_0\|_V/\|u_{R_n}\|_V \to 0$ and the above inequality contradicts the coercivity of A. Therefore, there exists a constant M such that $\|u_R\|_V \leq M$ for all R for which u_R exists.

Let us now take $R_0 = M + 1$. We are going to show that u_{R_0} is a solution of the full variational inequality (8.3), no $R \to +\infty$ limit needed. For all $v \in C$ and all $\lambda \in [0, 1]$, $\lambda v + (1-\lambda)u_{R_0}$ belongs to C. Moreover,

$$\|\lambda v + (1-\lambda)u_{R_0}\|_V \leq \lambda\|v\|_V + (1-\lambda)\|u_{R_0}\|_V \leq \lambda\|v\|_V + M.$$

In particular, if we take $0 < \lambda \leq 1/\|v\|_V$, then we see that $\lambda v + (1-\lambda)u_{R_0}$ belongs to C_{R_0}. Hence, by the variational inequality on C_{R_0},

$$\langle A(u_{R_0}) - f, \lambda v + (1-\lambda)u_{R_0} - u_{R_0}\rangle \geq 0,$$

so that

$$\lambda\langle A(u_{R_0}) - f, v - u_{R_0}\rangle \geq 0.$$

The result follows by dividing by $\lambda > 0$. □

Corollary 8.1. *Let $A\colon V \to V'$ be an operator that is bounded, hemicontinuous, monotone and coercive. Then A is surjective.*

Proof. We take $C = V$. For all $f \in V'$, there thus exists $u \in V$ such that, for all v in V, there holds

$$\langle A(u) - f, v - u\rangle \geq 0.$$

Setting $v = u + w$, we deduce that for all w in V

$$\langle A(u) - f, w\rangle \geq 0.$$

Now this inequality is also valid for $-w$, hence

$$\langle A(u) - f, w\rangle \leq 0,$$

so that $A(u) = f$ and A is surjective. □

Remark 8.5. Note that in the one-dimensional case, this is nothing but saying that a continuous nondecreasing function that tends to $-\infty$ at $-\infty$ and $+\infty$ at $+\infty$, is surjective. □

8.4 Examples of Variational Inequalities

Let us go back to the example of Proposition 8.1. Let Ω be a bounded open subset of $\mathbb{R}^d$, $p \in \,]1, +\infty[$, $V = W_0^{1,p}(\Omega)$, $F\colon \mathbb{R}^d \to \mathbb{R}^d$ continuous and monotone, satisfying the growth condition

$$\forall \xi \in \mathbb{R}^d, \quad |F(\xi)| \leq C(1 + |\xi|^{p-1}),$$

for a certain constant C independent of ξ. We assume in addition that there exists a constant $\alpha > 0$ such that

$$\forall \xi \in \mathbb{R}^d, \quad F(\xi) \cdot \xi \geq \alpha |\xi|^p.$$

Theorem 8.4. *For all $f \in W^{-1,p'}(\Omega)$, there exists $u \in W_0^{1,p}(\Omega)$ such that*

$$-\mathrm{div}\,(F(\nabla u)) = f \textit{ in the sense of } \mathscr{D}'(\Omega).$$

If F is strictly monotone, then this solution is unique.

Proof. We only need to show that the operator $A(v) = -\mathrm{div}\,(F(\nabla v))$ is coercive. We take the L^p norm of the gradient as a norm on $W_0^{1,p}(\Omega)$. There holds

$$\frac{\langle A(v), v\rangle}{\|v\|} = \frac{\int_\Omega F(\nabla v)\cdot \nabla v\,dx}{\|\nabla v\|_{L^p(\Omega;\mathbb{R}^d)}} \geq \alpha \|\nabla v\|^{p-1}_{L^p(\Omega;\mathbb{R}^d)} \longrightarrow +\infty$$

when $\|\nabla v\|_{L^p(\Omega;\mathbb{R}^d)} \to +\infty$. □

Remark 8.6. For $F(\xi) = |\xi|^{p-2}\xi$, we recover the existence and uniqueness result for the p-Laplacian already obtained by calculus of variations methods. □

We now give examples in which the convex set C is not the whole space. Consider first a C^1, convex function $J\colon V \to \mathbb{R}$ and assume that $u \in C$ minimizes J on the convex subset C of V. What is the analogue of the Euler-Lagrange equation in this case? It is a Euler-Lagrange variational inequality.

Proposition 8.2. *Let V be a reflexive Banach space, C a closed convex subset of V and $J\colon V \to \mathbb{R}$ a functional that is convex and of class C^1. Then $u \in C$ minimizes J on C if and only if*

$$\langle DJ(u), v - u\rangle \geq 0 \tag{8.10}$$

for all $v \in C$.

Proof. Indeed, given any $v \in C$, we define $j\colon [0,1] \to \mathbb{R}$ by

$$j(t) = J((1-t)u + tv).$$

Of course, j is convex and C^1, and has a minimum at $t = 0$. Now

$$j'(t) = \langle DJ((1-t)u + tv), v - u\rangle$$

and since j is minimum at $t = 0$, it follows that $j'(0) \geq 0$, or

$$\langle DJ(u), v - u\rangle \geq 0$$

for all $v \in C$, which is the expected variational inequality.

Conversely, assume that the above variational inequality is satisfied by u. Then, since $j(t) \geq j(0) + tj'(0)$ by convexity, so that in particular $j(1) \geq j(0) + j'(0)$, it follows that

$$J(v) \geq J(u) + \langle DJ(u), v - u\rangle \geq J(u),$$

hence u minimizes J over C. □

Let us apply this in a finite-dimensional, non-PDE example, the existence of a Nash equilibrium for a two-player game. Let C_1 be a compact convex subset of $\mathbb{R}^{p_1}$ and C_2 be a compact convex subset of $\mathbb{R}^{p_2}$. We consider a two-player game in which player i plays a vector $v_i \in C_i$. Let $C = C_1 \times C_2$. Each player has a cost function $f_i \colon C \to \mathbb{R}$, which we assume to be convex and C^1.

The goal of the game for each player is to minimize their own cost function. They however only control one of the two arguments, the other one being chosen by the other player. So an optimal strategy for player 1 is a function $s_1 \colon C_2 \to C_1$, such that

$$f_1(s_1(v_2), v_2) = \min_{v_1 \in C_1} f_1(v_1, v_2)$$

for all $v_2 \in C_2$, and likewise for player 2 with an optimal strategy s_2. A *Nash equilibrium* for the game is a point $(u_1, u_2) \in C$ such that

$$u_1 = s_1(u_2) \text{ and } u_2 = s_2(u_1).$$

In other words, if both players play a Nash equilibrium, neither one of the two has any incentive to change the current play unilaterally.

