MA4G6 Lecture Notes

Introduction to the Modern Calculus of Variations

Filip Rindler

Spring Term 2017

Filip Rindler
Mathematics Institute
University of Warwick
Coventry CV4 7AL
United Kingdom
F.Rindler@warwick.ac.uk
http://www.warwick.ac.uk/filiprindler

Preface

The calculus of variations has seen a sweeping renaissance since the 1970s, ignited by the discovery of powerful variational methods to investigate nonlinear elasticity and microstructure in materials. It was further fueled by the adaptation of sophisticated mathematical techniques from measure theory, geometric analysis, and modern PDE theory. With these fields now more closely intertwined than ever, the methods of the modern calculus of variations are now among the most powerful to study highly nonlinear problems.

These lecture notes, written for the *MA4G6 Calculus of Variations* course at the University of Warwick, intend to give a modern introduction to the Calculus of Variations. I have tried to cover different aspects of the field and to explain how they fit into the "big picture". I have tried to strike a balance between a pure introduction and a text that can be used for later revision of forgotten material.

The presentation is based around a few principles:

- I use modern techniques and present results which are perhaps not usually found in an introductory text. I have organized the material thematically, not necessarily in the order in which it was discovered.

- For most results, I try to use "reasonable" assumptions, not the most general ones.

- When presented with a choice of how to prove a result, I have usually chosen the (in my opinion) most conceptually clear approach over more "elementary" ones. For instance, since Young measures pervade large parts of the modern theory of microstructure, it is convenient to already use them for lower semicontinuity results, even though more elementary approaches exist. This comes at the expense of a higher initial burden of abstract theory, but it gives the reader a powerful "toolbox" that can easily be applied to a variety of problems.

- I do not attempt to trace every result to its original source.

- I consider vector-valued u right from the start since this situation has many applications and, in fact, much of the advanced theory was specifically developed for this situation.

- I occasionally refer to further results without giving a proof. The rationale behind this is that I want the reader to see the frontier of research, without compromising the

Comment...

coherence of the text. I hope the reader will take these "tasters" as motivation to read the original papers.

This text was strongly influenced by several previous works, I note in particular the lecture notes on microstructure by Müller [85], Dacorogna's treatise on the calculus of variations [29], Kirchheim's advanced lecture notes on differential inclusions [60], the book on Young measures by Pedregal [94], Giusti's more regularity theory-focused introduction to the calculus of variations [49], Dolzmann's book on modern aspects of microstructure [36], as well as lecture notes on several related courses by Ball, Kristensen, and Mielke.

These lecture notes are a living document and I would appreciate comments, corrections and remarks – however small. They can be sent back to me via e-mail at

`F.Rindler@warwick.ac.uk`

or anonymously via

`http://tinyurl.com/MA4G6-2017`

or by scanning the following QR-Code:

Moreover, every page contains its individual QR-code for immediate feedback. I encourage all readers to make use of them.

I would like to thank in particular my teachers Alexander Mielke and Jan Kristensen. Through their generosity and enthusiasm in sharing their knowledge they have provided me with the foundation for my study and research. Furthermore, I am grateful to Guido De Philippis, Richard Gratwick, Angkana Rüland, Giles Shaw, the participants of the MA4G6 course at Warwick in 2015, for many helpful comments, corrections, and remarks. Finally. Finally, I would like to acknowledge the support from an EPSRC Research Fellowship on "Singularities in Nonlinear PDEs" (EP/L018934/1).

January 2017 *Filip Rindler*

Comment...

Contents

Comment...

Comment...

Chapter 1

Introduction

A mathematical model of some aspect of reality needs to balance its *validity*, that is, its agreement with nature, and its *predictive capabilities*, with the *feasibility* of its mathematical treatment. In this quest to formulate a useful picture of an aspect of the world, it turns out on surprisingly many occasions that the formulation of the model becomes clearer, more compact, or more convincing if one introduces some form of *variational principle*. This is the case if one can define a quantity, such as energy or entropy, which obeys a minimization, maximization or saddle-point law.

How much we perceive such a quantity as "fundamental" depends on the situation. For example, in classical mechanics, one calls forces *conservative* if they are path-independent and hence originate from changing an "energy potential". It turns out that many forces in physics are conservative, which seems to imply that the concept of energy is fundamental. Furthermore, Einstein's famous formula $E = mc^2$ expresses the equivalence of mass and energy, positioning energy as the "most" fundamental quantity, with mass becoming "condensated" energy. On the other hand, the other ubiquitous scalar physical quantity, entropy, has a more "artificial" flavor as a measure of *missing information* in a model, as is now well understood in the field of statistical mechanics, where entropy becomes a derived quantity.

Our approach to variational quantities here is a pragmatic one: We see them as providing *structure* to a problem, which enables us to use powerful *variational* methods. For instance, in elasticity theory it is usually unrealistic that a system will tend to a *global* minimizer by itself, but this does not mean that – as an approximation – such a principle cannot be useful in practice. If we wait long enough, the inherent noise in a realistic system will move our system around and *with a high probability* it will be near a low-energy state.

In this text, we focus on minimization problems for integral functionals defined on maps from an open and bounded set $\Omega \subset \mathbb{R}^d$ to some $\mathbb{R}^m$, that is, we minimize

$$\int_\Omega f(x, u(x), \nabla u(x))\,\mathrm{d}x \quad \to \quad \min, \qquad u\colon \Omega \to \mathbb{R}^m,$$

possibly under conditions on the boundary values of u and further side constraints. These problems form the original core of the calculus of variations and are as relevant today as they have always been. The systematic understanding of these integral functionals starts in Euler's and Bernoulli's times in the late 1600s and the early 1700s, and their study was boosted

Comment...

into the modern age by Hilbert's 19th, 20th, 23rd problems, formulated in 1900 [55]. The excitement has never stopped since and new and important discoveries are made every year.

We start by looking at a parade of examples, which we treat at varying levels of detail. All these problems will be investigated further along the course once we have developed the necessary mathematical tools.

1.1 The brachistochrone problem

While the name "calculus of variations" is from Leonhard Euler's 1766 treatise *Elementa calculi variationum*, the beginning of the field can be traced back to June 1696, when Johann Bernoulli published a problem in the journal *Acta Eruditorum*. This "birth certificate" of the calculus of variations can be seen at

`http://tinyurl.com/CoVbirth`

or by scanning the following QR-Code:

Additionally, Bernoulli sent a letter containing the question to Gottfried Wilhelm Leibniz on 9 June 1696, who returned his solution only a few days later on 16 June, and commented that the problem tempted him "like the apple tempted Eve". Isaac Newton also published a solution (after the problem had reached him) without giving his identity, but Bernoulli identified him "ex ungue leonem" (from Latin, "by the lion's claw"). The problem was formulated as follows:

> *Given two points A and B in a vertical* [meaning "not horizontal"] *plane, one shall find a curve AMB for a movable point M, on which it travels from A to B in the shortest time, only driven by its own weight.*

The resulting curve is called the **brachistochrone** (from Ancient Greek, "shortest time") curve.

We set up the mathematical formulation as follows: We look for the curve y connecting the origin $(0,0)$ to the point $(\bar{x},\bar{y})$, where $\bar{x} > 0$, $\bar{y} < 0$, such that a point mass $m > 0$ slides from rest at $(0,0)$ to $(\bar{x},\bar{y})$ quickest among all such curves. See Figure 1.1 for several possible slide paths. We parametrize a point (x,y) on the curve by the time $t \geq 0$. The point mass has kinetic and potential energies

$$E_{\text{kin}} = \frac{m}{2}\left\{\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2\right\} = \frac{m}{2}\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2\left\{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2\right\},$$
$$E_{\text{pot}} = mgy,$$

Comment...

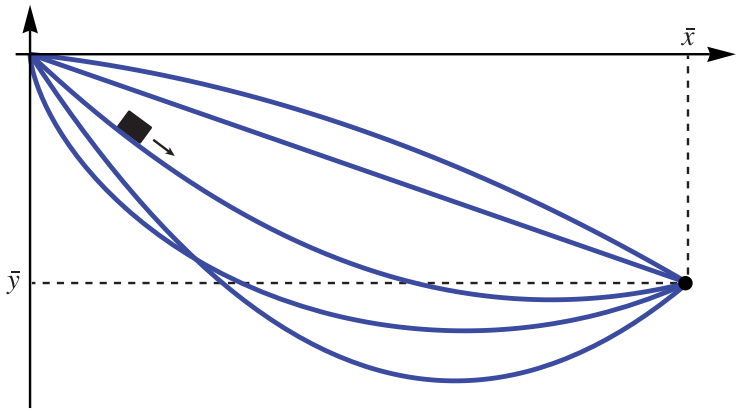

Figure 1.1: Several slide curves from the origin to $(\bar{x}, \bar{y})$.

where $g \approx 9.81\,\mathrm{m/s^2}$ is the (constant) gravitational acceleration on Earth. The total energy $E_{\text{kin}} + E_{\text{pot}}$ is zero at the beginning and conserved along the path. Hence, for all y we have

$$\frac{m}{2}\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 \left\{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2\right\} = -mgy.$$

We can solve this for $\mathrm{d}t/\mathrm{d}x$ (where $t = t(x)$ is the inverse of the x-parametrization) to get

$$\frac{\mathrm{d}t}{\mathrm{d}x} = \sqrt{\frac{1+(y')^2}{-2gy}} \qquad \left(\frac{\mathrm{d}t}{\mathrm{d}x} \geq 0\right),$$

where we wrote $y' = \frac{\mathrm{d}y}{\mathrm{d}x}$. Integrating over the whole x-length along the curve from 0 to $\bar{x}$, we get for the total time duration $T[y]$ that

$$T[y] = \frac{1}{\sqrt{2g}} \int_0^{\bar{x}} \sqrt{\frac{1+(y')^2}{-y}}\,\mathrm{d}x.$$

We may drop the constant in front of the integral, which does not influence the minimization problem, and set $\bar{x} = 1$ by reparametrization, to arrive at the problem

$$\begin{cases} \mathscr{F}[y] := \displaystyle\int_0^1 \sqrt{\frac{1+(y')^2}{-y}}\,\mathrm{d}x \quad \to \quad \min, \\ y(0) = 0,\ y(1) = \bar{y} < 0. \end{cases}$$

Clearly, the integrand is *convex* in y', which will be important for the solution theory. We will come back to this problem in Example 2.35.

A related problem concerns *Fermat's Principle* expressing that light (in vacuum or in a medium) always takes the fastest path between two points, from which many optical laws, e.g. about reflection and refraction, can be derived.

Comment...

1.2 Electrostatics

Consider an electric charge density $\rho\colon \mathbb{R}^3 \to \mathbb{R}$ (in units of C/m^3) in three-dimensional vacuum. Let $E\colon \mathbb{R}^3 \to \mathbb{R}^3$ (in V/m) and $B\colon \mathbb{R}^3 \to \mathbb{R}^3$ (in $\mathrm{T} = \mathrm{Vs}/\mathrm{m}^2$) be the electric and magnetic fields, respectively, which we assume to be constant in time (hence electro*statics*). Assuming that $\mathbb{R}^3$ is a linear, homogeneous, isotropic electric material, the **Gauss law** for electricity reads

$$\nabla \cdot E = \operatorname{div} E = \frac{\rho}{\varepsilon_0},$$

where $\varepsilon_0 \approx 8.854 \cdot 10^{-12}\,\mathrm{C}/(\mathrm{Vm})$. Moreover, we have the **Faraday law of induction**

$$\nabla \times E = \operatorname{curl} E = \frac{\mathrm{d}B}{\mathrm{d}t} = 0.$$

Thus, since E is curl-free, there exist an **electric potential** $\varphi\colon \mathbb{R}^3 \to \mathbb{R}$ (in V) such that

$$E = -\nabla\varphi.$$

Combining this with the Gauss law, we arrive at the Poisson equation,

$$\Delta\varphi = \nabla \cdot [\nabla\varphi] = -\frac{\rho}{\varepsilon_0}. \tag{1.1}$$

We can also look at electrostatics in a variational way: We use the norming condition $\varphi(0) = 0$. Then, the electric potential energy $U_E(x;q)$ of a point charge $q > 0$ (in C) at point $x \in \mathbb{R}^3$ in the electric field E is given by the path integral (which does not depend on the path chosen since E is a gradient)

$$U_E(x;q) = -\int_0^x qE \cdot \mathrm{d}s = -\int_0^1 qE(hx)\,\mathrm{d}h = q\varphi(x).$$

Thus, the total **electric energy** of our charge distribution ρ is (the factor $1/2$ is necessary to count mutual reaction forces correctly)

$$U_E := \frac{1}{2}\int_{\mathbb{R}^3} \rho\varphi\,\mathrm{d}x = \frac{\varepsilon_0}{2}\int_{\mathbb{R}^3} (\nabla \cdot E)\varphi\,\mathrm{d}x,$$

which has units of $\mathrm{CV} = \mathrm{J}$. Using $(\nabla \cdot E)\varphi = \nabla \cdot (E\varphi) - E \cdot (\nabla\varphi)$, the divergence theorem, and the natural assumption that φ vanishes at infinity, we get further

$$U_E = \frac{\varepsilon_0}{2}\int_{\mathbb{R}^3} \nabla \cdot (E\varphi) - E \cdot (\nabla\varphi)\,\mathrm{d}x = -\frac{\varepsilon_0}{2}\int_{\mathbb{R}^3} E \cdot (\nabla\varphi)\,\mathrm{d}x = \frac{\varepsilon_0}{2}\int_{\mathbb{R}^3} |\nabla\varphi|^2\,\mathrm{d}x.$$

The integral

$$\int_\Omega |\nabla\varphi|^2\,\mathrm{d}x$$

Comment...

is called **Dirichlet integral**.

In Example 2.15 we will see that (1.1) is equivalent to the minimization problem (over φ)

$$U_E - \int_{\mathbb{R}^3} \rho\varphi \,\mathrm{d}x = \int_{\mathbb{R}^3} \frac{\varepsilon_0}{2}|\nabla\varphi|^2 - \rho\varphi \,\mathrm{d}x \quad \to \quad \min.$$

The second term is the energy stored in the electric field caused by its interaction with the charge density ρ.

1.3 Stationary states in quantum mechanics

The non-relativistic evolution of a quantum mechanical system with N degrees of freedom in an electrical field is described completely through its **wave function** $\Psi\colon \mathbb{R}^N \times \mathbb{R} \to \mathbb{C}$ that satisfies the **Schrödinger equation**

$$\mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}t}\Psi(x,t) = \left[\frac{-\hbar^2}{2\mu}\Delta + V(x,t)\right]\Psi(x,t), \qquad x \in \mathbb{R}^N, t \in \mathbb{R},$$

where $\hbar \approx 1.05 \cdot 10^{-34}$ Js is the reduced Planck constant, $\mu > 0$ is the reduced mass, and $V = V(x,t) \in \mathbb{R}$ is the potential energy. The (self-adjoint) operator $H := -(2\mu)^{-1}\hbar\Delta + V$ is called the **Hamiltonian** of the system.

The value of the wave function itself at a given point in spacetime has no physical interpretation, but according to the *Copenhagen Interpretation* of quantum mechanics, $|\Psi(x)|^2$ is the probability density of finding a particle at the point x. In order for $|\Psi|^2$ to be a probability density, we need to impose the side constraint

$$\|\Psi\|_{\mathrm{L}^2} = 1.$$

In particular, $\Psi(x,t)$ has to decay as $|x| \to \infty$.

Of particular interest are the so-called *stationary states*, that is, solutions of the **stationary Schrödinger equation**

$$\left[\frac{-\hbar^2}{2\mu}\Delta + V(x)\right]\Psi(x) = E\Psi(x), \qquad x \in \mathbb{R}^N,$$

where $E > 0$ is an *energy level*. We then solve this *eigenvalue problem* (in a weak sense) for $\Psi \in \mathrm{W}^{1,2}(\mathbb{R}^N;\mathbb{C})$ with $\|\Psi\|_{\mathrm{L}^2} = 1$ (see Appendix A.4 for a collection of facts about Sobolev spaces like $\mathrm{W}^{1,2}$).

If we are just interested in the lowest-energy state, the so-called **ground state**, we can find minimizers of the energy functional

$$\mathscr{E}[\Psi] := \int_{\mathbb{R}^N} \frac{\hbar^2}{2\mu}|\nabla\Psi(x)|^2 + \frac{1}{2}V(x)|\Psi(x)|^2 \,\mathrm{d}x,$$

again under the side constraint

$$\|\Psi\|_{\mathrm{L}^2} = 1.$$

The two parts of the integral above correspond to kinetic and potential energy, respectively. We will continue this investigation in Example 2.40.

Comment...

1.4 Hyperelasticity

Elasticity theory is one of the most important theories of **continuum mechanics**, that is, the study of the mechanics of (idealized) continuous media. We will not go into much detail about elasticity modeling here and refer to the book [22] for a thorough introduction.

Consider a body of mass occupying a bounded Lipschitz **domain** $\Omega \subset \mathbb{R}^3$, that is an open, connected set such that $\partial\Omega$ is a Lipschitz manifold (the union of finitely many Lipschitz graphs). We call Ω the *reference configuration*. If we deform the body, any *material point* $x \in \Omega$ is mapped into a *spatial point* $y(x) \in \mathbb{R}^3$ and we call $y(\Omega)$ the *deformed configuration*. For a suitable continuum mechanics theory, we also need to require that $y\colon \Omega \to y(\Omega)$ is a differentiable bijection and that it is *orientation-preserving*, i.e. that

$$\det \nabla y(x) > 0 \qquad \text{for all } x \in \Omega.$$

For convenience one also introduces the *displacement*

$$u(x) := y(x) - x.$$

One can show (using the implicit function theorem and the mean value theorem) that the orientation-preserving condition and the invertibility are satisfied if

$$\sup_{x\in\Omega} |\nabla u(x)| < \delta(\Omega) \tag{1.2}$$

for a domain-dependent (small) constant $\delta(\Omega) > 0$.

Next, we need a measure of local "stretching", called a *strain tensor*, which should serve as the parameter to a local energy density. On physical grounds, **rigid body motions**, that is, deformations $u(x) = Rx + u_0$ with a rotation $R \in \mathbb{R}^{3\times 3}$ ($R^T = R^{-1}$ and $\det R = 1$) and $u_0 \in \mathbb{R}^3$, should not cause strain. In this sense, strain measures the deviation of the displacement from a rigid body motion. One popular choice is the **Green–St. Venant strain tensor**[1]

$$E := \frac{1}{2}\big(\nabla u + \nabla u^T + \nabla u^T \nabla u\big). \tag{1.3}$$

We first consider fully nonlinear ("finite strain") elasticity. For our purposes we simply postulate the existence of a stored-energy density $W\colon \mathbb{R}^{3\times 3} \to [0,\infty]$ and an external body force field $b\colon \Omega \to \mathbb{R}^3$ (e.g. gravity) such that

$$\mathscr{F}[y] := \int_\Omega W(\nabla y) - b \cdot y \,\mathrm{d}x$$

represents the total elastic energy stored in the system. If the elastic energy can be written in this way as

$$\int_\Omega W(\nabla y(x)) \,\mathrm{d}x,$$

we call the material **hyperelastic**. In applications, W is often given as depending on the Green–St. Venant strain tensor E instead of ∇y, but for our mathematical theory, the above form is more convenient. We require several properties of W for arguments $F \in \mathbb{R}^{3\times 3}$:

[1] There is an array of choices for the strain tensor, but they are mostly equivalent within *nonlinear* elasticity.

Comment...

(i) **Norming:** $W(\mathrm{Id}) = 0$ (the undeformed state costs no energy).

(ii) **Frame-invariance:** $W(RF) = W(F)$ for all $R \in \mathrm{SO}(3)$.

(iii) **Infinite compression costs infinite energy:** $W(F) \to +\infty$ as $\det F \downarrow 0$.

(iv) **Infinite stretching costs infinite energy:** $W(F) \to +\infty$ as $|F| \to \infty$.

The main problem of nonlinear hyperelasticity is to minimize $\mathscr{F}$ as above over all $y\colon \Omega \to \mathbb{R}^3$ with given boundary values. Of course, it is not a-priori clear in which space we should look for a solution. Indeed, this depends on the growth properties of W. For example, for the prototypical choice

$$W(F) := \mathrm{dist}(F, \mathrm{SO}(3))^2, \qquad \text{where} \qquad \mathrm{dist}(x,K) := \inf_{y \in K} |x-y|.$$

However, this W does not satisfy (3) from our list of requirements, we would look for square integrable functions. More realistic in applications are the **Mooney–Rivlin materials**, where W is of the form

$$W(F) := a|F|^2 + b|\operatorname{cof} F|^2 + \Gamma(\det F),$$

with $a, b > 0$ and $\Gamma(d) = \alpha d^2 - \beta \log d$ for $\alpha, \beta > 0$. If $b = 0$ the material is called **neo-Hookean**. An even larger class are the **Ogden materials**, for which

$$W(F) := \sum_{i=1}^{M} a_i \operatorname{tr}\big[(F^T F)^{\gamma_i/2}\big] + \sum_{j=1}^{N} b_j \operatorname{tr}\operatorname{cof}\big[(F^T F)^{\delta_j/2}\big] + \Gamma(\det F), \qquad F \in \mathbb{R}^{3\times 3},$$

where $a_i > 0$, $\gamma_i \geq 1$, $b_j > 0$, $\delta_j \geq 1$, and $\Gamma\colon \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is a convex function with $\Gamma(d) \to +\infty$ as $d \downarrow 0$, $\Gamma(d) = +\infty$ for $s \leq 0$. These materials occur in a wide range of applications.

In the setting of *linearized elasticity*, we make the "small strain" assumption that the quadratic term in (1.3) can be neglected and that (1.2) holds. In this case, we work with the **linearized strain tensor**

$$\mathscr{E} u := \frac{1}{2}\big(\nabla u + \nabla u^T\big).$$

In this infinitesimal deformation setting, the displacements that do not create strain are precisely the skew-affine maps[2] $u(x) = Qx + u_0$ with $Q^T = -Q$ and $u_0 \in \mathbb{R}$.

For linearized elasticity we consider an energy of the special "quadratic" form

$$\mathscr{F}[u] := \int_\Omega \frac{1}{2} \mathscr{E} u : \mathbf{C} \mathscr{E} u - b \cdot u \,\mathrm{d}x,$$

[2]This becomes more meaningful when considering a bit more algebra: The Lie group $\mathrm{SO}(3)$ of rotations has as its Lie algebra $\mathrm{Lie}(\mathrm{SO}(3)) = \mathfrak{so}(3)$ the space of all skew-symmetric matrices, which then can be seen as "infinitesimal rotations".

Comment...

where $\mathbf{C}(x) = \mathbf{C}_{ijkl}(x)$ is a symmetric, strictly positive definite ($A : \mathbf{C}(x)A > c|A|^2$ for some $c > 0$) fourth-order tensor, called the **elasticity/stiffness tensor**, and $b\colon \Omega \to \mathbb{R}^3$ is our external body force. Thus we will always look for solutions in $\mathrm{W}^{1,2}(\Omega;\mathbb{R}^3)$.

For homogeneous, isotropic media, $\mathbf{C}$ does not depend on x or the direction of strain, which translates into the additional condition

$$(AR) : \mathbf{C}(x)(AR) = A : \mathbf{C}(x)A \qquad \text{for all } x \in \Omega,\, A \in \mathbb{R}^{3\times 3} \text{ and } R \in \mathrm{SO}(3).$$

In this case, it can be shown that $\mathscr{F}$ simplifies to

$$\mathscr{F}[u] = \frac{1}{2}\int_\Omega 2\mu|\mathscr{E}u|^2 + \Big(\kappa - \frac{2}{3}\mu\Big)|\operatorname{tr}\mathscr{E}u|^2 - b\cdot u\,\mathrm{d}x$$

for $\mu > 0$ the *shear modulus* and $\kappa > 0$ the *bulk modulus*, which are material constants. For example, for cold-rolled steel $\mu \approx 75$ GPa and $\kappa \approx 160$ GPa.

1.5 Microstructure in crystals

Consider a material specimen with internal crystal structure, for instance a metal (like iron) or an alloy (like CuAlNi). We assume that all atoms are arranged in the *same* single crystal occupying an (open, bounded) reference domain $\Omega \subset \mathbb{R}^d$, where $d = 2$ (plates) or $d = 3$. If we subject our specimen to external forces or push it into a prescribed shape at the boundary, we want to determine the resulting shape. It turns out that on a microscopic scale crystals often display very fine oscillations between different *phases*, that is, the crystal exhibits locally periodic fine structures. It turns out that this *microstructure* has profound implications for the macroscopic behavior of the material. Usually, the phase oscillations do not reach atomic length scales, so we can still attempt to model our situation using continuum mechanics.

The fundamental Cauchy–Born hypothesis[3] postulates that for small linear displacements the crystal lattice atoms will follow this displacement. Assuming this, we can model the crystal as a *continuum* and assign the energy density $W(F) \geq 0$ to the linear deformation $x \mapsto Fx$. The crucial point here is that, thanks to the Cauchy–Born hypothesis, W depends only on F and no other "microscopic structure" of the crystal, at least for small to moderate crystal deformations. Then, in this approach, the total energy of a deformation $u\colon \Omega \to \mathbb{R}^3$ is given as

$$\mathscr{F}[u] := \int_\Omega W(\nabla u(x))\,\mathrm{d}x, \qquad u\colon \Omega \to \mathbb{R}^3.$$

Here, on $W\colon \mathbb{R}^{3\times 3} \to [0,\infty)$ we make the following assumptions:

(i) **Norming:** $W(\mathrm{Id}) = 0$ (the undeformed state costs no energy).

(ii) **Frame-invariance:** $W(RF) = W(F)$ for all $R \in \mathrm{SO}(3)$.

[3] This assumption is often made, but is not always justified, see [26, 38, 40, 46].

Comment...

(iii) **Symmetry-invariance:** $W(FS) = W(S)$ for all $S \in \mathscr{S}$, where $\mathscr{S} \subset \mathrm{SO}(3)$ is the compact symmetry point group of the crystal.

The basic variational postulate is now that the observed macroscopic deformation is a minimizer of $\mathscr{F}$ under the given boundary conditions. In fact, it is often experimentally observed that the observed deformation $u\colon \Omega \to \mathbb{R}^3$ is a *pointwise* minimizer of the integrand, at least in a very large portion of Ω. Thus, we are led to consider the *differential inclusion*

$$\nabla u(x) \in K := \left\{ A \in \mathbb{R}^{3\times 3} \ : \ W(A) = \min W \right\} = W^{-1}(0) \qquad \text{a.e.}$$

The set K is usually compact in the study of crystal, but other applications also lead to these inclusions for non-compact K.

For concrete crystals one often observes the following situation: Above a critical temperature, K is simply $\mathrm{SO}(d)$, i.e. the simplest possible set that is compatible with the frame-invariance (ii) above. This is called the *cubic* phase. Passing through the critical temperature, however, the material undergoes a solid–solid phase transition and K is now the union of several *wells*, that is,

$$K = \mathrm{SO}(d)U_1 \cup \cdots \cup \mathrm{SO}(d)U_N$$

for distinct matrices $U_1, \ldots, U_N \in \mathbb{R}^{d\times d}$ with $\det U_i > 0$ $(i = 1, \ldots, N)$. By the polar decomposition of matrices with positive determinants into the product of a rotation and a symmetric positive definite matrix, we can assume that

$$U_i \text{ is symmetric and positive definite, } i = 1, \ldots, N.$$

If $N \geq 2$ and other *compatibility conditions* between the U_j are satisfied, microstructure can be observed. It should be noted that while our model as formulated above may imply "infinitely fast" oscillations in the microstructure, in reality other (atomistic) effects limit the length scales that are observed.

As a concrete example, the CuAlNi alloy undergoes a *cubic-to-orthorhombic* phase transition and below the critical temperature we have

$$K = \mathrm{SO}(3)U_1 \cup \cdots \cup \mathrm{SO}(3)U_6 \subset \mathbb{R}^{3\times 3},$$

where

$$U_1 = \begin{pmatrix} \xi & \eta & 0 \\ \eta & \xi & 0 \\ 0 & 0 & \zeta \end{pmatrix}, \qquad U_2 = \begin{pmatrix} \xi & -\eta & 0 \\ -\eta & \xi & 0 \\ 0 & 0 & \zeta \end{pmatrix}, \qquad U_3 = \begin{pmatrix} \xi & 0 & \eta \\ 0 & \zeta & 0 \\ \eta & 0 & \xi \end{pmatrix},$$

$$U_4 = \begin{pmatrix} \xi & 0 & -\eta \\ 0 & \zeta & \\ -\eta & 0 & \xi \end{pmatrix}, \qquad U_5 = \begin{pmatrix} \zeta & 0 & 0 \\ 0 & \xi & \eta \\ 0 & \eta & \xi \end{pmatrix}, \qquad U_5 = \begin{pmatrix} \zeta & 0 & 0 \\ 0 & \xi & -\eta \\ 0 & -\eta & \xi \end{pmatrix},$$

where

$$\xi = \frac{\alpha+\gamma}{2} > \eta = \frac{\alpha-\gamma}{2} > 0, \qquad \zeta = \beta > 0.$$

Comment...

Depending on the choice of values of the parameters, a large variety of microstructure variation is observed.

One striking property of CuAlNi is the *shape-memory effect*, where a material specimen "remembers" the shape it had when it was hotter than a certain critical temperature. After cooling, the specimen can be freely deformed, but when it is again heated above the critical temperature, it "snaps back" into its orginal shape.

In this work we will only consider the basic principles underlying this problem, concrete applications are left to more specialized treatises like [36] (in particular, Section 5.3 in *loc. cit.* in relation to the above example of CuAlNi).

1.6 Optimal saving and consumption

Consider a capitalist worker earning a (constant) wage w per year, which he can either spend on consumption or save. Denote by $S(t)$ the savings at time t, where $t \in [0,T]$ is in years, with $t = 0$ denoting the beginning of his work life and $t = T$ his retirement. Further, let $C(t)$ be the consumption rate (consumption per time) at time t. On the saved capital, the worker earns interest, say with gross-continuous rate $\rho > 0$, meaning that a capital amount $m > 0$ grows as $\exp(\rho t)m$. If we were given an APR $\rho_1 > 0$ instead of ρ, we could compute $\rho = \ln(1+\rho_1)$. We further assume that salary is paid continuously, not in intervals, for simplicity. So, w is really the rate of pay, given in money per time. Then, the worker's savings evolve according to the differential equation

$$\dot{S}(t) = w + \rho S(t) - C(t). \tag{1.4}$$

Now, assume that our worker is mathematically inclined and wants to optimize the happiness due to consumption in his life by finding the optimal amount of consumption at any given time. Being a pure capitalist, the worker's happiness only depends on his consumption rate C. So, if we denote by $U(C)$ the **utility function**, that is, the "happiness" due to the consumption rate C, our worker wants to find $C\colon [0,T] \to \mathbb{R}$ such that

$$\mathscr{H}[C] := \int_0^T U(C(t))\,\mathrm{d}t$$

is maximized. The choice of U depends on our worker's personality, but it is probably realistic to assume that there is a *law of diminishing returns*, i.e. for twice as much consumption, our worker is happier, but not twice as happy. So, let us assume $U' > 0$ and $U'(C) \to 0$ as $C \to \infty$. Also, we should have $U(0) = -\infty$ (starvation). Moreover, it is realistic for U to be *concave*, which implies that there are no local maxima. One function that satisfies all of these requirements is the logarithm, and so we use

$$U(C) = \ln(C), \qquad C > 0.$$

Let us also assume that the worker starts with no savings, $S(0) = 0$, and at the end of his work life wants to retire with savings $S(T) = S_T \geq 0$. Rearranging (1.4) for $C(t)$

Comment...

and plugging this into the formula for $\mathscr{H}$, we therefore want to solve the **optimal saving problem**

$$\begin{cases} \mathscr{F}[S] := \displaystyle\int_0^T -\ln(w+\rho S(t)-\dot{S}(t))\,\mathrm{d}t \quad \to \quad \min, \\ S(0)=0,\ S(T)=S_T \geq 0. \end{cases}$$

This will be solved in Example 2.18.

1.7 Sailing against the wind

Every sailor knows how to sail against the wind by "beating": One has to sail at an angle of approximately 45° to the wind[4], then tack (turn the bow through the wind) and finally, after the sail has caught the wind on the other side, continue again at approximately 45° to the wind. Repeating this procedure makes the boat follow a zig-zag motion, which gives a net movement directly against the wind, see Figure 1.2. A mathematically inclined sailor might ask the question of "how often to tack". In an idealized model we can assume that tacking costs no time and that the forward sailing speed v_s of the boat depends on the angle α to the wind as follows (at least qualitatively):

$$v_s(\alpha) = -\cos(4\alpha),$$

which has maxima at $\alpha = \pm 45°$ (in real boats, the maximum might be at a lower angle, i.e. "closer to the wind"). Assume furthermore that our sailor is sailing along a straight river with the current. Now, the current is fastest in the middle of the river and goes down to zero at the banks. In fact, a good approximation would be the formula of *Poiseuille (channel) flow*, which can be derived from the flow equations of fluids: At distance r from the center of the river the current's flow speed is approximately

$$v_c(r) := v_{\max}\left(1-\frac{r^2}{R^2}\right),$$

where $R > 0$ is half the width of the river.

If we denote by $r(t)$ the distance of our boat from the middle of the channel at time $t \in [0,T]$, then the total speed (called the "velocity made good" in sailing parlance) is

$$\begin{aligned} v(t) &:= v_s(\arctan r'(t)) + v_c(r(t)) \\ &= -\cos(4\arctan r'(t)) + v_{\max}\left(1-\frac{r(t)^2}{R^2}\right). \end{aligned}$$

The key to understand this problem is now the fact that $a \mapsto -\cos(4\arctan a)$ has two maxima at $a = \pm 1$. We say that this function is a **double-well potential**.

[4]The optimal angle depends on the boat. Ask a sailor about this and you will get an evening-filling lecture.

Comment...

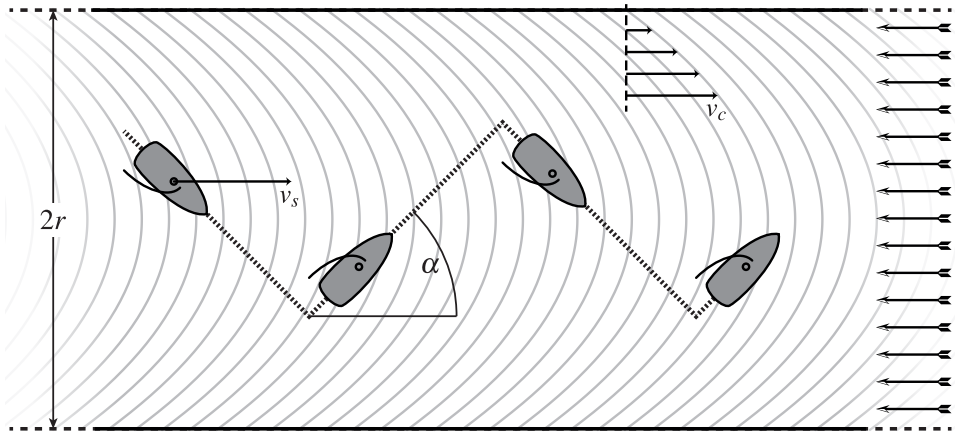

Figure 1.2: Sailing against the wind in a channel.

The total forward distance traveled over the time interval $[0,T]$ is

$$\int_0^T v(t)\,\mathrm{d}t = \int_0^T -\cos(4\arctan r'(t)) + v_{\max}\left(1-\frac{r(t)^2}{R^2}\right)\mathrm{d}t.$$

If we also require the initial and terminal conditions $r(0)=r(T)=0$, we arrive at the **optimal beating problem**:

$$\begin{cases} \mathscr{F}[r] := \displaystyle\int_0^T \cos(4\arctan r'(t)) - v_{\max}\left(1-\frac{r(t)^2}{R^2}\right)\mathrm{d}t \quad \to \quad \min, \\ r(0)=r(T)=0, \qquad |r(t)|\le R. \end{cases}$$

Our intuition tells us that in this idealized model, where tacking costs no time, we should be tacking as "infinitely fast" in order to stay in the middle of the river. Later, once we have advanced tools at our disposal, we will make this idea precise, see Example 5.7.

Comment...

Chapter 2

Convexity

In the introduction we got a glimpse of the many applications of minimization problems in physics, technology, and economics. In this chapter we start to develop the mathematical theory that will allow us to investigate these problems. Consider a minimization problem of the form

$$\begin{cases} \mathscr{F}[u] := \displaystyle\int_\Omega f(x,u(x),\nabla u(x))\,\mathrm{d}x \quad \to \quad \min, \\ \text{over all } u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m) \text{ with } u|_{\partial\Omega} = g. \end{cases}$$

Here, and throughout the text (if not otherwise stated) we will make the standard assumption that $\Omega \subset \mathbb{R}^d$ is a **bounded Lipschitz domain**, that is, Ω is open, bounded, and has a boundary that is the union of finitely many Lipschitz manifolds. The function

$$f\colon \Omega \times \mathbb{R}^m \times \mathbb{R}^{m\times d} \to \mathbb{R},$$

is required to be measurable in the first and continuous in the second and third arguments, hence f is a so-called **Carathéodory integrand**. Furthermore, in this chapter we let $1 < p < \infty$ and on the prescribed boundary values g we assume

$$g \in \mathrm{W}^{1-1/p,p}(\partial\Omega;\mathbb{R}^m).$$

In this context recall that $\mathrm{W}^{1-1/p,p}(\partial\Omega;\mathbb{R}^m)$ is the space of traces of all Sobolev functions $u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$, see Appendix A.4 for some background on Sobolev spaces. Furthermore, to make the above integral finite, we assume the p**-growth bound**

$$|f(x,v,A)| \le M(1+|v|^p+|A|^p) \qquad (x,v,A) \in \Omega \times \mathbb{R}^m \times \mathbb{R}^{m\times d},$$

for some $M > 0$.

Below, we will investigate the solvability and regularity properties of this minimization problem; in particular, we will take a close look at the way in which **convexity** properties of f in its gradient (third) argument determine whether $\mathscr{F}$ is lower semicontinuous, which is the decisive attribute in the fundamental Direct Method of the calculus of variations.

Comment...

We will also consider the partial differential equation associated with the minimization problem, the so-called Euler–Lagrange equation, which constitutes a necessary, but not sufficient, condition for minimizers. This leads naturally to the question of regularity of solutions. Finally, we also treat invariances, leading to the famous Noether theorem before having a glimpse at side constraints.

2.1 The Direct Method

Fundamental to all of the existence theorems in this text is the conceptionally simple **Direct Method** of the calculus of variations, which is "direct" since we prove the existence of solutions to minimization problems without the detour through a differential equation. Let X be a complete metric space (e.g. a Banach space with the norm topology or a closed, convex and bounded subset of a Banach space with the weak or weak* topology) and let $\mathscr{F} : X \to \mathbb{R} \cup \{+\infty\}$ be our **objective functional** satisfying the following two assumptions:

(H1) **Coercivity:** For all $\Lambda \in \mathbb{R}$, the sublevel set $\{\, u \in X \;:\; \mathscr{F}[u] \leq \Lambda \,\}$ is sequentially relatively compact, that is, if $\mathscr{F}[u_j] \leq \Lambda$ for a sequence $(u_j) \subset X$ and some $\Lambda \in \mathbb{R}$, then (u_j) has a convergent subsequence in X.

(H2) **Lower semicontinuity:** For all sequences $(u_j) \subset X$ with $u_j \to u$ it holds that

$$\mathscr{F}[u] \leq \liminf_{j\to\infty} \mathscr{F}[u_j].$$

The Direct Method for the abstract minimization problem

$$\mathscr{F}[u] \to \min \quad \text{over all } u \in X. \tag{2.1}$$

is encapsulated in the following simple result:

Theorem 2.1 (Direct Method). *Assume that $\mathscr{F}$ is both coercive and lower semicontinuous. Then, the abstract minimization problem* (2.1) *has at least one solution, that is, there exists $u_* \in X$ with $\mathscr{F}[u_*] = \min\{\, \mathscr{F}[u] \;:\; u \in X \,\}$.*

Proof. Let us assume that there exists at least one $u \in X$ such that $\mathscr{F}[u] < +\infty$; otherwise, any $u \in X$ is a "solution" to the (degenerate) minimization problem.

To construct a minimizer we take a **minimizing sequence** $(u_j) \subset X$ such that

$$\lim_{j\to\infty} \mathscr{F}[u_j] \to \alpha := \inf\big\{\, \mathscr{F}[u] \;:\; u \in X \,\big\} < +\infty.$$

Then, since convergent sequences are bounded, there exists $\Lambda \in \mathbb{R}$ such that $\mathscr{F}[u_j] \leq \Lambda$ for all $j \in \mathbb{N}$. Hence, by the coercivity (H1), the sequence (u_j) is contained in a compact subset of X. Now select a subsequence, which here as in the following we do not renumber, such that

$$u_j \to u_* \in X.$$

Comment...

By the lower semicontinuity (H2), we immediately conclude

$$\alpha \leq \mathscr{F}[u_*] \leq \liminf_{j\to\infty} \mathscr{F}[u_j] = \alpha.$$

Thus, $\mathscr{F}[u_*] = \alpha$ and u_* is the sought minimizer. □

Example 2.2. Using the Direct Method, one can easily see that

$$h(t) := \begin{cases} 1-t & \text{if } t < 0, \\ t & \text{if } t \geq 0, \end{cases} \qquad t \in \mathbb{R}\ (= X),$$

has the minimizer $t = 0$, despite not even being continuous there.

Despite its nearly trivial proof, the Direct Method is very useful and flexible in applications. Indeed, it pushes the difficulty in proving existence of a minimizer toward establishing coercivity and lower semicontinuity. This, however, is a big advantage, since we have many tools at our disposal to establish these two hypotheses separately. In particular, for integral functionals, lower semicontinuity is tightly linked to *convexity* properties of the integrand, as we will see throughout this book.

At this point it is crucial to observe how coercivity and lower semicontinuity interact with the *topology* in X: If we choose a stronger topology, i.e. one for which there are fewer convergent sequences, then it becomes easier for $\mathscr{F}$ to be lower semicontinuous, but harder for $\mathscr{F}$ to be coercive. The opposite holds if choosing a weaker topology. In the mathematical treatment of a problem we are most likely in a situation where $\mathscr{F}$ and X are given by the concrete problem. We then need to find a suitable topology in which we can establish both coercivity and lower semicontinuity.

In this text, X will always be an infinite-dimensional Banach space and we have a real choice between using the strong or weak convergence. Usually, it turns out that coercivity with respect to the strong convergence is *false* since strongly compact sets in infinite-dimensional spaces are very restricted, whereas coercivity with respect to the weak convergence is true under reasonable assumptions. However, whereas strong lower semicontinuity poses few challenges, lower semicontinuity with respect to weakly converging sequences is a delicate matter and we will spend considerable time on this topic. As a result of this discussion, we will always use the Direct Method in the following version:

Theorem 2.3 (Direct Method for weak convergence). *Let X be a Banach space and let $\mathscr{F} : X \to \mathbb{R} \cup \{+\infty\}$. Assume the following:*

(WH1) ***Weak Coercivity:*** *For all $\Lambda \in \mathbb{R}$, the sublevel set*

$$\{u \in X \ : \ \mathscr{F}[u] \leq \Lambda\} \qquad \textit{is sequentially weakly relatively compact,}$$

that is, if $\mathscr{F}[u_j] \leq \Lambda$ for a sequence $(u_j) \subset X$ and some $\Lambda \in \mathbb{R}$, then (u_j) has a weakly convergent subsequence.

Comment...

(WH2) ***Weak lower semicontinuity:*** *For all sequences $(u_j) \subset X$ with $u_j \rightharpoonup u$ (weak convergence X) it holds that*

$$\mathscr{F}[u] \leq \liminf_{j\to\infty} \mathscr{F}[u_j].$$

Then, the minimization problem

$$\mathscr{F}[u] \to \min \quad \textit{over all } u \in X,$$

has at least one solution.