Now the existence of a Nash equilibrium can be formulated as a variational inequality. Indeed, the optimal strategy for player 1 is characterized by the variational inequality

$$\big(\nabla_1 f_1(s_1(v_2), v_2)\big| v_1 - s_1(v_2)\big) \geq 0,$$

for all $v_1 \in C_1$, where we use the canonical inner product in $\mathbb{R}^{p_1}$, and likewise for player 2. If we now define $F\colon C \to \mathbb{R}^{p_1} \times \mathbb{R}^{p_2}$ by

$$F(v_1, v_2) = ((\nabla_1 f_1(v_1, v_2), \nabla_2 f_2(v_1, v_2)),$$

we see that a Nash equilibrium $u = (u_1, u_2)$ is characterized by the variational inequality

$$(F(u)|v - u) \geq 0$$

for all $v \in C$, where the inner product is now the canonical inner product on $\mathbb{R}^{p_1} \times \mathbb{R}^{p_2}$. This shows in particular that a Nash equilibrium exists under the above hypotheses.[5]

Let us now turn to PDE examples. The so-called *obstacle* problems fall under this category. A prototypical case is defined by a measurable function g on Ω with values in $\mathbb{R} \cup \{\pm\infty\}$. We set

$$C = \{v \in H_0^1(\Omega);\, v \geq g \text{ almost everywhere in } \Omega\}.$$

This set is obviously convex. Let us show that it is closed. Let $v_n \in C$ be a sequence such that $v_n \to v$ in $H_0^1(\Omega)$. We can extract a subsequence that converges almost everywhere, so that $v \geq g$ almost everywhere, i.e., $v \in C$. A typical application is the following theorem.

Theorem 8.5. *Assume that g is such that C is nonempty. Then, for all $f \in H^{-1}(\Omega)$, there exists a unique $u \in H_0^1(\Omega)$, $u \geq g$ almost everywhere, such that*

$$\langle -\Delta u - f, v - u\rangle \geq 0$$

for all $v \in H_0^1(\Omega)$ such that $v \geq g$ almost everywhere.

Remark 8.7. Note that even though the monotone operator is linear, this is nonetheless a nonlinear problem. □

To make sense of this variational inequality, let us consider the one-dimensional case $d = 1$ in more detail. We take g such that $g_+ \in H_0^1(]0, 1[)$ so that the convex set C is nonempty. Due to the Sobolev embeddings, all intervening functions are continuous and the set

$$E = \{x \in]0, 1[;\, u(x) > g(x)\}$$

is an open set, hence an at most countable union of disjoint open intervals. Let I be one such interval. For all $\varphi \in \mathscr{D}(I)$, there exists ε such that $v = u \pm \varepsilon\varphi \in C$. Using

[5] Since we are working in a finite dimensional, compact setting, we do not have to really bother about monotonicity.

these test-functions in the variational inequality, we obtain that $-u'' = f$ on I, and thus on E. Outside of E, $u = g$. In the case when $f = 0$, we thus see that u is affine on each connected component of E and equal to g elsewhere.

Let us see informally that the graph of u assumes the shape that an elastic string affixed at both extremities takes when it is stretched above an obstacle described by the graph of g, which is where such problems take their name from. We have just seen that the graph of u is composed of straight segments wherever is does not touch the obstacle. What happens at points where it leaves the obstacle? For simplicity, we assume that g is of class C^2, and that we have a point $x_0 \in]0, 1[$ such that $u(x) = g(x)$ for $x \leq x_0$ and $u(x) > g(x)$ for $x > x_0$, $|x - x_0|$ being small enough. Then u is affine to the right of x_0, with derivative α. Of course, $\alpha \geq g'(x_0)$. Let us show that $\alpha = g'(x_0)$, that is to say that the string leaves the obstacle tangentially.

There holds

$$u'(x) = g'(x) \text{ for } x \leq x_0, \text{ and } u'(x) = \alpha \text{ for } x > x_0.$$

Taking the distributional derivative of this function, we obtain

$$u'' = g''\mathbf{1}_{x\leq x_0} + (\alpha - g'(x_0))\delta_{x_0},$$

in a neighborhood of x_0. Consequently, for any test-function $v \in C$ such that $v - u$ has support in this neighborhood, we have

$$0 \leq \langle -u'', v - u\rangle = \langle -g''\mathbf{1}_{x\leq x_0}, v - g\rangle - (\alpha - g'(x_0))(v(x_0) - g(x_0)).$$

Let

$$h_n(x) = \begin{cases} 0 & \text{if } x < -\frac{1}{n}, \\ n(x + \frac{1}{n}) & \text{if } -\frac{1}{n} \leq x < 0, \\ 1 & \text{if } x \geq 0. \end{cases}$$

We take $v(x) = g(x) + h_n(x - x_0)$ in a neighborhood of x_0, so that $v(x_0) - g(x_0) = 1$ and

$$|\langle g''\mathbf{1}_{x\leq x_0}, v - g\rangle| = \left|\int_{x_0-\frac{1}{n}}^{x_0} g''(x)h_n(x - x_0)\,dx\right| \leq \frac{\max|g''|}{n} \to 0$$

when $n \to +\infty$. We thus obtain in the limit

$$0 \leq -(\alpha - g'(x_0)),$$

hence the result. Of course, this depends on the regularity of the obstacle. If the latter is rough, for instance a piecewise affine function, then we can clearly have cases in which the string does not leave the obstacle tangentially, see Fig. 8.2.

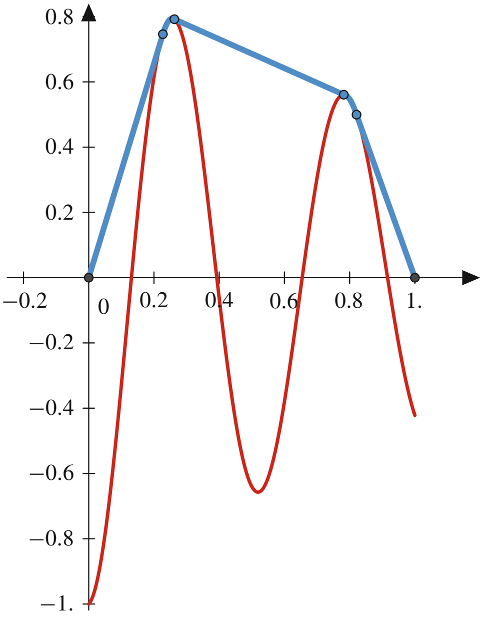


Fig. 8.2 The solution of a 1-d obstacle problem, the free boundary is figured by dots

More generally, for all positive test-functions $\varphi \in \mathscr{D}(]0,1[)$, there holds $v = u + \varphi \in C$. Consequently, the variational inequality implies that the distribution u'' is negative. Therefore, it is a measure, which is supported on the complement of E, and the function u' is decreasing, which in turn shows that the function u is concave. It actually can be shown that

$$u = \inf(v \in C, v \text{ is concave}).$$

From what we have seen above in Proposition 8.2, the obstacle problem may also be formulated as a minimization problem,

$$J(u) = \min_{v \in C} J(v), \text{ with } J(v) = \frac{1}{2}\int_\Omega |\nabla v|^2\,dx.$$

This is actually a kind of linearization of the *minimal surface problem* with an obstacle, where

$$J(v) = \frac{1}{2}\int_\Omega \sqrt{1 + |\nabla v|^2}\,dx,$$

which describes the shape of a soap film stretched over the obstacle and attached to the boundary of Ω. Now this is a much more difficult problem, since the integrand has linear growth at infinity, which seems to indicate that the function space should be $W_0^{1,1}(\Omega)$. Unfortunately, $W_0^{1,1}(\Omega)$ is not a nice space. In particular, it is not reflexive. Another function space is required, which makes for a much more difficult theory, see [20] for a general overview. In the one-dimensional case, the minimal surface problem becomes a minimal length problem,

$$J(u) = \min_{v \in C} J(v), \text{ with } J(v) = \frac{1}{2}\int_0^1 \sqrt{1 + (v')^2}\, dx.$$

See [14] for questions of regularity of the solution of the obstacle problem and also regularity of the set on which it coincides with the obstacle, the boundary of which is an unknown set, called a *free boundary*.