Before we move on to the more concrete theory, let us record one further remark: While it might appear as if "nature does not give us the topology" and it is up to mathematicians to "invent" a suitable one, it is remarkable that the topology that turns out to be *mathematically* convenient is also often *physically* relevant. This is for instance expressed in the following observation: The only phenomena which are present in weakly but not strongly converging sequences of Sobolev functions are *oscillations* and *concentrations*, as can be seen in the classical Vitali Convergence Theorem A.7.

2.2 Functionals with convex integrands

As a first instance of the theory of integral functionals to be developed in this text, we first consider the minimization problem for the simpler functional

$$\mathscr{F}[u] := \int_\Omega f(x, \nabla u(x))\,\mathrm{d}x$$

over all $u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ for some $1 < p < \infty$ to be chosen later (depending on lower growth properties of f). The following lemma together with (2.3) below allows us to conclude that $\mathscr{F}[u]$ is indeed well-defined.

Lemma 2.4. *Let $f\colon \Omega \times \mathbb{R}^N \to \mathbb{R}$ be a* ***Carathéodory integrand****, that is,*

(i) $x \mapsto f(x,A)$ is Lebesgue-measurable for all fixed $A \in \mathbb{R}^N$.

(ii) $A \mapsto f(x,A)$ is continuous for all fixed $x \in \Omega$.

Then, for any Lebesgue-measurable function $V\colon \Omega \to \mathbb{R}^N$ the composition $x \mapsto f(x,V(x))$ is Lebesgue-measurable.

We will prove this lemma via the following "Lusin-type" theorem for Carathéodory functions:

Theorem 2.5 (Scorza Dragoni 1948). *Let $f\colon \Omega \times \mathbb{R}^N \to \mathbb{R}$ be Carathéodory. Then, there exists an increasing sequence of compact sets $S_k \subset \Omega$ ($k \in \mathbb{N}$) with $|\Omega \setminus S_k| \downarrow 0$ such that $f|_{S_k \times \mathbb{R}^N}$ is continuous.*

Comment...

See Theorem 6.35 in [43] for a proof, which is rather measure-theoretic and similar to the proof of the Lusin theorem, cf. Theorem 2.24 in [97].

Proof of Lemma 2.4. Let $S_k \subset \Omega$ ($k \in \mathbb{N}$) be as in the Scorza Dragoni Theorem. Set $f_k := f|_{S_k \times \mathbb{R}^N}$ and

$$g_k(x) := f_k(x, V(x)), \qquad x \in S_k.$$

Then, for any open set $U \subset \mathbb{R}$, the pre-image of U under g_k is the set of all $x \in S_k$ such that $f_k(x, V(x)) \in U$. As f_k is continuous, $f_k^{-1}(U)$ is open and from the Lebesgue measurability of the product function $x \mapsto (x, V(x))$ we infer that $g_k^{-1}(U)$ is a Lebesgue-measurable subset of S_k. We can extend the definition of g_k to all of Ω by setting $g_k(x) := 0$ whenever $x \in \Omega \setminus S_k$. This function is still Lebesgue-measurable as S_k is compact, hence measurable. The conclusion then follows from the fact that $g_k(x) \to f(x, V(x))$ and pointwise limits of Lebesgue-measurable functions are Lebesgue-measurable. □

We next investigate coercivity: Let $1 < p < \infty$. The most basic assumption to guarantee coercivity, and the only one we want to consider here, is the following **lower p-coercivity bound**:

$$\mu |A|^p - \mu^{-1} \leq f(x, A) \qquad (x, A) \in \Omega \times \mathbb{R}^{m \times d}, \tag{2.2}$$

for some $\mu > 0$. This coercivity also suggests the exponent p for the definition of the function spaces where we look for solutions. We further assume the **upper p-growth bound**

$$|f(x, A)| \leq M(1 + |A|^p) \qquad (x, A) \in \Omega \times \mathbb{R}^{m \times d}, \tag{2.3}$$

for some $M > 0$, which gives finiteness of $\mathscr{F}[u]$ for all $u \in \mathrm{W}^{1,p}(\Omega; \mathbb{R}^m)$.

Proposition 2.6. *If f satisfies the lower growth bound* (2.2)*, then $\mathscr{F}$ is weakly coercive on the space* $\mathrm{W}^{1,p}_g(\Omega; \mathbb{R}^m) = \left\{ u \in \mathrm{W}^{1,p}(\Omega; \mathbb{R}^m) \ : \ u|_{\partial\Omega} = g \right\}$.

Proof. We need to show that any sequence $(u_j) \subset \mathrm{W}^{1,p}_g(\Omega; \mathbb{R}^m)$ with

$$\sup_j \mathscr{F}[u_j] < \infty$$

is weakly relatively compact. From (2.2) we get

$$\mu \cdot \sup_j \int_\Omega |\nabla u_j|^p \, \mathrm{d}x - \frac{|\Omega|}{\mu} \leq \sup_j \mathscr{F}[u_j] < \infty,$$

whereby $\sup_j \|\nabla u_j\|_{\mathrm{L}^p} < \infty$. By the Poincaré–Friedrichs inequality, Theorem A.21 (i), in conjunction with the fact that we fix the boundary values, we therefore get $\sup_j \|u_j\|_{\mathrm{W}^{1,p}} < \infty$. Bounded sets in reflexive Banach spaces, like $\mathrm{W}^{1,p}(\Omega; \mathbb{R}^m)$ for $1 < p < \infty$, are sequentially weakly relatively compact, see Theorem A.15, which gives the conclusion. □

Comment...

Having settled the question of (weak) coercivity, we can now turn to investigate the weak lower semicontinuity.

Theorem 2.7. *If $A \mapsto f(x,A)$ is convex for all $x \in \Omega$ and $f(x,A) \geq \kappa$ for some $\kappa \in \mathbb{R}$ (which follows in particular from* (2.2)*), then $\mathscr{F}$ is weakly lower semicontinuous on* $\mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$.

Proof. Step 1. We first establish that $\mathscr{F}$ is strongly lower semicontinuous, so let $u_j \to u$ in $\mathrm{W}^{1,p}$ and $\nabla u_j \to \nabla u$ almost everywhere, which holds after selecting a subsequence (not relabeled), see Appendix A.2. By assumption we have

$$f(x,\nabla u_j(x)) - \kappa \geq 0.$$

Applying Fatou's Lemma,

$$\liminf_{j\to\infty} \big(\mathscr{F}[u_j] - \kappa|\Omega|\big) \geq \mathscr{F}[u] - \kappa|\Omega|.$$

Thus, $\mathscr{F}[u] \leq \liminf_{j\to\infty} \mathscr{F}[u_j]$. Since this holds for all subsequences, it also follows for our original sequence.

Step 2. To prove the claimed weak lower semicontinuity take $(u_j) \subset \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ with $u_j \rightharpoonup u$. We need to show that

$$\mathscr{F}[u] \leq \liminf_{j\to\infty} \mathscr{F}[u_j] =: \alpha. \tag{2.4}$$

Taking a subsequence (not relabeled), we can in fact assume that $\mathscr{F}[u_j]$ converges to α.

By the Mazur Lemma A.18, we may find convex combinations

$$v_j = \sum_{n=j}^{N(j)} \theta_{j,n} u_n, \qquad \text{where} \qquad \theta_{j,n} \in [0,1] \quad \text{and} \quad \sum_{n=j}^{N(j)} \theta_{j,n} = 1,$$

such that $v_j \to u$ strongly. Since $f(x,\cdot)$ is convex,

$$\mathscr{F}[v_j] = \int_\Omega f\left(x, \sum_{n=j}^{N(j)} \theta_{j,n} \nabla u_n(x)\right) \mathrm{d}x \leq \sum_{n=j}^{N(j)} \theta_{j,n} \mathscr{F}[u_n].$$

Now, $\mathscr{F}[u_n] \to \alpha$ as $n \to \infty$. Thus,

$$\liminf_{j\to\infty} \mathscr{F}[v_j] \leq \liminf_{j\to\infty} \mathscr{F}[u_j] = \alpha.$$

On the other hand, from the first step and $v_j \to u$ strongly, we have $\mathscr{F}[u] \leq \liminf_{j\to\infty} \mathscr{F}[v_j]$. Thus, (2.4) follows and the proof is finished. □

We can summarize our findings in the following theorem:

Theorem 2.8. *Let $f\colon \Omega \times \mathbb{R}^{m\times d}$ be a Carathéodory integrand such that f satisfies the lower growth bound* (2.2)*, the upper growth bound* (2.3) *and is convex in its second argument. Then, the associated functional $\mathscr{F}$ has a minimizer over the space* $\mathrm{W}^{1,p}_g(\Omega;\mathbb{R}^m)$.

Comment...

Proof. This follows immediately from the Direct Method for the weak convence, Theorem 2.3, together with Proposition 2.6 and Theorem 2.7. □

Example 2.9. The **Dirichlet integral (functional)** is

$$\mathscr{F}[u] := \frac{1}{2}\int_\Omega |\nabla u(x)|^2 \,\mathrm{d}x, \qquad u \in \mathrm{W}^{1,2}(\Omega).$$

We already encountered this integral functional when considering electrostatics in Section 1.2. It is easy to see that this functional satisfies all requirements of Theorem 2.8 and so there exists a minimizer for any prescribed boundary values $g \in \mathrm{W}^{1/2,2}(\partial\Omega)$.

Example 2.10. In the prototypical problem of linearized elasticity from Section 1.4, we are tasked with the minimization problem

$$\begin{cases} \mathscr{F}[u] := \dfrac{1}{2}\displaystyle\int_\Omega 2\mu|\mathscr{E}u|^2 + \Big(\kappa - \frac{2}{3}\mu\Big)|\operatorname{tr}\mathscr{E}u|^2 - f\cdot u \,\mathrm{d}x \quad \to \quad \min, \\ \text{over all } u \in \mathrm{W}^{1,2}(\Omega;\mathbb{R}^3) \text{ with } u|_{\partial\Omega} = g, \end{cases}$$

where $\mu, \kappa > 0$, $f \in \mathrm{L}^2(\Omega;\mathbb{R}^3)$, and $g \in \mathrm{W}^{1/2,2}(\partial\Omega;\mathbb{R}^m)$. It is clear that $\mathscr{F}$ has quadratic growth. Let us first consider the coercivity: For $u \in \mathrm{W}^{1,2}(\Omega;\mathbb{R}^3)$ with $u|_{\partial\Omega} = g$ we have the L^2**-Korn inequality**

$$\|u\|^2_{\mathrm{W}^{1,2}} \le C_K\big(\|\mathscr{E}u\|^2_{\mathrm{L}^2} + \|g\|^2_{\mathrm{W}^{1/2,2}}\big). \tag{2.5}$$

Here, $C_K = C_K(\Omega) > 0$ is a constant. Let us for simplicity assume that $\kappa - \frac{2}{3}\mu \ge 0$ (this is not necessary, but shortens the argument). Then, also using the Young inequality (see Appendix A.1), we get for any $\delta > 0$,

$$\begin{aligned} \mathscr{F}[u] &\ge \mu\|\mathscr{E}u\|^2_{\mathrm{L}^2} - \|f\|_{\mathrm{L}^2}\|u\|_{\mathrm{L}^2} \\ &\ge \mu\big(\|\mathscr{E}u\|^2_{\mathrm{L}^2} + \|g\|^2_{\mathrm{W}^{1/2,2}}\big) - \frac{1}{2\delta}\|f\|^2_{\mathrm{L}^2} - \frac{\delta}{2}\|u\|^2_{\mathrm{L}^2} - \mu\|g\|^2_{\mathrm{W}^{1/2,2}} \\ &\ge \left(\frac{\mu}{C_K} - \frac{\delta}{2}\right)\|u\|^2_{\mathrm{W}^{1,2}} - \frac{1}{2\delta}\|f\|^2_{\mathrm{L}^2} - \mu\|g\|^2_{\mathrm{W}^{1/2,2}}. \end{aligned}$$

Choosing $\delta = \mu/C_K$, we obtain the coercivity estimate

$$\mathscr{F}[u] \ge \frac{\mu}{2C_K}\|u\|^2_{\mathrm{W}^{1,2}} - \frac{C_K}{2\mu}\|f\|^2_{\mathrm{L}^2} - \mu\|g\|^2_{\mathrm{W}^{1/2,2}}.$$

Hence, $\mathscr{F}[u]$ controls $\|u\|_{\mathrm{W}^{1,2}}$ and our functional is weakly coercive. Moreover, it is clear that our integrand is convex in in the $\mathscr{E}u$-argument (note that the trace function is linear). Hence, Theorem 2.8 yields the existence of a solution $u_* \in \mathrm{W}^{1,2}(\Omega;\mathbb{R}^3)$ to our minimization problem of linearized elasticity.

We finish this section with the following converse to Theorem 2.7:

Comment...

Proposition 2.11. *Let $\mathscr{F}$ be an integral functional with a Carathéodory integrand $f\colon \Omega \times \mathbb{R}^{m\times d} \to \mathbb{R}$ satisfying the upper growth bound* (2.3)*. Assume furthermore that $\mathscr{F}$ is weakly lower semicontinuous on $\mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$. If either $m = 1$ or $d = 1$ (the **scalar case**), then $A \mapsto f(x,A)$ is convex for almost every $x \in \Omega$.*

We will not prove this result here, since Proposition 3.31 together with Corollary 3.4 in the next chapter will give a stronger version.

In the **vectorial case**, i.e. $m \neq 1$ or $d \neq 1$, the necessity of convexity is far from being true and there is indeed another "canonical" condition ensuring weak lower semicontinuity, which is weaker than (ordinary) convexity; we will explore this in the next chapter.

2.3 Integrands with u-dependence

If we try to extend the results in the previous section to more general functionals

$$\mathscr{F}[u] := \int_\Omega f(x,u(x),\nabla u(x))\,\mathrm{d}x \quad \to \quad \min,$$

we discover that our proof strategy of using the Mazur Lemma runs into difficulties: We cannot "pull out" the convex combination inside

$$\int_\Omega f\left(x, \sum_{n=j}^{N(j)} \theta_{j,n} u_n(x), \sum_{n=j}^{N(j)} \theta_{j,n} \nabla u_n(x)\right) \mathrm{d}x$$

anymore. Nevertheless, a lower semicontinuity result analogous to the one for the scalar case turns out to be true:

Theorem 2.12. *Let $f\colon \Omega \times \mathbb{R}^m \times \mathbb{R}^{m\times d} \to \mathbb{R}$ satisfy the growth bound*

$$|f(x,v,A)| \leq M(1+|v|^p+|A|^p) \qquad \textit{for some } M>0,\ 1<p<\infty,$$

and the convexity property

$$A \mapsto f(x,v,A) \textit{ is convex for all } (x,v) \in \Omega \times \mathbb{R}^m.$$

Then, the functional

$$\mathscr{F}[u] := \int_\Omega f(x,u(x),\nabla u(x))\,\mathrm{d}x, \qquad u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m),$$

is weakly lower semicontinuous.

While it would be possible to give an elementary proof of this theorem here, we postpone the verification until the next chapter. There, using more advanced and elegant techniques (in particular Young measures), we will establish a much more general result and see that, when viewed from the right perspective, u-dependence does not pose any additional complications.

Comment...

2.4 The Euler–Lagrange equation

In analogy to the elementary fact that the derivative of a function at its minimizers is zero, we will now derive a necessary condition for a function to be a minimizer. This condition furnishes the connection between the calculus of variations and PDE theory and is of great importance for *computing* particular solutions.

Theorem 2.13 (Euler–Lagrange equation). *Assume that $f\colon \Omega \times \mathbb{R}^m \times \mathbb{R}^{m\times d} \to \mathbb{R}$ is a Carathéodory integrand that is continuously differentiable in v and A and satisfies the growth bounds*

$$|\mathrm{D}_v f(x,v,A)|, |\mathrm{D}_A f(x,v,A)| \le M(1+|v|^p+|A|^p), \qquad (x,v,A) \in \Omega \times \mathbb{R}^m \times \mathbb{R}^{m\times d},$$

for some $M > 0$ and $1 \le p < \infty$. If $u_* \in \mathrm{W}^{1,p}_g(\Omega;\mathbb{R}^m)$, where $g \in \mathrm{W}^{1-1/p,p}(\partial\Omega;\mathbb{R}^m)$, minimizes the functional

$$\mathscr{F}[u] := \int_\Omega f(x,u(x),\nabla u(x))\,\mathrm{d}x, \qquad u \in \mathrm{W}^{1,p}_g(\Omega;\mathbb{R}^m),$$

*then u_* is a weak solution of the following system of PDEs, called the **Euler–Lagrange equation**,*

$$\begin{cases} -\operatorname{div}\big[\mathrm{D}_A f(x,u,\nabla u)\big] + \mathrm{D}_v f(x,u,\nabla u) = 0 & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega. \end{cases} \tag{2.6}$$

Here we used the common convention to omit the x-arguments whenever this does not cause any confusion in order to curtail the proliferation of x's, for example in $f(x,u,\nabla u) = f(x,u(x),\nabla u(x))$. Note that our Euler–Lagrange "equation" is actually a *system* of PDEs.

Recall that u_* is a **weak solution** of (2.6) if

$$\int_\Omega \mathrm{D}_A f(x,u_*,\nabla u_*) : \nabla\psi + \mathrm{D}_v f(x,u_*,\nabla u_*)\cdot\psi\,\mathrm{d}x = 0$$

for all $\psi \in \mathrm{C}^\infty_c(\Omega;\mathbb{R}^m)$, where $\mathrm{D}_A f(x,v,A)$ is the matrix in $\mathbb{R}^{m\times d}$ such that

$$\mathrm{D}_A f(x,v,A) : B = \lim_{h\downarrow 0} \frac{f(x,v,A+hB) - f(x,v,A)}{h}, \qquad A,B \in \mathbb{R}^{m\times d},$$

called the **directional derivative** of $f(x,v,\cdot)$ at A in direction B; here we use the Frobenius matrix vector product ":" (see the appendix). A similar definition applies for $\mathrm{D}_v f(x,v,A)\cdot w$. More precisely, $\mathrm{D}_A f(x,v,A)$ and $\mathrm{D}_v f(x,v,A)$ are given as

$$\mathrm{D}_A f(x,v,A) := \big(\partial_{A^j_k} f(x,v,A)\big)^j_k, \qquad \mathrm{D}_v f(x,v,A) := \big(\partial_{v^j} f(x,v,A)\big)^j.$$

The boundary condition $u = g$ on $\partial\Omega$ is to be understood in the sense of trace.

Comment...

Proof. For all $\psi \in \mathrm{C}_c^\infty(\Omega;\mathbb{R}^m)$ and all $h > 0$ we have

$$\mathscr{F}[u_*] \leq \mathscr{F}[u_* + h\psi]$$

since $u_* + h\psi \in \mathrm{W}_g^{1,p}(\Omega;\mathbb{R}^m)$ is admissible in the minimization. Thus,

$$\begin{aligned}
0 &\leq \int_\Omega \frac{f(x,u_*+h\psi,\nabla u_*+h\nabla\psi) - f(x,u_*,\nabla u_*)}{h}\,\mathrm{d}x \\
&= \int_\Omega \int_0^1 \frac{1}{h}\frac{\mathrm{d}}{\mathrm{d}t}\big[f(x,u_*+th\psi,\nabla u_*+th\nabla\psi)\big]\,\mathrm{d}t\,\mathrm{d}x \\
&= \int_\Omega \int_0^1 \mathrm{D}_A f(x,u_*+th\psi,\nabla u_*+th\nabla\psi) : \nabla\psi \\
&\qquad\qquad + \mathrm{D}_v f(x,u_*+th\psi,\nabla u_*+th\nabla\psi)\cdot\psi\,\mathrm{d}t\,\mathrm{d}x.
\end{aligned}$$

By the growth bounds on the derivative, the integrand can be seen to have an h-uniform L^1-majorant, namely $C(1+|u_*|^p+|\psi|^p+|\nabla u_*|^p+|\nabla\psi|^p)$ (if we additionally assume $h \leq 1$) and so we may apply the Lebesgue dominated convergence theorem to let $h \downarrow 0$ under the double integral. This yields

$$0 \leq \int_\Omega \mathrm{D}_A f(x,u_*,\nabla u_*) : \nabla\psi + \mathrm{D}_v f(x,u_*,\nabla u_*)\cdot\psi\,\mathrm{d}x$$

and we conclude by taking ψ and $-\psi$ in this inequality. □

Remark 2.14. If we want to allow $\psi \in \mathrm{W}_0^{1,p}(\Omega;\mathbb{R}^m)$ in the weak formulation of (2.6), then we need to assume the stronger growth conditions

$$|\mathrm{D}_v f(x,v,A)|, |\mathrm{D}_A f(x,v,A)| \leq M(1+|v|^{p-1}+|A|^{p-1})$$

for some $M > 0$ and $1 \leq p < \infty$.

Example 2.15. Returning to the Dirichlet integral from Example 2.9, we see that the associated Euler–Lagrange equation is the **Laplace equation**

$$-\Delta u = 0,$$

where $\Delta := \partial_{x_1}^2 + \cdots + \partial_{x_d}^2$ is the **Laplace operator**. Solutions u are called **harmonic functions**. Since the Dirichlet integral is convex, all solutions of the Laplace equation are in fact minimizers. In particular, it can be seen that solutions of the Laplace equation are unique for given boundary values. The same results apply to the functional

$$\mathscr{F}[u] := \int_\Omega \frac{1}{2}|\nabla u(x)|^2 - g(x)\cdot u(x)\,\mathrm{d}x, \qquad u \in \mathrm{W}^{1,2}(\Omega),$$

where $g \in \mathrm{L}^2(\Omega)$. Here, the Euler–Lagrange equation is the **Poisson equation**

$$-\Delta u = g.$$

Comment...

Example 2.16. In the linearized elasticity problem from Example 2.10, we may compute the Euler–Lagrange equation as

$$\begin{cases} -\operatorname{div}\left[2\mu\,\mathscr{E}u + \left(\kappa - \frac{2}{3}\mu\right)(\operatorname{tr}\mathscr{E}u)I\right] = f & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega. \end{cases}$$

Sometimes, one also defines the **first variation** $\delta\mathscr{F}[u]$ of $\mathscr{F}$ at $u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ as the linear map $\delta\mathscr{F}[u]\colon \mathrm{W}^{1,p}_0(\Omega;\mathbb{R}^m) \to \mathbb{R}$ given as

$$\delta\mathscr{F}[u][\psi] := \lim_{h\downarrow 0} \frac{\mathscr{F}[u+h\psi] - \mathscr{F}[u]}{h}, \qquad \psi \in \mathrm{W}^{1,p}_0(\Omega;\mathbb{R}^m),$$

assuming this limit exists. Then, the assertion of the previous theorem can equivalently be formulated as

$$\delta\mathscr{F}[u_*] = 0 \quad \text{if } u_* \text{ minimizes } \mathscr{F} \text{ over } \mathrm{W}^{1,p}_g(\Omega;\mathbb{R}^m).$$

Of course, this condition is only *necessary* for u to be a minimizer. Indeed, any solution of the Euler–Lagrange equation is called a **critical point** of $\mathscr{F}$, which could be a minimizer, maximizer, or saddle point.

One crucial consequence of the results in this section is that usually we can use all available PDE methods to study minimizers. Immediately, one can ask about the type of PDE we are dealing with. In this respect we have the following result:

Proposition 2.17. *Let f be twice continuously differentiable in v and A. Also let $A \mapsto f(x,v,A)$ be convex for all $(x,v) \in \Omega \times \mathbb{R}^m$. Then, the Euler–Lagrange equation is an* ***elliptic*** *PDE, that is,*

$$0 \le \mathrm{D}^2_A f(x,v,A)[B,B] := \frac{\mathrm{d}^2}{\mathrm{d}t^2} f(x,v,A+tB)\bigg|_{t=0}$$

for all $x \in \Omega$, $v \in \mathbb{R}^m$, and $A, B \in \mathbb{R}^{m\times d}$.

The proof is immediate from the convexity.

One main use of the Euler–Lagrange equation is to find concrete solutions of variational problems:

Example 2.18. Recall the optimal saving problem from Section 1.6,

$$\begin{cases} \mathscr{F}[S] := \int_0^T -\ln(w + \rho S(t) - \dot{S}(t))\,\mathrm{d}t \quad \to \quad \min, \\ S(0) = 0,\ S(T) = S_T \ge 0. \end{cases}$$

The Euler–Lagrange equation is

$$-\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{1}{w + \rho S(t) - \dot{S}(t)}\right] = -\frac{\rho}{w + \rho S(t) - \dot{S}(t)},$$

Comment...

which resolves to

$$\frac{\rho\dot{S}(t)-\ddot{S}(t)}{w+\rho S(t)-\dot{S}(t)}=\rho.$$

With the consumption rate $C(t) := w+\rho S(t)-\dot{S}(t)$, this is equivalent to

$$\frac{\dot{C}(t)}{C(t)}=\rho,$$

and so, if $C(0)=C_0$ (to be determined later),

$$w+\rho S(t)-\dot{S}(t)=C(t)=\mathrm{e}^{\rho t}C_0.$$

This differential equation for $S(t)$ can be solved, for example via the *Duhamel formula*, which yields

$$\begin{aligned} S(t) &= \mathrm{e}^{\rho t}\cdot 0+\int_0^t \mathrm{e}^{\rho(t-s)}(w-\mathrm{e}^{\rho s}C_0)\,\mathrm{d}s \\ &= \frac{\mathrm{e}^{\rho t}-1}{\rho}w-t\mathrm{e}^{\rho t}C_0. \end{aligned}$$

and C_0 can now be chosen to satisfy the terminal condition $S(T)=S_T$, in fact, $C_0=(1-\mathrm{e}^{-\rho T})w/(\rho T)-\mathrm{e}^{-\rho T}S_T/T$.

Since $C\mapsto -\ln C$ is convex, we know that $S(t)$ as above is a minimizer of our problem. In Figure 2.1 we see the optimal savings strategy for a worker earning a (constant) salary of $w=£30,000$ per year and having a savings goal of $S_T=£100,000$. The worker has to save for approximately 27 years, reaching savings of just over £168,000, and then starts withdrawing his savings ($\dot{S}(t)<0$) for the last 13 years. The worker's consumption $C(t)=\mathrm{e}^{\rho t}C_0$ goes up continuously during the whole working life, making him/her (materially) happy.

Example 2.19. For functions $u=u(t,x)\colon \mathbb{R}\times\mathbb{R}^d\to\mathbb{R}$, consider the *action functional*

$$\mathscr{A}[u] := \frac{1}{2}\int_{\mathbb{R}\times\mathbb{R}^d} -|\partial_t u|^2+|\nabla_x u|^2\,\mathrm{d}(t,x),$$

where $\nabla_x u$ is the gradient of u with respect to $x\in\mathbb{R}^d$. This functional should be interpreted as the usual Dirichlet integral, see Example 2.9, where, however, we use the Lorentz metric. Then, the Euler–Lagrange equation is the **wave equation**

$$\partial_t^2 u-\Delta u=0.$$

Notice that $\mathscr{A}$ is not convex and consequently the wave equation is *hyperbolic* instead of elliptic.

Comment...

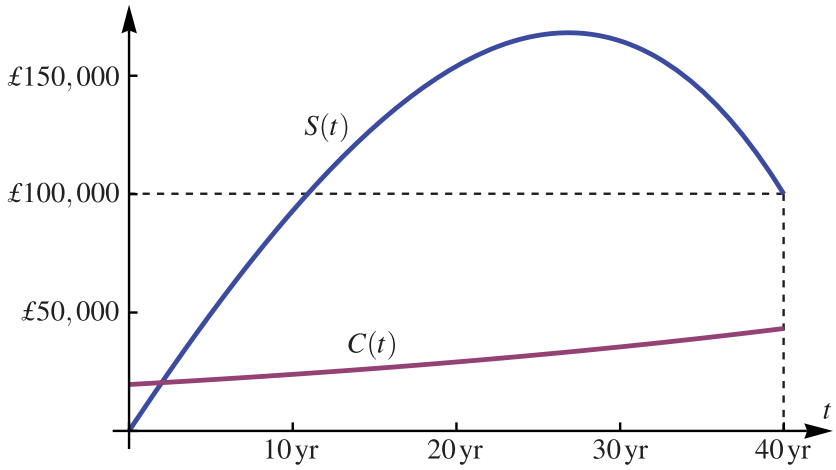

Figure 2.1: The solution to the optimal saving problem.

It is an important question whether a weak solution of the Euler–Lagrange equation (2.6) is also a **strong solution**, that is, whether $u \in \mathrm{W}^{2,2}(\Omega;\mathbb{R}^m)$ and

$$\begin{cases} -\operatorname{div}\big[\mathrm{D}_A f(x,u(x),\nabla u(x))\big] + \mathrm{D}_v f(x,u(x),\nabla u(x)) = 0 & \text{for a.e. } x \in \Omega, \\ u = g & \text{on } \partial\Omega. \end{cases} \tag{2.7}$$

If $u \in \mathrm{C}^2(\Omega;\mathbb{R}^m) \cap \mathrm{C}(\overline{\Omega};\mathbb{R}^m)$ satisfies this PDE for *every* $x \in \Omega$, then we call u a **classical solution**.

Multiplying (2.7) by a test function $\psi \in \mathrm{C}_c^\infty(\Omega;\mathbb{R}^m)$, integrating over Ω, and using the Gauss–Green Theorem, it follows that any solution of (2.7) also solves (2.6). The converse is also true whenever u is sufficiently regular:

Proposition 2.20. *Let the integrand f be twice continuously differentiable and let $u \in \mathrm{W}^{2,2}(\Omega;\mathbb{R}^m)$ be a weak solution of the Euler–Lagrange equation* (2.6)*. Then, u solves the Euler–Lagrange equation* (2.7) *in the strong sense.*

Proof. If $u \in \mathrm{W}^{2,2}(\Omega;\mathbb{R}^m)$ is a weak solution, then

$$\int_\Omega \mathrm{D}_A f(x,u,\nabla u) : \nabla\psi + \mathrm{D}_v f(x,u,\nabla u) \cdot \psi \,\mathrm{d}x = 0$$

for all $\psi \in \mathrm{C}_c^\infty(\Omega;\mathbb{R}^m)$. Integration by parts (more precisely, the Gauss–Green Theorem) gives

$$\int_\Omega \big(-\operatorname{div}\big[\mathrm{D}_A f(x,u,\nabla u)\big] + \mathrm{D}_v f(x,u,\nabla u)\big) \cdot \psi \,\mathrm{d}x = 0,$$

again for all ψ as before. We conclude using the following important lemma. □

Comment...

Lemma 2.21 (Fundamental lemma of the calculus of variations). *Let $\Omega \subset \mathbb{R}^d$ be open. If $g \in \mathrm{L}^1(\Omega)$ satisfies*

$$\int_\Omega g\psi \,\mathrm{d}x = 0 \qquad \textit{for all } \psi \in \mathrm{C}_c^\infty(\Omega),$$

then $g = 0$ almost everywhere.

Proof. We can assume that Ω is bounded by considering subdomains if necessary. Also, let g be extended by zero to all of $\mathbb{R}^d$. Fix $\varepsilon > 0$ and let $(\eta_\delta)_{\delta>0}$ be a family of mollifiers, see Appendix A.2 for details. Then, since $\eta_\delta \star g \to g$ in L^1, there is $h \in (\mathrm{L}^1 \cap \mathrm{C}^\infty)(\mathbb{R}^d)$ with the properties

$$\|g-h\|_{\mathrm{L}^1} \le \frac{\varepsilon}{4} \qquad \text{and} \qquad \|h\|_\infty < \infty.$$

Set $\varphi(x) := h(x)/|h(x)|$ for $h(x) \neq 0$ and $\varphi(x) := 0$ for $h(x) = 0$, so that $h\varphi = |h|$. Then take $\psi = \eta_{\bar\delta} \star \varphi \in (\mathrm{L}^1 \cap \mathrm{C}^\infty)(\mathbb{R}^d)$ for some $\bar\delta > 0$ such that

$$\|\varphi - \psi\|_{\mathrm{L}^1} \le \frac{\varepsilon}{2\|h\|_\infty}.$$

Since $|\psi| \le 1$ (this follows from the definition of the convolution),

$$\begin{aligned}
\|g\|_{\mathrm{L}^1} &\le \|g-h\|_{\mathrm{L}^1} + \int_\Omega h\varphi \,\mathrm{d}x \\
&\le \|g-h\|_{\mathrm{L}^1} + \int_\Omega h(\varphi-\psi) + (h-g)\psi + g\psi \,\mathrm{d}x \\
&\le 2\|g-h\|_{\mathrm{L}^1} + \|h\|_\infty \cdot \|\varphi-\psi\|_{\mathrm{L}^1} + 0 \\
&\le \varepsilon.
\end{aligned}$$

We conclude by letting $\varepsilon \downarrow 0$. □

2.5 Regularity of minimizers

As we saw at the end of the last section, whether a weak solution of the Euler–Lagrange equation is also a strong or even classical solution, depends on its *regularity*, that is, on the amount of differentiability. More generally, one would like to know *how much* regularity we can expect from solutions of a variational problem. Such a question also was the content of David Hilbert's 19th problem [55]:

> *Does every Lagrangian partial differential equation of a regular variational problem have the property of exclusively admitting analytic integrals?*[1]

[1] The German original asks "ob jede Lagrangesche partielle Differentialgleichung eines regulären Variationsproblems die Eigenschaft hat, daß sie nur analytische Integrale zuläßt."

Comment...

In modern language, Hilbert asked whether "regular" variational problems (defined below) admit only analytic solutions, i.e. ones that have a local power series representation.

In this section, we will prove some basic regularity assertions, but we will only sketch the solution of Hilbert's 19th problem, as the techniques needed are quite involved. We remark from the outset that regularity results are very sensitive to the dimensions involved, in particular the behavior of the scalar case ($m = 1$) and the vector case ($m > 1$) is fundamentally different. We also only consider the quadratic ($p = 2$) case for reasons of simplicity.

In the spirit of Hilbert's 19th problem, call

$$\mathscr{F}[u] := \int_\Omega f(\nabla u(x))\,\mathrm{d}x, \qquad u \in \mathrm{W}^{1,2}(\Omega;\mathbb{R}^m),$$

a **regular variational integral** if $f \in \mathrm{C}^2(\mathbb{R}^{m\times d})$ and there are constants $0 < \mu \le M$ with

$$\mu|B|^2 \le \mathrm{D}^2_A f(A)[B,B] \le M|B|^2 \qquad \text{for all } A,B \in \mathbb{R}^{m\times d}.$$

In this context,

$$\mathrm{D}^2_A f(A)[B_1,B_2] = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}s} f(A + sB_1 + tB_2)\bigg|_{s,t=0} \qquad \text{for all } A,B_1,B_2 \in \mathbb{R}^{m\times d},$$

which is a bilinear form. Clearly, regular variational problems are convex. In fact, integrands f that satisfy the above lower bound $\mu|B|^2 \le \mathrm{D}^2_A f(A)[B,B]$ are called **strongly convex**. For example, the Dirichlet integral from Example 2.9 is regular.

The basic regularity theorem is the following:

Theorem 2.22 ($\mathrm{W}^{2,2}_{\mathrm{loc}}$-regularity). *Let $\mathscr{F}$ be a regular variational integral. Then, for any minimizer $u \in \mathrm{W}^{1,2}(\Omega;\mathbb{R}^m)$ of $\mathscr{F}$, it holds that $u \in \mathrm{W}^{2,2}_{\mathrm{loc}}(\Omega;\mathbb{R}^m)$. Moreover, for any $B(x_0,3r) \subset \Omega$ ($x_0 \in \Omega$, $r > 0$) the **Caccioppoli inequality***

$$\int_{B(x_0,r)} |\nabla^2 u(x)|^2\,\mathrm{d}x \le \left(\frac{2M}{\mu}\right)^2 \int_{B(x_0,3r)} \frac{|\nabla u(x) - [\nabla u]_{B(x_0,3r)}|^2}{r^2}\,\mathrm{d}x \tag{2.8}$$

holds, where $[\nabla u]_{B(x_0,3r)} := \fint_{B(x_0,3r)} \nabla u\,\mathrm{d}x$. Consequently, the Euler–Lagrange equation holds strongly,

$$-\operatorname{div} \mathrm{D}f(\nabla u) = 0, \qquad \textit{a.e. in } \Omega.$$

For the proof we employ the *difference quotient method*, which is fundamental in regularity theory. Let $u\colon \Omega \to \mathbb{R}^m$, $x \in \Omega$, $k \in \{1,\dots,d\}$, and $h \in \mathbb{R}$. Then, define the **difference quotients**

$$\mathrm{D}^h_k u(x) := \frac{u(x + h\mathrm{e}_k) - u(x)}{h}, \qquad \mathrm{D}^h u := (\mathrm{D}^h_1 u, \dots, \mathrm{D}^h_d u),$$

where $\{\mathrm{e}_1,\dots,\mathrm{e}_d\}$ is the standard basis of $\mathbb{R}^d$. The key is the following characterization of Sobolev spaces in terms of difference quotients:

Comment...

Lemma 2.23. *Let $1 < p < \infty$, $D \Subset \Omega \subset \mathbb{R}^d$ be open, and $u \in \mathrm{L}^p(\Omega;\mathbb{R}^m)$.*

(i) If $u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$, then

$$\|\mathrm{D}_k^h u\|_{\mathrm{L}^p(D)} \leq \|\partial_k u\|_{\mathrm{L}^p(\Omega)} \qquad \text{for all } k \in \{1,\dots,d\},\ |h| < \operatorname{dist}(D,\partial\Omega).$$

(ii) If for some $0 < \delta < \operatorname{dist}(D,\partial\Omega)$ it holds that

$$\|\mathrm{D}_k^h u\|_{\mathrm{L}^p(D)} \leq C \qquad \text{for all } k \in \{1,\dots,d\} \text{ and all } |h| < \delta,$$

then $u \in \mathrm{W}^{1,p}_{\mathrm{loc}}(\Omega;\mathbb{R}^m)$ and $\|\partial_k u\|_{\mathrm{L}^p(D)} \leq C$ for all $k \in \{1,\dots,d\}$.

Proof. For (i) assume first that $u \in (\mathrm{L}^p \cap \mathrm{C}^1)(\Omega;\mathbb{R}^m)$. In this case, by the Fundamental Theorem of Calculus, at $x \in \Omega$ it holds that

$$\mathrm{D}_k^h u(x) = \frac{1}{h}\int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} u(x+the_k)\,\mathrm{d}t = \int_0^1 \partial_k u(x+the_k)\,\mathrm{d}t.$$

Thus,

$$\int_D |\mathrm{D}_k^h u|^p\,\mathrm{d}x \leq \int_\Omega |\partial_k u|^p\,\mathrm{d}x,$$

from which the assertion follows. The general case follows from the density of $(\mathrm{L}^p \cap \mathrm{C}^1)(\Omega;\mathbb{R}^m)$ in $\mathrm{L}^p(\Omega;\mathbb{R}^m)$.

For (ii), we observe that for fixed $k \in \{1,\dots,d\}$ by assumption $(\mathrm{D}_k^h u)_{0<h<\delta}$ is uniformly L^p-bounded. Thus, for an arbitrary fixed sequence of h's tending to zero, there exists a subsequence $h_j \downarrow 0$ with

$$\mathrm{D}_k^{h_j} u \rightharpoonup v_k \quad \text{in } \mathrm{L}^p$$

for some $v_k \in \mathrm{L}^p(\Omega;\mathbb{R}^m)$. Let $\psi \in \mathrm{C}_c^\infty(\Omega;\mathbb{R}^m)$. Using an "integration-by-parts" rule for difference quotients, which is elementary to check, we get

$$\int_D v_k \cdot \psi\,\mathrm{d}x = \lim_{j\to\infty}\int_D \mathrm{D}_k^{h_j} u \cdot \psi\,\mathrm{d}x = -\lim_{j\to\infty}\int_D u \cdot \mathrm{D}_k^{-h_j}\psi\,\mathrm{d}x = -\int_D u \cdot \partial_k \psi\,\mathrm{d}x.$$

Thus, $u \in \mathrm{W}^{1,p}(D;\mathbb{R}^m)$ and $v_k = \partial_k u$. The norm estimate follows from the lower semicontinuity of the norm under weak convergence. □

Proof of Theorem 2.22. The idea is to emulate the a-priori non-existent second derivatives using difference quotients and to derive estimates which allow one to conclude that these difference quotients are in L^2. Then we can conclude by the preceding lemma.

Let $u \in \mathrm{W}^{1,2}(\Omega;\mathbb{R}^m)$ be a minimizer of $\mathscr{F}$. By Theorem 2.13,

$$0 = \int_\Omega \mathrm{D}f(\nabla u) : \nabla\psi\,\mathrm{d}x \tag{2.9}$$

Comment...

holds for all $\psi \in \mathrm{C}_c^\infty(\Omega;\mathbb{R}^m)$, and then by density also for all $\psi \in \mathrm{W}^{1,2}(\Omega;\mathbb{R}^m)$. As before we have used the notation $\mathrm{D}f(A) : B$ for the directional derivative of f at A in direction B. Fix a ball $B(x_0, 3r) \subset \Omega$ and take a Lipschitz cut-off function $\rho \in \mathrm{W}^{1,\infty}(\Omega)$ such that

$$\mathbb{1}_{B(x_0,r)} \le \rho \le \mathbb{1}_{B(x_0,2r)} \qquad \text{and} \qquad |\nabla\rho| \le \frac{1}{r}.$$

Then, for any $k = 1,\dots,d$ and $|h| < r$, we let

$$\psi := \mathrm{D}_k^{-h}\big[\rho^2 \mathrm{D}_k^h(u-a)\big] \quad \in \mathrm{W}^{1,2}(\Omega;\mathbb{R}^m),$$

where a is an affine function to be chosen later. We may plug ψ into (2.9) to get

$$0 = \int_\Omega \mathrm{D}_k^h(\mathrm{D}f(\nabla u)) : \big[\rho^2 \mathrm{D}_k^h \nabla u + \mathrm{D}_k^h(u-a)\otimes\nabla(\rho^2)\big]\,\mathrm{d}x. \tag{2.10}$$

Here, we used the "integration-by-parts" formula for difference quotients again (also recall $a\otimes b := ab^T$).

Next, we estimate, using the assumptions on f,

$$\begin{aligned}
\mu|\mathrm{D}_k^h\nabla u|^2 &\le \int_0^1 \mathrm{D}^2 f(\nabla u + th\mathrm{D}_k^h\nabla u)[\mathrm{D}_k^h\nabla u, \mathrm{D}_k^h\nabla u]\,\mathrm{d}t \\
&= \frac{1}{h}\mathrm{D}f(\nabla u + th\mathrm{D}_k^h\nabla u) : \mathrm{D}_k^h\nabla u\Big|_{t=0}^1 \\
&= \mathrm{D}_k^h(\mathrm{D}f(\nabla u)) : \mathrm{D}_k^h\nabla u.
\end{aligned}$$

We get, using the Cauchy–Schwarz and Young inequalities,

$$\begin{aligned}
&\big|\mathrm{D}_k^h(\mathrm{D}f(\nabla u)) : \big[\mathrm{D}_k^h(u-a)\otimes\nabla(\rho^2)\big]\big| \\
&\quad\le 2\rho|\mathrm{D}_k^h(\mathrm{D}f(\nabla u))|\cdot|\mathrm{D}_k^h(u-a)|\cdot|\nabla\rho| \\
&\quad\le \frac{\mu}{2M^2}\rho^2|\mathrm{D}_k^h(\mathrm{D}f(\nabla u))|^2 + \frac{2M^2}{\mu}|\mathrm{D}_k^h(u-a)|^2\cdot|\nabla\rho|^2.
\end{aligned}$$

Using the last two estimates, (2.10), the mean-value theorem, and the regularity of the integrand, we get

$$\begin{aligned}
\mu\int_\Omega |\mathrm{D}_k^h\nabla u|^2\rho^2\,\mathrm{d}x &\le \int_\Omega \mathrm{D}_k^h(\mathrm{D}f(\nabla u)) : [\rho^2\mathrm{D}_k^h\nabla u]\,\mathrm{d}x \\
&= -\int_\Omega \mathrm{D}_k^h(\mathrm{D}f(\nabla u)) : \big[\mathrm{D}_k^h(u-a)\otimes\nabla(\rho^2)\big]\,\mathrm{d}x \\
&\le \int_\Omega \frac{\mu}{2M^2}\rho^2|\mathrm{D}_k^h(\mathrm{D}f(\nabla u))|^2 + \frac{2M^2}{\mu}|\mathrm{D}_k^h(u-a)|^2\cdot|\nabla\rho|^2\,\mathrm{d}x \\
&\le \int_\Omega \frac{\mu}{2}\rho^2|\partial_k\nabla u|^2 + \frac{2M^2}{\mu}|\mathrm{D}_k^h(u-a)|^2\cdot|\nabla\rho|^2\,\mathrm{d}x.
\end{aligned}$$

Comment...

Absorbing the first term in the right-hand side in the left-hand side and using the properties of ρ,

$$\int_{B(x_0,r)} |\mathrm{D}_k^h \nabla u|^2 \,\mathrm{d}x \le \left(\frac{2M}{\mu}\right)^2 \int_{B(x_0,2r)} \frac{|\mathrm{D}_k^h (u-a)|^2}{r^2} \,\mathrm{d}x.$$

Now invoke the difference-quotient lemma, part (i), to arrive at

$$\int_{B(x_0,r)} |\mathrm{D}_k^h \nabla u|^2 \,\mathrm{d}x \le \left(\frac{2M}{\mu}\right)^2 \int_{B(x_0,3r)} \frac{|\partial_k (u-a)|^2}{r^2} \,\mathrm{d}x.$$

Applying the lemma again, this time part (ii), we get $u \in \mathrm{W}^{2,2}(B(x_0,r);\mathbb{R}^m)$. The Caccioppoli inequality (2.8) follows once we take $\nabla a := [\nabla u]_{B(x_0,3r)}$. □

With the $\mathrm{W}^{2,2}_{\mathrm{loc}}$-regularity result at hand, we may use $\psi = \partial_k \tilde{\psi}$, $k = 1,\dots,d$, as test function in the weak formulation of the Euler–Lagrange equation $-\operatorname{div}[\mathrm{D}f(\nabla u)] = 0$ to conclude that

$$-\operatorname{div}\big[\mathrm{D}^2 f(\nabla u) : \nabla(\partial_k u)\big] = 0 \tag{2.11}$$

holds in the weak sense, i.e.

$$0 = -\int_\Omega \operatorname{div} \mathrm{D}f(\nabla u) \cdot \partial_k \psi \,\mathrm{d}x = \int_\Omega \operatorname{div}\big[\mathrm{D}^2 f(\nabla u) : \nabla(\partial_k u)\big] \cdot \psi \,\mathrm{d}x$$

for all $\psi \in \mathrm{C}_c^\infty(\Omega;\mathbb{R}^m)$. In the special case that f is quadratic, $\mathrm{D}^2 f$ is constant and the above PDE has the same structure as the original Euler–Lagrange equation, but now with $\partial_k u$ as unknown function. Thus, it is itself an Euler–Lagrange equation and we can **bootstrap** the regularity theorem to get the following result:

Corollary 2.24. *Let $\mathscr{F}$ be a quadratic and regular variational integral. Then, for any minimizer $u \in \mathrm{W}^{1,2}(\Omega;\mathbb{R}^m)$ of $\mathscr{F}$, it holds that $u \in \mathrm{W}^{k,2}_{\mathrm{loc}}(\Omega;\mathbb{R}^m)$ for all $k \in \mathbb{N}$, hence also $u \in \mathrm{C}^\infty(\Omega;\mathbb{R}^m)$.*

Example 2.25. For a minimizer $u \in \mathrm{W}^{1,2}(\Omega)$ of the Dirichlet integral as in Example 2.9, the theory presented here immediately gives $u \in \mathrm{C}^\infty(\Omega)$. The same applies to the the functional

$$\mathscr{F}[u] := \int_\Omega \frac{1}{2}|\nabla u(x)|^2 - g(x) \cdot u(x) \,\mathrm{d}x, \qquad u \in \mathrm{W}^{1,2}(\Omega).$$

Thus, solutions $u \in \mathrm{W}^{1,2}(\Omega)$ of the Laplace equation

$$-\Delta u = 0$$

or the Poisson equation

$$-\Delta u = g$$

are smooth.

Comment...

Example 2.26. Similarly, for minimizers $u \in \mathrm{W}^{1,2}(\Omega;\mathbb{R}^3)$ to the problem of linearized elasticity from Section 1.4 and Example 2.10, we also get $u \in \mathrm{C}^\infty(\Omega;\mathbb{R}^3)$. The reasoning has to be slightly adjusted, however, to take care of the fact that we are dealing with $\mathscr{E}u$ instead of ∇u.

In the general case, however, our Euler–Lagrange equation is *nonlinear* and the above bootstrapping procedure fails. In particular, this includes the problem posed in Hilbert's 19th problem, which was only solved, in the scalar case $m = 1$, by Ennio De Giorgi in 1957 [31] and, using different methods, by John Nash in 1958 [91]. After the proof was improved by Jürgen Moser in 1960/1961 [81, 82], the results are now called *De Giorgi–Nash–Moser theory*.

The current standard solution (based on De Giorgi's approach) is based on the **De Giorgi regularity theorem** and the classical **Schauder estimates**, which establish Hölder regularity of weak solutions of a PDE.

Theorem 2.27 (De Giorgi 1957). *Let the function $A\colon \Omega \to \mathbb{R}^{d\times d}$ be* measurable*, symmetric, i.e. $A(x) = A(x)^T$ for $x \in \Omega$, and satisfy the ellipticity and boundedness estimates*

$$\mu|v|^2 \le v^T A(x)v \le M|v|^2, \qquad x \in \Omega, v \in \mathbb{R}^d, \tag{2.12}$$

for constants $0 < \mu \le M$. If $u \in \mathrm{W}^{1,2}(\Omega)$ is a weak solution of

$$-\operatorname{div}[A\nabla u] = 0, \tag{2.13}$$

then $u \in \mathrm{C}^{0,\alpha_0}_{\mathrm{loc}}(\Omega)$, that is, u is α_0-Hölder continuous, for some $\alpha_0 = \alpha_0(d, M/\mu) \in (0,1)$.

Theorem 2.28 (Schauder 1934/1937). *Let $A\colon \Omega \to \mathbb{R}^{d\times d}$ be as above but in addition assumed to be of class $\mathrm{C}^{s-1,\alpha}$ for some $s \in \mathbb{N}$ and $\alpha \in (0,1)$. If $u \in \mathrm{W}^{1,2}(\Omega)$ is a weak solution of*

$$-\operatorname{div}[A\nabla u] = 0,$$

then $u \in \mathrm{C}^{s,\alpha}_{\mathrm{loc}}(\Omega)$.

We remark that the Schauder estimates also hold for *systems* of PDEs ($u\colon \Omega \to \mathbb{R}^m$, $m > 1$), but the De Giorgi regularity theorem does not. The proofs of these results are quite involved and reserved for an advanced course on regularity theory, but we establish one of their most important consequences, namely the solution of Hilbert's 19th problem in the scalar case:

Theorem 2.29 (De Giorgi–Nash–Moser 1961). *Let $\mathscr{F}$ be a regular variational integral and $f \in \mathrm{C}^s(\mathbb{R}^d)$, $s \in \{2,3,\dots\}$. If $u \in \mathrm{W}^{1,2}(\Omega)$ minimizes $\mathscr{F}$, then it holds that $u \in \mathrm{C}^{s-1,\alpha}_{\mathrm{loc}}(\Omega)$ for some $\alpha \in (0,1)$. In particular, if $f \in \mathrm{C}^\infty(\mathbb{R}^d)$, then $u \in \mathrm{C}^\infty(\Omega)$.*

Proof. We saw in (2.11) that the partial derivative $\partial_k u$, $k = 1,\dots,d$, of a minimizer $u \in \mathrm{W}^{1,2}(\Omega)$ of $\mathscr{F}$ satisfy (2.13) for

$$A(x) := \mathrm{D}^2 f(\nabla u(x)), \qquad x \in \Omega.$$

Comment...

From general properties of the Hessian we conclude that $A(x)$ is symmetric and the upper and lower estimates (2.12) on A follow from the respective properties of $\mathrm{D}^2 f$. However, we cannot conclude (yet) any regularity of A beyond measurability. Nevertheless, we may apply the De Giorgi Regularity Theorem 2.27, whereby $\partial_k u \in \mathrm{C}^{0,\alpha_0}_{\mathrm{loc}}(\Omega)$ for some $\alpha_0 \in (0,1)$ and all $k = 1,\dots,d$. Hence $u \in \mathrm{C}^{1,\alpha_0}_{\mathrm{loc}}(\Omega)$. This is the assertion for $s = 2$.

If $s = 3$, our arguments so far in conjunction with the regularity assumptions on f imply $\mathrm{D}^2 f(\nabla u(x)) \in \mathrm{C}^{0,\alpha_0}_{\mathrm{loc}}(\Omega)$. Hence, the Schauder estimates from Theorem 2.28 apply and yield $\partial_k u \in \mathrm{C}^{1,\alpha_0}_{\mathrm{loc}}(\Omega)$, whereby $u \in \mathrm{C}^{2,\alpha_0}_{\mathrm{loc}}(\Omega) = \mathrm{C}^{s-1,\alpha_0}_{\mathrm{loc}}(\Omega)$. For higher s, this procedure can be iterated until we run out of f-derivatives and we have achieved $u \in \mathrm{C}^{s-1,\alpha_0}_{\mathrm{loc}}(\Omega)$. □

The above argument also yields analyticity of u if f is analytic. Another refinement shows that in the situation of the De Giorgi–Nash–Moser Theorem, we even have $u \in \mathrm{C}^{s-1,\alpha}_{\mathrm{loc}}(\Omega)$ for *all* $\alpha \in (0,1)$. Here one needs the additional Schauder–type result that weak solutions u to $-\operatorname{div}\big[A\nabla(\partial_k u)\big] = 0$ for A continuous are of class $\mathrm{C}^{0,\alpha}_{\mathrm{loc}}$ for *all* $\alpha \in (0,1)$. In the above proof we can apply this result in the case $s = 2$ since $A(x) = \mathrm{D}^2 f(\nabla u(x))$ is continuous. Thus, $u \in \mathrm{C}^{1,\alpha}_{\mathrm{loc}}(\Omega)$ for *any* $\alpha \in (0,1)$. The other parts of the proof are adapted accordingly. See [98] for details and other refinements.

We close our look at regularity theory by considering the vectorial case $m > 1$. It was shown again by De Giorgi that if $d = m > 2$, then his regularity theorem does not hold:

Example 2.30 (De Giorgi 1968). Let $d = m > 2$ and define

$$u(x) := |x|^{-\gamma} x, \qquad \gamma := \frac{d}{2}\left(1 - \frac{1}{\sqrt{(2d-2)^2+1}}\right), \qquad x \in B(0,1).$$

Note that $1 < \gamma < \frac{d}{2}$ and so $u \in \mathrm{W}^{1,2}(B(0,1);\mathbb{R}^d)$ but $u \notin \mathrm{L}^\infty_{\mathrm{loc}}(B(0,1);\mathbb{R}^d)$. It can be checked, though, that u solves (2.13) for

$$v^T A(x) v := |v|^2 + \left[\left((d-1)I + d \cdot \frac{x \otimes x}{|x|^2}\right) v\right]^2,$$

which satisfies all assumptions of the De Giorgi Theorem. Moreover, u is a minimizer of the quadratic variational integral

$$\int_{B(0,1)} \nabla u(x)^T A(x) \nabla u(x) \, \mathrm{d}x.$$

However, this is not a regular variational integral because the integrand does not depend smoothly on x.

In principle, this still leaves the possibility that the vectorial analogue of Hilbert's 19th problem has a positive solution, just that its proof would have to proceed along a different avenue than via the De Giorgi Theorem. However, after first counterexamples of Nečas in 1975 [92], the current state of the art was established by Šverák and Yan in 2000–2002 [104, 105]. They prove that there exist regular (and C^∞) variational integrals such that minimizers must be:

Comment...

- non-Lipschitz if $d \geq 3$, $m \geq 5$,
- unbounded if $d \geq 5$, $m \geq 14$.

On the other hand, there are also positive results: in particular, we have that for $d = 2$ minimizers to regular variational problems are always as regular as the data allows (as in Theorem 2.29), this is the Morrey regularity theorem from 1938, see [78]. If we confine ourselves to Sobolev-regularity, then using the difference quotient technique, we can prove $\mathrm{W}^{2,2+\delta}_{\mathrm{loc}}$-regularity for minimizers of regular variational problems for some *dimension-dependent* $\delta > 0$, this result is originally due to Campanato. By an embedding theorem this yields $\mathrm{C}^{0,\alpha}$-regularity for some $\alpha \in (0,1)$ when $d \leq 4$. In 2008 Kristensen and Melcher were able to establish $\mathrm{W}^{2,2+\delta}_{\mathrm{loc}}$-regularity for a *dimension-independent* $\delta > 0$ (in fact, $\delta = \mu/(50M)$), see [65]. We will learn about further regularity results in Section 3.8. Regularity theory is still an area of active research and new results are constantly discovered.

Finally, we mention that regularity of solutions might in general be extended up to and including the boundary if the boundary is smooth enough, see Section 6.3.2 in [41] and also [49] for results in this direction.

2.6 Lavrentiev gap phenomenon

So far we have chosen the function spaces in which we look for solutions of a minimization problem according to coercivity and growth assumptions. However, at first sight, classically differentiable functions may appear to be more appealing. So the question arises whether the minimum (infimum) value is actually the same when considering different function spaces. Formally, given two spaces $X \subset Y$ such that X is dense in Y, and a functional $\mathscr{F} \colon Y \to \mathbb{R} \cup \{+\infty\}$, we ask whether

$$\inf_Y \mathscr{F} = \inf_X \mathscr{F}.$$

This is certainly true if we assume good growth conditions:

Theorem 2.31. *Let $f \colon \Omega \times \mathbb{R}^m \times \mathbb{R}^{m\times d} \to \mathbb{R}$ be a Caratheéodory function that has **p-growth**, i.e.*

$$|f(x,v,A)| \leq M(1 + |v|^p + |A|^p), \qquad (x,v,A) \in \Omega \times \mathbb{R}^m \times \mathbb{R}^{m\times d},$$

for some $M > 0$, $1 < p < \infty$. Then, the functional

$$\mathscr{F}[u] := \int_\Omega f(x,u(x),\nabla u(x))\,\mathrm{d}x, \qquad u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$$

is strongly continuous. Consequently,

$$\inf_{\mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)} \mathscr{F} = \inf_{\mathrm{C}^\infty(\Omega;\mathbb{R}^m)} \mathscr{F}.$$

Comment...

Proof. Let $u_j \to u$ in $\mathrm{W}^{1,p}$, whereby $u_j \to u$ and $\nabla u_j \to \nabla u$ almost everywhere (which holds after selecting a subsequence). Then, from the growth assumption we get

$$\mathscr{F}[u_j] = \int_\Omega f(x, u_j, \nabla u_j)\,\mathrm{d}x \le \int_\Omega M(1 + |u_j|^p + |\nabla u_j|^p)\,\mathrm{d}x$$

and by the Pratt Theorem A.6,

$$\mathscr{F}[u_j] \to \mathscr{F}[u].$$

This shows the continuity of $\mathscr{F}$ with respect to strong convergence in $\mathrm{W}^{1,p}$.

The assertion about the equality of infima now follows readily since $\mathrm{C}^\infty(\Omega;\mathbb{R}^m)$ is (strongly) dense in $\mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$. □

If we dispense with the (upper) growth, however, the infimum over different spaces may indeed be different – this is the **Lavrentiev gap phenomenon**, discovered in 1926 by Mikhail Lavrentiev. We here give an example between the spaces $\mathrm{W}^{1,1}$ and $\mathrm{W}^{1,\infty}$ (with boundary conditions) due to Basilio Manià.

Example 2.32 (Manià 1934). Consider the minimization problem

$$\begin{cases} \mathscr{F}[u] := \displaystyle\int_0^1 (u(t)^3 - t)^2 \dot{u}(t)^6\,\mathrm{d}t & \to \quad \min, \\ u(0) = 0,\ u(1) = 1 \end{cases}$$

for u from either $\mathrm{W}^{1,1}(0,1)$ or $\mathrm{W}^{1,\infty}(0,1)$ (Lipschitz functions). We claim that

$$\inf_{\mathrm{W}^{1,1}(0,1)} \mathscr{F} < \inf_{\mathrm{W}^{1,\infty}(0,1)} \mathscr{F},$$

where here and in the following these infima are to be taken only over functions u with boundary values $u(0) = 0$, $u(1) = 1$.

Clearly, $\mathscr{F} \ge 0$, and for $u_*(t) := t^{1/3} \in (\mathrm{W}^{1,1} \setminus \mathrm{W}^{1,\infty})(0,1)$ we have $\mathscr{F}[u_*] = 0$. Thus,

$$\inf_{\mathrm{W}^{1,1}(0,1)} \mathscr{F} = 0.$$

On the other hand, for any $u \in \mathrm{W}^{1,\infty}(0,1)$ with $u(0) = 0$, $u(1) = 1$, the Lipschitz-continuity of u implies that

$$u(t) \le h(t) := \frac{t^{1/3}}{2} \qquad \text{for all } t \in [0,\tau] \text{ and some } \tau > 0 \text{ with } u(\tau) = h(\tau).$$

Then, $u(t)^3 - t \le h(t)^3 - t$ for $t \in [0,\tau]$ and, since both of these terms are negative,

$$(u(t)^3 - t)^2 \ge (h(t)^3 - t)^2 = \frac{7^2}{8^2} t^2 \qquad \text{for all } t \in [0,\tau].$$

Comment...

We then estimate

$$\mathscr{F}[u] \geq \int_0^\tau (u(t)^3 - t)^2 \, \dot{u}(t)^6 \, \mathrm{d}t \geq \frac{7^2}{8^2} \int_0^\tau t^2 \, \dot{u}(t)^6 \, \mathrm{d}t.$$

Further, by Hölder's inequality,

$$\begin{aligned}
\int_0^\tau \dot{u}(t) \, \mathrm{d}t &= \int_0^\tau t^{-1/3} \cdot t^{1/3} \, \dot{u}(t) \, \mathrm{d}t \\
&\leq \left(\int_0^\tau t^{-2/5} \, \mathrm{d}t \right)^{5/6} \cdot \left(\int_0^\tau t^2 \, \dot{u}(t)^6 \, \mathrm{d}t \right)^{1/6} \\
&= \frac{5^{5/6}}{3^{5/6}} \tau^{1/2} \left(\int_0^\tau t^2 \, \dot{u}(t)^6 \, \mathrm{d}t \right)^{1/6}.
\end{aligned}$$

Since also

$$\int_0^\tau \dot{u}(t) \, \mathrm{d}t = u(\tau) - u(0) = h(\tau) = \frac{\tau^{1/3}}{2},$$

we arrive at

$$\mathscr{F}[u] \geq \frac{7^2 3^5}{8^2 5^5 2^6 \tau} \geq \frac{7^2 3^5}{8^2 5^5 2^6} > 0.$$

Thus,

$$\inf_{\mathrm{W}^{1,\infty}(0,1)} \mathscr{F} > 0 = \inf_{\mathrm{W}^{1,1}(0,1)} \mathscr{F}.$$

In a more recent example, Ball & Mizel [12] showed that the problem

$$\begin{cases} \mathscr{F}[u] := \displaystyle\int_{-1}^1 (t^4 - u(t)^6)^2 |\dot{u}(t)|^{2m} + \varepsilon \dot{u}(t)^2 \, \mathrm{d}t \quad \to \quad \min, \\ u(-1) = \alpha, \; u(1) = \beta, \end{cases}$$

also exhibits a Lavrentiev gap phenomenon between the spaces $\mathrm{W}^{1,2}$ and $\mathrm{W}^{1,\infty}$ if $m \in \mathbb{N}$ satisfies $m > 13$, $\varepsilon > 0$ is sufficiently small, and $-1 \leq \alpha < 0 < \beta \leq 1$. This example is significant because the Ball–Mizel functional is *coercive* on $\mathrm{W}^{1,2}(-1,1)$.

Finally, we note that the Lavrentiev gap phenomenon is a major obstacle for the numerical approximation of minimization problems. For instance, standard piecewise affine finite elements are in $\mathrm{W}^{1,\infty}$ and hence in the presence of the Lavrentiev gap phenomenon we cannot approximate the true solution with such elements. Thus, one is forced to work with non-conforming elements and other advanced schemes. This issue does not only affect "academic" examples such as the ones above, but is also of great concern in applied problems, such as nonlinear elasticity theory.

Comment...

2.7 Invariances and the Noether theorem

In physics and other applications of the calculus of variations, we are often highly interested in *symmetries* of minimizers or, more generally, critical points. These symmetries manifest themselves in other differential or pointwise relations that are automatically satisfied for any critical point. They can be used to identify concrete solutions or are interesting in their own right. In this section, for simplicity we only consider "sufficiently smooth" ($\mathrm{W}^{2,2}_{\mathrm{loc}}$) critical points, which is the natural level of regularity, see Theorem 2.22.

As a first concrete example of a symmetry, consider the Dirichlet integral

$$\mathscr{F}[u] := \frac{1}{2}\int_\Omega |\nabla u(x)|^2 \,\mathrm{d}x, \qquad u \in \mathrm{W}^{1,2}(\Omega),$$

which was introduced in Example 2.9. First we notice that $\mathscr{F}$ is *invariant* under translations in space: Let $u \in (\mathrm{W}^{1,2} \cap \mathrm{W}^{2,2}_{\mathrm{loc}})(\Omega)$, $\tau \in \mathbb{R}$, $k \in \{1,\dots,d\}$, and set

$$x_\tau := x + \tau \mathrm{e}_k, \qquad u_\tau(x) := u(x + \tau \mathrm{e}_k).$$

Then, for any open $D \subset \mathbb{R}^d$ with $D_\tau := D + \tau \mathrm{e}_k \subset \Omega$, we have the *invariance*

$$\frac{1}{2}\int_D |\nabla u_\tau(x)|^2 \,\mathrm{d}x = \frac{1}{2}\int_{D_\tau} |\nabla u(x_\tau)|^2 \,\mathrm{d}x_\tau. \tag{2.14}$$

The Dirichlet integral also exhibits invariance with respect to *scaling*: For $\lambda > 0$ set

$$x_\lambda := \lambda x, \qquad u_\lambda(x) := \lambda^{(d-2)/2} u(\lambda x), \qquad D_\lambda := \lambda D. \tag{2.15}$$

Then it is not hard to see that (2.14) again holds if we set $\lambda = \mathrm{e}^\tau$ (to allow $\tau \in \mathbb{R}$ as before).

The main result of this section, the Noether theorem, roughly says that "differentiable invariances of a functional give rise to conservation laws". More concretely, the two invariances of the Dirichlet integral presented above will yield two additional PDEs that any minimizer of the Dirichlet integral must satisfy.

To make these statement precise in the general case, we need a bit of notation: Let $u \in (\mathrm{W}^{1,2} \cap \mathrm{W}^{2,2}_{\mathrm{loc}})(\Omega;\mathbb{R}^m)$ be a critical point of the functional

$$\mathscr{F}[u] := \int_\Omega f(x,u(x),\nabla u(x)) \,\mathrm{d}x,$$

where $f\colon \Omega \times \mathbb{R}^m \times \mathbb{R}^{m\times d} \to \mathbb{R}$ is twice continuously differentiable. Then, u satisfies the strong Euler–Lagrange equation

$$-\operatorname{div}\big[\mathrm{D}_A f(x,u,\nabla u)\big] + \mathrm{D}_v f(x,u,\nabla u) = 0 \quad \text{a.e. in } \Omega.$$

see Proposition 2.20. We consider u to be extended to all of $\mathbb{R}^d$ (it will not matter below how we extend u).

Our invariance is given through maps $g\colon \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ and $H\colon \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^m$, which will depend on u above, such that

$$g(x,0) = x \qquad \text{and} \qquad H(x,0) = u(x), \qquad x \in \mathbb{R}^d.$$

Comment...

We also require that g, H are *continuously differentiable* in their second argument for almost every $x \in \Omega$. Then set for $x \in \mathbb{R}^d$, $\tau \in \mathbb{R}$, $D \Subset \mathbb{R}^d$

$$x_\tau := g(x,\tau), \qquad u_\tau(x) := H(x,\tau), \qquad D_\tau := g(D,\tau).$$

Therefore, g, H can be thought of as a form of *homotopy*.

We call $\mathscr{F}$ **invariant** under the transformation defined by (g,H) if

$$\int_D f(x,u_\tau(x),\nabla u_\tau(x))\,\mathrm{d}x = \int_{D_\tau} f(x',u(x'),\nabla u(x'))\,\mathrm{d}x'. \tag{2.16}$$

for all $\tau \in \mathbb{R}$ sufficiently small and all open $D \subset \mathbb{R}^d$ such that $D_\tau \Subset \Omega$.

The main result of this section goes back to Emmy Noether[2] and is considered to be one of the most important mathematical theorems ever proved. Its pivotal idea of systematically generating conservation laws from invariances has proved to be immensely influential in modern physics.

Theorem 2.33 (Noether 1915). *Let $f\colon \Omega \times \mathbb{R}^m \times \mathbb{R}^{m\times d} \to \mathbb{R}$ be twice continuously differentiable and satisfy the growth bounds*

$$|\mathrm{D}_v f(x,v,A)|, |\mathrm{D}_A f(x,v,A)| \le M(1+|v|^p+|A|^p), \qquad (x,v,A) \in \Omega \times \mathbb{R}^m \times \mathbb{R}^{m\times d},$$

for some $M > 0$, $1 \le p < \infty$. Further, let the associated functional $\mathscr{F}$ be invariant under the transformation defined by (g,H) as above, and assume that there exists a majorant $g \in \mathrm{L}^p(\Omega)$ such that

$$|\partial_\tau H(x,\tau)|, |\partial_\tau g(x,\tau)| \le g(x) \qquad \textit{for a.e. } x \in \Omega \textit{ and all } \tau \in \mathbb{R}. \tag{2.17}$$

Then, for any critical point $u \in (\mathrm{W}^{1,2} \cap \mathrm{W}^{2,2}_{\mathrm{loc}})(\Omega;\mathbb{R}^m)$ of $\mathscr{F}$ the conservation law

$$\operatorname{div}\big[\mu(x)\cdot \mathrm{D}_A f(x,u,\nabla u) - \nu(x)\cdot f(x,u,\nabla u)\big] = 0 \qquad \textit{for a.e. } x \in \Omega. \tag{2.18}$$

holds, where

$$\mu(x) := \partial_\tau H(x,0) \in \mathbb{R}^m \qquad \textit{and} \qquad \nu(x) := \partial_\tau g(x,0) \in \mathbb{R}^d \qquad (x \in \Omega)$$

*are the **Noether multipliers**.*

Corollary 2.34. *If $f = f(v,A)\colon \mathbb{R}\times\mathbb{R}^d \to \mathbb{R}$ does not depend on x, then every critical point $u \in (\mathrm{W}^{1,2} \cap \mathrm{W}^{2,2}_{\mathrm{loc}})(\Omega)$ of $\mathscr{F}$ satisfies*

$$\sum_{i=1}^d \partial_i\Big[(\partial_k u)\cdot \partial_{A_k} f(u,\nabla u) - \delta_{ik} f(u,\nabla u)\Big] = 0 \tag{2.19}$$

for all $k = 1,\dots,d$.

[2]The German mathematician Emmy Noether (23 March 1882 – 14 April 1935) has been called the most important woman in the history of mathematics by Einstein and others.

Comment...

Proof. Differentiate (2.16) with respect to τ and then set $\tau = 0$. To be able to differentiate the left hand side under the integral, we use the growth assumptions on the derivatives $|\mathrm{D}_v f(x,v,A)|, |\mathrm{D}_A f(x,v,A)|$ and (2.17) to get a uniform (in τ) majorant, which allows to move the differentiation under the integral sign. For the right hand side, we need to employ the formula for the differentiation of an integral with respect to a moving domain (this is a special case of the *Reynolds transport theorem*),

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\int_{D_\tau} f(x,u,\nabla u)\,\mathrm{d}x = \int_{\partial D_\tau} f(x,u,\nabla u)\partial_\tau g(x,\tau)\cdot n\,\mathrm{d}\sigma,$$

where n is the unit outward normal; this formula can be checked in an elementary way. Abbreviating for readability $F := f(x,u(x),\nabla u(x))$, $\mathrm{D}_A F := \mathrm{D}_A f(x,u(x),\nabla u(x))$, and $\mathrm{D}_v F := \mathrm{D}_v f(x,u(x),\nabla u(x))$, we get

$$\int_D \mathrm{D}_A F : \nabla\mu + \mathrm{D}_v F\cdot\mu\,\mathrm{d}x = \int_{\partial D} F\nu\cdot n\,\mathrm{d}\sigma.$$

Applying the Gauss–Green theorem,

$$\begin{aligned}\int_D \big[-\operatorname{div}\mathrm{D}_A F + \mathrm{D}_v F\big]\cdot\mu\,\mathrm{d}x &= \int_{\partial D}\big[-\mu\cdot\mathrm{D}_A F + \nu F\big]\cdot n\,\mathrm{d}\sigma\\ &= \int_D \operatorname{div}\big[-\mu\cdot\mathrm{D}_A F + \nu F\big]\,\mathrm{d}x\end{aligned}$$

Now, the Euler–Lagrange equation $-\operatorname{div}\mathrm{D}_A F + \mathrm{D}_v F = 0$ immediately implies

$$\int_D \operatorname{div}\big[-\mu\cdot\mathrm{D}_A F + \nu F\big]\,\mathrm{d}x = 0.$$

and varying D we conclude that (2.18) holds.

The corollary follows by considering the invariance $x_\tau := x + \tau\mathrm{e}_k$, $u_\tau(x) := u(x+\tau\mathrm{e}_k)$, $k = 1,\dots,d$. □

Example 2.35 (Brachistochrone problem). In the brachistochrone problem presented in Section 1.1, we were tasked to minimize the functional

$$\mathscr{F}[y] := \int_0^1 \sqrt{\frac{1+(y')^2}{-y}}\,\mathrm{d}x$$

over all $y\colon [0,1]\to\mathbb{R}$ with $y(0)=0$, $y(1)=\bar{y}<0$. The integrand $f(v,a) = \sqrt{-(1+a^2)/v}$ is independent of x, hence from (2.19) we get

$$y'\cdot\mathrm{D}_a f(y,y') - f(y,y') = \text{const} = -\frac{1}{\sqrt{2r}}$$

for some $r>0$ (positive constants lead to inadmissible y). So,

$$\frac{(y')^2}{\sqrt{1+(y')^2}\cdot\sqrt{-y}} - \frac{\sqrt{1+(y')^2}}{\sqrt{-y}} = -\frac{1}{\sqrt{2r}},$$

Comment...

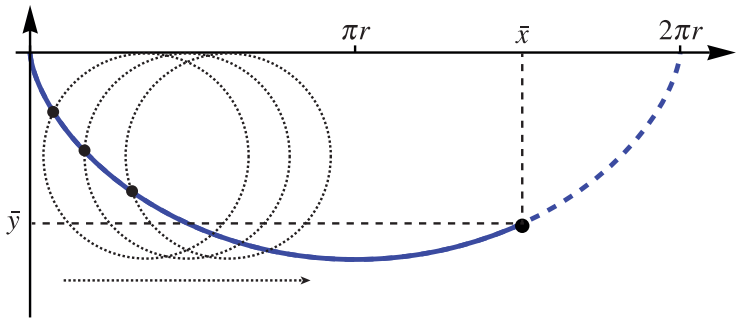

Figure 2.2: The brachistochrone curve (here, $\bar{x} = 1$).

which we transform into the differential equation

$$(y')^2 = -\frac{2r}{y} - 1.$$

The solution of this differential equation is called an *inverted cycloid* with radius $r > 0$, which is the curve traced out by a fixed point on a sphere of radius r (touching the y-axis at the beginning) that rolls to the right on the bottom of the x-axis, see Figure 2.2. It can be written in *parametric form* as

$$\begin{aligned} x(t) &= r(t - \sin t), \\ y(t) &= -r(1 - \cos t), \end{aligned} \qquad t \in \mathbb{R}.$$

The radius $r > 0$ has to be chosen as to satisfy the boundary conditions on *one* cycloid segment. This is the solution of the brachistochrone problem.

Example 2.36. We return to our canonical example, the Dirichlet integral, see Example 2.9. We know from Example 2.25 that critical points u are smooth. We noticed at the beginning of this section that the Dirichlet integral is invariant under the scaling transformation (2.15) with $\lambda = \mathrm{e}^{\tau}$. By the Noether theorem, this yields the (non-obvious) conservation law

$$\operatorname{div}\big[(2\nabla u(x) \cdot x + (d-2)u)\nabla u - |\nabla u|^2 x\big] = 0.$$

Integrating this over $B(0,r) \subset \Omega$, $r > 0$, and using the Gauss–Green theorem, we get

$$(d-2)\int_{B(0,r)} |\nabla u(x)|^2 \,\mathrm{d}x = r\int_{\partial B(0,r)} |\nabla u(x)|^2 - 2\left(\nabla u(x) \cdot \frac{x}{|x|}\right)^2 \mathrm{d}\sigma$$

and then, after some computations,

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{1}{r^{d-2}}\int_{B(0,r)} |\nabla u(x)|^2 \,\mathrm{d}x\right) = \frac{2}{r^{d-2}}\int_{\partial B(0,r)} \left(\nabla u(x) \cdot \frac{x}{|x|}\right)^2 \mathrm{d}\sigma \geq 0.$$

Comment...

This *monotonicity formula* implies that

$$\frac{1}{r^{d-2}} \int_{B(0,r)} |\nabla u(x)|^2 \, \mathrm{d}x \quad \text{is increasing in } r > 0.$$

Any harmonic function ($-\Delta u = 0$) that is defined on all of $\mathbb{R}^d$ satisfies this monotonicity formula. For example, this allows us to draw the conclusion that if $d \geq 3$, then u cannot be compactly supported. The formula also shows that the growth around a singularity at zero (or anywhere else by translation) has to behave "more smoothly" than $|x|^{-2}$ (in fact, we already know that solutions are C^∞). While these are not particularly strong remarks, they serve to illustrate how Noether's theorem restricts the candidates for solutions.

The last example exhibited a conservation law that was not obvious from the Euler–Lagrange equation. While in principle it could have been derived directly, the Noether theorem gave us a *systematic* way to find this conservation law from an invariance.

2.8 Side constraints

In many minimization problems, the class of functions over which we minimize our functional may be restricted to include one or several **side constraints**. To establish existence of a minimizer also in these cases, we first need to extend the Direct Method:

Theorem 2.37 (Direct Method with side constraint). *Let X be a Banach space and let $\mathscr{F}, \mathscr{G} \colon X \to \mathbb{R} \cup \{+\infty\}$. Assume the following:*

(WH1) ***Weak Coercivity of $\mathscr{F}$:*** *For all $\Lambda \in \mathbb{R}$, the sublevel set*

$$\{u \in X \;:\; \mathscr{F}[u] \leq \Lambda\} \qquad \textit{is sequentially weakly relatively compact,}$$

that is, if $\mathscr{F}[u_j] \leq \Lambda$ for a sequence $(u_j) \subset X$ and some $\Lambda \in \mathbb{R}$, then (u_j) has a weakly convergent subsequence.

(WH2) ***Weak lower semicontinuity $\mathscr{F}$:*** *For all sequences $(u_j) \subset X$ with $u_j \rightharpoonup u$ (weak convergence X) it holds that*

$$\mathscr{F}[u] \leq \liminf_{j\to\infty} \mathscr{F}[u_j].$$

(WH3) ***Weak continuity of $\mathscr{G}$:*** *For all sequences $(u_j) \subset X$ with $u_j \rightharpoonup x$ it holds that*

$$\mathscr{G}[u_j] \to \mathscr{G}[u].$$

Assume also that there exists at least one $u_0 \in X$ with $\mathscr{G}[u_0] = 0$. Then, the minimization problem

$$\mathscr{F}[u] \to \min \quad \textit{over all } u \in X \textit{ with } \mathscr{G}[u] = \alpha \in \mathbb{R},$$

has a solution.

Comment...

Proof. The proof is almost exactly the same as the one for the standard Direct Method in Theorem 2.3. However, we need to select the u_j for an infimizing sequence with $\mathscr{G}[u_j] = \alpha$. Then, by (WH3), this property also holds for any weak limit u_* of a subsequence of the u_j's. □

A large class of side constraints can be treated using the following simple result:

Lemma 2.38. *Let $g\colon \Omega \times \mathbb{R}^m \to \mathbb{R}$ be a Carathéodory integrand such that there exists $M > 0$ with*

$$|g(x,v)| \le M(1+|v|^p), \qquad (x,v) \in \Omega \times \mathbb{R}^m.$$

Then, $\mathscr{G}\colon \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m) \to \mathbb{R}$ defined through

$$\mathscr{G}[u] := \int_\Omega g(x,u(x))\,\mathrm{d}x.$$

is weakly continuous.

Proof. Let $u_j \rightharpoonup u$ in $\mathrm{W}^{1,p}$, whereby after selecting a subsequence and employing the Rellich–Kondrachov Theorem A.23 and Lemma A.4, $u_j \to u$ in L^p (strongly) and almost everywhere. By assumption we have

$$\pm g(x,v) + M(1+|v|^p) \ge 0.$$

Thus, applying Fatou's Lemma separately to these two integrands, we get

$$\liminf_{j\to\infty}\left(\pm\mathscr{G}[u_j] + \int_\Omega M(1+|u_j|^p)\,\mathrm{d}x\right) \ge \pm\mathscr{G}[u] + \int_\Omega M(1+|u|^p)\,\mathrm{d}x.$$

Since $\|u_j\|_{\mathrm{L}^p} \to \|u\|_{\mathrm{L}^p}$, we can combine these two assertions to get $\mathscr{G}[u_j] \to \mathscr{G}[u]$. Since this holds for all subsequences, it also follows for our original sequence. □

We will later see more complicated side constraints, in particular constraints involving the determinant of the gradient.

The following result is very important for the calculation of minimizers under side constraints. It effectively says that at a minimizer u_* of $\mathscr{F}$ subject to the side constraint $\mathscr{G}[u_*] = 0$, the derivatives of $\mathscr{F}$ and $\mathscr{G}$ have to be *parallel*, in analogy with the finite-dimensional situation.

Theorem 2.39 (Lagrange multipliers). *Let $f\colon \Omega \times \mathbb{R}^m \times \mathbb{R}^{m\times d} \to \mathbb{R}$, $g\colon \Omega \times \mathbb{R}^m \to \mathbb{R}$ be Carathéodory integrands and satisfy the growth bounds*

$$\begin{aligned} |f(x,v,A)| &\le M(1+|v|^p+|A|^p),\\ |g(x,v)| &\le M(1+|v|^p), \qquad\qquad (x,v,A) \in \Omega \times \mathbb{R}^m \times \mathbb{R}^{m\times d}, \end{aligned}$$

Comment...

for some $M > 0$, $1 \le p < \infty$, and let f, g be continuously differentiable in v and A. Suppose that $u_* \in \mathrm{W}^{1,p}_g(\Omega;\mathbb{R}^m)$, where $g \in \mathrm{W}^{1-1/p,p}(\partial\Omega;\mathbb{R}^m)$, minimizes the functional

$$\mathscr{F}[u] := \int_\Omega f(x,u(x),\nabla u(x))\,\mathrm{d}x, \qquad u \in \mathrm{W}^{1,p}_g(\Omega;\mathbb{R}^m),$$

under the side constraint

$$\mathscr{G}[u] := \int_\Omega g(x,u(x))\,\mathrm{d}x = \alpha \in \mathbb{R}.$$

Assume furthermore that the consistency condition

$$\delta\mathscr{G}[u_*][v] := \int_\Omega \mathrm{D}_v g(x,u_*(x)) \cdot v(x)\,\mathrm{d}x \neq 0 \tag{2.20}$$

holds for at least one $v \in \mathrm{W}^{1,p}_0(\Omega;\mathbb{R}^m)$. Then, there exists a ***Lagrange multiplier*** *$\lambda \in \mathbb{R}$ such that u_* is a weak solution of the system of PDEs*

$$\begin{cases} -\operatorname{div}\big[\mathrm{D}_A f(x,u,\nabla u)\big] + \mathrm{D}_v f(x,u,\nabla u) = \lambda \mathrm{D}_v g(x,u) & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega. \end{cases} \tag{2.21}$$

Proof. The proof is structurally similar to its finite-dimensional analogue.

Let $u_* \in \mathrm{W}^{1,p}_g(\Omega;\mathbb{R}^m)$ be a minimizer of $\mathscr{F}$ under the side constraint $\mathscr{G}[u_*] = 0$. From the consistency condition (2.20) we infer that there exists $w \in \mathrm{W}^{1,p}_0(\Omega;\mathbb{R}^m)$ such that

$$\delta\mathscr{G}[u_*][w] = \int_\Omega \mathrm{D}_v g(x,u_*) \cdot w\,\mathrm{d}x = 1.$$

Now fix any variation $v \in \mathrm{W}^{1,p}_0(\Omega;\mathbb{R}^m)$ and define for $s,t \in \mathbb{R}$,

$$H(s,t) := \mathscr{G}[u_* + sv + tw].$$

It is not difficult to see that H is continuously differentiable in both s and t (this uses the strong continuity of the integral, see Theorem 2.31) and

$$\begin{aligned} \partial_s H(s,t) &= \delta\mathscr{G}[u_* + sv + tw][v], \\ \partial_t H(s,t) &= \delta\mathscr{G}[u_* + sv + tw][w]. \end{aligned}$$

Thus, by definition of w,

$$H(0,0) = \alpha \qquad \text{and} \qquad \partial_t H(0,0) = 1.$$

By the implicit function theorem, there exists a function $\tau\colon \mathbb{R} \to \mathbb{R}$ such that $\tau(0) = 0$ and

$$H(s,\tau(s)) = \alpha \qquad \text{for small } |s|.$$

The chain rule yields for such s,

$$0 = \partial_s[H(s,\tau(s))] = \partial_s H(s,\tau(s)) + \partial_t H(s,\tau(s))\tau'(s),$$

Comment...

whereby

$$\tau'(0) = -\partial_s H(0,0) = -\int_\Omega \mathrm{D}_v g(x,u_*)\cdot v \,\mathrm{d}x. \tag{2.22}$$

Now define for small $|s|$ as above,

$$J(s) := \mathscr{F}[u_* + sv + \tau(s)w].$$

We have

$$\mathscr{G}[u_* + sv + \tau(s)w] = H(s,\tau(s)) = \alpha$$

and the continuously differentiable function J has a minimum at $s = 0$ by the minimization property of u_*. Thus, with the shorthand notations $\mathrm{D}_A f = \mathrm{D}_A f(x,u_*,\nabla u_*)$ and $\mathrm{D}_v f = \mathrm{D}_v f(x,u_*,\nabla u_*)$,

$$0 = J'(0) = \int_\Omega \mathrm{D}_A f : (\nabla v + \tau'(0)\nabla w) + \mathrm{D}_v f\cdot(v + \tau'(0)w)\,\mathrm{d}x.$$

Rearranging and using (2.22), we get

$$\begin{aligned}\int_\Omega \mathrm{D}_A f : \nabla v + \mathrm{D}_v f\cdot v\,\mathrm{d}x &= -\tau'(0)\int_\Omega \mathrm{D}_A f : \nabla w + \mathrm{D}_v f\cdot w\,\mathrm{d}x \\ &= \lambda\int_\Omega \mathrm{D}_v g(x,u_*)\cdot v\,\mathrm{d}x,\end{aligned} \tag{2.23}$$

where we have defined

$$\lambda := \int_\Omega \mathrm{D}_A f : \nabla w + \mathrm{D}_v f\cdot w\,\mathrm{d}x.$$

Since (2.23) shows that u_* is a weak solution of (2.21), the proof is finished. □

Example 2.40 (Stationary Schrödinger equation). When looking for ground states in quantum mechanics as in Section 1.3, we have to minimize

$$\mathscr{E}[\Psi] := \int_{\mathbb{R}^N} \frac{\hbar^2}{2\mu}|\nabla\Psi|^2 + \frac{1}{2}V(x)|\Psi|^2\,\mathrm{d}x$$

over all $\Psi \in \mathrm{W}^{1,2}(\mathbb{R}^N;\mathbb{C})$ under the side constraint

$$\|\Psi\|_{\mathrm{L}^2} = 1.$$

From Theorem 2.37 in conjunction with Lemma 2.38 (slightly extended to also apply to the whole space $\Omega = \mathbb{R}^N$) we see that this problem always has at least one solution. Theorem 2.39 (likewise extended) yields that this Ψ satisfies

$$\left[\frac{-\hbar^2}{2\mu}\Delta + V(x)\right]\Psi(x) = E\Psi(x), \qquad x \in \Omega,$$

for some Lagrange multiplier $E \in \mathbb{R}$, which is precisely the *stationary Schrödinger equation*. One can also show that $E > 0$ is the smallest eigenvalue of the operator $\Psi \mapsto [\frac{-\hbar^2}{2\mu}\Delta + V(x)]\Psi$.