Let us give another example of variational inequality, which is not an obstacle problem. This example is about the elasto-plastic torsion of a beam. We are given an open subset Ω of $\mathbb{R}^2$ representing the cross-section of the beam and consider the closed bounded convex set

$$C = \{v \in H_0^1(\Omega); |\nabla v| \leq 1 \text{ almost everywhere } \Omega\},$$

with the monotone operator $A(v) = -\Delta v$ and $f = 1$. The solution u of the problem represents some stress components associated with a torsion angle.[6] It presents two regions, that where $|\nabla u| < 1$, in which the material remains elastic, and that where $|\nabla u| = 1$, in which plasticity phenomena appear. The interface between the elastic and the plastic regions is also an unknown free boundary. For free boundary value problems, more generally, see [32]. See also [23] for other examples of variational inequalities stemming from mechanics and physics.

8.5 Pseudo-Monotone Operators

Monotone operators are generalisations of the differentials of convex functionals. Nevertheless, by observing the existence proofs for a variational inequality, we can realise that it is possible to go one step further in abstraction, hence in generality, by singling out the ingredients that are actually essential in the convergence proofs for the Galerkin method. We have for instance already seen that existence in finite dimension has essentially nothing to do with monotonicity. Of course, such a

[6] All material and mechanical constants have been set to 1.

generalisation is only useful if it has real applications, otherwise it would be vain. We will return to this point later on. We thus introduce the following definitions, see [9, 48].

Definition 8.3. i) We say that $A\colon V \to V'$ is of M type if

$$\left.\begin{array}{r} u_n \rightharpoonup u \text{ in } V \text{ weak} \\ A(u_n) \rightharpoonup \psi \text{ in } V' \text{ weak} \\ \limsup\limits_{n\to+\infty} \langle A(u_n), u_n\rangle \le \langle \psi, u\rangle \end{array}\right\} \Longrightarrow \psi = A(u). \tag{8.11}$$

ii) We say that A is sense 1 pseudo-monotone if

$$\left.\begin{array}{r} u_n \rightharpoonup u \text{ in } V \text{ weak} \\ A(u_n) \rightharpoonup \psi \text{ in } V' \text{ weak} \\ \limsup\limits_{n\to+\infty} \langle A(u_n), u_n\rangle \le \langle \psi, u\rangle \end{array}\right\} \Longrightarrow \psi = A(u) \text{ and } \langle A(u_n), u_n\rangle \to \langle A(u), u\rangle. \tag{8.12}$$

iii) We say that A is sense 2 pseudo-monotone if

$$\left.\begin{array}{r} u_n \rightharpoonup u \text{ in } V \text{ weak} \\ \limsup\limits_{n\to+\infty} \langle A(u_n), u_n - u\rangle \le 0 \end{array}\right\} \Longrightarrow \forall v \in V,\ \liminf_{n\to+\infty} \langle A(u_n), u_n - v\rangle \ge \langle A(u), u - v\rangle. \tag{8.13}$$

Remark 8.8. Definition i) is useful for equations ($C = V$) and definitions ii) and iii) for inequalities ($C \ne V$). Definition iii) may be restricted to $v \in C$. □

The two definitions of pseudo-monotonicity are basically equivalent. More precisely,

Proposition 8.3. *If A is bounded, then it is sense 1 pseudo-monotone if and only if it is sense 2 pseudo-monotone.*

Proof. Let us first assume that A is sense 2 pseudo-monotone (but not necessarily bounded). Let u_n be a sequence such that

$$u_n \rightharpoonup u,\ \ A(u_n) \rightharpoonup \psi \text{ and } \limsup_{n\to+\infty} \langle A(u_n), u_n\rangle \le \langle \psi, u\rangle.$$

It follows that

$$\limsup_{n\to+\infty} \langle A(u_n), u_n - u\rangle = \limsup_{n\to+\infty} (\langle A(u_n), u_n\rangle - \langle A(u_n), u\rangle) \le 0.$$

Thus, by sense 2 pseudo-monotonicity, there holds for all $v \in V$,

$$\liminf_{n\to+\infty}\langle A(u_n), u_n - v\rangle \geq \langle A(u), u - v\rangle,$$

so that by expanding the left-hand side bracket, we obtain

$$\liminf_{n\to+\infty}\langle A(u_n), u_n\rangle - \langle \psi, v\rangle \geq \langle A(u), u - v\rangle.$$

Taking $v = u$, we deduce that $\liminf_{n\to+\infty}\langle A(u_n), u_n\rangle \geq \langle \psi, v\rangle$, and consequently, that $\langle A(u_n), u_n\rangle \to \langle \psi, u\rangle$. We replace this in the above equality and obtain

$$\langle \psi, u - v\rangle \geq \langle A(u), u - v\rangle.$$

The choice $v = u + w$ then shows that $\psi = A(u)$ and that A is sense 1 pseudo-monotone.

Let us now assume that A is sense 1 pseudo-monotone and bounded. Let u_n we a sequence such that $u_n \rightharpoonup u$ and $\limsup_{n\to+\infty}\langle A(u_n), u_n - u\rangle \leq 0$. Let $v \in V$ be arbitrary. Since A is bounded, we can extract a subsequence such that

$$A(u_{n'}) \rightharpoonup \psi \quad \text{and} \quad \langle A(u_{n'}), u_{n'} - v\rangle \to \liminf_{n\to+\infty}\langle A(u_n), u_n - v\rangle.$$

Now $\langle A(u_{n'}), u\rangle \to \langle \psi, u\rangle$, so we first see that $\limsup_{n'\to+\infty}\langle A(u_{n'}), u_{n'}\rangle \leq \langle \psi, u\rangle$. By sense 1 pseudo-monotonicity, it follows that $\psi = A(u)$ and $\langle A(u_{n'}), u_{n'}\rangle \to \langle A(u), u\rangle$. Consequently

$$\langle A(u_{n'}), u_{n'} - v\rangle \to \langle A(u), u - v\rangle,$$

and A is sense 2 pseudo-monotone. □

Remark 8.9. Notice that if A is sense 1 pseudo-monotone, then it is obviously of M type. □

Pseudo-monotonicity is a generalisation of monotonicity.

Theorem 8.6. *If A is hemicontinuous and monotone, then A is sense* 1 *pseudo-monotone.*

Proof. Let A be a monotone hemicontinuous operator and u_n a sequence such that $u_n \rightharpoonup u$, $A(u_n) \rightharpoonup \psi$ and $\limsup_{n\to+\infty}\langle A(u_n), u_n\rangle \leq \langle \psi, u\rangle$. Since

$$\langle A(u_n) - A(u), u_n - u\rangle \geq 0,$$

expansion of the bracket yields

$$\langle A(u_n), u_n\rangle \geq \langle A(u_n), u\rangle + \langle A(u), u_n - u\rangle.$$

The right-hand side converges to $\langle \psi, u\rangle$. Therefore, the inferior limit is such that

$$\liminf_{n\to+\infty}\langle A(u_n), u_n\rangle \geq \langle \psi, u\rangle,$$

from which it follows that

$$\langle A(u_n), u_n\rangle \longrightarrow \langle \psi, u\rangle.$$

To show that $\psi = A(u)$, we use Minty's trick again. For all $w \in V$,

$$\langle A(u_n) - A(w), u_n - w\rangle \geq 0,$$

so that expanding and then passing to the limit, we obtain

$$\langle \psi - A(w), u - w\rangle \geq 0.$$

We take $w = u + tv$ with $t > 0$, so that

$$-t\langle \psi, v\rangle + t\langle A(u + tv), v\rangle \geq 0.$$

We divide by t, then let t tend to 0, so that by hemicontinuity

$$\langle A(u), v\rangle \geq \langle \psi, v\rangle,$$

for all v in V. This implies $A(u) = \psi$. □

In the next section, we will see that there are pseudo-monotone operators that are not monotone. Let us first give a few existence results.

Theorem 8.7. *Any type M operator A that is bounded and coercive, is surjective.*

Proof. We once more apply the Galerkin method. Let us first show that A is continuous from V strong into V' weak. Let us thus be given a sequence $v_n \to v$ strongly. Since $A(v_n)$ is bounded, we can extract a subsequence $v_{n'}$ such that $A(v_{n'}) \rightharpoonup \psi$ weakly. There thus holds $\langle A(v_{n'}), v_{n'}\rangle \to \langle \psi, v\rangle$ and since A is of M type, we deduce that $\psi = A(v)$. We conclude by uniqueness of the limit.

This continuity and the coercivity easily imply existence in finite dimension by the Brouwer theorem variant 2.7, i.e., for all $f \in V'$, there exists $u_m \in V_m$ such that for all $v \in V_m$,

$$\langle A(u_m), v\rangle = \langle f, v\rangle. \tag{8.14}$$

In particular,

$$\langle A(u_m), u_m\rangle = \langle f, u_m\rangle.$$

Due to coercivity, it follows that the sequence u_m is bounded. The sequence $A(u_m)$ is thus also bounded, and we can assume that $u_m \rightharpoonup u$ and $A(u_m) \rightharpoonup \psi$. Passing to the limit in the last equality, we see that

$$\langle A(u_m), u_m \rangle \to \langle f, u \rangle. \tag{8.15}$$

Furthermore, according to (8.14), for all v in $\bigcup V_m$, there holds

$$\langle f, v \rangle = \langle A(u_m), v \rangle \to \langle \psi, v \rangle.$$

Thus, by density, for all v in V, there also holds

$$\langle f, v \rangle = \langle \psi, v \rangle,$$

so that $\psi = f$. Combining this relation with (8.15), we obtain

$$\langle A(u_m), u_m \rangle \to \langle \psi, u \rangle,$$

and finally A is a type M operator, thus it follows that $\psi = A(u) = f$. □

We show a similar result for variational inequalities.

Theorem 8.8. *Let C be a nonempty, closed convex subset of V and $A\colon V \to V'$ a bounded pseudo-monotone operator* (*coercive in case C is unbounded*)*. Then, for all $f \in V'$, the variational inequality* (8.3) *admits at least one solution.*

Proof. We first note that if A is pseudo-monotone bounded, then it is of M type, and we have just seen that a M type bounded operator is continuous from V strong to V' weak.

We start with the case of a bounded convex C, which we approximate by a sequence C_m of closed, finite dimensional convex subsets as in Lemma 8.2. The variational inequality on C_m, (8.4), thus admits a solution u_m exactly as in Lemma 8.3, since the only ingredient needed is the continuity of A from V strong to V' weak.

Now C is bounded, we therefore extract a subsequence such that $u_m \rightharpoonup u$ in V weak and $A(u_m) \rightharpoonup \psi$ in V' weak, with of course $u \in C$, since the latter is a closed convex.

Using the finite dimensional variational inequality, we then deduce that there holds $\limsup_{m\to+\infty} \langle A(u_m), u_m \rangle \leq \langle \psi, u \rangle$, exactly as in the proof of Lemma 8.6.

By the very definition of pseudo-monotonicity, it follows that $\psi = A(u)$ and that $\langle A(u_m), u_m \rangle \to \langle A(u), u \rangle$ when $m \to +\infty$.[7]

[7] This precisely why pseudo-monotonicity was defined this way. The whole difficulty being displaced to the question of how to show that such and such an operator is actually pseudo-monotone, see Sect. 8.6.

Passing to the limit in the variational inequality (8.4) when $m \to +\infty$, we immediately deduce that $u \in C$ is a solution of the variational inequality on C, i.e., (8.3).

The C unbounded case follows from the bounded case with the exact same proof as that of Theorem 8.3. □

8.6 Leray-Lions Operators

This is a class of operators introduced in [47]. We work in $W_0^{1,p}(\Omega)$ spaces, where Ω is a bounded open subset of $\mathbb{R}^d$. Let $F : \Omega \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ be a function such that

i) F is measurable with respect to $x \in \Omega$, continuous with respect to $(s, \xi) \in \mathbb{R} \times \mathbb{R}^d$,

ii) There exists $k \in L^{p'}(\Omega)$ and a constant C such that

$$\forall (x, s, \xi) \in \Omega \times \mathbb{R} \times \mathbb{R}^d, \quad |F(x, s, \xi)| \leq k(x) + C(|s|^{p-1} + |\xi|^{p-1}),$$

iii) For x and s fixed, F is monotone with respect to ξ,

iv) There exists a constant $\alpha > 0$ such that

$$\forall (x, s, \xi) \in \Omega \times \mathbb{R} \times \mathbb{R}^d, \quad F(x, s, \xi) \cdot \xi \geq \alpha |\xi|^p.$$

Theorem 8.9. *The operator*

$$A : u \mapsto -\mathrm{div}\big(F(x, u, \nabla u)\big)$$

is bounded, pseudo-monotone and coercive from $W_0^{1,p}(\Omega)$ into $W^{-1,p'}(\Omega)$.

Such operators are called Leray-Lions operators.

Proof. Due to the growth condition ii), the operator A is well defined in these spaces, using the duality bracket

$$\langle A(u), v \rangle = \int_\Omega F(x, u(x), \nabla u(x)) \cdot \nabla v(x)\, dx.$$

Furthermore, by Carathéodory's theorem, it is continuous from $W_0^{1,p}(\Omega)$ strong into $W^{-1,p'}(\Omega)$ strong, and it is bounded. Finally, it is visibly coercive by condition iv).