Comment...

Notes and historical remarks

For the u-dependent variational integrals, the growth in the u-variable can be improved up to q-growth, where $q < p/(p-d)$ (the Sobolev embedding exponent for $\mathrm{W}^{1,p}$). Moreover, we can work with the more general growth bounds $|f(x,v,A)| \leq M(g(x)+|v|^q+|A|^p)$, with $g \in \mathrm{L}^1(\Omega;[0,\infty))$ and $q < p/(p-d)$. For reasons of simplicity, we have omitted these generalization here.

The Fundamental Lemma of calculus of variations 2.21 is due to Du Bois-Raymond and is sometimes named after him.

Difference quotients were already considered by Newton, their application to regularity theory is due to Nirenberg in the 1940s and 1950s. Many of the fundamental results of regularity theory (albeit in a non-variational context) can be found in [48]. The book by Giusti [49] contains theory relevant for variational questions. A nice framework for Schauder estimates is [98]. De Giorgi's 1968 counterexample is from [32].

The Lavrentiev phenomenon was discovered in [68], our Example 2.32 is due to Basilio Manià [70]. Tonelli's Regularity Theorem [114] gives regularity for some integral functionals with superlinear growth; also see [52,53] for some recent developments in this direction.

Noether's theorem has many ramifications and can be put into a very general form in Hamiltonian systems and Lie-group theory. The idea is to study groups of symmetries and their actions. For an introduction into this diverse field see [93].

Example 2.36 about the monotonicity formula is from [42].

Side constraints can also take the form of inequalities, which is often the case in *Optimization Theory* or *Mathematical Programming*. This leads to Karush–Kuhn–Tucker conditions or obstacle problems.

A recent, very accesible introduction to the regularity for variational problems is in [13], also see the entertaining survey [77].

Comment...

Chapter 3

Quasiconvexity

We saw in the last chapter that convexity of the integrand implies lower semicontinuity for integral functionals. Moreover, we also mentioned in Proposition 2.11 that in the scalar case $d = 1$ or $m = 1$ convexity is also necessary for weak lower semicontinuity. In the vectorial case ($d, m > 1$), however, it turns out that one can find weakly lower semicontinuous integral functionals whose integrand is non-convex. The following is the most prominent one: Let $p \geq d$ ($\Omega \subset \mathbb{R}^d$ a bounded Lipschitz domain as usual) and define

$$\mathscr{F}[u] := \int_\Omega \det \nabla u(x) \, \mathrm{d}x, \qquad u \in \mathrm{W}_0^{1,p}(\Omega; \mathbb{R}^d).$$

Then, we can argue using the Stokes theorem (this can also be computed in a more elementary way, see Lemma 3.8 below):

$$\mathscr{F}[u] = \int_\Omega \det \nabla u \, \mathrm{d}x = \int_\Omega du^1 \wedge \cdots \wedge du^d = \int_{\partial\Omega} u^1 \wedge du^2 \wedge \cdots \wedge du^d = 0,$$

because $u \in \mathrm{W}_0^{1,d}(\Omega; \mathbb{R}^d)$ is zero near the boundary $\partial\Omega$. Thus, $\mathscr{F}$ is in fact constant on $\mathrm{W}_0^{1,p}(\Omega; \mathbb{R}^d)$, hence trivially weakly lower semicontinuous. However, the determinant function is far from being convex if $d \geq 2$; for instance we can easily convex-combine a matrix with positive determinant from two singular ones. But we can also find examples not involving singular matrices: For

$$A := \begin{pmatrix} -1 & -2 \\ 2 & 1 \end{pmatrix}, \qquad B := \begin{pmatrix} 1 & -2 \\ 2 & -1 \end{pmatrix}, \qquad \frac{1}{2}A + \frac{1}{2}B = \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix},$$

we have $\det A = \det B = 3$, but $\det(A/2 + B/2) = 4$.

Furthermore, convexity of the integrand is not compatible with one of the most fundamental principles of continuum mechanics: Assume that our integrand $f = f(A)$ is **frame-indifferent**, that is,

$$f(RA) = f(A) \qquad \text{for all } A \in \mathbb{R}^{d \times d}, R \in \mathrm{SO}(d),$$

Comment...

where $SO(d)$ is the set of $(d \times d)$-orthogonal matrices with determinant 1, i.e. rotations. Furthermore, assume that every purely compressive or purely expansive deformation costs energy, i.e.

$$f(\alpha \,\mathrm{Id}) > f(\mathrm{Id}) \qquad \text{for all } \alpha \neq 1, \tag{3.1}$$

which is very reasonable in applications. Then, f cannot be convex: Let us for simplicity assume $d = 2$. Set, for a fixed $\gamma \in (0, 2\pi)$,

$$R := \begin{pmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{pmatrix} \in \mathrm{SO}(2).$$

Then, if f was convex, for $M := (R + R^T)/2 = (\cos\gamma)\,\mathrm{Id}$ we would get

$$f((\cos\gamma)\,\mathrm{Id}) = f(M) \leq \frac{1}{2}\big(f(R) + f(R^T)\big) = f(\mathrm{Id}),$$

contradicting (3.1). Sharper arguments are available, but the essential conclusion is the same: convexity is not suitable for many variational problems originating from continuum mechanics.

3.1 Quasiconvexity

We are now ready to replace convexity with a more general notion, which has become one of the cornerstones of the modern calculus of variations. This is the notion of quasiconvexity, introduced by Charles B. Morrey in the 1950s: A locally bounded Borel-measurable function $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ is called **quasiconvex** if

$$h(A) \leq \fint_{B(0,1)} h(A + \nabla\psi(z))\,\mathrm{d}z \tag{3.2}$$

for all $A \in \mathbb{R}^{m\times d}$ and all $\psi \in \mathrm{W}^{1,\infty}_0(B(0,1);\mathbb{R}^m)$. Here, $\fint_{B(0,1)} := \omega_d^{-1}\int_{B(0,1)}$, where $B(0,1)$ is the unit ball in $\mathbb{R}^d$ and its volume is $\omega_d := |B(0,1)|$.

Let us give a physical interpretation of quasiconvexity: Let $d = m = 3$ and assume that

$$\mathscr{F}[y] := \int_{B(0,1)} h(\nabla y(x))\,\mathrm{d}x, \qquad y \in \mathrm{W}^{1,\infty}(B(0,1);\mathbb{R}^3),$$

models the physical energy of an elastically deformed body, whose deformation from the reference configuration is given as $y\colon \Omega = B(0,1) \to \mathbb{R}^3$ (see Section 1.4 for more details on this modeling). A special class of deformations are the *affine* ones, $a(x) = y_0 + Ax$ for some $y_0 \in \mathbb{R}^3$, $A \in \mathbb{R}^{3\times 3}$. Then, quasiconvexity of f entails that

$$\mathscr{F}[a] = \int_{B(0,1)} h(A)\,\mathrm{d}x \leq \int_{B(0,1)} h(A + \nabla\psi(z))\,\mathrm{d}z = \mathscr{F}[a + \psi]$$

for all $\psi \in \mathrm{W}^{1,\infty}_0(B(0,1);\mathbb{R}^3)$. This means that the affine deformation a is always energetically favorable over the *internally distorted* deformation $a + \psi$; this is very often a

Comment...

reasonable assumption for real materials. This argument also holds on any other domain by Lemma 3.2 below.

To justify its name, we also need to convince ourselves that quasiconvexity is indeed a notion of *convexity*: For $A \in \mathbb{R}^{m\times d}$ and $V \in \mathrm{L}^1(B(0,1);\mathbb{R}^{m\times d})$ with $\int_{B(0,1)} V(x)\,\mathrm{d}x = 0$ define the probability measure $\mu \in \mathscr{M}_1(\mathbb{R}^{m\times d})$ via its *action* as follows (recall that $\mathscr{M}(\mathbb{R}^{m\times d}) \cong \mathrm{C}_0(\mathbb{R}^{m\times d})^*$ by the Riesz Representation Theorem A.12):

$$\langle h, \mu\rangle := \fint_{B(0,1)} h(A+V(x))\,\mathrm{d}x \qquad \text{for } h \in \mathrm{C}_0(\mathbb{R}^{m\times d}).$$

This μ is easily seen to be an element of the dual space to $\mathrm{C}_0(\mathbb{R}^{m\times d})$, and in fact μ is a probability measure: For the boundedness we observe $|\langle h,\mu\rangle| \le \|h\|_\infty$, whereas the positivity $\langle h,\mu\rangle \ge 0$ for $h \ge 0$ and the normalization $\langle \mathbb{1},\mu\rangle = 1$ follow at once. The barycenter $[\mu]$ of μ is

$$[\mu] := \langle \mathrm{id},\mu\rangle = A + \fint_{B(0,1)} V(x)\,\mathrm{d}x = A.$$

Therefore, if h is convex, we get from then Jensen inequality, Lemma A.11,

$$h(A) = h([\mu]) \le \langle h,\mu\rangle = \fint_{B(0,1)} h(A+V(x))\,\mathrm{d}x.$$

In particular, (3.2) holds if we set $V(x) := \nabla\psi(x)$ for any $\psi \in \mathrm{W}_0^{1,\infty}(B(0,1);\mathbb{R}^m)$. Thus, we have shown:

Proposition 3.1. *All convex functions $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ are quasiconvex.*

Some basic properties of quasiconvexity are collected in the following lemma.

Lemma 3.2. *(i) In the definition of quasiconvexity we can replace the domain $B(0,1)$ by any bounded Lipschitz domain $\Omega' \subset \mathbb{R}^d$.*

*(ii) If h has **p-growth**, i.e. $|h(A)| \le M(1+|A|^p)$ for some $1 \le p < \infty$, $M > 0$, then in the definition of quasiconvexity we can replace testing with all $\psi \in \mathrm{W}_0^{1,\infty}$ by testing with all $\psi \in \mathrm{W}_0^{1,p}$.*

Proof. To see the first statement, we will prove the following claim: If $\psi \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$, where $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain, satisfies $\psi(x)|_{\partial\Omega} = Ax$ for some $A \in \mathbb{R}^{m\times d}$ (in the sense of trace), then there exists $\tilde\psi \in \mathrm{W}^{1,p}(\Omega';\mathbb{R}^m)$ with $\tilde\psi(x)|_{\partial\Omega'} = Ax$ and, if one of the following integrals exists and is finite, then

$$\fint_\Omega h(\nabla\psi)\,\mathrm{d}x = \fint_{\Omega'} h(\nabla\tilde\psi)\,\mathrm{d}y \qquad \text{for all measurable } h\colon \mathbb{R}^{m\times d} \to \mathbb{R}. \tag{3.3}$$

In particular, $\|\nabla\psi\|_{\mathrm{L}^p} = \|\nabla\tilde\psi\|_{\mathrm{L}^p}$. Clearly, this implies that the definition of quasiconvexity is independent of the domain.

Comment...

To see (3.3), take a Vitali cover of Ω' with re-scaled versions of Ω, see Theorem A.10,

$$\Omega' = Z \cup \bigcup_{k=1}^{N} \Omega(a_k, r_k), \qquad |Z| = 0,$$

with $a_k \in \Omega$, $r_k > 0$, $\Omega(a_k, r_k) := a_k + r_k\Omega$, where $k = 1, \dots, N \in \mathbb{N}$. Then let

$$\tilde{\psi}(y) := \begin{cases} r_k \psi\left(\dfrac{y - a_k}{r_k}\right) + Aa_k & \text{if } y \in \Omega(a_k, r_k), \\ 0 & \text{otherwise,} \end{cases} \qquad y \in \Omega'.$$

It holds that $\tilde{\psi} \in \mathrm{W}^{1,\infty}(\Omega'; \mathbb{R}^m)$. We compute for any measurable $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$,

$$\begin{aligned} \int_{\Omega'} h(\nabla\tilde{\psi})\,\mathrm{d}y &= \sum_k \int_{\Omega(a_k,r_k)} h\left(\nabla\psi\left(\frac{y-a_k}{r_k}\right)\right)\mathrm{d}y \\ &= \sum_k r_k^d \int_\Omega h(\nabla\psi)\,\mathrm{d}x \\ &= \frac{|\Omega'|}{|\Omega|}\int_\Omega h(\nabla\psi)\,\mathrm{d}x \end{aligned}$$

since $\sum_k r_k^d = |\Omega'|/|\Omega|$. This shows (3.3).

The second assertion follows since $\mathrm{W}_0^{1,\infty}(B(0,1);\mathbb{R}^m)$ is dense in $\mathrm{W}_0^{1,p}(B(0,1);\mathbb{R}^m)$ and under a p-growth assumption, the integral functional

$$\psi \mapsto \fint_\Omega h(\nabla\psi)\,\mathrm{d}x, \qquad \psi \in \mathrm{W}_0^{1,p}(B(0,1);\mathbb{R}^m),$$

is well-defined and continuous, the latter for example by Pratt's Theorem A.6, see the proof of Theorem 2.31 for a similar argument. □

An even weaker notion of convexity is the following one: A locally bounded Borel-measurable function $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ is called **rank-one convex** if it is convex along any rank-one line, that is,

$$h\big(\theta A + (1-\theta)B\big) \le \theta h(A) + (1-\theta)h(B)$$

for all $A, B \in \mathbb{R}^{m\times d}$ with $\operatorname{rank}(A-B) \le 1$ and all $\theta \in (0,1)$. In this context, recall that a matrix $M \in \mathbb{R}^{m\times d}$ has rank one if and only if $M = a \otimes b = ab^T$ for some $a \in \mathbb{R}^m$, $b \in \mathbb{R}^d$.

Proposition 3.3. *If $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ is quasiconvex, then it is rank-one convex.*

Proof of Proposition 3.3. Let $A, B \in \mathbb{R}^{m\times d}$ with $B - A = a \otimes n$ for $a \in \mathbb{R}^m$ and $n \in \mathbb{R}^d$ with $|n| = 1$. Denote by Q_n a unit cube ($|Q_n| = 1$) centered at the origin and with one face normal n. By Lemma 3.2 we can require h to satisfy

$$h(M) \le \fint_{Q_n} h(\nabla\psi(z))\,\mathrm{d}z \tag{3.4}$$

Comment...

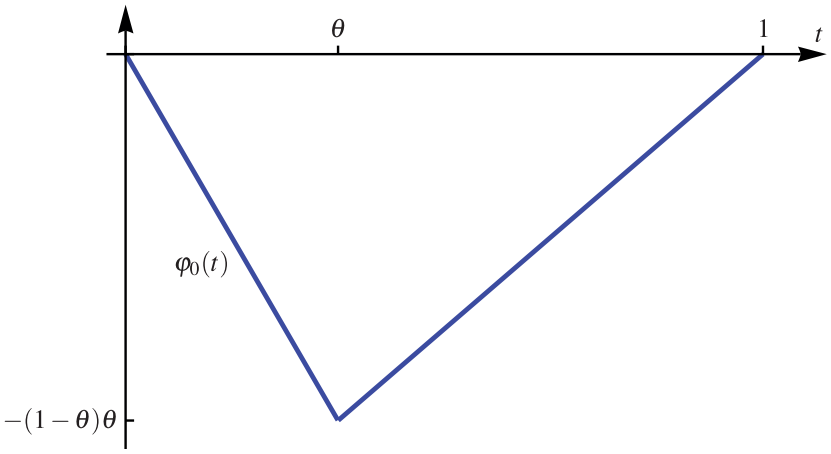

Figure 3.1: The function φ_0.

for all $M \in \mathbb{R}^{m\times d}$ and all $\psi \in \mathrm{W}^{1,\infty}(Q_n;\mathbb{R}^m)$ with $\psi(x)|_{\partial Q} = Mx$ (in the sense of trace).

For fixed $\theta \in (0,1)$, we now let $M := \theta A + (1-\theta)B$ and define the sequence of test functions $\varphi_j \in \mathrm{W}^{1,\infty}(Q_n;\mathbb{R}^m)$ as follows:

$$\varphi_j(x) := Mx + \frac{1}{j}\varphi_0\big(jx\cdot n - \lfloor jx\cdot n\rfloor\big)a, \qquad x \in Q_n,$$

where $\lfloor s\rfloor$ denotes the largest integer below or equal to $s \in \mathbb{R}$, and

$$\varphi_0(t) := \begin{cases} -(1-\theta)t & \text{if } t \in [0,\theta], \\ \theta t - \theta & \text{if } t \in (\theta,1], \end{cases} \qquad t \in [0,1).$$

Such φ_j are called **laminations** in direction n. See Figures 3.1 and 3.2 for φ_0 and φ_j, respectively.

We calculate

$$\nabla\varphi_j(x) = \begin{cases} M - (1-\theta)a\otimes n = A & \text{if } jx\cdot n - \lfloor jx\cdot n\rfloor \in (0,\theta), \\ M + \theta a\otimes n = B & \text{if } jx\cdot n - \lfloor jx\cdot n\rfloor \in (\theta,1). \end{cases}$$

Thus,

$$\lim_{j\to\infty} \fint_{Q_n} h(\nabla\varphi_j(x))\,\mathrm{d}x = \theta h(A) + (1-\theta)h(B).$$

Also notice that $\varphi_j \overset{*}{\rightharpoonup} Mx$ in $\mathrm{W}^{1,\infty}$ since φ_0 is uniformly bounded. We will show below that we may replace the sequence (φ_j) with a sequence $(\psi_j) \subset \mathrm{W}^{1,\infty}(Q_n;\mathbb{R}^m)$ with the additional property that $\psi_j(x)|_{\partial Q} = Mx$, but such that still

$$\lim_{j\to\infty} \fint_{Q_n} h(\nabla\psi_j(x))\,\mathrm{d}x = \theta h(A) + (1-\theta)h(B).$$

Comment...

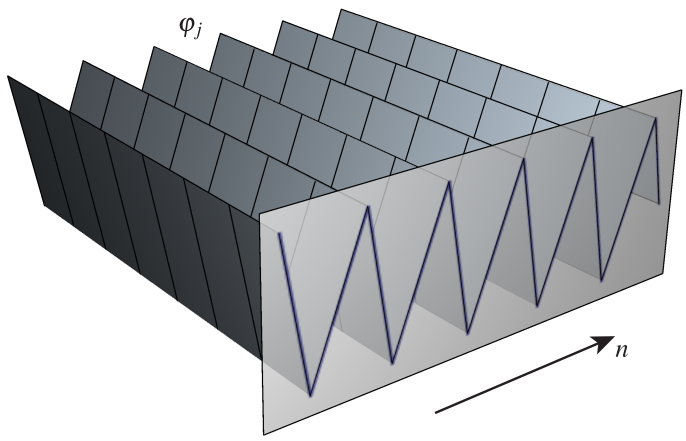

Figure 3.2: The laminate φ_j.

Combining this with (3.4) allows us to conclude

$$h\big(\theta A + (1-\theta)B\big) \leq \theta h(A) + (1-\theta)h(B)$$

and h is indeed rank-one convex.

It remains to construct the sequence (ψ_j) from (φ_j), for which we employ a standard **cut-off construction**: Take a sequence $(\rho_j) \subset \mathrm{C}_c^\infty(Q_n;[0,1])$ of cut-off functions such that for $G_j := \big\{ x \in \Omega \ : \ \rho_j(x) = 1 \big\}$ it holds that $|Q_n \setminus G_j| \to 0$ as $j \to \infty$. Set

$$\psi_{j,k}(x) := \rho_j \varphi_k(x) + (1-\rho_j)Mx, \qquad x \in \Omega,$$

which lies in $\mathrm{W}^{1,\infty}(Q_n;\mathbb{R}^m)$ and satisfies $\psi_{j,k}(x) = Mx$ near ∂Q_n. Also,

$$\nabla \psi_{j,k} = \rho_j \nabla \varphi_k + (1-\rho_j)M + (\varphi_k - a) \otimes \nabla \rho_j.$$

Since $\mathrm{W}^{1,\infty}(Q_n;\mathbb{R}^m)$ embeds compactly into $\mathrm{L}^\infty(Q_n;\mathbb{R}^m)$ by the Rellich-Kondrachov theorem (or the classical Arzéla–Ascoli theorem), we have $\varphi_k \to Mx$ uniformly. Thus, in

$$\|\nabla \psi_{j,k}\|_{\mathrm{L}^\infty} \leq \|\nabla \varphi_k\|_{\mathrm{L}^\infty} + |M| + \|(\varphi_k - a) \otimes \nabla \rho_j\|_{\mathrm{L}^\infty}$$

the last term vanishes as $k \to \infty$. As the $\nabla \varphi_k$ furthermore are uniformly L^∞-bounded, we can for every $j \in \mathbb{N}$ choose $k(j) \in \mathbb{N}$ such that $\|\nabla \psi_{j,k(j)}\|_{\mathrm{L}^\infty}$ is bounded by a constant that is independent of j. Then, as h is assumed to be locally bounded, this implies that there exists $C > 0$ (again independent of j) with

$$\|h(\nabla \varphi_j)\|_{\mathrm{L}^\infty} + \|h(\nabla \psi_{j,k(j)})\|_{\mathrm{L}^\infty} \leq C.$$

Comment...

Hence, for $\psi_j := \psi_{j,k(j)}$, we may estimate

$$\lim_{j\to\infty}\int_{Q_n}|h(\nabla\varphi_j)-h(\nabla\psi_j)|\,\mathrm{d}x = \lim_{j\to\infty}\int_{Q_n\setminus G_j}|h(\nabla\varphi_j)|+|h(\nabla\psi_{j,k(j)})|\,\mathrm{d}x$$
$$\leq \lim_{j\to\infty} C|Q_n\setminus G_j| = 0.$$

This shows that above we may replace (φ_j) by (ψ_j). □

Since for $d=1$ or $m=1$, rank-one convexity obviously is equivalent to convexity, we have for the *scalar* case:

Corollary 3.4. *If $h\colon \mathbb{R}^{m\times d}\to\mathbb{R}$ is quasiconvex and $d=1$ or $m=1$, then h is convex.*

However, quasiconvexity is weaker than classical convexity if $d,m\geq 2$. The determinant function and, more generally, minors are quasiconvex, as will be proved in the next section, but these minors (except for (1×1)-minors) are not convex. The following is a standard example of a quasiconvex function.

Example 3.5 (Alibert–Dacorogna–Marcellini). For $d=m=2$ and $\gamma\in\mathbb{R}$ define

$$h_\gamma(A) := |A|^2\big(|A|^2-2\gamma\det A\big), \qquad A\in\mathbb{R}^{2\times 2}.$$

For this function it is known that

- h_γ is convex if and only if $|\gamma|\leq \dfrac{2\sqrt{2}}{3}\approx 0.94$,
- h_γ is rank-one convex if and only if $|\gamma|\leq \dfrac{2}{\sqrt{3}}\approx 1.15$,
- h_γ is quasiconvex if and only if $|\gamma|\leq\gamma_{\mathrm{QC}}$ for some $\gamma_{\mathrm{QC}}\in\left(1,\dfrac{2}{\sqrt{3}}\right]$.

It is currently unknown whether $\gamma_{\mathrm{QC}}=2/\sqrt{3}$. We do not prove these statements here, see Section 5.3.8 in [29] for the details.

We end this section with the following regularity theorem for rank-one convex (and hence also for quasiconvex) functions:

Proposition 3.6. *If $h\colon\mathbb{R}^{m\times d}\to\mathbb{R}$ is rank-one convex, then it is locally Lipschitz continuous. If additionally h has p-growth, then*

$$|h(A)-h(B)|\leq C(1+|A|^{p-1}+|B|^{p-1})|A-B| \qquad \textit{for all } A,B\in\mathbb{R}^{m\times d} \tag{3.5}$$

and a dimensional constant $C=C(d,m)$. In particular, a rank-one convex h with linear growth is (globally) Lipschitz continuous.

Comment...

Proof. For any $A_0 \in \mathbb{R}^{m\times d}$ and $r > 0$, we will prove the stronger *quantitative bound*

$$\operatorname{lip}(h; B(A_0, r)) \leq \sqrt{\min(d,m)} \cdot \frac{\operatorname{osc}(h; \overline{B(A_0, 6r)})}{3r}, \tag{3.6}$$

where

$$\operatorname{lip}(h; B(A_0, r)) := \sup_{A,B\in B(A_0,r)} \frac{|h(A) - h(B)|}{|A - B|}$$

is the Lipschitz constant of h on the ball $B(A_0, r) \subset \mathbb{R}^{m\times d}$, and

$$\operatorname{osc}(h; B(A_0, r)) := \sup_{A,B\in B(A_0,r)} |h(A) - h(B)|$$

is the **oscillation** of h on $B(A_0, r)$. By the local boundedness of h, which is part of our definition of rank-one convexity, the oscillation is bounded on every ball. Thus the Lipschitz constant is locally finite.

To show (3.6), let $A, B \in B(x_0, r)$ and assume first that $\operatorname{rank}(A - B) \leq 1$. Define the point $C \in \mathbb{R}^{m\times d}$ as the intersection of $\partial B(A_0, 2r)$ with the ray starting at B and going through A. Then, because h is convex along this ray,

$$\frac{|h(A) - h(B)|}{|A - B|} \leq \frac{|h(C) - h(B)|}{|C - B|} \leq \frac{\operatorname{osc}(h; \overline{B(A_0, 2r)})}{r} =: \alpha(2r). \tag{3.7}$$

For general $A, B \in B(A_0, r)$, use the (real) singular value decomposition to write

$$B - A = \sum_{i=1}^{\min(d,m)} \sigma_i P(\mathrm{e}_i \otimes \mathrm{e}_i) Q^T,$$

where $\sigma_i \geq 0$ is the i'th singular value, and $P \in \mathbb{R}^{m\times m}$, $Q \in \mathbb{R}^{d\times d}$ are orthogonal matrices. Set

$$A_k := A + \sum_{i=1}^{k-1} \sigma_i P(\mathrm{e}_i \otimes \mathrm{e}_i) Q^T, \qquad k = 1, \ldots, \min(d,m) + 1,$$

for which we have $A_1 = A$, $A_{\min(d,m)+1} = B$. We estimate (recall that we are employing the Frobenius norm $|M| := \sqrt{\sum_{j,k} (M^j_k)^2} = \sqrt{\sum_i \sigma_i(M)^2}$)

$$|A_k - A_0| \leq |A - A_0| + \sqrt{\sum_{i=1}^{k-1} \sigma_i^2} \leq |A - A_0| + |B - A| < 3r$$

and

$$\sum_{k=1}^{\min(d,m)} |A_k - A_{k+1}|^2 = \sum_{k=1}^{\min(d,m)} \sigma_i^2 = |A - B|^2.$$

Comment...

Applying (3.7) to $A_k, A_{k+1} \in B(A_0, 3r)$, $k = 1, \dots, \min(d,m)$, we get

$$\begin{aligned}
|h(A) - h(B)| &\le \sum_{k=1}^{\min(d,m)} |h(A_k) - h(A_{k+1})| \\
&\le \alpha(6r) \sum_{k=1}^{\min(d,m)} |A_k - A_{k+1}| \\
&\le \alpha(6r)\sqrt{\min(d,m)} \cdot \left[\sum_{k=1}^{\min(d,m)} |A_k - A_{k+1}|^2 \right]^{1/2} \\
&= \alpha(6r)\sqrt{\min(d,m)} \cdot |A - B|.
\end{aligned}$$

This yields (3.6).

If we additionally assume that h has p-growth, then

$$\operatorname{osc}(h; B(0,r)) \le C(1 + r^p),$$

and so, with $r := \max\{|A|, |B|\}$, the estimate (3.5) follows from (3.6). □

Remark 3.7. An improved argument (see Lemma 2.2 in [11]), where one orders the singular values in a favorable way, allows to establish the better estimate

$$\operatorname{lip}(h; B(A_0, r)) \le \sqrt{\min(d,m)} \cdot \frac{\operatorname{osc}(h; \overline{B(A_0, 2r)})}{r}.$$

3.2 Null-Lagrangians

The determinant is quasiconvex, but it is only one representative of a larger class of canonical examples of quasiconvex, but not convex, functions: In this section, we will investigate the properties of minors (subdeterminants) as integrands. Let for $r \in \{1, 2, \dots, \min(d,m)\}$,

$$I \in P(m,r) := \left\{ (i_1, i_2, \dots, i_r) \in \{1, \dots, m\}^r \ : \ i_1 < i_2 < \dots < i_r \right\}$$

and $J \in P(d,r)$ be **ordered multi-indices**. Then, a $(r \times r)$**-minor** $M \colon \mathbb{R}^{m \times d} \to \mathbb{R}$ is a function of the form

$$M(A) = M^I_J(A) := \det\big(A^I_J\big),$$

where A^I_J is the (square) matrix consisting of the I-rows and J-columns of A. The number $r \in \{1, \dots, \min(d,m)\}$ is called the **rank** of the minor M.

The first result shows that all minors are **null-Lagrangians**, which by definition is the class of integrands $h \colon \mathbb{R}^{m \times d}$ such that $\int_\Omega h(\nabla u)\, \mathrm{d}x$ only depends on the boundary values of u.

Lemma 3.8. *Let $M \colon \mathbb{R}^{m \times d} \to \mathbb{R}$ be a $(r \times r)$-minor, $r \in \{1, \dots, \min(d,m)\}$. If $u, v \in \mathrm{W}^{1,p}(\Omega; \mathbb{R}^m)$, $p \ge r$, with $u - v \in \mathrm{W}^{1,p}_0(\Omega; \mathbb{R}^m)$, then*

$$\int_\Omega M(\nabla u)\, \mathrm{d}x = \int_\Omega M(\nabla v)\, \mathrm{d}x.$$

Comment...

Proof. In all of the following, we will assume that u, v are smooth and $\operatorname{supp}(u-v) \Subset \Omega$, which can be achieved by approximation and a cut-off procedure, see Theorem A.24. We also need that taking the minor M of the gradient commutes with strong convergence, i.e. the strong continuity of $u \mapsto M(\nabla u)$ in $\mathrm{W}^{1,p}$ for $p \geq r$; this follows by Hadamard's inequality $|M(A)| \leq |A|^r$ and Pratt's Theorem A.6 (see the proof of Lemma 2.38).

All first-order minors are just the entries of $\nabla u, \nabla v$ and the result follows from

$$\int_\Omega \nabla u \,\mathrm{d}x = \int_{\partial\Omega} u \cdot n \,\mathrm{d}\sigma = \int_{\partial\Omega} v \cdot n \,\mathrm{d}\sigma = \int_\Omega \nabla v \,\mathrm{d}x$$

since $\operatorname{supp}(u-v) \Subset \Omega$.

For higher-order minors, the crucial observation is that minors of gradients can be written as *divergences*, which we will establish below. So, if $M(\nabla u) = \operatorname{div} G(u, \nabla u)$, then, since $\operatorname{supp}(u-v) \Subset \Omega$,

$$\int_\Omega M(\nabla u) \,\mathrm{d}x = \int_{\partial\Omega} G(u, \nabla u) \cdot n \,\mathrm{d}\sigma = \int_{\partial\Omega} G(v, \nabla v) \cdot n \,\mathrm{d}\sigma = \int_\Omega M(\nabla v) \,\mathrm{d}x$$

and the result follows.

We first consider the physically most relevant cases $d = m \in \{2, 3\}$.

For $d = m = 2$ and $u = (u^1, u^2)^T$, the only second-order minor is the Jacobian determinant and we easily see from the fact that second derivatives of smooth functions commute that

$$\begin{aligned}
\det \nabla u &= \partial_1 u^1 \partial_2 u^2 - \partial_2 u^1 \partial_1 u^2 \\
&= \partial_1 (u^1 \partial_2 u^2) - \partial_2 (u^1 \partial_1 u^2) \\
&= \operatorname{div}\big(u^1 \partial_2 u^2, -u^1 \partial_1 u^2\big).
\end{aligned}$$

For $d = m = 3$, consider a second-order minor $M_{\neg l}^{\neg k}(A)$, i.e. the determinant of A after deleting the k'th row and l'th column. Then, analogously to the situation in dimension two, we get, using *cyclic* indices $k, l \in \{1, 2, 3\}$,

$$\begin{aligned}
M_{\neg l}^{\neg k}(\nabla u) &= (-1)^{k+l} \big[\partial_{l+1} u^{k+1} \partial_{l+2} u^{k+2} - \partial_{l+2} u^{k+1} \partial_{l+1} u^{k+2}\big] \\
&= (-1)^{k+l} \big[\partial_{l+1} (u^{k+1} \partial_{l+2} u^{k+2}) - \partial_{l+2} (u^{k+1} \partial_{l+1} u^{k+2})\big].
\end{aligned} \tag{3.8}$$

For the three-dimensional Jacobian determinant, we will show

$$\det \nabla u = \sum_{l=1}^{3} \partial_l \big(u^1 (\operatorname{cof} \nabla u)_l^1\big), \tag{3.9}$$

where we recall that $(\operatorname{cof} A)_l^k = (-1)^{k+l} M_{\neg l}^{\neg k}(A)$. To see this, use the Cramer formula $(\det A) I = A (\operatorname{cof} A)^T$, which holds for any matrix $A \in \mathbb{R}^{n \times n}$, to get

$$\det \nabla u = \sum_{l=1}^{3} \partial_l u^1 \cdot (\operatorname{cof} \nabla u)_l^1.$$

Comment...

Then, our formula (3.9) follows from the **Piola identity**

$$\operatorname{div}\operatorname{cof}\nabla u = 0, \tag{3.10}$$

which can be verified directly from the expression (3.8) for $M_{\neg l}^{\neg k}(\nabla u)$. Hence, (3.9) holds. For general dimensions d,m we use the notation of differential geometry to tame the multilinear algebra involved in the proof. So, let M be a $(r\times r)$-minor. Reordering $x^1,\dots,x^d$ and $u^1,\dots,u^m$, we can assume without loss of generality that M is a *principal* minor, i.e. M is the determinant of the top-left $(r\times r)$-submatrix. Then,

$$\begin{aligned} M(\nabla u)\,dx^1\wedge\cdots\wedge dx^d &= du^1\wedge\cdots\wedge du^r\wedge dx^{r+1}\wedge\cdots\wedge dx^d \\ &= d(u^1\wedge du^2\wedge\cdots\wedge du^r\wedge dx^{r+1}\wedge\cdots\wedge dx^d). \end{aligned}$$

Thus, the general Stokes theorem gives

$$\begin{aligned} \int_\Omega M(\nabla u)\,dx^1\wedge\cdots\wedge dx^d &= \int_\Omega d(u^1\wedge du^2\wedge\cdots\wedge du^r\wedge dx^{r+1}\wedge\cdots\wedge dx^d) \\ &= \int_{\partial\Omega} u^1\wedge du^2\wedge\cdots\wedge du^r\wedge dx^{r+1}\wedge\cdots\wedge dx^d. \end{aligned}$$

Therefore, $\int_\Omega M(\nabla u)\,dx^1\wedge\cdots\wedge dx^d$ only depends on the values of u around $\partial\Omega$. □

Consequently, we also have:

Corollary 3.9. *All $(r\times r)$-minors $M\colon \mathbb{R}^{m\times d}\to\mathbb{R}$ are **quasiaffine**, that is, both M and $-M$ are quasiconvex.*

We will later use this property to define a large and useful class of quasiconvex functions, namely the *polyconvex* ones; this is the topic of the next chapter.

Further, minors enjoy a surprising continuity property:

Lemma 3.10 (Weak continuity of minors). *Let $M\colon \mathbb{R}^{m\times d}\to\mathbb{R}$ be an $(r\times r)$-minor, $r\in\{1,\dots,\min(d,m)\}$, and $(u_j)\subset \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ where $r<p\le\infty$. If*

$$u_j \rightharpoonup u \quad \text{in } \mathrm{W}^{1,p} \qquad (\text{or } u_j \overset{*}{\rightharpoonup} u \text{ in } \mathrm{W}^{1,\infty}).$$

then

$$M(\nabla u_j) \rightharpoonup M(\nabla u) \quad \text{in } \mathrm{L}^{p/r} \qquad (\text{or } M(\nabla u_j) \overset{*}{\rightharpoonup} M(\nabla u) \text{ in } \mathrm{W}^{1,\infty}).$$

Proof. For $d=m\in\{2,3\}$, we use the special structure of minors as divergences, as exhibited in the proof of Lemma 3.8. For a (2×2)-minor $M_{\neg l}^{\neg k}$ in three dimensions (that is, deleting the k'th row and l'th column; in two dimensions there is only one (2×2)-minor, the determinant, but we still use the same notation), we rely on (3.8) to observe that with cyclic indices $k,l\in\{1,2,3\}$,

$$\int_\Omega M_{\neg l}^{\neg k}(\nabla u_j)\psi\,\mathrm{d}x = -(-1)^{k+l}\int_\Omega (u_j^{k+1}\partial_{l+2}u_j^{k+2})\partial_{l+1}\psi - (u_j^{k+1}\partial_{l+1}u_j^{k+2})\partial_{l+2}\psi\,\mathrm{d}x$$

Comment...

for all $\psi \in C_c^\infty(\Omega)$ and then by density also for all $\psi \in L^{p/r}(\Omega)^* \cong L^{p/(p-r)}(\Omega)$. Since $u_j \rightharpoonup u$ in $W^{1,p}$, we have $u_j \to u$ strongly in L^p. The above expression consists of products of one L^p-weakly and one L^p-strongly convergent factor as well as a uniformly bounded function under the integral. Hence the integral is weakly continuous and as $j \to \infty$ converges to

$$\int_\Omega M_{\neg l}^{\neg k}(\nabla u)\psi \,\mathrm{d}x.$$

This finishes the proof for $d = m = 2$.

For $d = m = 3$, we additionally need to consider the determinant. However, as a consequence of the two-dimensional reasoning, $\operatorname{cof}\nabla u_j \rightharpoonup \operatorname{cof}\nabla u$ in $L^{p/2}$. Then, (3.9) implies

$$\int_\Omega \det\nabla u_j\,\psi \,\mathrm{d}x = -\sum_{l=1}^{3}\int_\Omega \left[u_j^1(\operatorname{cof}\nabla u_j)_l^1\right]\partial_l\psi \,\mathrm{d}x$$

and by a similar reasoning as above, this converges to

$$-\sum_{l=1}^{3}\int_\Omega \left[u^1(\operatorname{cof}\nabla u)_l^1\right]\partial_l\psi \,\mathrm{d}x = \int_\Omega \det\nabla u\,\psi \,\mathrm{d}x.$$

In the general case, one proceeds by induction. □

It can also be shown that *any* quasiaffine function can be written as an affine function of all the minors of the argument. This characterization of quasiaffine functions is due to Ball [4], a different proof (also including further characterizing statements) can be found in Theorem 5.20 of [29].

3.3 Young measures

We now introduce a very versatile tool for the study of minimization problems, the *Young measure*, named after its inventor Laurence Chisholm Young. In this chapter, we will only use it as a device to organize the proof of lower semicontinuity for quasiconvex integral functions, see Theorems 3.28, 3.33, but later it will also be important as an object in its own right.

Let us motivate this device through the following central question: Assume that we have a weakly convergent sequence $v_j \rightharpoonup v$ in $L^2(\Omega)$, where $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain, and an integral functional

$$\mathscr{F}[v] := \int_\Omega f(x, v(x))\,\mathrm{d}x$$

with $f\colon \Omega \times \mathbb{R} \to \mathbb{R}$ continuous and bounded, say. Then, we may want to know the limit of $\mathscr{F}[v_j]$ as $j \to \infty$, for instance if (v_j) is a minimizing sequence. In fact, this will be cornerstone of the proof of weak lower semicontinuity for integral functionals with quasiconvex integrands.

Comment...

Equivalently, we could ask for the weak* limit in L^∞ of the sequence of compound functions $g_j(x) := f(x, v_j(x))$. It is easy to see that the weak* limit in general is *not* equal to $x \mapsto f(x, v(x))$. For example, in $\Omega = (0,1)$, consider the sequence

$$v_j(x) := \begin{cases} a & \text{if } jx - \lfloor jx \rfloor \le \theta, \\ b & \text{otherwise,} \end{cases} \qquad x \in \Omega,$$

where $a, b \in \mathbb{R}$, $\theta \in (0,1)$, and $\lfloor s \rfloor$ is the largest integer below or equal to $s \in \mathbb{R}$. Then, if $f(x,a) = \alpha \in \mathbb{R}$ and $f(x,b) = \beta \in \mathbb{R}$ (and smooth and bounded elsewhere), we see $v_j \overset{*}{\rightharpoonup} \theta a + (1-\theta) b$ and

$$f(x, v_j(x)) \overset{*}{\rightharpoonup} h(x) = \theta\alpha + (1-\theta)\beta.$$

The right-hand side, however, is not necessarily $f(x, \theta a + (1-\theta) b)$. Instead, we could write

$$h(x) = \langle f(x,\cdot), \nu_x \rangle = \int f(x,v) \, \mathrm{d}\nu_x(v)$$

with a x-parametrized family of probability measures $\nu_x \in \mathscr{M}_1(\mathbb{R})$, namely

$$\nu_x = \theta \delta_\alpha + (1-\theta)\delta_\beta, \qquad x \in (0,1).$$

This family $(\nu_x)_{x\in\Omega}$ is the "Young measure generated by the sequence (v_j)".

We start with the **fundamental theorem of Young measure theory**:

Theorem 3.11 (Young 1937). *Let $(V_j) \subset \mathrm{L}^p(\Omega; \mathbb{R}^N)$ be a bounded sequence, where $1 \le p \le \infty$. Then, there exists a subsequence (not relabeled) and a family $(\nu_x)_{x\in\Omega} \subset \mathscr{M}_1(\mathbb{R}^N)$ of probability measures, called the* ***(L^p-)Young measure generated by the sequence*** *(V_j), such that the following assertions are true:*

(i) The family (ν_x) is ***weakly* measurable****, i.e., for all Carathéodory integrands $f\colon \Omega \times \mathbb{R}^N \to \mathbb{R}$, the compound function*

$$x \mapsto \langle f(x,\cdot), \nu_x \rangle = \int f(x,A) \, \mathrm{d}\nu_x(A), \qquad x \in \Omega,$$

is Lebesgue-measurable.

(ii) If $1 \le p < \infty$, it holds that

$$\langle\!\langle |\cdot|^p, \nu \rangle\!\rangle := \int_\Omega \int |A|^p \, \mathrm{d}\nu_x(A) \, \mathrm{d}x < \infty,$$

or, if $p = \infty$, there exists a compact set $K \subset \mathbb{R}^N$ such that

$$\operatorname{supp} \nu_x \subset K \quad \text{for a.e. } x \in \Omega.$$

Comment...

(iii) For all Carathéodory integrands $f\colon \Omega \times \mathbb{R}^N \to \mathbb{R}$ such that the family $(f(x,V_j(x)))_j$ is uniformly L^1-bounded and equiintegrable, it holds that

$$f(x,V_j(x)) \rightharpoonup \left(x \mapsto \int f(x,A)\,\mathrm{d}\nu_x(A)\right) \quad \textit{in } \mathrm{L}^1(\Omega). \tag{3.11}$$

For parametrized measures $\nu = (\nu_x)_{x\in\Omega}$ that satisfy (i) and (ii) above, we write $\nu = (\nu_x) \in \mathbf{Y}^p(\Omega;\mathbb{R}^N)$, the latter being called the **set of L^p-Young measures**. In symbols we express the generation of a Young measure ν by as sequence (V_j) as

$$V_j \overset{\mathrm{Y}}{\to} \nu.$$

Furthermore, the convergence (3.11) can equivalently be expressed as (absorb a test function for weak convergence into f)

$$\int_\Omega f(x,V_j(x))\,\mathrm{d}x \;\to\; \int_\Omega \int f(x,A)\,\mathrm{d}\nu_x(A)\,\mathrm{d}x =: \langle\!\langle f,\nu \rangle\!\rangle.$$

We call $\langle\!\langle f,\nu \rangle\!\rangle$ the **duality pairing** between f and ν. In this context, recall from the Dunford–Pettis Theorem A.17 that $(f(x,V_j(x)))_j$ is equiintegrable if and only it is weakly relatively compact in $\mathrm{L}^1(\Omega)$.