Let us show that A is sense 1 pseudo-monotone. We thus consider a sequence $u_n \in W_0^{1,p}(\Omega)$ such that

$$\begin{cases} u_n \rightharpoonup u, \\ -\mathrm{div}\,(F(x,u_n,\nabla u_n)) \rightharpoonup \psi, \\ \limsup\limits_{n\to+\infty} \displaystyle\int_\Omega F(x,u_n,\nabla u_n)\cdot\nabla u_n\,dx \le \langle\psi,u\rangle. \end{cases}$$

By the growth condition ii), $F(x,u_n,\nabla u_n)$ is bounded in $L^{p'}(\Omega;\mathbb{R}^d)$. We can thus extract a subsequence, still denoted u_n, such that

$$F(x,u_n,\nabla u_n) = g_n \rightharpoonup g \text{ dans } L^{p'}(\Omega;\mathbb{R}^d),$$

and it immediately follows that

$$\psi = -\mathrm{div}\,g.$$

By hypothesis, we thus have

$$\limsup_{n\to+\infty}\int_\Omega g_n\cdot\nabla u_n\,dx \le \int_\Omega g\cdot\nabla u\,dx. \tag{8.16}$$

We now use the monotonicity of F with respect to its third variable. For all ϕ in $L^p(\Omega;\mathbb{R}^d)$—which is not necessarily a gradient—there holds

$$\int_\Omega \big(F(x,u_n,\nabla u_n) - F(x,u_n,\phi)\big)\cdot(\nabla u_n-\phi)\,dx \ge 0,$$

or

$$\int_\Omega g_n\cdot\nabla u_n\,dx - \int_\Omega g_n\cdot\phi\,dx - \int_\Omega F(x,u_n,\phi)\cdot(\nabla u_n-\phi)\,dx \ge 0. \tag{8.17}$$

By the Rellich-Kondrašov theorem, $u_n \to u$ strongly in $L^p(\Omega)$ and we can thus extract a further subsequence such that $u_n \to u$ almost everywhere in Ω and is dominated by a function of $L^p(\Omega)$. Since F is continuous with respect to s and satisfies the growth condition ii), it follows from the Lebesgue dominated convergence theorem that

$$F(x,u_n,\phi) \longrightarrow F(x,u,\phi) \text{ in } L^{p'}(\Omega;\mathbb{R}^d) \text{ strong.}$$

Consequently

$$\int_\Omega F(x, u_n, \phi) \cdot (\nabla u_n - \phi)\, dx \longrightarrow \int_\Omega F(x, u, \phi) \cdot (\nabla u - \phi)\, dx,$$

and besides,

$$\int_\Omega g_n \cdot \phi\, dx \longrightarrow \int_\Omega g \cdot \phi\, dx,$$

so that, passing to the superior limit in (8.17) and using inequality (8.16), we obtain

$$\int_\Omega g \cdot \nabla u\, dx - \int_\Omega g \cdot \phi\, dx - \int_\Omega F(x, u, \phi) \cdot (\nabla u - \phi)\, dx \geq 0,$$

or in other words

$$\int_\Omega \big(g - F(x, u, \phi)\big) \cdot (\nabla u - \phi)\, dx \geq 0.$$

It is now time for Minty's trick again. For any $\varphi \in \mathscr{D}(\Omega; \mathbb{R}^d)$, we take $\phi = \nabla u + t\varphi$ with $t > 0$. It follows that

$$\int_\Omega \big(g - F(x, u, \nabla u + t\varphi)\big) \cdot \varphi\, dx \geq 0,$$

hence, letting t tend to 0 and using the dominated convergence theorem in conjunction with the growth condition ii),

$$\int_\Omega \big(g - F(x, u, \nabla u)\big) \cdot \varphi\, dx \geq 0.$$

We deduce from this that $g = F(x, u, \nabla u)$, that is to say $\psi = A(u)$.

We now go back to the monotonicity of F with respect to its third argument,

$$\int_\Omega \big(F(x, u_n, \nabla u_n) - F(x, u_n, \nabla u)\big) \cdot (\nabla u_n - \nabla u)\, dx \geq 0,$$

which expands as

$$\int_\Omega F(x, u_n, \nabla u_n) \cdot \nabla u_n\, dx \geq \int_\Omega F(x, u_n, \nabla u_n) \cdot \nabla u\, dx \\ + \int_\Omega F(x, u_n, \nabla u) \cdot \nabla(u_n - u)\, dx. \quad (8.18)$$

Just as before,

$$\int_\Omega F(x,u_n,\nabla u)\cdot\nabla(u_n-u)\,dx \longrightarrow 0,$$

$$\int_\Omega F(x,u_n,\nabla u_n)\cdot\nabla u\,dx \longrightarrow \int_\Omega g\cdot\nabla u\,dx = \int_\Omega F(x,u,\nabla u)\cdot\nabla u\,dx,$$

consequently, passing to the inferior limit in (8.18), we obtain

$$\liminf_{n\to+\infty}\int_\Omega F(x,u_n,\nabla u_n)\cdot\nabla u_n\,dx \geq \int_\Omega F(x,u,\nabla u)\cdot\nabla u\,dx = \langle A(u),u\rangle.$$

By hypothesis,

$$\limsup_{n\to+\infty}\int_\Omega F(x,u_n,\nabla u_n)\cdot\nabla u_n\,dx \leq \langle\psi,u\rangle = \langle A(u),u\rangle,$$

since by the previous step, $\psi = A(u)$. We have thus shown that

$$\int_\Omega F(x,u_n,\nabla u_n)\cdot\nabla u_n\,dx = \langle A(u_n),u_n\rangle \longrightarrow \langle A(u),u\rangle,$$

hence that A is sense 1 pseudo-monotone. □

Remark 8.10. i) The equation $-\operatorname{div}(F(x,u,\nabla u)) = f$ reads, at least formally,

$$-\frac{\partial F_i}{\partial\xi_k}(x,u,\nabla u)\partial_{ik}u - \frac{\partial F_i}{\partial s}(x,u,\nabla u)\partial_i u - \frac{\partial F_i}{\partial x_i}(x,u,\nabla u) = f.$$

A Leray-Lions operator is therefore quasilinear in general. It is also elliptic, with the same argument as that already used in the monotone case.

ii) Let us give an example of a pseudo-monotone operator that is not monotone. We take a Leray-Lions operator, that does not however satisfy condition iv) of coercivity, for simplicity. Indeed, coercivity does not play any role in the above pseudo-monotonicity proof. Let thus $d = 1$, $\Omega =]0,1[$, $p = 2$ and $F(x,s,\xi) = 4\pi^2 xs+\xi$. It is clear that conditions i), ii) and iii) are satisfied, hence the underlying differential operator $u \mapsto -u'' - 4\pi^2 xu' - 4\pi^2 u$ is pseudo-monotone from $H_0^1(]0,1[)$ to $H^{-1}(]0,1[)$.

Let us compute $I = \int_0^1 F\big(x,u(x),u'(x)\big)u'(x)\,dx$ for $u(x) = \sin(\pi x)$. There holds

$$\begin{aligned} I &= \int_0^1 \big(4\pi^2 xu(x)u'(x) + (u'(x))^2\big)\,dx \\ &= 4\pi^3\int_0^1 x\sin(\pi x)\cos(\pi x)\,dx + \pi^2\int_0^1 \cos^2(\pi x)\,dx \end{aligned}$$

$$= 2\pi^3 \int_0^1 x \sin(2\pi x)\, dx + \frac{\pi^2}{2} \int_0^1 \big(1 + \cos(2\pi x)\big)\, dx$$
$$= -\pi^2 + \frac{\pi^2}{2} < 0.$$

The bilinear form associated with the operator is not positive, hence the operator is not monotone.