Remark 3.12. Some assumptions in the fundamental theorem can be weakened: For the existence of a Young measure, we only need to require

$$\lim_{K\uparrow\infty} \sup_j |\{|V_j| \geq K\}| = 0,$$

which for example follows from $\sup_j \int_\Omega |V_j|^r\,\mathrm{d}x < \infty$ for some $r > 0$. Of course, in this case (ii) needs to be suitably adapted.

Proof. We associate with each V_j an **elementary Young measure** $\delta[V_j] \in \mathbf{Y}^p(\Omega;\mathbb{R}^N)$ as follows:

$$(\delta[V_j])_x := \delta_{V_j(x)} \qquad \text{for a.e. } x \in \Omega. \tag{3.12}$$

Below we will show the following **Young measure compactness principle**:

Lemma 3.13. *Let $1 \leq p \leq \infty$ and let $(\nu^{(j)})_j \subset \mathbf{Y}^p(\Omega;\mathbb{R}^N)$ be a sequence of L^p-Young measures such that, if $1 \leq p < \infty$,*

$$\sup_j \langle\!\langle |\cdot|^p, \nu^{(j)} \rangle\!\rangle = \sup_j \int_\Omega \int |A|^p\,\mathrm{d}\nu_x^{(j)}(A)\,\mathrm{d}x < \infty \tag{3.13}$$

or, if $p = \infty$, $\operatorname{supp} \nu_x^{(j)} \subset K$ for all $j \in \mathbb{N}$, almost every $x \in \Omega$, and a compact set $K \subset \mathbb{R}^{m\times d}$. Then, there exists a subsequence of the (ν_j) (not relabeled) and $\nu \in \mathbf{Y}^p(\Omega;\mathbb{R}^N)$ such that

$$\lim_{j\to\infty} \langle\!\langle f,\nu^{(j)} \rangle\!\rangle = \langle\!\langle f,\nu \rangle\!\rangle \tag{3.14}$$

Comment...

for all Carathéodory $f\colon \Omega\times\mathbb{R}^N\to\mathbb{R}$ *for which the sequence of functions* $x\mapsto\langle f(x,\bullet),\nu_x^{(j)}\rangle$ *is uniformly* L^1*-bounded and the equiintegrability condition*

$$\sup_j \left\langle\!\!\left\langle f(x,A)\mathbb{1}_{\{|f(x,A)|\geq m\}},\nu^{(j)}\right\rangle\!\!\right\rangle\to 0 \qquad \textit{as} \qquad m\to\infty. \tag{3.15}$$

holds. Moreover,

$$\left\langle\!\!\left\langle |\bullet|^p,\nu\right\rangle\!\!\right\rangle \leq \liminf_{j\to\infty}\left\langle\!\!\left\langle |\bullet|^p,\nu^{(j)}\right\rangle\!\!\right\rangle.$$

In the theorem, our assumption of L^p-boundedness of (V_j) corresponds to (3.13), so (i)–(III) follow immediately from the compactness principle applied to our sequence $(\delta[V_j])_j$ of elementary Young measures from (3.12) above once we realize that for the sequence of Young measures $(\delta[V_j])_j$ the condition (3.15) corresponds precisely to the equiintegrability of $(f(x,V_j(x)))_j$. □

Proof of Lemma 3.13. Define the measures

$$\mu^{(j)} := \mathscr{L}^d_x \llcorner \Omega\otimes\nu_x^{(j)},$$

which is just a shorthand notation for the Radon measures $\mu^{(j)}\in\mathscr{M}(\Omega\times\mathbb{R}^N)\cong\mathrm{C}_0(\Omega\times\mathbb{R}^N)^*$ defined through their action

$$\left\langle f,\mu^{(j)}\right\rangle = \int_\Omega\int f(x,A)\,\mathrm{d}\nu_x^{(j)}(A)\,\mathrm{d}x \qquad \text{for all } f\in\mathrm{C}_0(\Omega\times\mathbb{R}^N).$$

So, for the $\nu^{(j)}=\delta[V_j]$ from (3.12), we recover the familiar form

$$\left\langle f,\mu^{(j)}\right\rangle = \int_\Omega f(x,V_j(x))\,\mathrm{d}x \qquad \text{for all } f\in\mathrm{C}_0(\Omega\times\mathbb{R}^N),$$

but in the compactness principle we might be in a more general situation.

Clearly, every so-defined $\mu^{(j)}$ is a *positive* measure and

$$\left|\left\langle f,\mu^{(j)}\right\rangle\right| \leq |\Omega|\cdot\|f\|_\infty,$$

whereby the $(\mu^{(j)})$ constitute a uniformly bounded sequence in the dual space $\mathrm{C}_0(\Omega\times\mathbb{R}^N)^*$. Thus, by the Sequential Banach–Alaoglu Theorem A.16, we can select a (non-relabeled) subsequence such that with a $\mu\in\mathrm{C}_0(\Omega\times\mathbb{R}^N)^*$ it holds that

$$\left\langle f,\mu^{(j)}\right\rangle\to\left\langle f,\mu\right\rangle \qquad \text{for all } f\in\mathrm{C}_0(\Omega\times\mathbb{R}^N). \tag{3.16}$$

Next we will show that μ can be written as $\mu=\mathscr{L}^d_x\llcorner\Omega\otimes\nu_x$ for a weakly* measurable parametrized family $\nu=(\nu_x)_{x\in\Omega}\subset\mathscr{M}_1(\mathbb{R}^N)$ of probability measures.

Theorem 3.14 (Disintegration). *Let* $\mu\in\mathscr{M}(\Omega\times\mathbb{R}^N)$ *be a finite Radon measure. Then, there exists a weakly* measurable family* $(\nu_x)_{x\in\Omega}\subset\mathscr{M}_1(\mathbb{R}^N)$ *of probability measures such that with* $\kappa(A):=\mu(A\times\mathbb{R}^N)$*,*

$$\mu=\kappa(\mathrm{d}x)\otimes\nu_x,$$

that is,

$$\int f\,\mathrm{d}\mu = \int_\Omega\int f(x,A)\,\mathrm{d}\nu_x(A)\,\mathrm{d}\kappa(x) \qquad \textit{for all } f\in\mathrm{C}_0(\Omega\times\mathbb{R}^N).$$

Comment...

A proof of this measure-theoretic result can be found in Theorem 2.28 of [1].

In our situation, we have for $f(x,v) := \varphi(x)$ with arbitrary $\varphi \in \mathrm{C}_0(\Omega)$ that

$$\int \varphi \, \mathrm{d}\mu = \lim_{j\to\infty} \int \varphi \, \mathrm{d}\mu^{(j)} = \int_\Omega \varphi(x) \, \mathrm{d}x,$$

and so $\kappa = \mathscr{L}^d \llcorner \Omega$. Hence, the disintegration theorem immediately yields the weak* measurability of (ν_x) and thus $\lim_{j\to\infty} \langle\!\langle f, \nu^{(j)} \rangle\!\rangle = \langle\!\langle f, \nu \rangle\!\rangle$ for $f \in \mathrm{C}_0(\Omega \times \mathbb{R}^N)$, which is (3.14).

Next, we show (3.14) for the case of f merely Carathéodory and bounded and such that there exists a compact set $K \subset \mathbb{R}^N$ with $\operatorname{supp} f \subset \Omega \times K$. We invoke the Scorza Dragoni Theorem 2.5 to get an increasing sequence of compact sets $S_k \Subset \Omega$ $(k \in \mathbb{N})$ with $|\Omega \setminus S_k| \downarrow 0$ such that $f|_{S_k \times \mathbb{R}^N}$ is continuous. Let $f_k \in \mathrm{C}(\Omega \times \mathbb{R}^N)$ be an extension of $f|_{S_k \times \mathbb{R}^N}$ to all of $\Omega \times \mathbb{R}^N$ with the property that the f_k are uniformly bounded. Indeed, since K is bounded and $|f_k(x,A)| \le M(1+|A|^p)$, f is bounded and hence we may require uniform boundedness of the f_k.

Now, since $f_k \in \mathrm{C}_0(\Omega \times \mathbb{R}^N)$, (3.16) implies

$$\big\langle f_k(x,\cdot), \nu_x^{(j)} \big\rangle \rightharpoonup \big\langle f_k(x,\cdot), \nu_x \big\rangle \quad \text{in } \mathrm{L}^1 \text{ (as functions of } x) \qquad \text{as } j \to \infty.$$

Thus, we also get

$$\big\langle f(x,\cdot), \nu_x^{(j)} \big\rangle \rightharpoonup \big\langle f(x,\cdot), \nu_x \big\rangle \quad \text{in } \mathrm{L}^1(S_k) \text{ (as functions of } x) \qquad \text{as } j \to \infty.$$

On the other hand,

$$\int_\Omega \big| \big\langle f(x,\cdot), \nu_x^{(j)} \big\rangle - \mathbb{1}_{S_k} \big\langle f(x,\cdot), \nu_x^{(j)} \big\rangle \big| \, \mathrm{d}x \le \int_{\Omega \setminus S_k} \big| \big\langle f(x,\cdot), \nu_x^{(j)} \big\rangle \big| \, \mathrm{d}x$$

and this converges to zero as $k \to \infty$, uniformly in j, by the uniform boundedness of f_k; the same estimate holds with ν in place of $\nu^{(j)}$. Therefore, we may conclude

$$\big\langle f(x,\cdot), \nu_x^{(j)} \big\rangle \rightharpoonup \big\langle f(x,\cdot), \nu_x \big\rangle \quad \text{in } \mathrm{L}^1(\Omega) \text{ (as functions of } x) \qquad \text{as } j \to \infty,$$

which directly implies (3.14) (for f with $\operatorname{supp} f \subset \Omega \times K$).

Finally, to remove the restriction of compact support in A, we remark that it suffices to show the assertion under the additional constraint that $f \ge 0$ by considering the positive and negative part separately. Then, in case f positive, choose for any $m \in \mathbb{N}$ a function $\rho_m \in \mathrm{C}_c^\infty(\mathbb{R}^N; [0,1])$ with $\rho_m = 1$ on $B(0,m)$ and $\operatorname{supp} \rho_m \subset B(0,2m)$. Set

$$f^m(x,A) := \rho_m(|A|^{p/2}) \rho_m(f(x,A)) f(x,A).$$

Comment...

Then,

$$
\begin{aligned}
E_{j,m} &:= \int_\Omega \int f(x,\bullet) - f^m(x,\bullet)\,\mathrm{d}\nu_x^{(j)}\,\mathrm{d}x \\
&\le \int_\Omega \int \big[1-\rho_m(|A|^{p/2})\rho_m(f(x,A))\big] f(x,A)\,\mathrm{d}\nu_x^{(j)}(A)\,\mathrm{d}x \\
&\le \iint_{\{|A|\,\ge\, m^{2/p} \text{ or } f(x,A)\,\ge\, m\}} f(x,A)\,\mathrm{d}\nu_x^{(j)}(A)\,\mathrm{d}x \\
&\le M\int_\Omega \int_{\{|A|\ge m^{2/p}\}} m\,\mathrm{d}\nu_x^{(j)}(A)\,\mathrm{d}x + \int_{\{f(x,A)\ge m\}} f(x,A)\,\mathrm{d}\nu_x^{(j)}(A)\,\mathrm{d}x \\
&\le \frac{M}{m}\sup_j \langle\!\langle |A|^p, \nu^{(j)}\rangle\!\rangle + \sup_j \langle\!\langle f(x,A)\mathbb{1}_{\{|f(x,A)|\ge m\}}, \nu^{(j)}\rangle\!\rangle.
\end{aligned}
$$

Both of these terms converge to zero as $m\to\infty$ by the assumption from the compactness principle, in particular (3.13) and the equiintegrability condition (3.15).

As the Young measure convergence holds for f^m by the previous proof step, we get

$$
\begin{aligned}
\limsup_{j\to\infty}\big(\langle\!\langle f,\nu^{(j)}\rangle\!\rangle - \langle\!\langle f^m,\nu\rangle\!\rangle\big) &= \limsup_{j\to\infty}\big(\langle\!\langle f^m,\nu^{(j)}\rangle\!\rangle - \langle\!\langle f^m,\nu\rangle\!\rangle\big) + E_{j,m}\big) \\
&\le \sup_j E_{j,m}.
\end{aligned}
$$

Then, since $f^m \uparrow f$ and $\sup_j E_{j,m}$ vanishes as $m\to\infty$ and $f\ge 0$, we get that indeed

$$
\langle\!\langle f,\nu^{(j)}\rangle\!\rangle \to \langle\!\langle f,\nu\rangle\!\rangle \qquad \text{as} \qquad j\to\infty.
$$

The last assertion to be shown in the compactness principle is, for $1\le p<\infty$,

$$
\langle\!\langle |\bullet|^p,\nu\rangle\!\rangle \le \liminf_{j\to\infty}\langle\!\langle |\bullet|^p,\nu^{(j)}\rangle\!\rangle.
$$

For $K\in\mathbb{N}$ define $|A|_K := \min\{|A|,K\}\le |A|$. Then, by the Young measure representation for equiintegrable integrands,

$$
\liminf_{j\to\infty}\langle\!\langle |\bullet|^p,\nu^{(j)}\rangle\!\rangle \ge \lim_{j\to\infty}\langle\!\langle |\bullet|_K^p,\nu^{(j)}\rangle\!\rangle = \langle\!\langle |\bullet|_K^p,\nu\rangle\!\rangle \qquad \text{for all } K\in\mathbb{N}.
$$

We conclude by letting $K\uparrow\infty$ and using the monotone convergence lemma. □

Remark 3.15. The Fundamental Theorem can also be proved in a more functional analytic way as follows: Let $\mathrm{L}^\infty_{w*}(\Omega;\mathscr{M}(\mathbb{R}^N))$ be the set of essentially bounded weakly* measurable functions defined on Ω with values in the Radon measures $\mathscr{M}(\mathbb{R}^N)$. It turns out (see for example [6] or [33, 34]) that $\mathrm{L}^\infty_{w*}(\Omega;\mathscr{M}(\mathbb{R}^N))$ is the dual space to $\mathrm{L}^1(\Omega;\mathrm{C}_0(\mathbb{R}^N))$. One can show that the maps $\nu^{(j)} = (x\mapsto \nu_x^{(j)})$ form a uniformly bounded set in $\mathrm{L}^\infty_{w*}(\Omega;\mathscr{M}(\mathbb{R}^N))$ and by the Sequential Banach–Alaoglu Theorem A.16 we can again conclude the existence of a weak* limit point ν of the $\nu^{(j)}$'s. The extended representation of limits for Carathéodory f follows as before.

Comment...

It is known that *every* weakly* measurable parametrized measure $(\nu_x)_{x\in\Omega} \subset \mathscr{M}_1(\mathbb{R}^N)$ with $(x \mapsto \langle |\bullet|^p, \nu_x\rangle \in \mathrm{L}^1(\Omega)$ $(1 \le p < \infty)$ can be generated by a sequence $(u_j) \subset \mathrm{L}^p(\Omega;\mathbb{R}^N)$ (one starts first by approximating Dirac masses and then uses the approximation of a general measure by linear combinations of Dirac masses, a spatial gluing argument completes the reasoning). Thus, in the definition of $\mathbf{Y}^p(\Omega;\mathbb{R}^N)$ we did not need to include (III) from the Fundamental Theorem.

A Young measure $\nu = (\nu_x) \in \mathbf{Y}^p(\Omega;\mathbb{R}^N)$ is called **homogeneous** if ν_x is almost everywhere constant in $x \in \Omega$; we then simply write ν in place of ν_x.

Before we come to further properties of Young measures, let us consider a few examples:

Example 3.16. In $\Omega := (0,1)$ define $u := \mathbb{1}_{(0,1/2)} - \mathbb{1}_{(1/2,1)}$ and extend this function periodically to all of $\mathbb{R}$. Then, the functions $u_j(x) := u(jx)$ for $j \in \mathbb{N}$ (see Figure 3.3) generate the homogeneous Young measure $\nu \in \mathbf{Y}^\infty((0,1);\mathbb{R})$ with

$$\nu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1}.$$

Indeed, for $\varphi \otimes h \in \mathrm{C}_0((0,1)\times\mathbb{R})$ we have that φ is uniformly continuous, say $|\varphi(x) - \varphi(y)| \le \omega(|x-y|)$ with a modulus of continuity $\omega\colon [0,\infty) \to [0,\infty)$ (that is, ω is continuous, increasing, and $\omega(0) = 0$). Then,

$$\begin{aligned}
&\lim_{j\to\infty} \int_0^1 \varphi(x) h(u_j(x))\,\mathrm{d}x \\
&\quad = \lim_{j\to\infty} \sum_{k=0}^{j-1} \left(\int_{k/j}^{(k+1)/j} \varphi(k/j) h(u_j(x))\,\mathrm{d}x + \frac{\omega(1/j)\|h\|_\infty}{j} \right) \\
&\quad = \lim_{j\to\infty} \sum_{k=0}^{j-1} \frac{1}{j} \varphi(k/j) \int_0^1 h(u(y))\,\mathrm{d}y \\
&\quad = \int_0^1 \varphi(x)\,\mathrm{d}x \cdot \left(\frac{1}{2}h(-1) + \frac{1}{2}h(+1) \right)
\end{aligned}$$

since the Riemann sums converge to the integral of φ. The last line implies the assertion.

Example 3.17. Take $\Omega := (0,1)$ again and let $u_j(x) = \sin(2\pi jx)$ for $j \in \mathbb{N}$ (see Figure 3.4). The sequence (u_j) generates the homogeneous Young measure $\nu \in \mathbf{Y}^\infty((0,1);\mathbb{R})$ with

$$\nu = \frac{1}{\pi\sqrt{1-y^2}}\,\mathscr{L}^1_y \llcorner (-1,1).$$

as should be plausible from the oscillating sequence (there is more mass close to the horizontal axis than farther away).

Comment...

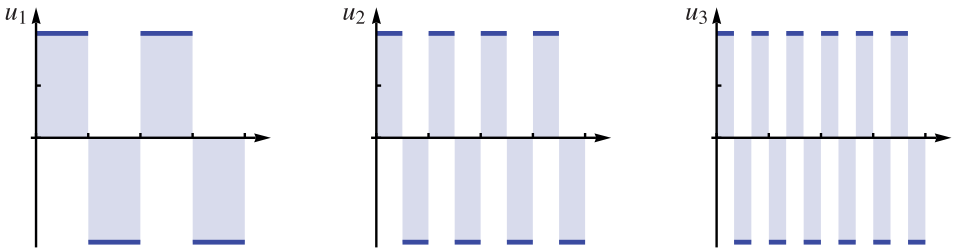

Figure 3.3: An oscillating sequence.

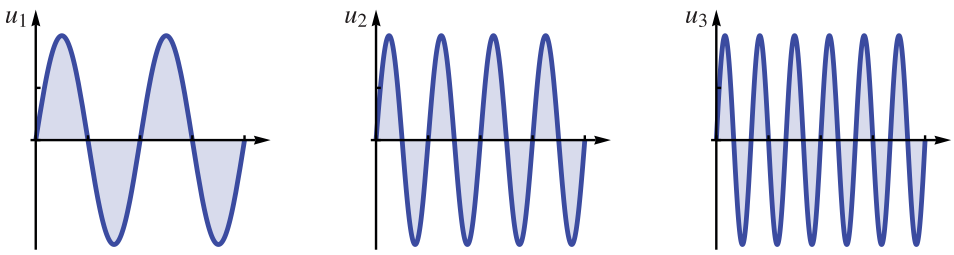

Figure 3.4: Another oscillating sequence.

Example 3.18. Take a bounded Lipschitz domain $\Omega \subset \mathbb{R}^2$, let $A, B \in \mathbb{R}^{2\times 2}$ with $B - A = a \otimes n$ (this is equivalent to $\operatorname{rank}(A-B) \leq 1$), where $a, n \in \mathbb{R}^2$, and for $\theta \in (0,1)$ define

$$u(x) := Ax + \left(\int_0^{x\cdot n} \chi(t)\,\mathrm{d}t \right) a, \qquad x \in \mathbb{R}^2,$$

where $\chi := \mathbb{1}_{\bigcup_{z\in\mathbb{Z}}[z,z+1-\theta)}$. If we let $u_j(x) := u(jx)/j$, $x \in \Omega$, then (∇u_j) (restricted to $(0,1)^2$) generates the homogeneous Young measure $\nu \in \mathbf{Y}^\infty((0,1)^2;\mathbb{R}^2)$ with

$$\nu = \theta\delta_A + (1-\theta)\delta_B.$$

This example can also be extended to include multiple scales, cf. [85].

To determine a generated Young measure it suffices to verify (3.11) for a countable family of integrands:

Lemma 3.19. *There exists a countable family $\{\varphi_k \otimes h_l\}_{k,l} \subset \mathrm{C}_0(\Omega) \times \mathrm{C}_0(\mathbb{R}^N)$ with the following property: If $(V_j) \subset \mathrm{L}^p(\Omega;\mathbb{R}^N)$ is uniformly norm-bounded and $\nu \in \mathbf{Y}^p(\Omega;\mathbb{R}^N)$ such that*

$$\lim_{j\to\infty} \int_\Omega \varphi_k(x) h_l(V_j(x))\,\mathrm{d}x = \int_\Omega \varphi_k(x) \int h_l\,\mathrm{d}\nu_x\,\mathrm{d}x \qquad \textit{for all } k,l \in \mathbb{N},$$

then $V_j \overset{\mathrm{Y}}{\to} \nu$.

Comment...

Proof. Let $\{\varphi_k\}_k$ and $\{h_l\}_l$ be countable dense subsets of $\mathrm{C}_0(\Omega)$ and $\mathrm{C}_0(\mathbb{R}^N)$, respectively. The assertion of the lemma is immediatei once we realize that the set of linear combinations of the functions $f_{k,l} := \varphi_k \otimes h_l$ (that is, $f_{k,l}(x,A) := \varphi_k(x)h_l(A)$) is dense in $\mathrm{C}_0(\Omega \times \mathbb{R}^N)$ and testing with such functions determines Young measure convergence, see the proof of the Fundamental Theorem 3.11. □

Before we move to use Young measures to investigate integral functionals, let us first investigate how Young measure generation interacts with other notions of convergence.

Lemma 3.20. *Let $(V_j) \subset \mathrm{L}^p(\Omega;\mathbb{R}^N)$, $1 < p \leq \infty$, be a bounded sequence generating the Young measure $\nu \in \mathbf{Y}^p(\Omega;\mathbb{R}^N)$. Then,*

$$V_j \rightharpoonup V \quad \text{in } \mathrm{L}^p(\Omega;\mathbb{R}^N) \qquad \text{for} \qquad V(x) := [\nu_x].$$

Proof. Bounded sequences in L^p with $p > 1$ are weakly relatively compact and it suffices to identify the limit of any weakly convergent subsequence. From the Dunford–Pettis Theorem A.17 it follows that any such sequence is in fact equiintegrable (in L^1). Now simply apply assertion (III) of the Fundamental Theorem for the integrand $f(x,A) := A$ (or, more pedantically, $f^i_j(A) := A^i_j$ for $i = 1,\dots,d$, $m = 1,\dots,m$). □

The preceding lemma does not hold for $p = 1$. One example is $V_j := j\mathbb{1}_{(0,1/j)}$, which *concentrates*.

Another important feature of Young measures is that they allow us to read off whether or not our sequence converges in measure:

Lemma 3.21. *Let $(\nu_x) \subset \mathscr{M}_1(\mathbb{R}^N)$ be the Young measure generated by a bounded sequence $(V_j) \subset \mathrm{L}^p(\Omega;\mathbb{R}^N)$, $1 \leq p \leq \infty$. Let furthermore $K \subset \mathbb{R}^N$ be compact. Then,*

$$\operatorname{dist}(V_j(x),K) \to 0 \text{ in measure} \qquad \Longleftrightarrow \qquad \operatorname{supp} \nu_x \subset K \text{ for a.e. } x \in \Omega.$$

Moreover,

$$V_j \to V \text{ in measure} \qquad \Longleftrightarrow \qquad \nu_x = \delta_{V(x)} \text{ for a.e. } x \in \Omega.$$

Proof. We only show the second assertion, the proof of the first one is similar. Take the bounded, positive Carathéodory integrand

$$f(x,A) := \frac{|A - V(x)|}{1 + |A - V(x)|}, \qquad x \in \Omega, A \in \mathbb{R}^N.$$

Then, for all $\delta \in (0,1)$, the Markov inequality and the Fundamental Theorem on Young measures yield

$$\begin{aligned}\limsup_{j\to\infty} \big|\{\, x \in \Omega \;:\; f(x,V_j(x)) \geq \delta \,\}\big| &\leq \lim_{j\to\infty} \frac{1}{\delta} \int_\Omega f(x,V_j(x))\,\mathrm{d}x \\ &= \frac{1}{\delta} \int_\Omega \int f(x,\cdot)\,\mathrm{d}\nu_x\,\mathrm{d}x.\end{aligned}$$

Comment...

On the other hand,

$$\int_\Omega \int f(x,\cdot)\,\mathrm{d}\nu_x\,\mathrm{d}x = \lim_{j\to\infty}\int_\Omega f(x,V_j(x))\,\mathrm{d}x$$
$$\leq \delta|\Omega| + \limsup_{j\to\infty}\big|\big\{\,x\in\Omega\ :\ f(x,V_j(x))\geq\delta\,\big\}\big|.$$

The preceding estimates show that $f(x,V_j(x))$ converges to zero in measure if and only if $\langle f(x,\cdot),\nu_x\rangle = 0$ for $\mathscr{L}^d$-almost every $x\in\Omega$, which is the case if and only if $\nu_x = \delta_{V(x)}$ almost everywhere. We conclude by observing that $f(x,V_j(x))$ converges to zero in measure if and only if V_j converges to V in measure. □

3.4 Gradient Young measures

The most important subclass of Young measures for our purposes is that of Young measures that can be generated by a sequence of *gradients*: Let $\nu\in\mathbf{Y}^p(\Omega;\mathbb{R}^{m\times d})$ (use $\mathbb{R}^N=\mathbb{R}^{m\times d}$ in the theory of the last section). We say that ν is a **p-gradient Young measure**, in symbols $\nu\in\mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$, if there exists a bounded sequence $(u_j)\subset \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ such that $\nabla u_j \overset{\mathrm{Y}}{\to} \nu$, i.e. the sequence (∇u_j) generates ν. Note that it is *not* necessary that every sequence that generates ν is a sequence of gradients (which, in fact, is *never* the case). We have already considered examples of gradient Young measures in the previous section, see in particular Example 3.18.

Is it the case that every Young measure is a gradient Young measure? The answer is no, as can be easily seen: Let $V\in(\mathrm{L}^p\cap\mathrm{C}^\infty)(\Omega;\mathbb{R}^{m\times d})$ with $\operatorname{curl}V(x)\neq 0$ for some $x\in\Omega$. Then, $\nu_x := \delta_{V(x)}$ cannot be a gradient Young measure since for such measures it must hold that $[\nu]$ is a gradient itself: if there existed a sequence $(u_j)\subset\mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ with $\nabla u_j\overset{\mathrm{Y}}{\to}\nu$, then $\nabla u_j \rightharpoonup V$ by Lemma 3.20, and it would follow that $\operatorname{curl}V=0$. Consequently, for every $\nu\in\mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$ there must exist $u\in\mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ with $[\nu]=\nabla u$. In this case we call any such u an **underlying deformation** of ν.

We first show the following technical, but immensely useful, result about gradient Young measures. Recall that we assume Ω to be open, bounded, and to have a Lipschitz boundary (to make the trace well-defined).

Lemma 3.22. *Let $\nu\in\mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$, $1<p\leq\infty$, and let $u\in\mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ be an underlying deformation of ν, i.e. $[\nu]=\nabla u$. Then, there exists a sequence $(u_j)\subset\mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ such that*

$$u_j|_{\partial\Omega} = u|_{\partial\Omega} \qquad \textit{and} \qquad \nabla u_j\overset{\mathrm{Y}}{\to}\nu.$$

Furthermore, if $1<p<\infty$, in addition we can ensure that (∇u_j) is p-equiintegrable.

Proof. Step 1. Since Ω is assumed to have a Lipschitz boundary, we can extend a generating sequence (∇v_j) for ν to all of $\mathbb{R}^d$, so we assume $(v_j)\subset\mathrm{W}^{1,p}(\mathbb{R}^d;\mathbb{R}^m)$ with $\sup_j\|v_j\|_{\mathrm{W}^{1,p}}\leq C<\infty$.

Comment...

In the following, we need the maximal function $Mf\colon \mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$ of $f\colon \mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$, see Section A.5. With this tool at hand, consider the sequence

$$V_j := M\big(|v_j| + |\nabla v_j|\big), \qquad j \in \mathbb{N}.$$

By Theorem A.29, (V_j) is uniformly bounded in $\mathrm{L}^p(\Omega)$ and we may select a subsequence (not relabeled) such that V_j generates a Young measure $\mu \in \mathbf{Y}^p(\Omega;\mathbb{R}^{m\times d})$.

Next, define for $h \in \mathbb{N}$ the (nonlinear) *truncation operator* τ_h,

$$\tau_h v := \begin{cases} v & \text{if } |v| \le h, \\ h\frac{v}{|v|} & \text{if } |v| > h, \end{cases} \qquad v \in \mathbb{R},$$

and let $(\varphi_l)_l \subset \mathrm{C}^\infty(\Omega)$ be a countable family that is dense in $\mathrm{C}(\Omega)$.

Step 2. Let $1 < p < \infty$. Then, for fixed $h \in \mathbb{N}$, the sequence $(\tau_h V_j)_j$ is uniformly bounded in $\mathrm{L}^\infty(\Omega)$ and so, by Young measure representation,

$$\begin{aligned}\lim_{h\to\infty}\lim_{j\to\infty}\int_\Omega \varphi_l(x)|\tau_h V_j(x)|^p \,\mathrm{d}x &= \lim_{h\to\infty}\int_\Omega \varphi_l(x)\int |\tau_h v|^p \,\mathrm{d}\mu_x(v)\,\mathrm{d}x \\ &= \int_\Omega \varphi_l(x)\int |\cdot|^p \,\mathrm{d}\mu_x\,\mathrm{d}x,\end{aligned}$$

where in the last step we used the monotone convergence lemma. Thus, for some diagonal sequence $W_k := \tau_k V_{j(k)}$, we have that

$$\lim_{k\to\infty}\int_\Omega \varphi_l(x)\,|W_k(x)|^p \,\mathrm{d}x = \int_\Omega \varphi_l(x)\int |\cdot|^p \,\mathrm{d}\mu_x\,\mathrm{d}x \qquad \text{for all } l \in \mathbb{N}.$$

Thus, $|W_k|^p$ converges weakly to $x \mapsto \langle |\cdot|^p, \mu_x\rangle$ in L^1 and $\{|W_k|^p\}_k$ is an equiintegrable family by the Dunford–Pettis Theorem. We can see also the equiintegrability in a more elementary way as follows: For a Borel subset $A \subset \Omega$ take $\varphi \in \mathrm{C}^\infty(\Omega)$ with $\mathbb{1}_A \le \varphi$ and estimate

$$\limsup_{k\to\infty}\int_A |W_k(x)|^p \,\mathrm{d}x \le \lim_{k\to\infty}\int_\Omega \varphi(x)\,|W_k(x)|^p \,\mathrm{d}x = \int_\Omega \varphi(x)\int |v|^p \,\mathrm{d}\mu_x(v)\,\mathrm{d}x$$

by the L^1-convergence of $|W_k|^p$ to $x \mapsto \langle |\cdot|^p, \mu_x\rangle$. The last integral tends to

$$\int_A \int |\cdot|^p \,\mathrm{d}\mu_x\,\mathrm{d}x$$

as $\varphi \downarrow \mathbb{1}_A$, which vanishes if $|A| \to 0$. Thus the family $\{W_k\}_k$ is p-equiintegrable.

Theorem A.29 implies that $v_k = v_{j(k)}$ is Ck-Lipschitz on the set $G_k := \{W_k \le k\}$ and by the Kirszbraun Theorem A.27, we may extend v_k to a function $w_k\colon \mathbb{R}^d \to \mathbb{R}^m$ that is *globally* Lipschitz continuous with the same Lipschitz constant Ck and $w_k = v_k$ in G_k. We have

$$|\nabla w_k| \le CW_k \quad \text{in } \Omega$$

and thus $\{|\nabla w_k|\}_k$ inherits the p-equiintegrable from $\{W_k\}_k$.

Comment...

Moreover, by the Markov inequality,

$$|\Omega \setminus G_k| \leq \frac{\|W_k\|_{\mathrm{L}^p}^p}{k^p} \to 0 \qquad \text{as } k \to \infty.$$

Therefore, for all $\varphi \in \mathrm{C}_0(\Omega)$ and $h \in \mathrm{C}_0(\mathbb{R}^m)$,

$$\int_\Omega |\varphi h(\nabla w_k) - \varphi h(\nabla v_k)| \,\mathrm{d}x \leq \|\varphi \otimes h\|_\infty \cdot |\Omega \setminus G_k| \;\; \to \;\; 0.$$

Thus, since all such φ, h determine Young measure convergence (see the proof of the Fundamental Theorem 3.11), we have shown that (∇w_k) generates the same Young measure ν as (∇v_j) and additionally the family $\{\nabla w_k\}_k$ is p-equiintegrable.

Step 3. Since $\mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ embeds compactly into the space $\mathrm{L}^p(\Omega;\mathbb{R}^m)$ by the Rellich-Kondrachov Theorem A.23, we have $w_k \to u$ in L^p. Let moreover $(\rho_j) \subset \mathrm{C}_c^\infty(\Omega;[0,1])$ be a sequence of cut-off functions with the property that for the sets $G_j := \big\{ x \in \Omega \;:\; \rho_j(x) = 1 \big\}$ it holds that $|\Omega \setminus G_j| \to 0$ as $j \to \infty$. For

$$u_{j,k} := \rho_j w_k + (1-\rho_j)u \quad \in \mathrm{W}_u^{1,p}(\Omega;\mathbb{R}^m)$$

we observe $\nabla u_{j,k} = \rho_j \nabla w_k + (1-\rho_j)\nabla u + (w_k - u) \otimes \nabla \rho_j$.

We only consider the case $p < \infty$ in the following; the proof for the case $p = \infty$ is in fact easier. For every Carathéodory function $f\colon \Omega \times \mathbb{R}^{m\times d} \to \mathbb{R}$ with $|f(x,A)| \leq M(1+|A|^p)$ we have

$$\begin{aligned}
&\limsup_{k\to\infty} \int_\Omega \big| f(x,\nabla w_k(x)) - f(x,\nabla u_{j,k}(x)) \big| \,\mathrm{d}x \\
&\quad \leq \limsup_{k\to\infty} CM \int_{\Omega\setminus G_j} 2 + 2|\nabla w_k|^p + |\nabla u|^p + |w_k - u|^p \cdot |\nabla \rho_j|^p \,\mathrm{d}x \\
&\quad \leq \limsup_{k\to\infty} CM \int_{\Omega\setminus G_j} 2 + 2|\nabla w_k|^p + |\nabla u|^p \,\mathrm{d}x,
\end{aligned}$$

where $C > 0$ is a constant. However, by assumption, the last integrand is equiintegrable and hence

$$\lim_{j\to\infty} \limsup_{k\to\infty} \int_\Omega \big| f(x,\nabla w_k(x)) - f(x,\nabla u_{j,k}(x)) \big| \,\mathrm{d}x = 0.$$

We can now select a diagonal sequence $u_j = u_{j,k(j)}$ such that (∇u_j) generates ν and satisfies all requirements from the statement of the lemma. Note that the equiintegrability is not affected by the cut-off procedure. □

3.5 Homogeneous gradient Young measures

We next discuss some properties of **homogeneous gradient Young measures**, that is, $\nu \in \mathbf{GY}^p(B(0,1);\mathbb{R}^{m\times d})$ and ν_x is a.e. constant in $x \in B(0,1)$. We simply write ν for any ν_x and $[\nu] = [\nu_x]$.

According to the following lemma the domain in the definition of homogeneous gradient Young measures can be chosen as any bounded Lipschitz domain $D \subset \mathbb{R}^d$ (notice that for homogeneous Young measures the averaged integral in (3.17) can be omitted).

Comment...

Lemma 3.23 (Averaging). *Let $\nu = (\nu_x) \in \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$, where $1 \le p \le \infty$, such that $[\nu] = \nabla u$ for some $u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ with affine boundary values. Then, for any bounded Lipschitz domain $D \subset \mathbb{R}^d$ there exists a (unique) homogeneous gradient Young measure $\overline{\nu} \in \mathbf{GY}^p(D;\mathbb{R}^{m\times d})$, i.e. $\overline{\nu} = \overline{\nu}_y$ is constant in $y \in D$, such that*

$$\int h \,\mathrm{d}\overline{\nu} = \fint_\Omega \int h \,\mathrm{d}\nu_x \,\mathrm{d}x \tag{3.17}$$

for all continuous $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ with p-growth (just continuity if $p = \infty$). This result remains valid if $\Omega = Q^d = (-1/2,1/2)^d$, the d-dimensional unit cube and u has periodic boundary values.

Proof. We only explicitly treat the case that $1 \le p < \infty$, the case $p = \infty$ is in fact easier.

Let $(u_j) \subset \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ with $u_j(x)|_{\partial\Omega} = Mx$ (in the sense of trace) for a fixed matrix $M \in \mathbb{R}^{m\times d}$ such that $\nabla u_j \overset{\mathrm{Y}}{\to} \nu$, in particular $\sup_j \|\nabla u_j\|_{\mathrm{L}^p} < \infty$. Now choose for every $j \in \mathbb{N}$ a disjoint Vitali cover of D, see Theorem A.10,

$$D = Z \cup \bigcup_{k=1}^{N(j)} \Omega(a_k^{(j)}, r_k^{(j)}), \qquad |Z| = 0,$$

with $a_k^{(j)} \in D$, $r_k^{(j)} \le 1/j$ $(k = 1,\dots,N(j))$, and $\Omega(a,r) := a + r\Omega$. Then define

$$v_j(y) := \begin{cases} r_k^{(j)} u_j\left(\dfrac{y - a_k^{(j)}}{r_k^{(j)}}\right) + Ma_k^{(j)} & \text{if } y \in \Omega(a_k^{(j)}, r_k^{(j)}), \\ 0 & \text{otherwise,} \end{cases} \qquad y \in D.$$

We have $v_j \in \mathrm{W}^{1,p}(D;\mathbb{R}^m)$ (it is easy to see that there are no jumps over the gluing boundaries) and

$$\nabla v_j(y) = \nabla u_j\left(\frac{y - a_k^{(j)}}{r_k^{(j)}}\right) \qquad \text{if } y \in \Omega(a_k^{(j)}, r_k^{(j)}).$$

We can then use a change of variables to compute for all $\varphi \in \mathrm{C}_0(D)$ and $h \in \mathrm{C}(\mathbb{R}^{m\times d})$ with p-growth,

$$\begin{aligned} \int_D \varphi(y) h(\nabla v_j(y)) \,\mathrm{d}y &= \sum_{k=1}^{N(j)} \int_{\Omega(a_k^{(j)}, r_k^{(j)})} \varphi(y)\, h\left(\nabla u_j\left(\frac{y - a_k^{(j)}}{r_k^{(j)}}\right)\right) \mathrm{d}y \\ &= \sum_{k=1}^{N(j)} (r_k^{(j)})^d \varphi(a_k^{(j)}) \int_\Omega h(\nabla u_j(x)) \,\mathrm{d}x + \mathrm{O}\left(\frac{1}{j}\right)|D| \end{aligned}$$

where we also used that φ is uniformly continuous. Letting $j \to \infty$ and using that (Riemann sums)

$$\lim_{j\to\infty} \sum_{k=1}^{N(j)} (r_k^{(j)})^d \varphi(a_k^{(j)}) = \frac{1}{|\Omega|} \int_D \varphi(x) \,\mathrm{d}x,$$

Comment...

we arrive at

$$\lim_{j\to\infty}\int_D \varphi(y)h(\nabla v_j(y))\,\mathrm{d}y = \int_D \varphi(x)\,\mathrm{d}x \cdot \fint_\Omega \int h(A)\,\mathrm{d}\nu_x\,\mathrm{d}x. \tag{3.18}$$

For $\varphi = \mathbb{1}$ and $h(A) := |A|^p$, this gives

$$\sup_j \|\nabla v_j\|_{\mathrm{L}^p}^p = \sup_j \|\nabla u_j\|_{\mathrm{L}^p}^p < \infty.$$

Thus, there exists $\overline{\nu} \in \mathbf{GY}^p(D;\mathbb{R}^m)$ such that $\nabla v_j \overset{\mathrm{Y}}{\to} \overline{\nu}$.

Using Lemma 3.22 we may moreover assume that (∇u_j), and hence also (∇v_j), is p-equiintegrable. Then, (3.18) implies

$$\int_D \int \varphi(y)h(A)\,\mathrm{d}\overline{\nu}_y(A)\,\mathrm{d}y = \int_D \varphi(x)\,\mathrm{d}x \cdot \fint_\Omega \int h(A)\,\mathrm{d}\nu_x(A)\,\mathrm{d}x$$

for all φ, h as above. This implies in particular that $(\overline{\nu}_y) = \overline{\nu}$ is homogeneous. For $\varphi = \mathbb{1}$, we get (3.17).

The additional claim about $\Omega = Q^d$ and an underlying deformation u with periodic boundary values follows analogously, since also in this case we can "glue" generating functions on a (special) covering. □

Applying the preceding homogenization lemma to a gradient Young measure consisting only of Dirac masses, we get the following result:

Lemma 3.24 (Riemann–Lebesgue lemma). *Let $u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ have affine boundary values, where $1 \le p \le \infty$. Then, there exists a homogeneous gradient Young measure $\overline{\delta[\nabla u]} \in \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$ ($\Omega \subset \mathbb{R}^d$ any bounded Lipschitz domain) such that*

$$\int h\,\mathrm{d}\overline{\delta[\nabla u]} = \fint_\Omega h(\nabla u(x))\,\mathrm{d}x \tag{3.19}$$

for all continuous $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ with p-growth (just continuity if $p = \infty$). Furthermore, this result remains valid if $\Omega = Q^d = (-1/2,1/2)^d$, the d-dimensional unit cube and u has periodic boundary values.

The following result is now easy to prove:

Lemma 3.25 (Jensen-type inequality). *Let $1 < p \le \infty$ and let $\nu \in \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$ be a homogeneous gradient Young measure. Then, for all quasiconvex functions $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ with p-growth it holds that*

$$h([\nu]) \le \int h\,\mathrm{d}\nu. \tag{3.20}$$

Notice that the conclusion of this lemma is trivial (and also holds for general Young measures, not just the gradient ones) if h is *convex*. Thus (3.20) expresses the generalized convexity aspect of quasiconvexity.

Comment...

Proof. Set $A_0 := [\nu]$ and let $(u_j) \subset \mathrm{W}^{1,p}(B(0,1);\mathbb{R}^m)$ with $\nabla u_j \xrightarrow{\mathrm{Y}} \nu$, $u_j|_{\partial B(0,1)}(x) = A_0 x$ for $x \in \partial B(0,1)$ (in the sense of trace) and (∇u_j) p-equiintegrable, the latter two conditions being realizable by Lemma 3.22. Then, from the definition of quasiconvexity, we get

$$h(A_0) \leq \fint_\Omega h(\nabla u_j)\,\mathrm{d}x.$$

Passing to the Young measure limit as $j \to \infty$ on the right-hand side, for which we note that $h(\nabla u_j)$ is equiintegrable by the growth assumption on h, we arrive at

$$h(A_0) \leq \fint_\Omega \int h\,\mathrm{d}\nu\,\mathrm{d}x = \int h\,\mathrm{d}\nu,$$

which already is the sought inequality. □

The last result will be of crucial importance in proving weak lower semicontinuity in the next section. It is remarkable that the converse also holds, i.e. the validity of (3.20) for *all* quasiconvex h with p-growth (no condition if $p = \infty$) precisely characterizes the class of (homogeneous) gradient Young measures in the class of all (homogeneous) Young measures; this is the content of the Kinderlehrer–Pedregal Theorem 5.12, which we will meet in Chapter 5.