Now of course, this also prevents it from being coercive, since it is a linear operator. Nonetheless, existence for associated variational inequalities in bounded convex subsets still holds. More complicated, nonlinear examples would be needed to either satisfy condition iv) or perhaps less demanding inequalities that still imply coercivity, see Exercise 6.

iii) The same definitions can be applied in $W_0^{1,p}(\Omega; \mathbb{R}^m)$ to deal with systems. There is also a concept of quasimonotone function that generalizes the gradients of quasiconvex functions. It is not clear that an analogue of polyconvexity is known, that could be used to construct nontrivial quasimonotone functions. □

8.7 Exercises of Chap. 8

1. Solve the elasto-plastic torsion problem.

2. Prove Theorem 8.8.

3. Let V be a reflexive, separable Banach space and C a nonempty closed convex subset of V. All operators are from V to V'. We say that an operator β which is monotone, hemicontinuous and bounded is a penalization operator associated with C if

$$v \in C \Longleftrightarrow \beta(v) = 0.$$

3.1. Let A_1 be a bounded pseudo-monotone operator and A_2 a bounded, hemicontinuous, pseudo-monotone operator. Show that $A_1 + A_2$ is bounded pseudo-monotone.

3.2. Let A be a bounded, coercive pseudo-monotone operator and $f \in V'$. For all $\varepsilon > 0$, we consider the problem:

$$A(u^\varepsilon) + \frac{1}{\varepsilon}\beta(u^\varepsilon) = f.$$

Show that this problem admits at least one solution and that there exists a subsequence ε' such that $u^{\varepsilon'}$ converges weakly in V to a solution $u \in C$ of the variational inequality:

$$\forall\, v \in C, \quad \langle A(u) - f, v - u\rangle \geq 0.$$

3.3. We assume that V is a Hilbert space, identified with its dual via its inner product. Let P be the orthogonal projection onto C. Show that the operator β defined by $\beta(v) = v - P(v)$ is a penalization operator associated with C.

4. Let Ω be an open bounded subset of $\mathbb{R}^d$, equipped with the standard Euclidean inner product.

4.1. Let A_1 and A_2 be two monotone operators from $H_0^1(\Omega)$ into $H^{-1}(\Omega)$. Show that $A = A_1 + A_2$ is monotone.

4.2. Let us be given a mapping $B\colon \mathbb{R}^d \to \mathbb{R}^d$ which is continuous, strictly monotone for the Euclidean inner product, and such that there exists $C \geq 0$ and $\alpha > 0$ with

$$\forall\, \xi \in \mathbb{R}^d, \quad |B(\xi)| \leq C(1 + |\xi|) \quad \text{and} \quad B(\xi)\cdot\xi \geq \alpha|\xi|^2.$$

Let $\psi \in H_0^1(\Omega)$. For all $\varepsilon > 0$ and all $v \in H_0^1(\Omega)$, we set

$$\mathscr{B}_\varepsilon(v) = -\operatorname{div}(B(\nabla v)) - \frac{1}{\varepsilon}(v - \psi)_-.$$

Show that for all $f \in H^{-1}(\Omega)$, there exists a unique $u^\varepsilon \in H_0^1(\Omega)$ such that

$$\mathscr{B}_\varepsilon(u^\varepsilon) = f.$$

4.3. Show that u^ε is bounded in $H_0^1(\Omega)$ independently of ε. We extract a subsequence (still denoted u^ε) such that $u^\varepsilon \rightharpoonup u$ in $H_0^1(\Omega)$ weak when $\varepsilon \to 0$.

4.4. Show that $u \geq \psi$ almost everywhere in Ω.

4.5. Let $C = \{v \in H_0^1(\Omega);\ v \geq \psi \text{ a.e. in } \Omega\}$. Show that C is a nonempty closed convex set and that u satisfies

$$\forall\, v \in C, \langle \mathscr{B}(u) - f, v - u\rangle_{H^{-1}(\Omega), H_0^1(\Omega)} \geq 0$$

where

$$\mathscr{B}(v) = -\operatorname{div}(B(\nabla v)).$$

Deduce from this that the whole sequence converges.

4.6. We set $\mu = f + \operatorname{div}(B(\nabla v)) \in H^{-1}(\Omega)$. Show that μ is a nonpositive Radon measure that satisfies $\langle \mu, u - \psi\rangle_{H^{-1}(\Omega), H_0^1(\Omega)} = 0$.

4.7. We assume that both u and ψ are continuous on $\overline{\Omega}$. Let $E = \{x \in \Omega;\ u(x) > \psi(x)\}$. Show that $-\operatorname{div}(B(\nabla u)) = f$ in E (in which sense?). What happens if $d = 1$?

4.8. In this question, we assume that there exists a constant $\gamma > 0$ such that for all $\xi, \xi' \in \mathbb{R}^d$, $(B(\xi) - B(\xi')) \cdot (\xi - \xi') \geq \gamma|\xi - \xi'|^2$. Show that u^ε tends to u strongly in $H_0^1(\Omega)$.

4.9. We no longer assume the existence of any such γ. Show that the function

$$g^\varepsilon = (B(\nabla u^\varepsilon) - B(\nabla u)) \cdot (\nabla u^\varepsilon - \nabla u)$$

tends to 0 in $L^1(\Omega)$ strong. Deduce from this that there exists a subsequence such that $\nabla u^{\varepsilon'}$ converges almost everywhere, then that $u^\varepsilon \to u$ in $H_0^1(\Omega)$ strong.

4.10. We assume $B(\xi) = \xi$ and $\psi = 0$. Using appropriate test-functions, show that there exists a constant $C > 0$ such that $\|(u^\varepsilon)_-\|_{H_0^1(\Omega)} \leq C\sqrt{\varepsilon}$ and that $\|(u^\varepsilon)_+ - u\|_{H_0^1(\Omega)} \leq C\sqrt{\varepsilon}$. Deduce that $\|u^\varepsilon - u\|_{H_0^1(\Omega)} \leq C\sqrt{\varepsilon}$.

5. Let Ω be a bounded open subset of $\mathbb{R}^d$, $d \leq 3$, with the usual Euclidean structure, $s = 2$ if $d = 1$, $1 \leq s < 2$ if $d = 2$, $s = 3/2$ if $d = 3$, $f \in H^{-1}(\Omega)$, F a continuous monotone function from $\mathbb{R}^d$ into $\mathbb{R}^d$ such that there exists $C \geq 0$ and $\alpha > 0$ with

$$\forall \xi \in \mathbb{R}^d, \quad \begin{cases} |F(\xi)| \leq C(1 + |\xi|), \\ F(\xi) \cdot \xi \geq \alpha|\xi|^2. \end{cases}$$

We are interested in the problem:

$$\begin{cases} u \in H_0^1(\Omega), \\ -\operatorname{div}(F(\nabla u)) + u\partial_1 u = f \text{ in } \mathscr{D}'(\Omega). \end{cases} \tag{8.19}$$

5.1. Show that problem (8.19) is equivalent to the variational formulation

$$\forall v \in H_0^1(\Omega) \cap L^{s'}(\Omega), \quad \int_\Omega F(\nabla u) \cdot \nabla v \, dx + \int_\Omega u\partial_1 u v \, dx = \langle f, v \rangle.$$

5.2. Let $w_i \in \mathscr{D}(\Omega)$ be a countable family whose linear combinations are dense in $H_0^1(\Omega) \cap L^{s'}(\Omega)$. Let $V_m = \operatorname{vect}\{w_1, w_2, \dots, w_m\}$ be the vector space spanned by the first m vectors. Show that problem: Find $u_m \in V_m$ tel que

$$\forall v \in V_m, \quad \int_\Omega F(\nabla u_m) \cdot \nabla v \, dx + \int_\Omega u_m \partial_1 u_m v \, dx = \langle f, v \rangle,$$

admits at least one solution.