Corollary 3.26. *Let $1 < p \leq \infty$ and let $\nu \in \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$ be a homogeneous gradient Young measure. Then, for all quasiaffine functions $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ with p-growth it holds that*

$$h([\nu]) = \int h\,\mathrm{d}\nu.$$

Proof. Apply the preceding lemma to h and $-h$. □

In particular, the preceding corollary applies to the determinant and, more generally, minors, see Corollary 3.9.

3.6 Lower semicontinuity

After our detour through Young measure theory, we now return to the central subject of this chapter, namely to minimization problems of the form

$$\begin{cases} \mathscr{F}[u] := \displaystyle\int_\Omega f(x,\nabla u(x))\,\mathrm{d}x \quad \to \quad \min, \\ \text{over all } u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m) \text{ with } u|_{\partial\Omega} = g, \end{cases}$$

where $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain, $1 < p < \infty$, the integrand $f\colon \Omega \times \mathbb{R}^m \times \mathbb{R}^{m\times d} \to \mathbb{R}$ has p-growth,

$$|f(x,A)| \leq M(1+|A|^p) \qquad \text{for some } M > 0,$$

Comment...

and $g \in \mathrm{W}^{1-1/p,p}(\partial\Omega;\mathbb{R}^m)$ specifies the boundary values. In Chapter 2 we solved this problem via the Direct Method of the calculus of variations, a coercivity result, and, crucially, the Lower Semicontinuity Theorem 2.7. In this section, we recycle the Direct Method and the coercivity result, but extend lower semicontinuity to *quasi*convex integrands; some motivation for this was given at the beginning of the chapter.

Let us first consider how we could approach the proof of lower semicontinuity (it should be clear that the proof via Mazur's Lemma that we used for the convex lower semicontinuity theorem, does not extend). Assume we have a sequence $(u_j) \subset \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ such that $u_j \rightharpoonup u$ in $\mathrm{W}^{1,p}$ and we want to show lower semicontinuity of our functional $\mathscr{F}$. If we assume that our (norm-bounded) sequence ∇u_j also generates the gradient Young measure $\nu \in \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$, which is true up to a subsequence, and that the sequence of integrands $(f(x,\nabla u_j(x)))_j$ is equiintegrable, then we already have a limit:

$$\mathscr{F}[u_j] \to \int_\Omega \int f(x,A)\,\mathrm{d}\nu_x(A)\,\mathrm{d}x.$$

This is useful, because it now suffices to show the inequality

$$\int f(x,\bullet)\,\mathrm{d}\nu_x \geq f(x,\nabla u(x))\,\mathrm{d}x$$

for almost every $x \in \Omega$, to conclude lower semicontinuity. Now, this inequality is close to the Jensen-type inequality (3.20) for gradient Young measures from Lemma 3.25, and we already explained how this is related to (quasi)convexity. The only difference in comparison to (3.20) is that here we need to "localize" in $x \in \Omega$ and make ν_x a gradient Young measure in its own right. This is accomplished via the fundamental **blow-up technique** (also called "localization"):

Lemma 3.27 (Blow-up technique). *Let $1 \leq p < \infty$ and let $\nu = (\nu_x) \in \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$ be a gradient Young measure. Then, for almost every $x_0 \in \Omega$, $\nu_{x_0} \in \mathbf{GY}^p(B(0,1);\mathbb{R}^{m\times d})$ is a homogeneous gradient Young measure in its own right.*

Proof. Take a countable collection $\{\varphi_k \otimes h_l\}_{k,l}$ as in Lemma 3.19. Let $x_0 \in \Omega$ be a Lebesgue point of all the functions $x \mapsto \langle h_l, \nu_x\rangle$ ($l \in \mathbb{N}$), that is,

$$\lim_{r\downarrow 0} \int_{B(0,1)} \left|\langle h_l, \nu_{x_0+ry}\rangle - \langle h_l, \nu_{x_0}\rangle\right| \mathrm{d}y = 0.$$

By Theorem A.9, almost every $x \in \Omega$ has this property. Then, at such a Lebesgue point x_0, set

$$v_j^{(r)}(y) := \frac{u_j(x_0+ry) - [u_j]_{B(x_0,r)}}{r}, \qquad y \in B(0,1),$$

where $[u]_{B(x_0,r)} := \fint_{B(x_0,r)} u\,\mathrm{d}x$.

For φ_k, h_l we get

$$\begin{aligned}\int_{B(0,1)} \varphi_k(y) h_l(\nabla v_j^{(r)}(y))\,\mathrm{d}y &= \int_{B(0,1)} \varphi_k(y) h_l(\nabla u_j(x_0+ry))\,\mathrm{d}y \\ &= \frac{1}{r^d}\int_{B(x_0,r)} \varphi_k\Big(\frac{x-x_0}{r}\Big) h_l(\nabla u_j(x))\,\mathrm{d}x\end{aligned}$$

Comment...

after a change of variables. Letting $j \to \infty$ and then $r \downarrow 0$, we get

$$\begin{aligned}\lim_{r\downarrow 0}\lim_{j\to\infty}\int_{B(0,1)} \varphi_k(y)h_l(\nabla v_j^{(r)}(y))\,\mathrm{d}y &= \lim_{r\downarrow 0}\frac{1}{r^d}\int_{B(x_0,r)} \varphi_k\Big(\frac{x-x_0}{r}\Big)\big\langle h_l, \nu_x\big\rangle\,\mathrm{d}x \\ &= \lim_{r\downarrow 0}\int_{B(0,1)} \varphi_k(y)\big\langle h_l, \nu_{x_0+ry}\big\rangle\,\mathrm{d}x \\ &= \int_{B(0,1)} \varphi_k(y)\big\langle h_l, \nu_{x_0}\big\rangle\,\mathrm{d}y,\end{aligned}$$

the last convergence following from the Lebesgue point property of x_0. Moreover,

$$\int_{B(0,1)} |\nabla v_j^{(r)}|^p\,\mathrm{d}y = \int_{B(0,1)} |\nabla u_j(x_0+ry)|^p\,\mathrm{d}y = \frac{1}{r^d}\int_{B(x_0,r)} |\nabla u_j(x)|^p\,\mathrm{d}x$$

and the last integral is uniformly bounded in j (for fixed r). Denote by $\lambda \in \mathscr{M}(\overline{\Omega};\mathbb{R}^m)$ the weak* limit of the measures $|\nabla u_j|^p\mathscr{L}^d \llcorner \Omega$, which exists after taking a subsequence. If we require of x_0 additionally that

$$\limsup_{r\downarrow 0}\frac{\lambda(B(x_0,r))}{r^d} < \infty,$$

which hold at $\mathscr{L}^d$-almost every $x_0 \in \Omega$, then

$$\limsup_{r\downarrow 0}\lim_{j\to\infty}\int_{B(0,1)} |\nabla v_j^{(r)}|^p\,\mathrm{d}y < \infty.$$

Since also $[v_j^{(r)}]_{B(0,1)} = 0$, the Poincaré inequality from Theorem A.21 yields that there exists a diagonal sequence $w_j := v_{J(j)}^{R(j)}$ that is uniformly bounded in $\mathrm{W}^{1,p}(B(0,1);\mathbb{R}^m)$ and such that for all $k,l \in \mathbb{N}$,

$$\lim_{j\to\infty}\int_{B(0,1)} \varphi_k(y)h_l(\nabla w_j(y))\,\mathrm{d}y = \int_{B(0,1)} \varphi_k(y)\big\langle h_l, \nu_{x_0}\big\rangle\,\mathrm{d}y.$$

Therefore, $\nabla w_j \overset{\mathrm{Y}}{\to} \nu_{x_0}$ by Lemma 3.19, where we now understand ν_{x_0} as a homogeneous (gradient) Young measure on $B(0,1)$. □

Our main weak lower semicontinuity is then a straightforward application of the theory developed so far. The first result of this type is due to Morrey from 1952 (for integral functionals defined on $\mathrm{W}^{1,\infty}$ and under additional technical assumptions), but our Young measure approach allows us to prove a fairly general result, which was first established by Acerbi–Fusco (using different methods, however).

Theorem 3.28 (Acerbi–Fusco 1984). *Let $1 < p < \infty$ and let $f\colon \Omega \times \mathbb{R}^{m\times d} \to [0,\infty)$ be a Carathéodory integrand with p-growth and such that $f(x,\cdot)$ is quasiconvex for every fixed $x \in \Omega$. Then, the corresponding functional $\mathscr{F}$ is weakly lower semicontinuous on $\mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$.*

Comment...

We first need the following result about Young measures:

Proposition 3.29. *Let $(V_j) \subset \mathrm{L}^p(\Omega;\mathbb{R}^N)$, $1 \le p \le \infty$, be a bounded sequence generating the Young measure $\nu \in \mathbf{Y}^p(\Omega;\mathbb{R}^N)$, and let $f\colon \Omega \times \mathbb{R}^N \to [0,\infty)$ be any Carathéodory integrand (not necessarily satisfying the equiintegrability property in (iii) of the Fundamental Theorem). Then, it holds that*

$$\liminf_{j\to\infty} \int_\Omega f(x,V_j(x))\,\mathrm{d}x \ge \langle\!\langle f,\nu \rangle\!\rangle.$$

Proof of Theorem 3.28. For $M \in \mathbb{N}$ define $f_M(x,A) := \min(f(x,A),M)$. Then, (III) from the Fundamental Theorem is applicable with f_M and we get

$$\int_\Omega f_M(x,V_j(x))\,\mathrm{d}x \to \langle\!\langle f_M,\nu \rangle\!\rangle = \int_\Omega \int f_M(x,A)\,\mathrm{d}\nu_x(A)\,\mathrm{d}x.$$

Since $f \ge f_M$, we have

$$\liminf_{j\to\infty} \int_\Omega f(x,V_j(x))\,\mathrm{d}x \;\ge\; \langle\!\langle f_M,\nu \rangle\!\rangle.$$

We conclude by letting $M \uparrow \infty$ and using the monotone convergence lemma. □

Proof. Let $(\nabla u_j) \subset \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ with $u_j \rightharpoonup u$ in $\mathrm{W}^{1,p}$. Assume that (∇u_j) generates the gradient Young measure $\nu = (\nu_x) \in \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$, for which it holds that $[\nu] = \nabla u$. This is only true up to a (non-relabeled) subsequence, but if we can establish the lower semicontinuity for every such subsequence it follows that the result also holds for the original sequence.

From Proposition 3.29 we get

$$\liminf_{j\to\infty} \int_\Omega f(x,\nabla u_j(x))\,\mathrm{d}x \ge \langle\!\langle f,\nu \rangle\!\rangle = \int_\Omega \int f(x,A)\,\mathrm{d}\nu_x(A)\,\mathrm{d}x.$$

Now, for almost every $x \in \Omega$, we can consider ν_x as a homogeneous Young measure in $\mathbf{GY}^p(B(0,1);\mathbb{R}^{m\times d})$ by the Blow-up Lemma 3.27. Thus, the Jensen-type inequality from Lemma 3.25 reads

$$\int f(x,A)\,\mathrm{d}\nu_x(A) \ge f(x,\nabla u(x)) \qquad \text{for a.e. } x \in \Omega.$$

Combining, we arrive at

$$\liminf_{j\to\infty} \mathscr{F}[u_j] \ge \mathscr{F}[u],$$

which is what we wanted to show. □

At this point is is worthwhile to reflect on the role of Young measures in the proof of the preceding result, namely that they allowed us to split the argument into two parts: First, we passed to the limit in the functional via the Young measure. Second, we established a Jensen-type inequality, which then yielded the lower semicontinuity inequality. It is remarkable that the Young measure preserves exactly the "right" amount of information to serve as an intermediate object.

We can now sum up and prove existence to our minimization problem:

Comment...

Theorem 3.30. *Let $f\colon \Omega\times\mathbb{R}^{m\times d}\to[0,\infty)$ be a Carathéodory integrand such that f has p-growth, satisfies the* ***p-coercivity estimate***

$$\mu|A|^p-\mu^{-1}\le f(x,A)\qquad \text{for some } \mu>0,$$

and is quasiconvex in its second argument. Then, the associated functional $\mathscr{F}$ has a minimizer over the space $\mathrm{W}^{1,p}_g(\Omega;\mathbb{R}^m)$, where $g\in\mathrm{W}^{1-1/p,p}(\partial\Omega;\mathbb{R}^m)$.

Proof. This follows directly by combining the Direct Method from Theorem 2.3 with the coercivity result in Proposition 2.6 and the Lower Semicontinuity Theorem 3.28. □

The following result shows that quasiconvexity is also necessary for weak lower semicontinuity; we only state and show this for x-independent integrands, but we note that it also holds for x-dependent ones by a localization technique.

Proposition 3.31. *Let $f\colon\mathbb{R}^{m\times d}\to\mathbb{R}$ have p-growth. If the associated functional*

$$\mathscr{F}[u]:=\int_\Omega f(\nabla u(x))\,\mathrm{d}x,\qquad u\in\mathrm{W}^{1,p}(\Omega;\mathbb{R}^m),$$

is weakly lower semicontinuous (with or without fixed boundary values), then f is quasiconvex.

Proof. We may assume that $B(0,1)\Subset\Omega$, otherwise we can translate and rescale the domain. Let $A\in\mathbb{R}^{m\times d}$ and $\psi\in\mathrm{W}^{1,\infty}_0(B(0,1);\mathbb{R}^m)$. We need to show

$$h(A)\le \fint_{B(0,1)} h(A+\nabla\psi(z))\,\mathrm{d}z.$$

Take for every $j\in\mathbb{N}$ a Vitali cover of $B(0,1)$, see Theorem A.10,

$$B(0,1)=Z\cup\bigcup_{k=1}^{N(j)}B(a_k^{(j)},r_k^{(j)}),\qquad |Z|=0,$$

with $a_k^{(j)}\in B(0,1)$, $r_k^{(j)}\le 1/j$ $(k=1,\dots,N(j))$, and fix a smooth function $h\colon\Omega\setminus B(0,1)\to\mathbb{R}^m$ with $h(x)=Ax$ for $x\in\partial B(0,1)$ (and $h|_{\partial\Omega}$ equal to the prescribed boundary values if there are any). Define

$$u_j(x):=Ax+r_k^{(j)}\psi\left(\frac{x-a_k^{(j)}}{r_k^{(j)}}\right)\quad \text{if } x\in B(a_k^{(j)},r_k^{(j)}),$$

and $u_j:=h$ on $\Omega\setminus B(0,1)$. Then, it is not hard to see that $u_j\rightharpoonup u$ in $\mathrm{W}^{1,p}$ for

$$u(x)=\begin{cases}Ax & \text{if } x\in B(0,1),\\ h(x) & \text{if } x\in\Omega\setminus B(0,1),\end{cases}\qquad x\in\Omega.$$

Comment...

Thus, the lower semicontinuity yields, after eliminating the constant part of the functional on $\Omega \setminus B(0,1)$,

$$\begin{aligned}
\int_{B(0,1)} f(A)\,\mathrm{d}x &\le \liminf_{j\to\infty} \int_{B(0,1)} f(\nabla u_j(x))\,\mathrm{d}x \\
&= \liminf_{j\to\infty} \sum_k \int_{B(a_k^{(j)},r_k^{(j)})} f\left(A + \nabla\psi\left(\frac{x-a_k^{(j)}}{r_k^{(j)}}\right)\right)\mathrm{d}x \\
&= \liminf_{j\to\infty} \sum_k r_k^{(j)} \int_{B(0,1)} f(A+\nabla\psi(y))\,\mathrm{d}y \\
&= \int_{B(0,1)} f(A+\nabla\psi(y))\,\mathrm{d}y
\end{aligned}$$

since $\sum_k r_k^{(j)} = 1$. This is nothing else than quasiconvexity. □

We note that also for quasiconvex integrands, our results about the Euler–Lagrange equations hold just the same (if we assume the same growth conditions).

3.7 Integrands with u-dependence

One very useful feature of our Young measure approach is that it allows to derive a lower semicontinuity result for u-dependent integrands with minimal additional effort. So consider

$$\begin{cases} \mathscr{F}[u] := \displaystyle\int_\Omega f(x,u(x),\nabla u(x))\,\mathrm{d}x \quad \to \quad \min, \\ \text{over all } u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m) \text{ with } u|_{\partial\Omega} = g, \end{cases}$$

where $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain, $1 < p < \infty$, and the Carathéodory integrand $f\colon \Omega \times \mathbb{R}^m \times \mathbb{R}^{m\times d} \to \mathbb{R}$ satisfies the p-growth bound

$$|f(x,v,A)| \le M(1+|v|^p+|A|^p) \qquad \text{for some } M > 0, \tag{3.21}$$

and $g \in \mathrm{W}^{1-1/p,p}(\partial\Omega;\mathbb{R}^m)$.

The main idea of our approach is to consider Young measures generated by the pairs $(u_j,\nabla u_j) \in \mathbb{R}^{m+md}$. We have the following result:

Lemma 3.32. *Let $(u_j) \subset \mathrm{L}^p(\Omega;\mathbb{R}^M)$ and $(V_j) \subset \mathrm{L}^p(\Omega;\mathbb{R}^N)$ be norm-bounded sequences such that*

$$u_j \to u \text{ pointwise a.e.} \qquad \text{and} \qquad V_j \overset{\mathrm{Y}}{\to} \nu = (\nu_x) \in \mathbf{Y}^p(\Omega;\mathbb{R}^N).$$

Then, $(u_j,V_j) \overset{\mathrm{Y}}{\to} \mu = (\mu_x) \in \mathbf{Y}^p(\Omega;\mathbb{R}^{M+N})$ with

$$\mu_x = \delta_{u(x)} \otimes \nu_x \qquad \text{for a.e. } x \in \Omega,$$

Comment...

that is,

$$\int_\Omega f(x,u_j(x),V_j(x))\,\mathrm{d}x \to \int_\Omega \int f(x,u(x),A)\,\mathrm{d}\nu_x(A)\,\mathrm{d}x \tag{3.22}$$

for all Carathéodory $f\colon \Omega\times\mathbb{R}^M\times\mathbb{R}^N\to\mathbb{R}$ *satisfying the p-growth bound* (3.21).

Proof. By a similar density argument as in the proof of Lemma 3.27, it suffices to show the convergence (3.22) for $f(x,v,A)=\varphi(x)\psi(v)h(A)$, where $\varphi\in \mathrm{C}_0(\Omega)$, $\psi\in\mathrm{C}_0(\mathbb{R}^M)$, $h\in\mathrm{C}_0(\mathbb{R}^N)$. We already know by assumption

$$h(V_j)\overset{*}{\rightharpoonup}\big(x\mapsto\langle h,\nu_x\rangle\big)\quad\text{in } \mathrm{L}^\infty.$$

Furthermore, $\psi(u_j)\to\psi(u)$ almost everywhere and thus strongly in L^1 since ψ is bounded. Thus, since the product of a L^∞-weakly* converging sequence and a L^1-strongly converging sequence converges itself weakly* in the sense of measures,

$$\int_\Omega \varphi(x)\psi(u_j(x))h(V_j(x))\,\mathrm{d}x \to \int_\Omega \varphi(x)\psi(u(x))\langle h,\nu_x\rangle\,\mathrm{d}x,$$

which is (3.22) for our special f. This already finishes the proof. □

The trick of the preceding lemma is that in our situation, where $V_j=\nabla u_j\overset{\mathrm{Y}}{\to}\nu\in\mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$, it allows us to "freeze" $u(x)$ in the integrand. Then, we can apply the Jensen-type inequality from Lemma 3.25 just as we did in Theorem 3.28 to get:

Theorem 3.33. *Let* $f\colon\Omega\times\mathbb{R}^m\times\mathbb{R}^{m\times d}\to\mathbb{R}$ *be a Carathéodory integrand with p-growth (in v and A), i.e.* $|f(x,v,A)|\le M(1+|v|^p+|A|^p)$ *for some* $M>0$, $1<p<\infty$. *Assume furthermore that* $f(x,v,\cdot)$ *is quasiconvex for every fixed* $(x,v)\in\Omega\times\mathbb{R}^m$. *Then, the corresponding functional* $\mathscr{F}$ *is weakly lower semicontinuous on* $\mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$.

Remark 3.34. It is not too hard to extend the previous theorem to integrands f satisfying the more general growth condition

$$|f(x,v,A)|\le M(1+|v|^q+|A|^p)$$

for a $1\le q<p/(d-p)$. By the Sobolev Embedding Theorem A.22, $u_j\to u$ in L^q (strongly) for all such q and thus Lemma 3.32 and then Theorem 3.33 can be suitably generalized.

3.8 Regularity of minimizers

At the end of Section 2.5 we discussed the failure of regularity for vector-valued problems. In particular, we mentioned various counterexamples to regularity. Upon closer inspection, however, it turns out that in all counterexamples, the points where regularity fails form a closed set. This is not a coincidence.

Another issue was that all regularity theorems discussed so far require (strong) *convexity* of the integrand and our discussion so far has shown that convexity is not a good

Comment...

notion for vector-valued problems (however, much of the early regularity theory for vector-valued problems focused on convex problems, we will not elaborate on this aspect here). To remedy this, we need a new notion, which will take over from strong convexity in regularity theory: A locally bounded Borel-measurable function $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ is called **strongly (2-)quasiconvex** if there exists $\mu > 0$ such that

$$A \mapsto h(A) - \mu|A|^2 \quad \text{is quasiconvex.}$$

Equivalently, we may require that

$$\mu \int_{B(0,1)} |\nabla\psi(z)|^2 \,\mathrm{d}z \le \int_{B(0,1)} h(A+\nabla\psi(z)) - h(A) \,\mathrm{d}z$$

for all $A \in \mathbb{R}^{m\times d}$ and all $\psi \in \mathrm{W}^{1,\infty}_0(B(0,1);\mathbb{R}^m)$.

The most well-known regularity result in this situation is by Evans from 1986 (there is significant overlap with work by Acerbi & Fusco in the same year):

Theorem 3.35 (Evans 1986). *Let $f\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ be of class C^2, strongly quasiconvex, and assume that there exists $M > 0$ such that*

$$\mathrm{D}^2_A f(A)[B,B] \le M|B|^2 \qquad \textit{for all } A,B \in \mathbb{R}^{m\times d}.$$

*Let $u \in \mathrm{W}^{1,2}_g(\Omega;\mathbb{R}^m)$ be a minimizer of $\mathscr{F}$ over $\mathrm{W}^{1,2}_g(\Omega;\mathbb{R}^m)$ where $g \in \mathrm{W}^{1/2,2}(\partial\Omega;\mathbb{R}^m)$. Then, there exists a relatively closed **singular set** $\Sigma_u \subset \Omega$ with $|\Sigma_u| = 0$ and it holds that $u \in \mathrm{C}^{1,\alpha}_{\mathrm{loc}}(\Omega\setminus\Sigma_u)$ for all $\alpha \in (0,1)$.*

The proof is long and builds on earlier work of Almgren and De Giorgi in Geometric Measure Theory.

The result is called a "partial regularity" result because the regularity does not hold everywhere. It should be noted that while the scalar regularity theory was essentially a theory for PDEs, and hence applies to *all* solutions of the Euler–Lagrange equations, this is not the case for the present result: There is *no* regularity theory for critical points of the Euler–Lagrange equation for a quasiconvex or even polyconvex (see next chapter) integral functional. This was shown by Müller & Švérak in 2003 [87] (for the quasiconvex case) and Székelyhidi Jr. in 2004 [107] (for the polyconvex case). We finally remark that sometimes better estimates on the "smallness" of Σ_u than merely $|\Sigma_u| = 0$ are available. In fact, for strongly convex integrands it can be shown that the (Hausdorff-)dimension of the singular set is at most $d-2$, this is actually within reach of our discussion in Section 2.5, because we showed $\mathrm{W}^{2,2}_{\mathrm{loc}}$-regularity, and general properties of such functions imply the conclusion, see Chapter 2 of [49]. In the strongly quasiconvex case, much less is known, but at least for minimizers that happen to be $\mathrm{W}^{1,\infty}$ it was established in 2007 by Kristensen & Mingione that the dimension of the singular set is strictly less than d, see [66]. Many other questions are open.

Comment...

Notes and historical remarks

The notion of quasiconvexity was first introduced in Morrey's fundamental paper [79] and later refined by Meyers in [76]. The results about null-Lagrangians, in particular Lemma 3.8 and Lemma 3.10, go back to Morrey [80], Reshetnyak [95] and Ball [4]. Our proof is based on the idea that certain combinations of derivatives might have good convergence properties even if the individual derivatives do not. This is also pivotal for so-called *compensated compactness*, which originated in work of Tartar [109, 110] and Murat [88, 89], in particular in their celebrated *Div–Curl Lemma*, this is detailed in [41]. It is quite remarkable that one can prove Brouwer's fixed point theorem using the fact that minors are null-Lagrangians, see Theorem 8.1.3 in [42]. Proposition 3.6 is originally due to Morrey [80], we follow the presentation in [11]. We also remark that if only $r = p$, then a weaker version of Lemma 3.10 is true where we only have convergence of the minors in the sense of distributions, see Theorem 8.20 in [29].

The convexity properties of the quadratic function $A \mapsto \mathbf{M}A : A$, where $\mathbf{M}$ is a fourth-order tensor (or, equivalently, a matrix in $\mathbb{R}^{md\times md}$) has received considerable attention because it corresponds to a *linear* Euler–Lagrange equation. In this case, quasiconvexity and rank-one convexity are the same. Moreover, even polyconvexity (see next chapter) is equivalent to quasiconvexity (and rank-one convexity) if $d = 2$ or $m = 2$, but this does not hold anymore for $d, m \geq 3$. These results together with pointers to the vast literature can be found in Section 5.3.2 in [29].

A more general result on why convexity is inadmissible for realistic problems in nonlinear elasticity can be found in Section 4.8 of [22].

The result that rank-one convex functions are locally Lipschitz continuous is well-known for convex functions, see for example Corollary 2.4 in [39] and an adapted version for rank-one convex (even separately convex) functions is in Theorem 2.31 in [29]. Our proof with a quantitative bound is from Lemma 2.2 in [11].

The Fundamental Theorem of Young Measure Theory 3.11 has seen considerable evolution since its original proof by Young in the 1930s and 1940s [115–117]. It was popularized in the calculus of variations by Tartar [110] and Ball [6]. Further, much more general, versions have been contributed by Balder [3], also see [21] for probabilistic applications.

The Ball–James rigidity theorem 5.11 is from [9].

Lemma 3.22 is an expression of the well-known *decomposition lemma* from [44], another version is in [63]. The blow-up technique is related to the use of tangent measures in Geometric Measure Theory, see for example [72].

Many results in this chapter are formulated for Carathéodory integrands continue to hold for $f\colon \Omega \times \mathbb{R}^N \to \mathbb{R}$ that are Borel measurable and lower semicontinuous in the second argument, so called **normal integrands**, see [16] and [43].

It is possible to prove the Lower Semicontinuity Theorem 3.28 for quasiconvex integrands without the use of Young measures, see for instance Chapter 8 in [29] for such an approach. However, many of the ideas are essentially the same, they are just carried out directly without the intermediate steps of Young measures. However, Young measures allow one to very clearly organize and compartmentalize these ideas, which aids the understand-

Comment...

ing of the material. We will also use Young measures later for the relaxation of non-lower semicontinuous functionals. More on lower semicontinuity and Young measures can be found in the book [94]. The first use of Young measures to investigate questions of lower semicontinuity for quasiconvex integral functionals seems to be [58].

Kružík [67] showed the curious property that for a quasiconvex $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ with $m \geq 3$, $d \geq 2$ the function $A \mapsto h(A^T)$ may not be quasiconvex. The proof is based on Šverák's example of a rank-one convex function that is not quasiconvex (for the same dimensions), which we will present in Example 5.15 in Chapter 5.

If the integrand has critical negative growth, then lower semicontinuity only holds if the boundary values are fixed along a sequence or if one imposes *quasiconvexity at the boundary*, see [14] for a recent survey article discussing this topic.

Laurence Chisholm Young originally introduced the objects that are now called Young measures as "generalized curves/surfaces" in the late 1930s and early 1940s [115–117] to treat problems in the calculus of variations and optimal control theory that could not be solved using classical methods. His book [118] explains these objects and their applications in great detai (in particular, our "sailing against the wind" example from Section 1.7 is from there).

In Chapter 5 we will consider relaxation problems formulated using Young measures. Further, Young measures have provided a convenient framework to describe fine phase mixtures in the theory of microstructure. A second avenue of development—somewhat different from Young's original intention—is to use Young measures as a *technical tool* only (that is without them appearing in the final statement of a result). This approach is in fact quite old and was probably first adopted in a series of articles by McShane from the 1940s [73–75]. There, the author first finds a generalized Young measure solution to a variational problem, then proves additional properties of the obtained Young measures, and finally concludes that these properties entail that the generalized solution is in fact classical. This idea exhibits some parallels to the hugely successful approach of first proving existence of a weak solution to a PDE and then showing that this weak solution has (in some cases) additional regularity.

Several people contributed to Young measure theory from the 1970s onward, including Berliocchi & Lasry [16], Balder [3], Ball [6] and Kristensen [64], among many others. An important breakthrough in this respect was the characterization of the class of Young measures generated by sequences of gradients in the early 1990s by Kinderlehrer and Pedregal [57, 59], see Theorem 5.12. Their result places gradient Young measures in duality with quasiconvex functions via Jensen-type inequalities (another work in this direction is Sychev's article [106]). Young measures can also be used to show regularity, see the recent work by Dolzmann & Kristensen [37]. Carstensen & Roubíček [20] considered numerical approximations.

Another branch of Young measure theory started in the late 1970s and early 1980s, when Tartar [109, 110, 112] and Murat [88–90] developed the theory of compensated compactness and were able to settle many open problems in the theory of hyperbolic conservation laws; another important contributor here was DiPerna, see for example [35]. A key point of this strategy is to use the good compactness properties of Young measures to pass to limits easily

Comment...

in nonlinear quantities and then to deduce from pointwise and differential constraints on the generating sequences that the Young measure collapses to a point mass, corresponding to a classical function (so no oscillation phenomena occurred). Moreover, in this situation weak convergence improves to convergence in measure (or even in norm), hence the name compensated *compactness*. Extensions of Young measures that allow to pass to the limit in quadratic expressions have been developed by Tartar [111] under the name "H-measures" and, independently, by Gérard [47], who called them "micro-local defect measures", cf. the survey articles [45, 113].

The theory of Young measures is now very mature and there are several monographs [21, 94, 96] that give overviews over the theory from different points of view.

Comment...

Chapter 4

Polyconvexity

At the beginning of Chapter 3 we saw that convexity cannot hold concurrently with frame-indifference (and a mild positivity condition). Thus, we were lead to consider quasiconvex integrands. However, while quasiconvexity is of huge importance in the theory of the calculus of variations, our Lower Semicontinuity Theorem 3.28 has one major drawback: we needed to require the p-growth bound

$$|f(x,A)| \leq M(1+|A|^p), \qquad x \in \Omega,\ A \in \mathbb{R}^{d\times d},$$

for some $M > 0$ and $1 < p < \infty$. Unfortunately, this is not a realistic assumption for nonlinear elasticity because it ignores the fact that infinite compressions should cost infinite energy. Indeed, realistic integrands have the property that

$$f(A) \to +\infty \quad \text{as} \quad \det A \downarrow 0$$

and

$$f(A) = +\infty \quad \text{if} \quad \det A \leq 0.$$

For example, the family of matrices

$$A_\alpha := \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix}, \qquad \alpha > 0,$$

satisfies $\det A_\alpha \downarrow 0$ as $\alpha \downarrow 0$, but $|A_\alpha|$ remains uniformly bounded, so the above p-growth bound cannot hold.

The question whether the Lower Semicontinuity Theorem 3.28 for quasiconvex integrands can be extended to integrands with the above growth is currently a major unsolved problem, see [7]. For the time being, we have to confine ourself to a more restrictive notion of convexity if we want to allow for the above "elastic" growth. This type of convexity is called *polyconvexity* by John Ball in [4], who proved the corresponding existence theorem for minimization problems and enabled a mature mathematical treatment of nonlinear elasticity theory, including the realistic Mooney–Rivlin and Ogden materials, see Section 1.4 and Examples 4.3, 4.4.

Comment...

We will only look at the three-dimensional theory because it is by far the most physically important case and because this restriction eases the notational burden; however, other dimensions can be treated as well, we refer to the comprehensive treatment in [29].

4.1 Polyconvexity

A continuous integrand $h\colon \mathbb{R}^{3\times 3} \to \mathbb{R}\cup\{+\infty\}$ (now we allow the value $+\infty$) is called **polyconvex** if it can be written in the form

$$h(A) = H(A, \operatorname{cof} A, \det A), \qquad A \in \mathbb{R}^{3\times 3},$$

where $H\colon \mathbb{R}^{3\times 3}\times\mathbb{R}^{3\times 3}\times\mathbb{R} \to \mathbb{R}\cup\{+\infty\}$ is *convex*. Here, $\operatorname{cof} A$ denotes the cofactor matrix as defined in Appendix A.1.

While convexity obviously implies polyconvexity, the converse is clearly false. However, we have:

Proposition 4.1. *A polyconvex $h\colon \mathbb{R}^{3\times 3} \to \mathbb{R}$ (not taking the value $+\infty$) is quasiconvex.*

Proof. Let h be as stated and let $\psi \in \mathrm{W}^{1,\infty}(B(0,1);\mathbb{R}^3)$ with $\psi(x)|_{\partial B(0,1)} = Ax$. Then, using Jensen's inequality

$$\begin{aligned}
\fint_{B(0,1)} h(\nabla\psi)\,\mathrm{d}x &= \fint_{B(0,1)} H(\nabla\psi, \operatorname{cof}\nabla\psi, \det\nabla\psi)\,\mathrm{d}x \\
&\geq H\left(\fint_{B(0,1)} \nabla\psi\,\mathrm{d}x, \fint_{B(0,1)} \operatorname{cof}\nabla\psi\,\mathrm{d}x, \fint_{B(0,1)} \det\nabla\psi\,\mathrm{d}x\right) \\
&= H(A, \operatorname{cof} A, \det A) = h(A),
\end{aligned}$$

where for the last line we used the fact that minors are null-Lagrangians as proved in Lemma 3.8. □

At this point we have seen all four major convexity conditions that play a role in the modern calculus of variations. They can be put into a linear order by implications:

$$\text{convexity} \;\Rightarrow\; \text{polyconvexity} \;\Rightarrow\; \text{quasiconvexity} \;\Rightarrow\; \text{rank-one convexity}$$

While in the scalar case ($d = 1$ or $m = 1$) all these notions are equivalent, in higher dimensions the reverse implications do not hold in general. Clearly, the determinant function is polyconvex but not convex. However, the fact that the notions of polyconvexity, quasiconvexity, and rank-one convexity are all different in higher dimensions, are non-trivial. The fact that rank-one convexity does not imply quasiconvexity, at least for $d \geq 2$, $m \geq 3$, is the subject of Švérak's famous counterexample, see Example 5.15. The case $d = m = 2$ is still open and one of the major unsolved problems in the field. Here we discuss an example of a quasiconvex, but not polyconvex function:

Comment...

Example 4.2 (Alibert–Dacorogna–Marcellini). Recall from Example 3.5 the Alibert–Dacorogna–Marcellini function

$$h_\gamma(A) := |A|^2\big(|A|^2 - 2\gamma \det A\big), \qquad A \in \mathbb{R}^{2\times 2}.$$

It can be shown that h_γ is polyconvex if and only if $|\gamma| \le 1$. Since it is quasiconvex if and only if $|\gamma| \le \gamma_{\mathrm{QC}}$, where $\gamma_{\mathrm{QC}} > 1$, we see that there are quasiconvex functions that are not polyconvex. See again Section 5.3.8 in [29] for details.

Example 4.3 (Mooney–Rivlin materials). Functions W of the form

$$W(F) := a|F|^2 + b|\operatorname{cof} F|^2 + \Gamma(\det F),$$

with $a, b > 0$ and $\Gamma(d) = \alpha d^2 - \beta \log d$ for $\alpha, \beta > 0$ are polyconvex. This is obvious once we realize that Γ is convex. Such energy functionals model Mooney–Rivlin materials or, if $b = 0$ neo-Hookean materials.

Example 4.4 (Ogden materials). Functions W of the form

$$W(F) := \sum_{i=1}^{M} a_i \operatorname{tr}\big[(F^T F)^{\gamma_i/2}\big] + \sum_{j=1}^{N} b_j \operatorname{tr}\operatorname{cof}\big[(F^T F)^{\delta_j/2}\big] + \Gamma(\det F), \qquad F \in \mathbb{R}^{3\times 3},$$

where $a_i > 0$, $\gamma_i \ge 1$, $b_j > 0$, $\delta_j \ge 1$, and $\Gamma\colon \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is a convex function with $\Gamma(d) \to +\infty$ as $d \downarrow 0$, $\Gamma(d) = +\infty$ for $s \le 0$ and good enough growth from below, can be shown to be polyconvex. These stored energy functionals correspond to so-called Ogden materials and occur in a wide range of elasticity applications, see [22] for details.

It can also be shown that in three dimensions convex functions of certain combinations of the singular values of a matrix (the singular values of A are the square roots of the eigenvalues of $A^T A$) are polyconvex, see Section 5.3.3 in [29] for details.

4.2 Existence of minimizers

Let $\Omega \subset \mathbb{R}^3$ be a *connected* bounded Lipschitz domain. In this section we will prove existence of a minimizer of the variational problem

$$\begin{cases} \mathscr{F}[u] := \displaystyle\int_\Omega W(x, \nabla u(x)) - b(x)\cdot u(x)\,\mathrm{d}x \quad \to \quad \min, \\ \text{over all } u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^3) \text{ with } \det \nabla u > 0 \text{ a.e. and } u|_{\partial\Omega} = g, \end{cases} \tag{4.1}$$

where $W\colon \Omega \times \mathbb{R}^{3\times 3} \to \mathbb{R} \cup \{+\infty\}$ is Carathéodory and $W(x,\cdot)$ is polyconvex for almost every $x \in \Omega$. Thus,

$$W(x,A) = \mathbb{W}(x, A, \operatorname{cof} A, \det A), \qquad (x,A) \in \Omega \times \mathbb{R}^{3\times 3},$$

Comment...

for $\mathbb{W}\colon \Omega \times \mathbb{R}^{3\times 3} \times \mathbb{R}^{3\times 3} \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ with $\mathbb{W}(x,\cdot,\cdot,\cdot)$ jointly convex and continuous for almost every $x \in \Omega$. As the only upper growth assumptions on W we impose

$$\begin{cases} W(x,A) \to +\infty & \text{as} \quad \det A \downarrow 0, \\ W(x,A) = +\infty & \text{if} \quad \det A \leq 0. \end{cases}$$

Furthermore, we assume as usual that $g \in \mathrm{W}^{1-1/p,p}(\partial\Omega;\mathbb{R}^3)$ and that $b \in \mathrm{L}^q(\Omega;\mathbb{R}^3)$ where $1/p + 1/q = 1$. The exponent $p > 1$ will remain unspecified for now. Later, when we impose conditions on the *coercivity* of W, we will also specify p.

The first lower semicontinuity result is relatively straightforward:

Theorem 4.5. *If in addition to the above assumptions it holds that*

$$\mu|A|^p - \mu^{-1} \leq W(x,A) \qquad \textit{for some } \mu > 0 \textit{ and } p > 3, \tag{4.2}$$

then the minimization problem (4.1) *has at least one solution in the space*

$$\mathscr{A} := \left\{ u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^3) \ : \ \det \nabla u > 0 \textit{ a.e. and } u|_{\partial\Omega} = g \right\},$$

whenever this set is non-empty.

Proof. We apply the usual Direct Method. For a minimizing sequence $(u_j) \subset \mathscr{A}$ with $\mathscr{F}[u_j] \to \min$, we get from the coercivity estimate (4.2) in conjunction with the Poincaré–Friedrichs inequality (and the fixed boundary values) that for any $\delta > 0$,

$$\begin{aligned} &\int_\Omega W(x,\nabla u_j) - b(x)\cdot u_j(x)\,\mathrm{d}x \\ &\qquad \geq \mu\|\nabla u_j\|_{\mathrm{L}^p}^p - \mu^{-1}|\Omega| - \|b\|_{\mathrm{L}^q}\cdot\|u_j\|_{\mathrm{L}^p} \\ &\qquad \geq \frac{\mu}{C}\|u_j\|_{\mathrm{W}^{1,p}}^p - \frac{\mu}{C}\|g\|_{\mathrm{W}^{1-1/p,p}}^p - \frac{|\Omega|}{\mu} - \frac{1}{\delta q}\|b\|_{\mathrm{L}^q}^q - \frac{\delta}{p}\|u_j\|_{\mathrm{W}^{1,p}}^p. \end{aligned}$$

Choosing $\delta = p\mu/(2C)$, we arrive at

$$\sup_{j\in\mathbb{N}} \|u_j\|_{\mathrm{W}^{1,p}} \leq C\Big(\sup_{j\in\mathbb{N}} \mathscr{F}[u_j] + 1\Big).$$

Thus, we may select a subsequence (not relabeled) such that $u_j \rightharpoonup u$ in $\mathrm{W}^{1,p}(\Omega;\mathbb{R}^3)$.

Now, by Lemma 3.10 and $p > 3$,

$$\begin{cases} \det\nabla u_j \rightharpoonup \det\nabla u & \text{in } \mathrm{L}^{p/3} \qquad \text{and} \\ \operatorname{cof}\nabla u_j \rightharpoonup \operatorname{cof}\nabla u & \text{in } \mathrm{L}^{p/2}. \end{cases}$$

Thus, an argument entirely analogous to the proof of Theorem 2.7 (note that in that theorem we did not need any upper growth condition) yields that the hyperelastic part of $\mathscr{F}$,

$$u \mapsto \int_\Omega W(x,\nabla u_j)\,\mathrm{d}x = \int_\Omega \mathbb{W}(x,\nabla u_j,\operatorname{cof}\nabla u_j,\det\nabla u_j)\,\mathrm{d}x$$

Comment...

is weakly lower semicontinuous in $\mathrm{W}^{1,p}$. The second part of $\mathscr{F}$ is weakly continuous by Lemma 2.38. Thus,

$$\mathscr{F}[u] \leq \liminf_{j\to\infty} \mathscr{F}[u_j] = \inf_{\mathscr{A}} \mathscr{F}.$$

In particular, $\det \nabla u > 0$ almost everywhere and $u|_{\partial\Omega} = g$ by the weak continuity of the trace. Hence $u \in \mathscr{A}$ and the proof is finished. □

Unfortunately, the preceding theorem has a major drawback in that $p > 3$ has to be assumed. In applications in elasticity theory, however, a more realistic form of W is

$$W(A) = \frac{\lambda}{2}(\operatorname{tr} E)^2 + \mu \operatorname{tr} E^2 + \mathrm{O}(|E|^2), \qquad E = \frac{1}{2}(A^T A - I), \tag{4.3}$$

where $\lambda, \mu > 0$ are the **Lamé constants**. Except for the last term $\mathrm{O}(|E|^2)$, which vanishes as $|E| \downarrow 0$, this energy corresponds to a so-called **St. Venant–Kirchhoff material**. It was shown by Ciarlet & Geymonat [23] (also see Theorem 4.10-2 in [22]) that for any such Lamé constants, there exists a polyconvex function W of the form

$$W(A) = a|A|^2 + b|\operatorname{cof} A|^2 + \gamma(\det A) + c,$$

where $a, b, c > 0$ and $\gamma\colon \mathbb{R} \to [0,\infty]$ is convex, such that the above expansion (4.3) holds. Clearly, such W has 2-growth in $|A|$. Thus, we need an existence theorem for functions with these growth properties.

The core of the refined argument will be an improvement of Lemma 3.10:

Lemma 4.6. *Let* $1 \leq p, q, r \leq \infty$ *with*

$$p \geq 2, \qquad \frac{1}{p} + \frac{1}{q} \leq 1, \qquad r \geq 1,$$

and assume that the sequence $(u_j) \subset \mathrm{W}^{1,p}(\Omega;\mathbb{R}^3)$ *satisfies for some* $H \in \mathrm{L}^q(\Omega;\mathbb{R}^{3\times 3})$, $d \in \mathrm{L}^r(\Omega)$ *that*

$$\begin{cases} u_j \rightharpoonup u & \text{in } \mathrm{W}^{1,p}, \\ \operatorname{cof}\nabla u_j \rightharpoonup H & \text{in } \mathrm{L}^q, \\ \det \nabla u_j \rightharpoonup d & \text{in } \mathrm{L}^r. \end{cases}$$

Then, $H = \operatorname{cof}\nabla u$ *and* $d = \det \nabla u$.

Proof. The idea is to use *distributional* versions of the cofactor matrix and Jacobian determinant and to show that they agree with the usual definitions for sufficiently regular functions. To this aim we will use the representation of minors as divergences already employed in Lemmas 3.8, 3.10.

Step 1. From (3.8) we get that for all $\varphi = (\varphi^1, \varphi^2, \varphi^3) \in \mathrm{C}^1(\Omega;\mathbb{R}^3)$,

$$(\operatorname{cof}\nabla\varphi)^k_l = \partial_{l+1}(\varphi^{k+1}\partial_{l+2}\varphi^{k+2}) - \partial_{l+2}(\varphi^{k+1}\partial_{l+1}\varphi^{k+2}),$$

Comment...

where $k,l \in \{1,2,3\}$ are cyclic indices. Thus, for all $\psi \in \mathrm{C}_c^\infty(\Omega)$,

$$\begin{aligned}\int_\Omega (\operatorname{cof}\nabla\varphi)_l^k \psi \,\mathrm{d}x &= -\int_\Omega (\varphi^{k+1}\partial_{l+2}\varphi^{k+2})\partial_{l+1}\psi - (\varphi^{k+1}\partial_{l+1}\varphi^{k+2})\partial_{l+2}\psi \,\mathrm{d}x \\ &=: \big\langle (\operatorname{Cof}\nabla\varphi)_l^k, \psi\big\rangle \end{aligned} \tag{4.4}$$

and we call the quantity "$\operatorname{Cof}\nabla\varphi$" the **distributional cofactors**.

To investigate the continuity properties of the distributional cofactors, we consider

$$G(\varphi) := \int_\Omega (\varphi^k \partial_l \varphi^m)\partial_n \psi \,\mathrm{d}x, \qquad \varphi \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^3),$$

for some $k,l,m,n \in \{1,2,3\}$ and fixed $\psi \in \mathrm{C}_c^\infty(\Omega)$. We can estimate this using the Hölder inequality as follows:

$$|G(\varphi)| \le \|\varphi\|_{\mathrm{L}^s}\|\nabla\varphi\|_{\mathrm{L}^p}\|\nabla\psi\|_\infty$$

whenever

$$\frac{1}{s} + \frac{1}{p} \le 1. \tag{4.5}$$

In particular this is true for $s = p \ge 2$. Thus, since $\mathrm{C}^1(\Omega;\mathbb{R}^3)$ is dense in $\mathrm{W}^{1,p}(\Omega;\mathbb{R}^3)$, we have that the equality (4.4) also holds for $\varphi \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^3)$ whenever $p \ge 2$.

Moreover, the above arguments yield for a sequence $(\varphi_j) \subset \mathrm{W}^{1,p}(\Omega;\mathbb{R}^3)$ that

$$G(\varphi_j) \to G(\varphi) \qquad \text{if} \qquad \begin{cases} \varphi_j \to \varphi & \text{in } \mathrm{L}^s, \\ \nabla\varphi_j \rightharpoonup \nabla\varphi & \text{in } \mathrm{L}^p. \end{cases}$$

The first convergence is automatically satisfied by the compact Rellich–Kondrachov Theorem A.23 if

$$s < \begin{cases} \frac{3p}{3-p} & \text{if } p < 3, \\ \infty & \text{if } p \ge 3. \end{cases} \tag{4.6}$$

We can always choose s such that it *simultaneously* satisfies (4.5) and (4.6), as a quick calculation shows. Thus,

$$\big\langle \operatorname{Cof}\nabla\varphi_j, \psi\big\rangle \to \big\langle \operatorname{Cof}\nabla\varphi, \psi\big\rangle \qquad \text{if} \qquad \varphi_j \rightharpoonup \varphi \quad \text{in } \mathrm{W}^{1,p},\ p \ge 2. \tag{4.7}$$

Step 2. First, if $\varphi = (\varphi^1,\varphi^2,\varphi^3) \in \mathrm{C}^1(\Omega;\mathbb{R}^3)$, then we know from (3.9) that

$$\det\nabla\varphi = \sum_{l=1}^3 \partial_l \varphi^1 (\operatorname{cof}\nabla\varphi)_l^1 = \sum_{l=1}^3 \partial_l\big(\varphi^1(\operatorname{cof}\nabla\varphi)_l^1\big).$$

Thus, in this case, we have for all $\psi \in \mathrm{C}_c^\infty(\Omega)$ that

$$\int_\Omega (\det\nabla\varphi)\psi \,\mathrm{d}x = \sum_{l=1}^3 \int_\Omega \partial_l\varphi^1(\operatorname{cof}\nabla\varphi)_l^1\psi \,\mathrm{d}x = -\sum_{l=1}^3 \int_\Omega \big(\varphi^1(\operatorname{cof}\nabla\varphi)_l^1\big)\partial_l\psi \,\mathrm{d}x. \tag{4.8}$$

Comment...

The key idea now is that by Hölder's inequality this quantity makes sense if only $\varphi \in \mathrm{L}^p(\Omega;\mathbb{R}^3)$ and $\operatorname{cof}\nabla\varphi \in \mathrm{L}^q(\Omega;\mathbb{R}^{3\times 3})$ with $\frac{1}{p}+\frac{1}{q}\le 1$. Analogously to the argument for the cofactor matrix, this motivates us to define for

$$\varphi \in \mathrm{L}^p(\Omega;\mathbb{R}^3) \quad \text{with} \quad \operatorname{cof}\nabla\varphi \in \mathrm{L}^q(\Omega;\mathbb{R}^3), \quad \text{where} \quad \frac{1}{p}+\frac{1}{q}\le 1,$$

the **distributional determinant**

$$\big\langle \operatorname{Det}\nabla\varphi, \psi\big\rangle := -\sum_{l=1}^{3}\int_\Omega \big(\varphi^1(\operatorname{cof}\nabla\varphi)^1_l\big)\partial_l\psi\,\mathrm{d}x, \qquad \psi\in \mathrm{C}^\infty_c(\Omega).$$

From (4.8) we see that if $\varphi\in\mathrm{C}^1(\Omega;\mathbb{R}^3)$, then "det = Det", i.e.

$$\int_\Omega (\det\nabla\varphi)\psi\,\mathrm{d}x = \big\langle \operatorname{Det}\nabla\varphi,\psi\big\rangle \tag{4.9}$$

for all $\psi\in\mathrm{C}^\infty_c(\Omega)$. We want to show that this equality remains to hold if merely $\varphi\in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^3)$ and $\operatorname{cof}\nabla\varphi\in\mathrm{L}^q(\Omega;\mathbb{R}^{3\times3})$ with p,q as in the statement of the lemma. For the moment fix $\psi\in\mathrm{C}^\infty_c(\Omega)$ and define for $\varphi\in\mathrm{C}^1(\Omega;\mathbb{R}^3)$ and $F\in\mathrm{C}^1(\Omega;\mathbb{R}^{3\times3})$,

$$Z(\varphi,F) := \sum_{l=1}^{3}\int_\Omega \partial_l\varphi^1 F^1_l\psi + \varphi^1 F^1_l\partial_l\psi\,\mathrm{d}x = \sum_{l=1}^{3}\int_\Omega F^1_l\partial_l(\varphi^1\psi)\,\mathrm{d}x.$$

To establish (4.9), we need to show $Z(\varphi,\operatorname{cof}\nabla\varphi)=0$.

Recall the Piola identity (3.10),

$$\operatorname{div}\operatorname{cof}\nabla v = 0 \qquad \text{if } v\in\mathrm{C}^1(\Omega;\mathbb{R}^3).$$

As a consequence,

$$Z(\varphi,\operatorname{cof}\nabla v)=0 \qquad \text{for } \varphi,v\in\mathrm{C}^1(\Omega;\mathbb{R}^3).$$

Furthermore, we have the estimate

$$|Z(\varphi,F)|\le \|\nabla\varphi\|_{\mathrm{L}^p}\|F\|_{\mathrm{L}^q}\|\psi\|_\infty \qquad \text{whenever} \qquad \frac{1}{p}+\frac{1}{q}\le 1.$$

Thus, we get, via two approximation procedures (first in $F=\operatorname{cof}\nabla v$, then in φ)

$$Z(\varphi,\operatorname{cof}\nabla v)=0 \qquad \text{for } \varphi\in\mathrm{W}^{1,p}(\Omega;\mathbb{R}^3),\ \operatorname{cof}\nabla v\in\mathrm{L}^q(\Omega;\mathbb{R}^{3\times3}).$$

In particular,

$$Z(\varphi,\operatorname{cof}\nabla\varphi)=0 \qquad \text{for } \varphi\in\mathrm{W}^{1,p}(\Omega;\mathbb{R}^3),\ \operatorname{cof}\nabla\varphi\in\mathrm{L}^q(\Omega;\mathbb{R}^{3\times3}).$$

Therefore, (4.9) holds also for $\varphi=u$ as in the statement of the lemma.

Comment...

Moreover, we see from the definition of the distributional determinant that for a sequence $(\varphi_j) \subset \mathrm{W}^{1,p}(\Omega;\mathbb{R}^3)$ we have

$$\big\langle \operatorname{Det}\nabla\varphi_j, \psi\big\rangle \to \big\langle \operatorname{Det}\nabla\varphi, \psi\big\rangle \qquad \text{if} \qquad \begin{cases} \varphi_j \to \varphi & \text{in } \mathrm{L}^s, \\ \operatorname{cof}\nabla\varphi_j \rightharpoonup \operatorname{cof}\nabla\varphi & \text{in } \mathrm{L}^q. \end{cases}$$

whenever

$$\frac{1}{s} + \frac{1}{q} \le 1. \tag{4.10}$$

The first convergence follows, as before, from the Rellich–Kondrachov Theorem A.23 if

$$s < \begin{cases} \frac{3p}{3-p} & \text{if } p < 3, \\ \infty & \text{if } p \ge 3. \end{cases} \tag{4.11}$$

However, since we assumed

$$\frac{1}{p} + \frac{1}{q} \le 1,$$

we can always choose s such that it satisfies (4.10) and (4.11) *simultaneously*, as can be seen by some elementary algebra. Thus,

$$\big\langle \operatorname{Det}\nabla\varphi_j, \psi\big\rangle \to \big\langle \operatorname{Det}\nabla\varphi, \psi\big\rangle \qquad \text{if} \qquad \begin{cases} \varphi_j \rightharpoonup \varphi & \text{in } \mathrm{W}^{1,p}, \\ \operatorname{cof}\nabla\varphi_j \rightharpoonup \operatorname{cof}\nabla\varphi & \text{in } \mathrm{L}^q, \end{cases} \tag{4.12}$$

where p, q satisfy the assumptions of the lemma.

Step 3. Assume that, as in the statement of the lemma, we are given a sequence $(u_j) \subset \mathrm{W}^{1,p}(\Omega;\mathbb{R}^3)$ such that for $H \in \mathrm{L}^q(\Omega;\mathbb{R}^{3\times 3})$, $d \in \mathrm{L}^r(\Omega)$ it holds that

$$\begin{cases} u_j \rightharpoonup u & \text{in } \mathrm{W}^{1,p}, \\ \operatorname{cof}\nabla u_j \rightharpoonup H & \text{in } \mathrm{L}^q, \\ \det\nabla u_j \rightharpoonup d & \text{in } \mathrm{L}^r. \end{cases}$$

Then, (4.4), (4.9) imply that for all $\psi \in \mathrm{C}_c^\infty(\Omega)$,

$$\big\langle \operatorname{Cof}\nabla u_j, \psi\big\rangle \to \big\langle H, \psi\big\rangle \qquad \text{and} \qquad \big\langle \operatorname{Det}\nabla u_j, \psi\big\rangle \to \big\langle d, \psi\big\rangle.$$

On the other hand, from (4.7), (4.12),

$$\big\langle \operatorname{Cof}\nabla u_j, \psi\big\rangle \to \big\langle \operatorname{Cof}\nabla u, \psi\big\rangle \qquad \text{and} \qquad \big\langle \operatorname{Det}\nabla u_j, \psi\big\rangle \to \big\langle \operatorname{Det}\nabla u, \psi\big\rangle.$$

Thus,

$$\int_\Omega (\operatorname{Cof}\nabla u - H)\psi \,\mathrm{d}x = 0 \qquad \text{and} \qquad \int_\Omega (\operatorname{Det}\nabla u - d)\psi \,\mathrm{d}x = 0$$

for all $\psi \in \mathrm{C}_c^\infty(\Omega)$. The Fundamental Lemma of the calculus of variations 2.21 then gives immediately

$$H = \operatorname{cof}\nabla u \qquad \text{and} \qquad d = \det\nabla u.$$

This finishes the proof. □

Comment...

With this tool at hand, we can now prove the improved existence result:

Theorem 4.7 (Ball 1976). *Let* $1 \leq p,q,r < \infty$

$$p \geq 2, \qquad \frac{1}{p} + \frac{1}{q} \leq 1, \qquad r > 1,$$

such that in addition to the assumptions at the beginning of this section, it holds that

$$W(x,A) \geq \mu\big(|A|^p + |\operatorname{cof} A|^q + |\det A|^r\big) - \mu^{-1} \qquad \textit{for some } \mu > 0. \tag{4.13}$$

Then, the minimization problem (4.1) *has at least one solution in the space*

$$\begin{aligned}\mathscr{A} := \big\{ u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^3) \ :\ & \operatorname{cof}\nabla u \in \mathrm{L}^q(\Omega;\mathbb{R}^{3\times 3}),\ \det\nabla u \in \mathrm{L}^r(\Omega), \\ & \det\nabla u > 0 \textit{ a.e. and } u|_{\partial\Omega} = g \big\},\end{aligned}$$

whenever this set is non-empty.

Proof. This follows in a completely analogous way to the proof of Theorem 4.5, but now we select a subsequence of a minimizing sequence $(u_j) \subset \mathscr{A}$ such that for some $H \in \mathrm{L}^q(\Omega;\mathbb{R}^{3\times 3})$, $d \in \mathrm{L}^r(\Omega)$ we have

$$\begin{cases} u_j \rightharpoonup u & \text{in } \mathrm{W}^{1,p}, \\ \operatorname{cof}\nabla u_j \rightharpoonup H & \text{in } \mathrm{L}^q, \\ \det\nabla u_j \rightharpoonup d & \text{in } \mathrm{L}^r, \end{cases}$$

which is possible by the usual weak compactness results in conjunction with the coercivity assumption (4.13). Thus, Lemma 4.6 yields

$$\begin{cases} u_j \rightharpoonup u & \text{in } \mathrm{W}^{1,p}, \\ \operatorname{cof}\nabla u_j \rightharpoonup \operatorname{cof}\nabla u & \text{in } \mathrm{L}^q, \\ \det\nabla u_j \rightharpoonup \det\nabla u & \text{in } \mathrm{L}^r, \end{cases}$$

and we may argue as in Theorem 4.5 to conclude that u is indeed a minimizer. The proof that $u \in \mathscr{A}$ is the same as in Theorem 4.5. □

4.3 Global injectivity

For reasons of *physical admissibility*, we need to additionally prove that we can find a minimizer that is globally injective, at least almost everywhere (since we work with Sobolev functions). There are several approaches to this issue, for example via the *topological degree*. We here present a classical argument by Ciarlet & Nečas [25]. It is important to notice that on physical grounds it is only realistic to expect this injectivity for $p > d$. For lower exponents, complex effects such as cavitation and (microscopic) fracture have to be considered. This is already expressed in the fact that Sobolev functions in $\mathrm{W}^{1,p}$ for $p \leq d$ are not necessarily continuous.

We will prove the following basic theorem:

Comment...

Theorem 4.8. *In the situation of Theorem 4.5, in particular $p > 3$, the minimization problem* (4.1) *has at least one solution in the space*

$$\mathscr{A} := \left\{ u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^d) \ : \ \det \nabla u > 0 \text{ a.e., } u \text{ is injective a.e., } u|_{\partial\Omega} = g \right\},$$

whenever this set is non-empty.

Proof. Let $(u_j) \subset \mathscr{A}$ be a minimizing sequence. The proof is completely analogous to the one of Theorem 4.5, we only need to show in addition that u is injective almost everywhere.

For our choice $p > 3$, the space $\mathrm{W}^{1,p}(\Omega;\mathbb{R}^3)$ embeds continuously into $\mathrm{C}(\Omega;\mathbb{R}^3)$, so we have that

$$u_j \to u \quad \text{uniformly.}$$

Let $U \subset \mathbb{R}^3$ be any relatively compact open set with $u(\Omega) \Subset U$ (note that $u(\Omega)$ is relatively compact). Then, $u_j(\Omega) \subset U$ for j sufficiently large by the uniform convergence. Hence, for such j,

$$|u_j(\Omega)| \leq |U|.$$

Furthermore, since $\det \nabla u_j$ converges weakly to $\det \nabla u$ in $\mathrm{L}^{p/3}$ by Lemma 3.10 and all u_j are injective almost everywhere by definition of the space $\mathscr{A}$,

$$\int_\Omega \det \nabla u \, \mathrm{d}x = \lim_{j\to\infty} \int_\Omega \det \nabla u_j \, \mathrm{d}x = \lim_{j\to\infty} |u_j(\Omega)| \leq |U|.$$

Letting $|U| \downarrow |u(\Omega)|$, we get the **Ciarlet–Nečas non-interpenetration condition**

$$\int_\Omega \det \nabla u \, \mathrm{d}x \leq |u(\Omega)|. \tag{4.14}$$

It is known (see for example [71] or [17]) that even if only $u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^3)$ it holds that

$$\int_\Omega |\det \nabla u| \, \mathrm{d}x = \int_{u(\Omega)} \mathscr{H}^0(u^{-1}(x')) \, \mathrm{d}x',$$

where we denote by $\mathscr{H}^0$ the counting measure. We estimate

$$|u(\Omega)| \leq \int_{u(\Omega)} \mathscr{H}^0(u^{-1}(x')) \, \mathrm{d}x' = \int_\Omega \det \nabla u \, \mathrm{d}x \leq |u(\Omega)|,$$

where we used that $\det \nabla u > 0$ almost everywhere and (4.14). Thus,

$$\mathscr{H}^0(u^{-1}(x')) = 1 \qquad \text{for a.e. } x' \in u(\Omega),$$

which is nothing else than the almost everywhere injectivity of u. □

Injectivity almost everywhere still allows for **self-contact**, where different parts of a specimen touch. Moreover, it still includes unphysical examples, where a countable dense set is mapped into the same point. Injectivity *everywhere* is a harder problem. A well-known result is the following:

Comment...

Theorem 4.9 (Ball 1981). *In the situation of Theorem 4.5, assume furthermore that*

$$\mu\left(|A|^p + \frac{|\operatorname{cof} A|^p}{(\det A)^{p-1}}\right) - \mu^{-1} \leq W(x,A) \qquad \text{for some } \mu > 0 \text{ and } p > 3,$$

and that there exists an injective $u_0 \in \mathrm{C}(\overline{\Omega};\mathbb{R}^3)$ such that $u_0(\Omega)$ is also a bounded Lipschitz domain. Then, the minimization problem (4.1) *has at least one solution u_* in the space*

$$\mathscr{A} := \left\{ u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^3) \ : \ \det \nabla u > 0 \text{ a.e. and } u|_{\partial\Omega} = u_0|_{\partial\Omega} \right\}$$

such that the following assertions hold:

(i) u_* *is a homeomorphism of* Ω *onto* $u_*(\Omega)$*;*

(ii) $u_*(\overline{\Omega}) = u_0(\overline{\Omega})$*;*

(iii) $u^{-1} \in \mathrm{W}^{1,p}(u_0(\Omega);\mathbb{R}^3)$*;*

(iv) $\nabla u^{-1}(v) = (\nabla u(x))^{-1}$ *for* $v = u(x)$ *and almost every* $x \in \Omega$*.*

Finer results are possible, but are best investigated within the context of more concrete applications.

Notes and historical remarks

Theorem 4.7 is a refined version by Ball, Currie & Olver [8] of Ball's original result in [4]. Also see [9, 10] for further reading.

Many questions about polyconvex integral functionals remain open to this day: In particular, the regularity of solutions and the validity of the Euler–Lagrange equations are largely open in the general case. Note that the regularity theory from the previous chapter is not in general applicable, at least if we do not assume the upper p-growth. These questions are even open for the more restricted situation of nonlinear elasticity theory. See [7] for a survey on the current state of the art and a collection of challenging open problems.

As for the almost injectivity (for $p > 3$), this is in fact sometimes automatic, as shown by Ball [5], but the arguments do not apply to all situations. Tang [108] extended this to $p > 2$, but since then we have to deal with non-continuous functions, it is not even clear how to define $u(\Omega)$. A study on the surprising (mis)behavior of the determinant for $p < d$ can be found in [61, 62].

Theorem 4.9 is from [5]. More general results can be found in [101]. The questions of injectivity, invertibility and regularity are intimately connected with *cavitation* and *fracture* phenomena, see for instance [86] and the recent [54], which also contains a large bibliography.

Let us also mention that uniqueness and regularity are difficult and partially unsolved questions for polyconvex functionals. There are counterexamples to uniqueness, see [99] for large boundary values, whereas for small boundary values uniqueness and regularity have been announced to hold [19].

Comment...

Finally, we also mention the alternative so-called intrinsic approach to elasticity, as pioneered by Ciarlet, see [24].

Comment...

Chapter 5

Relaxation

When trying to minimize an integral functional over a Sobolev space, it it a very real possibility that there is no minimizer if the integrand is not quasiconvex; of course, even functionals with non-quasiconvex integrands can have minimizers, but, generally speaking, we should expect non-quasiconvexity to be correlated with the non-existence of minimizers.

For instance, consider the functional

$$\mathscr{F}[u] := \int_0^1 |u(x)|^2 + \big(|u'(x)|^2 - 1\big)^2 \,\mathrm{d}x, \qquad u \in \mathrm{W}_0^{1,4}(0,1).$$

The gradient part of the integrand, $a \mapsto (a^2-1)^2$, is called a **double-well potential**, see Figure 5.1. Approximate minimizers of $\mathscr{F}$ try to satisfy $u' \in \{-1,+1\}$ as closely as possible, while at the same time staying close to zero because of the first term under the integral. These contradicting requirements lead to minimizing sequences that develop faster and faster oscillations similar to the ones shown in Figure 3.2 on page 50. It should be intuitively clear that no classical function can be a minimizer of $\mathscr{F}$.

In this situation, we have essentially two options: First, if we only care about the infimal *value* of $\mathscr{F}$, we can compute the *relaxation* $\mathscr{F}_*$ of $\mathscr{F}$, which (by definition) will be lower semicontinuous and then find its minimizer. This will give us the infimum of $\mathscr{F}$ over all admissible functions, but the minimizer of $\mathscr{F}_*$ says only very little about the minimizing sequence of our original $\mathscr{F}$ since all oscillations (and concentrations in some cases) have been "averaged out".

Second, we can focus on the minimizing sequences themselves and try to find a generalized limit object to a minimizing sequence encapsulating "interesting" information about the minimization. This limit object will be a Young measure and in fact this was the original motivation for introducing them. Effectively, Young measure theory allows one to replace a minimization problem over a Sobolev space by a generalized minimization problem over (gradient) Young measures that *always* has a solution. In applications, the emerging oscillations correspond to **microstructure**, which is very important for example in material science. In this chapter we will consider both approaches in turn.

Comment...

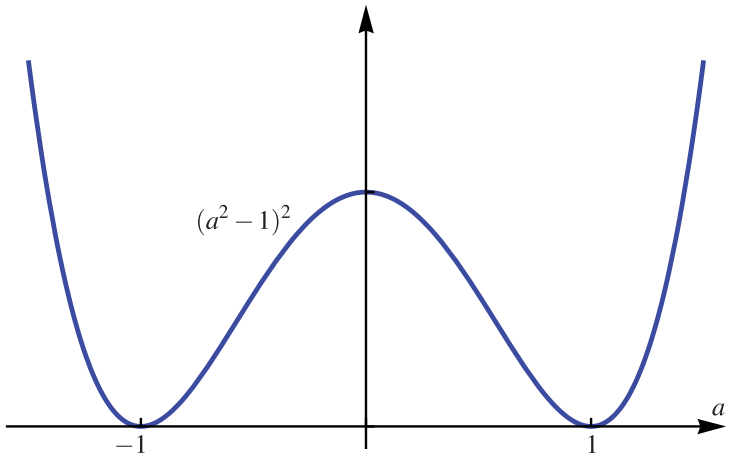

Figure 5.1: A double-well potential.

5.1 Quasiconvex envelopes

We have seen in Chapter 3 that integral functionals with quasiconvex integrands are weakly lower semicontinuous in $\mathrm{W}^{1,p}$, where the exponent $1 < p < \infty$ is determined by growth properties of the integrand. If the integrand is *not* quasiconvex, then we would like to compute the functional's **relaxation**, that for a generic functional $\mathscr{F}\colon X \to \mathbb{R} \cup \{+\infty\}$ is defined to be

$$\mathscr{F}_* := \sup\left\{\mathscr{G} \;:\; \mathscr{G} \leq \mathscr{F} \text{ and } \mathscr{G} \text{ is lower semicontinuous}\right\}.$$

We can expect that when passing from the original functional to the relaxed one, we should also pass from the original integrand to a related but quasiconvex integrand. For this, we define the **quasiconvex envelope** $Qh\colon \mathbb{R}^{m\times d} \to \mathbb{R} \cup \{-\infty\}$ of a locally bounded Borel-function $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ as

$$Qh(A) := \inf\left\{ \fint_\omega h(A + \nabla\psi(z))\,\mathrm{d}z \;:\; \psi \in \mathrm{W}^{1,\infty}_0(\omega;\mathbb{R}^m)\right\}, \qquad A \in \mathbb{R}^{m\times d}, \tag{5.1}$$

where $\omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain. Clearly, $Qh \leq h$. By a similar argument as in Lemma 3.2, one can see that this formula is independent of the choice of ω. If h has p-growth, we may replace the space $\mathrm{W}^{1,\infty}_0(\omega;\mathbb{R}^m)$ in the above formula by $\mathrm{W}^{1,p}_0(\omega;\mathbb{R}^m)$. Finally, by a covering argument as for instance in Proposition 3.31, we may furthermore restrict the class of ψ in the above infimum to those satisfying $\|\psi\|_{\mathrm{L}^\infty} \leq \varepsilon$ for any $\varepsilon > 0$.

It is known that for continuous $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ the quasiconvex envelope $Qh\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ is equal to

$$Qh = \sup\left\{ g \;:\; g \text{ quasiconvex and } g \leq h\right\}, \tag{5.2}$$

see Section 6.3 in [29]. Note that the equality also holds if one (hence both) of the two expressions is $-\infty$.

Comment...

We first need the following lemma:

Lemma 5.1. *For continuous $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ with p-growth, $|h(A)| \le M(1+|A|^p)$, the quasiconvex envelope Qh as defined in* (5.1) *is either identically $-\infty$ or real-valued with p-growth, and quasiconvex.*

Proof. Step 1. If Qh takes the value $-\infty$, say $Qh(B_0) = -\infty$, then for all $N \in \mathbb{N}$ there exists $\hat{\psi}_N \in \mathrm{W}_0^{1,\infty}(B(0,1/2);\mathbb{R}^m)$ such that

$$\int_{B(0,1/2)} h(A_0 + \nabla\hat{\psi}_N(z))\,\mathrm{d}z \le -N.$$

For any $A_0 \in \mathbb{R}^{m\times d}$ pick $\psi_* \in \mathrm{W}^{1,\infty}_{A_0x}(B(0,1);\mathbb{R}^m)$ that has trace B_0x on $\partial B(0,1/2)$ and set

$$\psi_N(z) := \begin{cases} \hat{\psi}_N(z) & \text{if } z \in B(0,1/2), \\ \psi_*(z & \text{otherwise,} \end{cases} \qquad z \in B(0,1).$$

Hence,

$$Qh(A_0) \le \fint_{B(0,1)} h(\nabla\psi_N(z))\,\mathrm{d}z \le \frac{-N+C}{B(0,1)}$$

for some fixed $C > 0$ (depending on our choice of ψ_*). Letting $N \to \infty$, we get $Qh(A_0) = -\infty$. Thus, Qh is *identically* $-\infty$.

Step 2. For any $\psi \in \mathrm{W}_0^{1,\infty}(B(0,1);\mathbb{R}^m)$ and any $A_0 \in \mathbb{R}^{m\times d}$ we need to show

$$\fint_{B(0,1)} Qh(A_0 + \nabla\psi(z))\,\mathrm{d}z \ge Qh(A_0). \tag{5.3}$$

It can be shown that also Qh has p-growth, see for instance Lemma 2.5 in [64]. Thus, we can use an approximation argument in conjunction with Theorem 2.31 to prove that it suffices to show the inequality (5.3) for piecewise affine $\psi \in \mathrm{W}_0^{1,\infty}(B(0,1);\mathbb{R}^m)$, say $\psi(x) = v_k + A_kx$ ($v_k \in \mathbb{R}^m$, $A_k \in \mathbb{R}^{m\times d}$) for $x \in G_k$ from a disjoint collection of Lipschitz subdomains $G_k \subset \Omega$ ($k = 1,\dots,N$) with $|\Omega \setminus \bigcup_k G_k| = 0$. Here we use the $\mathrm{W}^{1,p}$-density of piecewise affine functions in $\mathrm{W}_0^{1,\infty}$, which can be proved by mollification and an elementary approximation of smooth functions by piecewise affine ones (e.g. employing finite elements).

Fix $\varepsilon > 0$. By the definition of Qh, for every k we can find $\varphi_k \in \mathrm{W}_0^{1,\infty}(G_k;\mathbb{R}^m)$ such that

$$Qh(A_0 + A_k) \ge \fint_{G_k} f(A_0 + A_k + \nabla\varphi_k(z))\,\mathrm{d}z - \varepsilon.$$

Then,

$$\begin{aligned}
\int_{B(0,1)} Qh(A_0 + \nabla\psi(z))\,\mathrm{d}z &= \sum_{k=1}^N |G_k| Qh(A_0 + A_k) \\
&\ge \sum_{k=1}^N \int_{G_k} h(A_0 + A_k + \nabla\varphi_k(z))\,\mathrm{d}z - \varepsilon|G_k| \\
&= \int_{B(0,1)} h(A_0 + \nabla\varphi(z))\,\mathrm{d}z - \varepsilon|B(0,1)| \\
&\ge |B(0,1)|\big(Qh(A_0) - \varepsilon\big),
\end{aligned}$$

Comment...

where $\varphi \in \mathrm{W}_0^{1,\infty}(B(0,1);\mathbb{R}^m)$ is defined as

$$\varphi(x) := \begin{cases} v_k + A_k x + \varphi_k(x) & \text{if } x \in G_k, \\ 0 & \text{otherwise,} \end{cases} \qquad x \in B(0,1),$$

and the last step follows from the definition of Qh. Letting $\varepsilon \downarrow 0$, we have shown (5.3). □

We can also use the notion of the quasiconvex envelope to introduce a class of non-trivial quasiconvex functions with p-growth, $1 < p < \infty$ (the linear growth case is also possible using a refined technique).

Lemma 5.2. *Let $F \in \mathbb{R}^{2\times 2}$ with* $\operatorname{rank} F = 2$ *and let $1 < p < \infty$. Define*

$$h(A) := \operatorname{dist}(A, \{-F, F\})^p, \qquad A \in \mathbb{R}^{2\times 2}.$$

Then, the quasiconvex envelope Qh of h is not convex (at zero). Moreover, Qh has p-growth.

Remark 5.3. The result remains true for $p = 1$.

Proof. We first show $Qh(0) > 0$. Assume to the contrary that $Qh(0) = 0$. Then, by (5.1) there exists a sequence $(\psi_j) \subset \mathrm{W}_0^{1,\infty}(B(0,1);\mathbb{R}^2)$ with

$$\fint_{B(0,1)} h(\nabla \psi_j)\,\mathrm{d}z \to 0. \tag{5.4}$$

Let $P\colon \mathbb{R}^{2\times 2} \to \mathbb{R}^{2\times 2}$ be the projection onto the orthogonal complement of $\operatorname{span}\{F\}$. It is straightforward to see that $|P(A)|^p \le h(A)$ for all $A \in \mathbb{R}^{2\times 2}$. Therefore,

$$P(\nabla \psi_j) \to 0 \quad \text{in } \mathrm{L}^p(B(0,1);\mathbb{R}^{2\times 2}). \tag{5.5}$$

We will prove below that we may "invert" P in the sense that if $P(\nabla w) = R$ for some $w \in \mathrm{W}^{1,p}(\mathbb{R}^d;\mathbb{R}^m)$, then

$$\widehat{\nabla w}(\xi) = M(\xi)\hat{R}(\xi), \qquad \xi \in \mathbb{R}^d, \tag{5.6}$$

where $M(\xi)\colon \mathbb{R}^{2\times 2} \to \mathbb{R}^{2\times 2}$ is linear and furthermore depends smoothly and positively 0-homogeneously on ξ. Here, we identified P with its complexification (that is, $P(A+\mathrm{i}B) = P(A) + \mathrm{i}P(B)$ for $A, B \in \mathbb{R}^{m\times d}$).

For $p = 2$, Parseval's formula $\|g\|_{\mathrm{L}^2} = \|\hat{g}\|_{\mathrm{L}^2}$ together with (5.5), (5.6) then implies

$$\begin{aligned} \|\nabla \psi_j\|_{\mathrm{L}^2} &= \|\widehat{\nabla \psi_j}\|_{\mathrm{L}^2} = \|M(\xi)P(\hat{\psi}_j(\xi) \otimes \xi)\|_{\mathrm{L}^2} \\ &\le \|M\|_\infty \|P(\hat{\psi}_j(\xi) \otimes \xi)\|_{\mathrm{L}^2} = \|M\|_\infty \|P(\nabla \psi_j)\|_{\mathrm{L}^2} \to 0. \end{aligned}$$

But then $h(\nabla \psi_j) \to |F|$ in L^1, contradicting (5.4).

For $1 < p < \infty$, we may apply the Mihlin Multiplier Theorem A.28 to get analogously to the case $p = 2$ that

$$\|\nabla \psi_j\|_{\mathrm{L}^p} \le \|M\|_{\mathrm{C}^d} \|P(\nabla \psi_j)\|_{\mathrm{L}^p} \to 0,$$

Comment...

which is again at odds with (5.4). If Qh was convex at zero, we would have

$$Qh(0) \leq \frac{1}{2}\big(Qh(F)+Qh(-F)\big) \leq \frac{1}{2}\big(h(F)+h(-F)\big) = 0,$$

a contradiction.

It remains to show that if $P(\nabla w) = R$, i.e.

$$P(\widehat{\nabla w}(\xi)) = \hat{R}(\xi), \tag{5.7}$$

then (5.6) will follow. To see this, we first observe

$$\widehat{\nabla w}(\xi) = (2\pi \mathrm{i})\,\hat{w}(\xi)\otimes\xi.$$

Notice that $P(a\otimes\xi)\neq 0$ for any $a\in\mathbb{C}^m$, $\xi\in\mathbb{R}^d$ by the assumption that L does not contain a rank-one line. Thus, for some constant $C>0$ we have the *ellipticity* estimate

$$|a\otimes\xi| \leq C|P(a\otimes\xi)| \qquad \text{for all } a\in\mathbb{C}^m,\ \xi\in\mathbb{R}^d,$$

The (complexification of) the projection $P\colon \mathbb{C}^{m\times d}\to\mathbb{C}^{m\times d}$ has kernel $\mathbb{C}L$ (the complex span of L) and hence descends to the quotient

$$[P]\colon \mathbb{C}^{m\times d}/L \to \operatorname{ran} A,$$

and $[P]$ is an invertible linear map. Take a basis $\{V_1,\dots,V_k\}$ of L. For $\xi\in\mathbb{R}^d\setminus\{0\}$ then let

$$\big\{V_1,\dots,V_k,\mathrm{e}_1\otimes\xi,\dots,\mathrm{e}_d\otimes\xi,G_{d+1}(\xi),\dots,G_{md-k}(\xi)\big\}$$

be a $\mathbb{C}$-basis of $\mathbb{C}^{m\times d}$ with the property that the matrices $G_{d+1}(\xi),\dots,G_{md-k}(\xi)$ depend smoothly on ξ and are positively 1-homogeneous in ξ. For all $\xi\in\mathbb{R}^d\setminus\{0\}$, denote by $Q(\xi)\colon\mathbb{C}^{m\times d}\to\mathbb{C}^{m\times d}$ the (non-orthogonal) projection with

$$\begin{aligned}&\ker B(\xi) = L,\\ &\operatorname{ran} B(\xi) = \operatorname{span}\big\{\mathrm{e}_1\otimes\xi,\dots,\mathrm{e}_d\otimes\xi,G_{md-k}(\xi),\dots,G_{md-k}(\xi)\big\}.\end{aligned}$$

If we interpret $\mathrm{e}_1\otimes\xi,\dots,\mathrm{e}_d\otimes\xi,G_{d+1}(\xi),\dots,G_{md-1}(\xi)$ as vectors in $\mathbb{R}^{md}$, collect them into the columns of the matrix $X(\xi)\in\mathbb{R}^{md\times(md-1)}$, and if we further let $Y\in\mathbb{R}^{md\times(md-k)}$ be a matrix whose columns comprise an orthonormal basis of $L^\perp$, then $B(\xi)$ can be written explicitly as (it is elementary to see that $Y^TX(\xi)$ is invertible, see below)

$$B(\xi) = X(\xi)(Y^TX(\xi))^{-1}Y^T.$$

This implies that $B(\xi)$ is positively 0-homogeneous and using Cramer's Rule we also see that $B(\xi)$ depends smoothly on $\xi\in\mathbb{R}^d\setminus\{0\}$. Indeed, if $\det(Y^TX(\xi))$ was not bounded away from zero for $\xi\in\mathbb{S}^{d-1}$, then by compactness there would exist $\xi_0\in\mathbb{S}^{d-1}$ with $\det(Y^TX(\xi_0))=0$, a contradiction. Of course, also $B(\xi)$ descends to a quotient

$$[B(\xi)]\colon \mathbb{C}^{m\times d}/L \to \operatorname{ran} B(\xi),$$

Comment...

which is now invertible. It is not difficult to see that $\xi \mapsto [B(\xi)]$ is still positively 0-homogeneous and smooth in $\xi \neq 0$ (for example by utilizing the basis given above).

Since $\hat{w}(\xi) \otimes \xi \in \operatorname{ran} B(\xi)$, we notice that $[B(\xi)]^{-1}(\hat{w}(\xi) \otimes \xi) = [\hat{w}(\xi) \otimes \xi]$, the equivalence class of $\hat{w}(\xi) \otimes \xi$ in $\mathbb{C}^{m\times d}/L$. Thus, we may rewrite (5.7) as

$$(2\pi \mathrm{i})\,[A][B(\xi)]^{-1}(\hat{w}(\xi) \otimes \xi) = \hat{R}(\xi),$$

or equivalently as

$$\widehat{\nabla w}(\xi) = (2\pi \mathrm{i})\,\hat{w}(\xi) \otimes \xi = [B(\xi)][A]^{-1}\hat{R}(\xi).$$

The function $M\colon \mathbb{R}^d \setminus \{0\} \to \mathbb{R}^{md\times md}$ given by $M(\xi) := [B(\xi)][A]^{-1}$ is smooth and positively 0-homogeneous, and we conclude the multiplier equation (5.6).

For the addition it suffices to notice that there exists a constant $C > 0$ such that

$$|A|^p - C \leq h(A) \leq |A|^p + C \qquad \text{for all } A \in \mathbb{R}^{m\times d}.$$

This concludes the proof. □

5.2 Relaxation of integral functionals

We first consider the abstract principles of relaxation before moving on to more concrete integral functionals. Suppose we have a functional $\mathscr{F}\colon X \to \mathbb{R} \cup \{+\infty\}$, where X is a reflexive Banach space. Assume furthermore that our $\mathscr{F}$ is *not* weakly lower semicontinuous[1], i.e. there exists at least one sequence $u_j \rightharpoonup u$ in X with $\mathscr{F}[u] > \liminf_{j\to\infty} \mathscr{F}[u_j]$. Then, the **relaxation** $\mathscr{F}_*\colon X \to \mathbb{R} \cup \{-\infty, +\infty\}$ of F is defined as

$$\mathscr{F}_*[u] := \inf\left\{ \liminf_{j\to\infty} \mathscr{F}[u_j] \;:\; u_j \rightharpoonup u \text{ in } X \right\}.$$

Theorem 5.4 (Abstract relaxation theorem). *Let X be a separable, reflexive Banach space and let $\mathscr{F}\colon X \to \mathbb{R} \cup \{+\infty\}$ be a functional. Assume furthermore:*

(WH1) ***Weak Coercivity:*** *For all $\Lambda > 0$, the sublevel set*

$$\{u \in X \;:\; \mathscr{F}[u] \leq \Lambda\} \qquad \textit{is sequentially weakly relatively compact.}$$

Then, the relaxation $\mathscr{F}_*$ of $\mathscr{F}$ is weakly lower semicontinuous and $\min_X \mathscr{F}_* = \inf_X \mathscr{F}$.

Proof. Since X^* is also separable, we can find a countable set $\{f_l\}_{l\in\mathbb{N}} \subset X^*$ such that

$$u_j \rightharpoonup u \quad \text{in } X \qquad \Longleftrightarrow \qquad \sup_j \|u_j\| < \infty \text{ and } \langle u_j, f_l\rangle \to \langle u, f_l\rangle \text{ for all } l \in \mathbb{N}.$$

[1] In principle, if $\mathscr{F}$ nevertheless is weakly lower semicontinuous for a *minimizing* sequence, then no relaxation is necessary. It is, however, very difficult to establish such a property without establishing general weak lower semicontinuity.

Comment...

Step 1. We first show weak lower semicontinuity of $\mathscr{F}_*$: Let $u_j \rightharpoonup u$ in X such that $\liminf_{j\to\infty} \mathscr{F}_*[u_j] < \infty$ (if this is not satisfied, there is nothing to show). Selecting a (not relabeled) subsequence if necessary, we may further assume that $\alpha := \sup_j \mathscr{F}_*[u_j] < \infty$.

For every $j \in \mathbb{N}$, there exists a sequence $(u_{j,k})_k \subset X$ such that $u_{j,k} \rightharpoonup u_j$ in X as $k \to \infty$ and

$$\liminf_{k\to\infty} \mathscr{F}[u_{j,k}] \le \mathscr{F}_*[u_j] + \frac{1}{j}.$$

Now select a $v_j := u_{j,k(j)}$ such that

$$\big|\langle v_j, f_l\rangle - \langle u_j, f_l\rangle\big| \le \frac{1}{j} \qquad \text{for all } l = 1,\dots,j,$$

and

$$\mathscr{F}[v_j] \le \mathscr{F}_*[u_j] + \frac{2}{j} \le \alpha + \frac{2}{j}.$$

Then, from the weak coercivity (WH1) we get $\sup_j \|v_j\| < \infty$. Thus, it follows that the so-selected diagonal sequence (v_j) satisfies $v_j \rightharpoonup u$ and

$$\mathscr{F}_*[u] \le \liminf_{j\to\infty} \mathscr{F}[v_j] \le \liminf_{j\to\infty} \mathscr{F}_*[u_j]$$

and $\mathscr{F}_*$ is shown to be weakly lower semicontinuous.