5.3. Show that there is a subsequence still denoted u_m such that $u_m \rightharpoonup u$ in $H_0^1(\Omega)$ weak, $F(\nabla u_m) \rightharpoonup g$ in $L^2(\Omega; \mathbb{R}^d)$ weak and $-\operatorname{div}(F(\nabla u_m)) \rightharpoonup \xi$ in $H^{-1}(\Omega)$ weak.

5.4. Show that

$$-\operatorname{div} g + u\partial_1 u = f \text{ in } \mathscr{D}'(\Omega).$$

5.5. Show that

$$\liminf_{m\to+\infty}\int_\Omega F(\nabla u_m)\cdot\nabla u_m\,dx \geq \int_\Omega g\cdot\nabla u\,dx.$$

5.6. Show that

$$\int_\Omega u_m^2\partial_1 u_m\,dx \to \int_\Omega u^2\partial_1 u\,dx \text{ when } m\to+\infty,$$

(*Hint:* $d \leq 3$.) Deduce that

$$\limsup_{m\to+\infty}\int_\Omega F(\nabla u_m)\cdot\nabla u_m\,dx \leq \int_\Omega g\cdot\nabla u\,dx.$$

5.7. Show that $g = F(\nabla u)$ and that u is a solution of problem (8.19).

6. We want to construct a coercive, pseudo-monotone operator that is not monotone. We use the same setting as that of Remark 8.10, $d = 1$, $\Omega =]0,1[$ and $p = 2$.

6.1. Let $g\colon \mathbb{R}\to\mathbb{R}$ be a C^1 mapping that is not affine. Show that the mapping

$$\begin{aligned} J\colon \mathscr{D}(\Omega)\times\mathscr{D}(\Omega) &\to \mathbb{R} \\ (\varphi,\psi) &\mapsto \int_0^1 \big(g(\psi)\varphi' + g(\varphi)\psi'\big)\,dx \end{aligned}$$

is not identically 0. (*Hint:* Take $\varphi \neq 0$ and consider the mapping $\varepsilon \mapsto J(\varphi,\varepsilon\psi)$ around $\varepsilon = 0$ for an adequate choice of ψ.)

6.2. Let $g\colon \mathbb{R}\to\mathbb{R}$ be a C^1 mapping that is bounded by 1 in absolute value, and not constant. Define $F(x,s,\xi) = \lambda g(s) + \xi$ for some $\lambda\in\mathbb{R}$. Show that the operator $u\mapsto -u'' - \lambda g'(u)u'$ is pseudo-monotone from $H_0^1(\Omega)$ to $H^{-1}(\Omega)$.

6.3. Show that there exists $\lambda\in\mathbb{R}$ such that it is not monotone.

6.4. Show that it is coercive. (*Hint:* Condition iv) is not satisfied, work around it.)

References

The bibliography on the topics touched on in this book is extremely vast. The following list of references is a concise, and at times somewhat arbitrary, selection of both reference monographs and research articles.

1. Acerbi, E., Fusco, N.: Semicontinuity problems in the calculus of variations. Arch. Ration. Mech. Anal. **62**, 371–387 (1984)
2. Adams, R.A.: Sobolev Spaces. Academic Press, New York (1975)
3. Agmon, S., Douglis, A., Nirenberg, L.: Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions I. Commun. Pure Appl. Math. **12**, 623–727 (1959)
4. Agmon, S., Douglis, A., Nirenberg, L.: Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions II. Commun. Pure Appl. Math. **17**, 35–92 (1964)
5. Ball, J.M.: Convexity conditions and existence theorems in nonlinear elasticity. Arch. Ration. Mech. Anal. **63**, 337–403 (1977)
6. Ball, J.M.: A version of the fundamental theorem for Young measures. In: PDEs and Continuum Models of Phase Transitions (Nice, 1988). Lecture Notes in Physics, vol. 344, pp. 207–215. Springer, Berlin (1989)
7. Ball, J.M., Currie, J.C., Olver, P.J.: Null Lagrangians, weak continuity, and variational problems of arbitrary order. J. Funct. Anal. **41**, 135–174 (1981)
8. Bourbaki, N.: Espaces vectoriels topologiques. Hermann, Paris (1967)
9. Brezis, H.: Équations et inéquations non linéaires dans les espaces vectoriels en dualité. Ann. Inst. Fourier (Grenoble) **18**, 115–175 (1968)
10. Brezis, H.: Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert. North-Holland Mathematics Studies, vol. 5. Notas de Matemática, vol. 50. North-Holland Publishing Co., Amsterdam; American Elsevier Publishing Co., Inc., New York (1973)
11. Brezis, H.: Functional Analysis, Sobolev Spaces and Partial Differential Equations. Universitext. Springer, New York (2011)
12. Browder, F.: Problèmes non linéaires. Université de Montréal (1966)
13. Buttazzo, G., Giaquinta, M., Hildebrandt, S.: One-Dimensional Variational Problems, An Introduction. Oxford Lecture Series in Mathematics and Its Applications, vol. 15. Clarendon Press, Oxford (1998)
14. Caffarelli, L.A.: The obstacle problem revisited. J. Fourier Anal. Appl. **4**, 383–402 (1988)

H. Le Dret, *Nonlinear Elliptic Partial Differential Equations*, Universitext,
https://doi.org/10.1007/978-3-319-78390-1

15. Chazarain, J., Piriou, A.: Introduction à la théorie des équations aux dérivées partielles linéaires. Gauthier-Villars, Paris (1981)
16. Ciarlet, P.G.: Mathematical Elasticity. Vol. I. Three-dimensional Elasticity. Studies in Mathematics and Its Applications, vol. 20. North-Holland Publishing Co., Amsterdam (1988)
17. Ciarlet, P.G.: Linear and Nonlinear Functional Analysis with Applications. Society for Industrial and Applied Mathematics, Philadelphia (2013)
18. Dacorogna, B.: Direct Methods in the Calculus of Variations. Springer, Berlin (1989)
19. Dautray, R., Lions, J.-L.: Mathematical Analysis and Numerical Methods for Science and Technology. Vol. 5. Evolution Problems. Springer, Berlin (1992)
20. Dierkes, U., Hildebrandt, S., Sauvigny, F.: Minimal Surfaces. Revised and enlarged second edition. Grundlehren der Mathematischen Wissenschaften, vol. 339. Springer, Heidelberg (2010)
21. Dieudonné, J.: Éléments d'analyse, vol. 2. Gauthier-Villars, Paris (1974)
22. DiPerna, R.J.: Convergence of approximate solutions to conservation laws. Arch. Ration. Mech. Anal. **82**, 27–70 (1983)
23. Duvaut, G., Lions, J.-L.: Inequalities in Mechanics and Physics. Grundlehren der Mathematischen Wissenschaften, vol. 219. Springer, Berlin (1976)
24. Ekeland, I.: On the variational principle. J. Math. Anal. Appl. **47**, 324–353 (1974)
25. Ekeland, I., Temam, R.: Convex Analysis and Variational Problems. Classics in Applied Mathematics, vol. 28. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (1999). Corrected reprint of the 1976 English edition
26. Evans, L.C.: Quasiconvexity and partial regularity in the calculus of variations. Arch. Ration. Mech. Anal. **95**, 227–252 (1986)
27. Evans, L.C.: Weak Convergence Methods for Nonlinear Partial Differential Equations. Regional Conference Series in Mathematics, vol. 74. AMS, Providence (1990)
28. Evans, L.C., Gariepy, R.F.: Measure Theory and Fine Properties of Functions. Studies in Advanced Mathematics. CRC Press, Boca Raton (1992)
29. Fonseca, I.: The lower quasiconvex envelope of the stored energy function for an elastic crystal. J. Math. Pure Appl. **67**, 175–195 (1988)
30. Fonseca, I., Gangbo, W.: Degree Theory in Analysis and Applications. Oxford University Press, Oxford (1995)
31. Fonseca, I., Müller, S.: Quasiconvex integrals and lower semicontinuity in L^1. SIAM J. Math. Anal. **23**, 1081–1098 (1992)
32. Friedman, A.: Variational Principles and Free Boundary Value Problems. Wiley, New York (1982)
33. Geymonat, G.: Sui problemi ai limiti per i sistemi lineari ellittici. Ann. Mat. Pura Appl. **69**, 207–284 (1965)
34. Giaquinta, M.: Multiple Integrals in the Calculus of Variations and Nonlinear Elliptic Systems. Princeton University Press, Princeton (1983)
35. Giaquinta, M., Giusti, E.: Nonlinear elliptic systems with quadratic growth. Manuscr. Math. **148**, 323–349 (1978)
36. Gilbarg, D., Trudinger, N.S.: Elliptic Partial Differential Equations of Second Order, 2nd edn. Springer, Berlin (1983)
37. Giusti, E., Miranda, M.: Sulla regolarità delle soluzioni deboli di una classe di sistemi ellittici quasi-lineari. Arch. Ration. Mech. Anal. **115**, 329–365 (1991)
38. Grisvard, P.: Elliptic Problems in Nonsmooth Domains. Pitman, Marshfield (1985). Reprinted in Classics in Applied Mathematics. SIAM, Philadelphia (2011)
39. Hörmander, L.: The Analysis of Linear Partial Differential Operators. I. Distribution Theory and Fourier Analysis. Grundlehren der Mathematischen Wissenschaften, vol. 256. Springer, Berlin (1983)
40. Kavian, O.: Introduction à la théorie des points critiques et applications aux problèmes elliptiques. Springer, Paris (1993)
41. Kinderlehrer, D., Pedregal, P.: Characterizations of Young measures generated by gradients. Arch. Ration. Mech. Anal. **115**, 329–365 (1991)

42. Kinderlehrer, D., Pedregal, P.: Gradient Young measures generated by sequences in Sobolev spaces. J. Geom. Anal. **4**, 59–90 (1994)
43. Kinderlehrer, D., Stampacchia, G.: An Introduction to Variational Inequalities and Their Applications. Academic Press, New York (1980). Reprinted in Classics in Applied Mathematics. SIAM, Philadelphia (2000)
44. Kristensen, J.: On the non-locality of quasiconvexity. Ann. Inst. H. Poincaré Anal. Non Linéaire **16**, 1–13 (1999)
45. Le Dret, H., Raoult, A.: Variational convergence for nonlinear shell models with directors and related semicontinuity and relaxation results. Arch. Ration. Mech. Anal. **154**, 101–134 (2000)
46. Le Dret, H., Raoult, A.: Hexagonal lattices with three-point interactions. J. Math. Pure Appl. (2017). https://doi.org/10.1016/j.matpur.2017.05.008
47. Leray, J., Lions, J.-L.: Quelques résultats de Visik sur les problèmes elliptiques non linéaires par les méthodes de Minty-Browder. Bull. Soc. Math. France **93**, 97–107 (1965)
48. Lions, J.-L.: Quelques méthodes de résolution des problèmes aux limites non linéaires. Dunod, Gauthier-Villars, Paris (1969). Reprinted by Dunod, Paris (2002)
49. Lions, J.-L., Magenes, E.: Non-homogeneous Boundary Value Problems and Applications, vol. 1. Grundlehren der mathematischen Wissenschaften, Band 181. Springer, New York (1972)
50. Marcellini, P.: Approximation of quasiconvex functions, and lower semicontinuity of multiple integrals. Manuscr. Math. **51**, 1–28 (1985)
51. Meyer, Y., Coifman, R.R.: Opérateurs multilinéaires. Hermann, Paris (1991)
52. Meyers, N.: An L^p-estimate for the gradient of solutions of second order elliptic divergence equations. Ann. Sc. Norm. Super. Pisa **17**, 189–206 (1963)
53. Milnor, J.W.: Topology from the Differentiable Viewpoint. University Press of Virginia, Charlottesville (1965)
54. Minty, G.: On a monotonicity method for the solution of non linear equations in Banach spaces. Proc. Natl. Acad. Sci. USA **50**, 1038–1041 (1963)
55. Morrey, C.B. Jr.: Quasiconvexity and the lower semicontinuity of multiple integrals. Pac. J. Math. **2**, 25–53 (1952)
56. Morrey, C.B. Jr.: Multiple Integrals in the Calculus of Variations. Springer, Berlin (1966)
57. Nečas, J.: Les méthodes directes en théorie des équations elliptiques. Masson, Paris (1967)
58. Palais, R.S.: Critical point theory and the minimax principle. In: Global Analysis (Proceedings of Symposia in Pure Mathematics, Vol. XV, Berkeley, CA, 1968), pp. 185–212. American Mathematical Society, Providence (1970)
59. Rabinowitz, P.H.: Théorie du degré topologique et applications à des problèmes aux limites non linéaires. Lecture notes of the Université Pierre et Marie Curie, taken by H. Beresticky (1975)
60. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
61. Rudin, W.: Real and Complex Analysis, 3rd edn. McGraw-Hill Book Co., New York (1987)
62. Rudin, W.: Functional Analysis. International Series in Pure and Applied Mathematics, 2nd edn. McGraw-Hill, Inc., New York (1991)
63. Schaefer, H.H.: Topological Vector Spaces, 5th edn. Springer, New York (1986)
64. Schwartz, J.T.: Nonlinear Functional Analysis. Lecture Notes, 1963–1964. Courant Institute of Mathematical Sciences, New York (1965)
65. Stampacchia, G.: Équations elliptiques du second ordre à coefficients discontinus, Presses de l'Université de Montréal, série (Séminaires de Mathématiques Supérieures), 16, Montréal (1965)
66. Šverák, V.: Rank one convexity does not imply quasiconvexity. Proc. R. Soc. Edinb. A **120**, 185–189 (1992)
67. Tartar, L.: Compensated compactness and applications to partial differential equations. In: Nonlinear Analysis and Mechanics: Heriot-Watt Symposium, Volume IV. Research Notes in Mathematics, vol. 39, pp. 136–212. Pitman, Boston (1979)
68. Trèves, F.: Topological Vector Spaces, Distributions and Kernels. Academic Press, New York (1967). Reprinted by Dover Publications, Mineola (2006)

69. Young, L.C.: Lectures on the Calculus of Variations and Optimal Control Theory. W.B. Saunders, Philadelphia (1969)
70. Ziemer, W.P.: Weakly Differentiable Functions. Springer, Berlin (1989)

Index

H. Le Dret, *Nonlinear Elliptic Partial Differential Equations*, Universitext,
https://doi.org/10.1007/978-3-319-78390-1