Step 2. Next take a minimizing sequence $(u_j) \subset X$ such that

$$\lim_{j\to\infty} \mathscr{F}_*[u_j] = \inf_X \mathscr{F}_*.$$

Now, for every u_j select a $u_j \in X$ such that

$$\mathscr{F}[v_j] \le \mathscr{F}_*[u_j] + \frac{1}{j}.$$

From the coercivity assumption (WH1) we get that we may select a weakly converging subsequence (not relabeled) $v_j \rightharpoonup v$, where $v \in X$. Then, since $\mathscr{F}_*$ is weakly lower semicontinuous and $\mathscr{F}_* \le \mathscr{F}$,

$$\mathscr{F}_*[v] \le \liminf_{j\to\infty} \mathscr{F}_*[v_j] \le \liminf_{j\to\infty} \mathscr{F}[v_j] \le \lim_{j\to\infty} \mathscr{F}_*[u_j] = \inf_X \mathscr{F}_*,$$

so $\mathscr{F}_*[v] = \min_X \mathscr{F}_*$. Furthermore,

$$\inf_X \mathscr{F} \le \liminf_{j\to\infty} \mathscr{F}[v_j] \le \min_X \mathscr{F}_* \le \inf_X \mathscr{F}.$$

Thus, $\min_X \mathscr{F}_* = \inf_X \mathscr{F}$, completing the proof. □

Comment...

As usual, we here are most interested in the concrete case of integral functionals

$$\mathscr{F}[u] := \int_\Omega f(x, \nabla u(x))\,\mathrm{d}x, \qquad u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m),$$

where, as usual, $f\colon \Omega \times \mathbb{R}^{m\times d} \to \mathbb{R}$ is a Carathéodory integrand satisfying the p-growth and coercivity assumption

$$\mu|A|^p - \mu^{-1} \le f(x,A) \le M(1+|A|^p), \qquad (x,A) \in \Omega \times \mathbb{R}^{m\times d}, \tag{5.8}$$

for some $\mu, M > 0$, $1 < p < \infty$, and $\Omega \subset \mathbb{R}^d$ a bounded Lipschitz domain. The main result of this section is:

Theorem 5.5 (Relaxation). *Let $\mathscr{F}$ be as above and assume furthermore that there exists a* ***modulus of continuity*** *ω, that is, $\omega\colon [0,\infty) \to [0,\infty)$ is continuous, increasing, and $\omega(0) = 0$, such that*

$$|f(x,A) - f(y,A)| \le \omega(|x-y|)(1+|A|^p), \qquad x,y \in \Omega,\ A \in \mathbb{R}^{m\times d}. \tag{5.9}$$

Then, the relaxation $\mathscr{F}_*$ of $\mathscr{F}$ is

$$\mathscr{F}_*[u] = \int_\Omega Qf(x, \nabla u(x))\,\mathrm{d}x, \qquad u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m),$$

where $Qf(x,\bullet)$ is the quasiconvex envelope of $f(x,\bullet)$ for all $x \in \Omega$.

Proof. We define

$$\mathscr{G}[u] := \int_\Omega Qf(x, \nabla u(x))\,\mathrm{d}x, \qquad u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m).$$

As $Qf(x,\bullet)$ is quasiconvex for all $x \in \Omega$ by Lemma 5.1, it is continuous by Proposition 3.6. We will see below that Qf is continuous in its first argument, so Qf is Carathéodory and $\mathscr{G}$ is thus well-defined.

We will show in the following that (i) $\mathscr{G} \le \mathscr{F}_*$ and (ii) $\mathscr{F}_* \le \mathscr{G}$.

To see (i), it suffices to observe that $\mathscr{G}$ is weakly lower semicontinuous by Theorem 3.28 and that $\mathscr{G} \le \mathscr{F}$. Thus, from the definition of $\mathscr{F}_*$ we immediately get $\mathscr{G} \le \mathscr{F}_*$.

We will show (ii) in several steps.

Step 1. Fix $A \in \mathbb{R}^{m\times d}$. For $x \in \Omega$ let $\psi_x \in \mathrm{W}^{1,p}_0(B(0,1);\mathbb{R}^m)$ be such that

$$\fint_{B(0,1)} f(x, A + \nabla\psi_x(z))\,\mathrm{d}z \le Qf(x,A) + 1.$$

Then we use (5.8) to observe

$$\mu \fint_{B(0,1)} |A + \nabla\psi_x(z)|^p\,\mathrm{d}z - \mu^{-1} \le Qf(x,A) + 1 \le M(1+|A|^p) + 1.$$

Comment...

Let now $x, y \in \Omega$ and estimate using (5.9) and the definition of $Qf(x,\cdot)$,

$$\begin{aligned} Qf(y,A) - Qf(x,A) &\le \fint_{B(0,1)} \big| f(y, A + \nabla\psi_x(z)) - f(x, A + \nabla\psi_x(z)) \big| \, \mathrm{d}z + 1 \\ &\le \omega(|x-y|) \fint_{B(0,1)} 1 + |A + \nabla\psi_x(z)|^p \, \mathrm{d}z + 1 \\ &\le C\omega(|x-y|)(1 + |A|^p), \end{aligned}$$

where $C = C(\mu, M)$ is a constant. Exchanging the roles of x and y, we see

$$\big| Qf(x,A) - Qf(y,A) \big| \le C\omega(|x-y|)(1+|A|^p) \tag{5.10}$$

for all $x, y \in \Omega$ and $A \in \mathbb{R}^{m\times d}$. In particular, Qf is continuous in x.

Step 2. Next we show that it suffices to consider piecewise affine u. If $u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$, then there exists a sequence $(v_j) \subset \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ of piecewise affine functions such that $v_j \to u$ in $\mathrm{W}^{1,p}$. Since Qf is Carathéodory and has p-growth (as $Qf \le f$), Theorem 2.31 shows that $\mathscr{G}[v_j] \to \mathscr{G}[u]$ and, by lower semicontinuity, $\mathscr{F}_*[u] \le \liminf_{j\to\infty} \mathscr{F}_*[v_j]$. Thus, if (ii) holds for all piecewise affine u, it also holds for all $u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$.

Step 3. Fix $\varepsilon > 0$ and let $u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ be piecewise affine, say $u(x) = v_k + A_k x$ (where $v_k \in \mathbb{R}^m$, $A_k \in \mathbb{R}^{m\times d}$) on the set G_k from a disjoint collection of open sets $G_k \subset \Omega$ ($k = 1, \dots, N$) such that $|\Omega \setminus \bigcup_{k=1}^N G_k| = 0$. Fix such k and cover G_k up to a nullset with countably many balls $B_{k,l} := B(x_{k,l}, r_{k,l}) \subset G_k$, where $x_{k,l} \in G_k$, $0 < r_{k,l} < \varepsilon$ ($l \in \mathbb{N}$) such that

$$\left| \fint_{B_{k,l}} Qf(x, A_k) \, \mathrm{d}x - Qf(x_{k,l}, A_k) \right| \le C\omega(\varepsilon)(1 + |A_k|^p).$$

This is possible by the Vitali Covering Theorem A.10 in conjunction with (5.10).

Now, from the definition of Qf and the independence of this definition from the domain, we can find a function $\psi_{k,l} \in \mathrm{W}_0^{1,\infty}(B_{k,l};\mathbb{R}^m)$ with $\|\psi_{k,l}\|_{\mathrm{L}^\infty} \le \varepsilon$ and

$$\left| Qf(x_{k,l}, A_k) - \fint_{B_{k,l}} f\big(x_{k,l}, A_k + \nabla\psi_{k,l}(z)\big) \, \mathrm{d}z \right| \le \varepsilon.$$

Set

$$v_\varepsilon(x) := u(x) + \begin{cases} \psi_{k,l}(x) & \text{if } x \in B_{k,l}, \\ 0 & \text{otherwise,} \end{cases} \qquad x \in \Omega.$$

In a similar way to Step 1 we can show that

$$\mu \fint_{B_{k,l}} |A_k + \nabla\psi_{k,l}(z)|^p \, \mathrm{d}z - \mu^{-1} \le M(1 + |A_k|^p) + \varepsilon.$$

Thus,

$$\begin{aligned} \int_\Omega |\nabla v_\varepsilon|^p \, \mathrm{d}x &= \sum_{k,l} \int_{B_{k,l}} |A_k + \nabla\psi_{k,l}(x)|^p \, \mathrm{d}x \\ &\le \frac{M}{\mu} \int_\Omega 1 + |\nabla u(x)|^p \, \mathrm{d}x + \frac{|\Omega|}{\mu^2} + \frac{\varepsilon|\Omega|}{\mu} < \infty, \end{aligned}$$

Comment...

so, by the Poincaré–Friedrichs inequality, $(v_\varepsilon)_{\varepsilon>0} \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ is uniformly bounded in $\mathrm{W}^{1,p}$.

By the continuity assumption (5.9),

$$\left| \fint_{B_{k,l}} f(x_{k,l}, \nabla v_\varepsilon(x))\,\mathrm{d}x - \fint_{B_{k,l}} f(x, \nabla v_\varepsilon(x))\,\mathrm{d}x \right| \leq \omega(\varepsilon) \fint_{B_{k,l}} 1 + |\nabla v_\varepsilon(x)|^p\,\mathrm{d}x.$$

Then, we can combine all the above arguments to get

$$\begin{aligned} &\left| \fint_{B_{k,l}} Qf(x, \nabla u(x))\,\mathrm{d}x - \fint_{B_{k,l}} f(x, \nabla v_\varepsilon(x))\,\mathrm{d}x \right| \\ &\qquad \leq C\omega(\varepsilon) \fint_{B_{k,l}} 2 + |A_k|^p + |\nabla v_\varepsilon(x)|^p\,\mathrm{d}x + \varepsilon. \end{aligned}$$

Multiplying this by $|B_{k,l}|$ and summing over all k,l, we get

$$\left|\mathscr{G}[u] - \mathscr{F}[v_\varepsilon]\right| \leq C\omega(\varepsilon) \int_\Omega 2 + |\nabla u(x)|^p + |\nabla v_\varepsilon(x)|^p\,\mathrm{d}x + \varepsilon|\Omega|$$

and this vanishes as $\varepsilon \downarrow 0$. For $u_j := v_{1/j}$ we can convince ourselves that $u_j \rightharpoonup u$ since (u_j) is weakly relatively compact in $\mathrm{W}^{1,p}$ and $\|u_j - u\|_{\mathrm{L}^\infty} \leq 1/j \to 0$. Hence,

$$\mathscr{G}[u] = \lim_{j\to\infty} \mathscr{F}[u_j] \geq \mathscr{F}_*[u],$$

which is (ii) for piecewise affine u. □

5.3 Young measure relaxation

As discussed at the beginning of this chapter, the relaxation strategy as outlined in the preceding section has one serious drawback: While it allows to find the infimal value, the relaxed functional does not tell us anything about the structure (e.g. oscillations) of minimizing sequences. Often, however, in applications this structure of minimizing sequences is decisive as it represents the development of **microstructure**. For example, in material science, this microstructure can often be observed and greatly influences material properties. Therefore, in this section we implement the second strategy mentioned at the beginning of the chapter, that is, we extend the minimization problem to a larger space and look for solutions there.

Let us first, in an abstract fashion, collect a few properties that our extension should satisfy. Assume we are given a space X together with a notion of convergence "$\to$" and a functional $\mathscr{F}\colon X \to \mathbb{R} \cup \{+\infty\}$, which is not lower semicontinuous with respect to the convergence "$\to$" in X. Then, we need to extend X to a larger space $\overline{X}$ with a convergence "$\rightsquigarrow$" and $\mathscr{F}$ to a functional $\overline{\mathscr{F}}\colon \overline{X} \to \mathbb{R} \cup \{+\infty\}$, called the **extension–relaxation** such that the following conditions are satisfied:

(i) **Extension property:** $X \subset \overline{X}$ (in the sense of an embedding) and, using this embedding, $\overline{\mathscr{F}}|_X = \mathscr{F}$.

Comment...

(ii) **Lower bound:** If $x_j \to x$ in X, then, up to a subsequence, there exists $\bar{x} \in \overline{X}$ such that $x_j \rightsquigarrow \bar{x}$ in $\overline{X}$ and

$$\liminf_{j\to\infty} \mathscr{F}[u_j] \geq \overline{\mathscr{F}}[\bar{x}].$$

(iii) **Recovery sequence:** For all $\bar{x} \in \overline{X}$ there exists a **recovery sequence** $(u_j) \subset X$ with $x_j \rightsquigarrow \bar{x}$ in $\overline{X}$ and such that

$$\lim_{j\to\infty} \mathscr{F}[u_j] = \overline{\mathscr{F}}[\bar{x}].$$

Intuitively, (ii) entails that we can solve our minimization problem for $\mathscr{F}$ by passing to the extended space; in particular, if $(u_j) \subset X$ is minimizing and relatively compact (which, up to a subsequence, implies $x_j \to x$ in X), then (ii) tells us that $\bar{x}$ should be considered the "minimizer". On the other hand, (iii) ensures that the minimization problems for $\mathscr{F}$ and for $\overline{\mathscr{F}}$ are sufficiently related. In particular, it is easy to see that the infima agree and indeed $\overline{\mathscr{F}}$ attains its minimum (for this, we need to assume some coercivity of $\mathscr{F}$):

$$\min_{\overline{X}} \overline{\mathscr{F}} = \inf_{X} \mathscr{F}.$$

If we consider integral functionals defined on $X = \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$, we have already seen a very convenient extended space $\overline{X}$, namely the space of gradient L^p-Young measures. In fact, these generalized variational problems are the original *raison d'être* for Young measures. In the following we will show how this fits into the above abstract framework. This, we return to our prototypical integral functional

$$\mathscr{F}[u] := \int_\Omega f(x,\nabla u(x))\,\mathrm{d}x, \qquad u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m),$$

with $f\colon \Omega \times \mathbb{R}^{m\times d} \to \mathbb{R}$ Carathéodory and

$$\mu|A|^p - \mu^{-1} \leq f(x,A) \leq M(1+|A|^p), \qquad (x,A) \in \Omega \times \mathbb{R}^{m\times d}, \tag{5.11}$$

for some $\mu, M > 0$, $1 < p < \infty$, and $\Omega \subset \mathbb{R}^d$ a bounded Lipschitz domain.

Motivated by the considerations in Sections 3.3 and 3.4, we let for a gradient Young measure $\nu = (\nu_x) \in \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$ the **extension–relaxation** $\overline{\mathscr{F}}\colon \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d}) \to \mathbb{R}$ be defined as

$$\overline{\mathscr{F}}[\nu] := \langle\!\langle f, \nu \rangle\!\rangle = \int_\Omega \int f(x,A)\,\mathrm{d}\nu_x(A)\,\mathrm{d}x.$$

Then, our relaxation result takes the following form:

Theorem 5.6 (Relaxation with Young measures). *Let $\mathscr{F}, \overline{\mathscr{F}}$ be as above.*

Comment...

(i) ***Extension property:*** *For every* $u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ *define the* elementary Young measure $(\delta_{\nabla u}) \in \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$ *as*

$$(\delta_{\nabla u})_x := \delta_{\nabla u(x)} \qquad \text{for a.e. } x \in \Omega.$$

Then, $\mathscr{F}[u] = \overline{\mathscr{F}}[\delta_{\nabla u}]$.

(ii) ***Lower bound:*** *If* $u_j \rightharpoonup u$ *in* $\mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$*, then, up to a subsequence, there exists a Young measure* $\nu \in \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$ *such that* $\nabla u_j \xrightarrow{\mathrm{Y}} \nu$ *(i.e.* (∇u_j) *generates* ν*) with* $[\nu] = \nabla u$ *a.e., and*

$$\liminf_{j\to\infty} \mathscr{F}[u_j] \geq \overline{\mathscr{F}}[\nu]. \tag{5.12}$$

(iii) ***Recovery sequence:*** *For all* $\nu \in \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$ *there exists a recovery generating sequence* $(u_j) \subset \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ *with* $\nabla u_j \xrightarrow{\mathrm{Y}} \nu$ *and such that*

$$\lim_{j\to\infty} \mathscr{F}[u_j] = \overline{\mathscr{F}}[\nu]. \tag{5.13}$$

(iv) ***Equality of minima:*** $\overline{\mathscr{F}}$ *attains its minimum and*

$$\min_{\mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})} \overline{\mathscr{F}} = \inf_{\mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)} \mathscr{F}.$$

Furthermore, all these statements remain true if we prescribe boundary values. For a Young measure we here prescribe the boundary values for an underlying deformation (which is only determined up to a translation, of course).

Proof. Ad (i). This follows directly from the definition of $\overline{\mathscr{F}}$.

Ad (ii). The existence of ν follows from the Fundamental Theorem 3.11. The lower bound (5.12) is a consequence of Proposition 3.29.

Ad (iii). By virtue of Lemma 3.22 we may construct an (improved) sequence $(u_j) \subset \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ such that

(i) $u_j|_{\partial\Omega} = u$, where $u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ is an underlying deformation of ν, that is, $\nabla u = [\nu]$,

(ii) the family $\{\nabla u_j\}_j$ is p-equiintegrable, and

(iii) $\nabla u_j \xrightarrow{\mathrm{Y}} \nu$.

For this improved generating sequence we can now use the statement about representation of limits of integral functionals from the Fundamental Theorem 3.11 (here we need the p-equiintegrability) to get

$$\mathscr{F}[u_j] \to \langle\!\langle f, \nu \rangle\!\rangle = \overline{\mathscr{F}}[\nu],$$

which is nothing else than (5.13).

Ad (IV). This is not hard to see using (ii), (III) and the coercivity assumption, that is, the lower bound in (5.11). □

Comment...

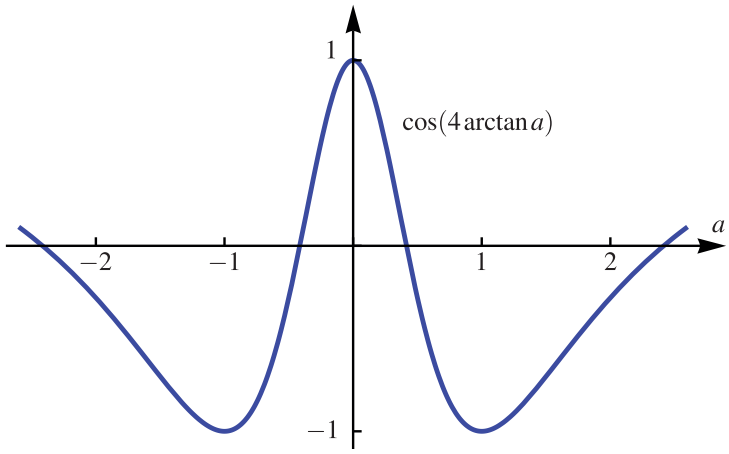

Figure 5.2: The double-well potential in the sailing example.

Example 5.7. In our sailing example from Section 1.7, we were tasked with solving

$$\begin{cases} \mathscr{F}[r] := \displaystyle\int_0^T \cos(4\arctan r'(t)) - v_{\max}\left(1 - \frac{r^2}{R^2}\right) \mathrm{d}t \quad \to \quad \min, \\ r(0) = r(T) = 0, \qquad |r(t)| \le R. \end{cases}$$

Here, because of the additional constraints we can work in any L^p-space that we want, even L^∞. Clearly, the integrand

$$W(r,a) := \cos(4\arctan a) - v_{\max}\left(1 - \frac{r^2}{R^2}\right), \qquad |r(t)| \le R,$$

is not convex in a, see Figure 5.2. Technically, the preceding theorem is not applicable, since W depends on r and a and $p = \infty$, but it is obvious that we can simply extend it to consider Young measures $(\delta_{r(x)} \otimes \nu_x)_x \in \mathbf{Y}^\infty(\Omega; \mathbb{R} \times \mathbb{R})$ with $\nu \in \mathbf{GY}^\infty(\Omega; \mathbb{R})$ and $[\nu] = r'$ as in Section 3.7 (we omit the details for $p = \infty$). We collect all such product Young measures in the set $\mathbf{GY}^\infty(\Omega; \mathbb{R} \times \mathbb{R})$.

Then, we consider the extended–relaxed variational problem

$$\begin{cases} \overline{\mathscr{F}}[\delta_r \otimes \nu] := \langle\!\langle W, \delta_r \otimes \nu \rangle\!\rangle = \displaystyle\int_0^T W(r(t), a)\, \mathrm{d}\nu_x(a)\, \mathrm{d}x \quad \to \quad \min, \\ \delta_r \otimes \nu = (\delta_{r(x)} \otimes \nu_x)_x \in \mathbf{GY}^\infty(\Omega; \mathbb{R} \times \mathbb{R}), \\ r(0) = r(T) = 0, \qquad |r(t)| \le R. \end{cases}$$

Let us also construct a sequence of solutions that generate the optimal Young measure solution. Some elementary analysis yields that $g(a) := \cos(4\arctan a)$ has two minima in $a = \pm 1$, where $g(a) = -1$ (see Figure 5.2). Moreover, $h(r) := -v_{\max}(1 - r^2/R^2)$ attains its minimum $-v_{\max}$ for $r = 0$. Thus,

$$W \ge -\big(1 + v_{\max}\big) =: W_{\min}.$$

Comment...

We let

$$h(s) := \begin{cases} s-1 & \text{if } s \in [0,2], \\ 3-s & \text{if } s \in (2,4], \end{cases} \qquad s \in [0,4],$$

and consider h to be extended to all $s \in \mathbb{R}$ by periodicity. Then set

$$r_j(t) := \frac{T}{4j} h\left(\frac{4j}{T}t + 1\right).$$

It is easy to see that $r_j' \in \{-1,+1\}$ and $r_j \to 0$ uniformly. Thus,

$$\mathscr{F}[r_j] \to T \cdot W_{\min} = \inf \mathscr{F}.$$

By the Fundamental Theorem 3.11 on Young measures, we know that we may select a subsequence of j's (not relabeled) such that (see Lemma 3.32)

$$(r_j, r_j') \overset{\mathrm{Y}}{\to} \delta_r \otimes \nu = (\delta_{r(x)} \otimes \nu_x)_x \in \mathbf{GY}^\infty(\Omega; \mathbb{R} \times \mathbb{R}).$$

From the construction of r_j we see

$$\delta_r \otimes \nu = \delta_0 \otimes \left(\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1}\right).$$

Clearly, this $\delta_r \otimes \nu$ is minimizing for $\overline{\mathscr{F}}$.

The following lemma creates a connection to the other relaxation approach from the last section:

Lemma 5.8. *Let $1 < p < \infty$ and let $h \colon \mathbb{R}^{m\times d} \to \mathbb{R}$ be continuous and satisfy*

$$\mu|A|^p - \mu^{-1} \le h(A) \le M(1+|A|^p) \qquad A \in \mathbb{R}^{m\times d},$$

for some $\mu, M > 0$. Then, for all $A \in \mathbb{R}^{m\times d}$ there exists a homogeneous gradient Young measure $\nu^A \in \mathbf{GY}^p(B(0,1); \mathbb{R}^{m\times d})$ with $[\nu^A] = A$ and such that

$$Qh(A) = \int h \, \mathrm{d}\nu^A.$$

Proof. According to (5.1) and the remarks following it,

$$Qh(A) := \inf\left\{ \fint_{B(0,1)} h(A + \nabla\psi(z)) \, \mathrm{d}z \ : \ \psi \in \mathrm{W}_0^{1,p}(B(0,1); \mathbb{R}^m) \right\}.$$

Let now $(\psi_j) \subset \mathrm{W}^{1,p}(B(0,1); \mathbb{R}^m)$ be a minimizing sequence in the above formula. This minimization problem is admissible in Theorem 5.6, which yields a Young measure minimizer $\nu \in \mathbf{GY}^p(B(0,1); \mathbb{R}^m)$ such that

$$Qh(A) = \fint_{B(0,1)} \int h \, \mathrm{d}\nu_x \, \mathrm{d}x.$$

It only remains to show that we can replace (ν_x) by a *homogeneous* Young measure ν^A. This, however, follows directly from the Averaging Lemma 3.23. □

Comment...

For the case $p = \infty$, we prove the following truncation result.

Lemma 5.9 (Zhang). *Let $\nu \in \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$, $1 < p \le \infty$ such that there exists a compact set $K \subset \mathbb{R}^{m\times d}$ with $\operatorname{supp}\nu_x \subset K$ for a.e. $x \in \Omega$. Then, $\nu \in \mathbf{GY}^\infty(\Omega;\mathbb{R}^{m\times d})$. More precisely, there exists a sequence $(u_j) \subset \mathrm{W}^{1,\infty}(\Omega;\mathbb{R}^m)$ with $\nabla u_j \overset{\mathrm{Y}}{\to} \nu$ and*

$$\|\nabla u_j\|_{\mathrm{L}^\infty} \le C|K|_\infty, \qquad \textit{where} \qquad |K|_\infty := \sup\big\{\, |A| \;:\; A \in K \,\big\}$$

and $C = C(d,m)$ is a dimensional constant.

Proof. We may without loss of generality assume that K is convex.

Step 1. We assume first that there is $(v_j) \subset \mathrm{W}^{1,p}(\mathbb{R}^d;\mathbb{R}^m)$ with $v_j \rightharpoonup 0$ in $\mathrm{W}^{1,p}$ and $\nabla v_j \overset{\mathrm{Y}}{\to} \nu$. Define, using the maximal function from Section A.5,

$$V_j := M\big(|v_j| + |\nabla v_j|\big), \qquad j \in \mathbb{N},$$

and

$$G_j := \big\{\, x \in \Omega \;:\; V_j(x) \le \lambda \,\big\}, \qquad \text{where} \qquad \lambda := 8|K|_\infty.$$

Theorem A.29 implies that v_j is Lipschitz continuous on G with Lipschitz constant $C\lambda$. Hence, by the Kirszbraun Theorem A.27, we may extend $v_j|_G$ to a Lipschitz function $u_j \in \mathrm{W}^{1,\infty}(\Omega;\mathbb{R}^m)$ without changing the Lipschitz constant.

By the weak-type estimate on the maximal function, see Theorem A.29, we get

$$\begin{aligned}
|\Omega \setminus G_j| &\le \big|\big\{\, x \in \Omega \;:\; |Mv_j| > \lambda/2 \,\big\}\big| + \big|\big\{\, x \in \Omega \;:\; |M(\nabla v_j)| > \lambda/2 \,\big\}\big| \\
&\le \frac{C}{\lambda}\int_\Omega |v_j|\,\mathrm{d}x + \frac{C}{\lambda}\int_{\{|\nabla v_j| \ge \lambda/4\}} |\nabla v_j|\,\mathrm{d}x \\
&\le C\int_\Omega |v_j|\,\mathrm{d}x + C\int_\Omega g(|\nabla v_j|)\,\mathrm{d}x, \qquad (5.14)
\end{aligned}$$

where $g\colon [0,\infty) \to [0,\infty)$ is given as

$$g(s) := \begin{cases} 0 & \text{if } s < |K|_\infty, \\ 2(s - |K|_\infty) & \text{if } |K|_\infty \le s < 2|K|_\infty, \\ s & \text{if } s \ge 2|K|_\infty. \end{cases}$$

The first term in (5.14) goes to zero since $v_j \to 0$ in L^1 by the compact embedding from $\mathrm{W}^{1,p}$ into L^1, see the Rellich–Kondrachov Theorem A.23. By the Young measure representation (note that $g(|\nabla v_j|)$ is p-equiintegrable), the second term converges to

$$\int_\Omega \int g(|A|)\,\mathrm{d}\nu_x(A)\,\mathrm{d}x = 0$$

since $\operatorname{supp}\nu_x \subset K$. Therefore, for all $\varphi \in \mathrm{C}_0(\Omega)$ and $h \in \mathrm{C}_0(\mathbb{R}^m)$,

$$\int_\Omega |\varphi h(\nabla u_k) - \varphi h(\nabla v_k)|\,\mathrm{d}x \le \|\varphi \otimes h\|_\infty \cdot |\Omega \setminus G_k| \;\to\; 0$$

Comment...

and we may conclude that also (∇u_j) generates ν, which therefore lies in $\mathbf{GY}^\infty(\Omega;\mathbb{R}^{m\times d})$.

Step 2. If $[\nu] = \nabla u \neq 0$ for some $u \in \mathrm{W}^{1,\infty}(\Omega;\mathbb{R}^m)$ (since $\nabla u(x) \in K$ a.e.), we consider the **shifted Young measure** $\hat{\nu} = (\hat{\nu}_x)_x \in \mathbf{Y}^p(\Omega;\mathbb{R}^{m\times d})$ defined via $\hat{\nu}_x := \nu_x \star \delta_{-\nabla u(x)}$, that is

$$\int h \,\mathrm{d}\hat{\nu}_x = \int h(A - \nabla u(x)) \,\mathrm{d}\nu_x, \qquad \text{for all } h \in \mathrm{C}_0(\mathbb{R}^{m\times d}).$$

Then, since $\nabla u(x) \in K$ (since we assumed that K is convex) we have

$$\operatorname{supp} \hat{\nu}_x \subset K - K := \left\{ A - B \;:\; A, B \in K \right\} \subset B(0, 2|K|_\infty).$$

Thus, the first step applies to $\hat{\nu}$ and yields a sequence $(\hat{u}_j) \subset \mathrm{W}^{1,\infty}(\Omega;\mathbb{R}^m)$ with $\nabla \hat{u}_j \xrightarrow{\mathrm{Y}} \hat{\nu} \in \mathbf{GY}^\infty(\Omega;\mathbb{R}^{m\times d})$ and $\|\nabla \hat{u}_j\|_{\mathrm{L}^\infty} \leq 2C|K|_\infty$. Thus, setting

$$u_j := \hat{u}_j + u,$$

we get $\|\nabla \hat{u}_j\|_{\mathrm{L}^\infty} \leq (2C+1)|K|_\infty$ and $\nabla u_j \xrightarrow{\mathrm{Y}} \nu \in \mathbf{GY}^\infty(\Omega;\mathbb{R}^{m\times d})$. This concludes the proof. □

Remark 5.10. The result as stated here is far from being sharp. A refined version (with a more complicated proof) shows that in fact the constant $C = C(d,m)$ can be chosen arbitrarily close to 1 and for the sequence (u_j) constructed in the proof one can achieve $\operatorname{dist}(\nabla u_j, K) \to 0$ in L^∞. This is proved in [84].

5.4 Rigidity for gradients

Let us return to the question, raised in Chapter 3, whether every Young measure is a gradient Young measure, but now with the additional requirement that the Young measure in question is *homogeneous*. Then, the barycenter is constant and hence trivially a gradient, so the simple reasoning from the beginning of the section no longer applies. Still, there are homogeneous Young measures that are not gradients, but proving that no generating sequence of gradients can be found is not so trivial. One possible way would be to show that there is a quasiconvex function $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ such that the Jensen-type inequality of Lemma 3.25 fails; this is, however, not easy.

Let us demonstrate this assertion instead through the so-called **(approximate) two-gradient problem**: Let $A, B \in \mathbb{R}^{m\times d}$, $\theta \in (0,1)$ and consider the homogeneous Young measure

$$\nu := \theta\delta_A + (1-\theta)\delta_B \quad \in \mathbf{Y}^\infty(B(0,1);\mathbb{R}^{m\times d}). \tag{5.15}$$

We know from Example 3.18 that for $\operatorname{rank}(A-B) \leq 1$, ν is a gradient Young measure. In any case, $\operatorname{supp}\nu = \{A, B\}$ and by Lemma 3.21 it follows that $\operatorname{dist}(V_j, \{A,B\}) \to 0$ in measure for any generating sequence $V_j \xrightarrow{\mathrm{Y}} \nu$.

The following is the first instance of so-called *rigidity* results, which play a prominent role in the field:

Comment...

Theorem 5.11 (Ball–James 1987). *Let $\Omega \subset \mathbb{R}^d$ be open and bounded. Also, let $A, B \in \mathbb{R}^{m\times d}$.*

(i) Let $u \in \mathrm{W}^{1,\infty}(\Omega;\mathbb{R}^m)$ satisfy the ***exact two-gradient inclusion***

$$\nabla u \in \{A, B\} \quad \text{a.e. in } \Omega.$$

(a) If $\operatorname{rank}(A-B) \geq 2$*, then $\nabla u = A$ a.e. or $\nabla u = B$ a.e.*

(b) If $B - A = a \otimes n$ ($a \in \mathbb{R}^m$, $n \in \mathbb{R}^d \setminus \{0\}$) and Ω additionally is convex*, then there exists a Lipschitz function $h\colon \mathbb{R} \to \mathbb{R}$ with $h' \in \{0,1\}$ almost everywhere, and a constant $v_0 \in \mathbb{R}^m$ such that*

$$u(x) = v_0 + Ax + h(x\cdot n)a.$$

(ii) Let $\operatorname{rank}(A-B) \geq 2$ *and assume that $(u_j) \subset \mathrm{W}^{1,\infty}(\Omega;\mathbb{R}^m)$ is uniformly bounded and satisfies the* ***approximate two-gradient inclusion***

$$\operatorname{dist}(\nabla u_j, \{A, B\}) \to 0 \quad \text{in measure.}$$

If also (u_j) converges weakly to a limit in $\mathrm{W}^{1,\infty}$, then*

$$\nabla u_j \to A \quad \text{in measure} \qquad \text{or} \qquad \nabla u_j \to B \quad \text{in measure.}$$

Proof. Ad (i) (a). Assume after a translation that $B = 0$ and thus $\operatorname{rank} A \geq 2$. Then, $\nabla u = Ag$ for a scalar function $g\colon \Omega \to \mathbb{R}$. Mollifying u (i.e. working with $u_\varepsilon := \rho_\varepsilon \star u$ instead of u and passing to the limit $\varepsilon \downarrow 0$), we may assume that in fact $g \in \mathrm{C}^1(\Omega)$.

The idea of the proof is that the curl of ∇u vanishes, expressed as follows: for all $i, j = 1, \dots, d$ and $k = 1, \dots, m$, it holds that

$$\partial_i (\nabla u)^k_j = \partial_{ij} u^k = \partial_{ji} u^k = \partial_j (\nabla u)^k_i$$

where M^k_j denotes the element in the k'th row and j'th column of the matrix M. For our special $\nabla u = Ag$, this reads as

$$A^k_j \partial_i g = A^k_i \partial_j g. \tag{5.16}$$

Under the assumptions of (i) (a), we claim that $\nabla g = 0$. If otherwise $\xi(x) := \nabla g(x) \neq 0$ for some $x \in \Omega$, then with $a_k(x) := A^k_j / \xi_j(x)$ ($k = 1, \dots, m$) for any j such that $\xi_j(x) \neq 0$, where $a_k(x)$ is well-defined by the relation (5.16), we have

$$A^k_j = a_k(x)\xi_j(x), \qquad \text{i.e.} \qquad A = a(x) \otimes \xi(x).$$

This, however, is impossible if $\operatorname{rank} A \geq 2$. Hence, $\nabla g = 0$ and u is an affine function. This property is also stable under mollification.

Comment...

Ad (i) (b). As in (i) (a), we assume $\nabla u = Ag$ ($B = 0$) and $A = a \otimes n$. Pick any $b \perp n$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t} u(x+tb)\bigg|_{t=0} = \nabla u(x) b = [an^T b] g(x) = 0.$$

This implies that u is constant in direction b. As $b \perp n$ was arbitrary and Ω is assumed convex, $u(x)$ can only depend on $x \cdot n$. This implies the claim.

Ad (ii). Assume once more that $B = 0$ and that there exists a (2×2)-minor M with $M(A) \neq 0$. By assumption there are sets $D_j \subset \Omega$ with

$$\nabla u_j - A \mathbb{1}_{D_j} \to 0 \quad \text{in measure.}$$

Let us also assume that we have selected a subsequence such that

$$u_j \overset{*}{\rightharpoonup} u \quad \text{in } \mathrm{W}^{1,\infty} \qquad \text{and} \qquad \mathbb{1}_{D_j} \overset{*}{\rightharpoonup} \chi \quad \text{in } \mathrm{L}^\infty.$$

In the following we use that under an assumption of uniform L^∞-boundedness convergence in measure implies weak* convergence in L^∞. Indeed, for any $\psi \in \mathrm{L}^1(\Omega)$ and any $\varepsilon > 0$ we have

$$\begin{aligned}
&\int_\Omega (\nabla u_j - A \mathbb{1}_{D_j}) \psi \, \mathrm{d}x \\
&\qquad \leq \left| \left\{ x \in \Omega \,:\, |\nabla u_j(x) - A \mathbb{1}_{D_j}(x)| > \varepsilon \right\} \right| \cdot \|\nabla u_j - A \mathbb{1}_{D_j}\|_{\mathrm{L}^\infty} \cdot \|\psi\|_{\mathrm{L}^1} + \varepsilon \|\psi\|_{\mathrm{L}^1} \\
&\qquad \to 0 + \varepsilon \|\psi\|_{\mathrm{L}^1}
\end{aligned}$$

As $\varepsilon > 0$ is arbitrary, we have indeed $\nabla u_j - A \mathbb{1}_{D_j} \overset{*}{\rightharpoonup} 0$ in L^∞.

Now,

$$\operatorname*{w*-lim}_{j\to\infty} M(\nabla u_j) = \operatorname*{w*-lim}_{j\to\infty} M(A \mathbb{1}_{D_j}) = M(A) \cdot \operatorname*{w*-lim}_{j\to\infty} \mathbb{1}_{D_j} = M(A)\chi.$$

By a similar argument, $\nabla u_j \overset{*}{\rightharpoonup} \nabla u = A\chi$, and so, by the weak* continuity of minors proved in Lemma 3.10,

$$\operatorname*{w*-lim}_{j\to\infty} M(\nabla u_j) = M(\nabla u) = M(A\chi) = M(A)\chi^2.$$

Thus, $\chi = \chi^2$ and we can conclude that there exists $D \subset \Omega$ such that $\chi = \mathbb{1}_D$ and $\nabla u = A \mathbb{1}_D$. Furthermore, combining the above convergence assertions,

$$\nabla u_j \to \nabla u = A \mathbb{1}_D \quad \text{in measure.}$$

Then, part (i) (a) implies that $\nabla u = A$ or $\nabla u = 0 = B$ a.e. in Ω. As we assumed weak* convergence of our original sequence (u_j), the limit of the selected subsequence is unique and we did not have to pass to a subsequence in the first place. □

With this result at hand it is easy to see that our example (5.15) cannot be a gradient Young measure if $\operatorname{rank}(A - B) \geq 2$: If there was a sequence of gradients $\nabla u_j \overset{\mathrm{Y}}{\to} \nu$, then by the reasoning above we would have $\operatorname{dist}(\nabla u_j, \{A, B\}) \to 0$, but from (ii) of the Ball–James rigidity theorem, we would get $\nabla u_j \to A$ or $\nabla u_j \to B$ in measure, either one of which yields a contradiction by Lemma 3.21.

Comment...

5.5 Characterization of gradient Young measures

In the previous section we replaced a minimization problem over a Sobolev space by its extension–relaxation, defined on the space of gradient Young measures. Unfortunately, this subset of the space of Young measures is so far defined only "extrinsically", i.e. through the existence of a generating sequence of gradients. The question arises whether there is also an "intrinsic" characterization of gradient Young measures. Intuitively, trying to understand gradient Young measures amounts to understanding the "structure" of sequences of gradients, which is a useful endeavor.

Recall that in Lemma 3.25 we showed that a homogeneous gradient Young measure $\nu \in \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$ satisfies the Jensen-type inequality

$$h([\nu]) \leq \int h \,\mathrm{d}\nu$$

for all *quasiconvex* functions $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ with p-growth. There, we interpreted this as an expression of the (generalized) convexity of h, in analogy with the classical Jensen-inequality.

However, we can also dualize the situation and see the validity of the above Jensen-type inequality for quasiconvex functions as a property of gradient Young measures. The following (perhaps surprising) result shows that this dual point of view is indeed valid and the Jensen-type inequality (essentially) characterizes gradient Young measures, a result due to Kinderlehrer and Pedregal from the early 1990s:

Theorem 5.12 (Kinderlehrer–Pedregal 1991 & 1994). *Let $\nu \in \mathbf{Y}^p(\Omega;\mathbb{R}^{m\times d})$, $1 < p \leq \infty$, be a Young measure such that $[\nu] = \nabla u$ for some $u \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$. Then, it holds that $\nu \in \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$, i.e. ν is a gradient L^p-Young measure, if and only if for almost every $x \in \Omega$ the Jensen-type inequality*

$$h(\nabla u(x)) \leq \int h \,\mathrm{d}\nu_x \tag{5.17}$$

holds for all quasiconvex $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ with p-growth (no condition if $p = \infty$).

The idea of the proof is to show that the set of gradient Young measures is *convex* and weakly* closed in the set of Young measures and the Jensen-type inequalities entail that ν cannot be separated from this set by a hyperplane (which turn out to be in one-to-one correspondence with integrands). Then, the Hahn–Banach Theorem implies that ν actually lies in this set and is hence a gradient Young measure.

Lemma 5.13. *The set $\mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$ is convex and weakly* closed in the topology of $\mathrm{C}_0(\mathbb{R}^{m\times d})^*$.*

Proof. Step 1: Convexity. Recall that we may define homogeneous Young measures on any bounded Lipschitz domain, see Lemma 3.23; we will use this fact several times in the sequel. So, assume $\Omega = B(0,1)$. Let $\mu_1, \mu_2 \in \mathbf{GY}^p(B(0,1);\mathbb{R}^{m\times d})$ and $\theta \in (0,1)$. Let

Comment...

$D_1 \subset B(0,1)$ be an open Lipschitz domain with

$$|D_1| = \theta |B(0,1)|$$

We assume that μ_1 is a Young measure on D_1 and μ_2 is a Young measure on $D_2 := B(0,1) \setminus D$. Let $(u_j) \subset \mathrm{W}^{1,p}(D_1;\mathbb{R}^m)$, $(v_j) \subset \mathrm{W}^{1,p}(D_2;\mathbb{R}^m)$ with $\nabla u_j \xrightarrow{\mathrm{Y}} \mu_1$, $\nabla v_j \xrightarrow{\mathrm{Y}} \mu_2$. We define $(w_j) \subset \mathrm{W}^{1,p}(B(0,1);\mathbb{R}^m)$ as

$$w_j(x) := \begin{cases} u_j(x) & \text{if } x \in D_1, \\ v_j(x) & \text{if } x \in D_2 \end{cases}$$

and we see that $(w_j) \subset \mathrm{W}^{1,p}(B(0,1);\mathbb{R}^m)$ is uniformly bounded. Thus, up to a subsequence, we may assume that $\nabla w_j \xrightarrow{\mathrm{Y}} \nu \in \mathbf{GY}^p(B(0,1);\mathbb{R}^m)$. For all continuous $h\colon \mathbb{R}^{m\times d} \to \mathbb{R}$ with p-growth we find

$$\begin{aligned} \fint_{B(0,1)} \int h \,\mathrm{d}\nu_x \,\mathrm{d}x &= \frac{1}{|B(0,1)|} \cdot \lim_{j\to\infty} \left[\int_{D_1} h(\nabla u_j(x)) \,\mathrm{d}x + \int_{D_2} h(\nabla v_j(y)) \,\mathrm{d}y \right] \\ &= \theta \int h \,\mathrm{d}\mu_1 + (1-\theta) \int h \,\mathrm{d}\mu_2. \end{aligned}$$

for all h as before. Then use the averaging principle from Lemma 3.23 to get a homogeneous gradient Young measure $\overline{\nu} \in \mathbf{GY}^p(B(0,1);\mathbb{R}^{m\times d})$ with

$$\int h \,\mathrm{d}\overline{\nu} = \fint_{B(0,1)} \int h \,\mathrm{d}\nu_x \,\mathrm{d}x = \theta \int h \,\mathrm{d}\mu_1 + (1-\theta) \int h \,\mathrm{d}\mu_2$$

and the convexity of $\mathbf{GY}^p(B(0,1);\mathbb{R}^{m\times d})$ follows.

Step 1: Weak-closedness.* Note that we need to show weak* *topological* closedness, not just the sequential closedness.

Take a countable collection $\{\varphi_k \otimes h_l\}_{k,l}$ as in Lemma 3.19. If $\nu \in \mathbf{Y}^p(B(0,1);\mathbb{R}^{m\times d})$ lies in the topological closure of $\mathbf{GY}^p(B(0,1);\mathbb{R}^{m\times d})$, then for all $j \in \mathbb{N}$ there exists $\nu_j \in \mathbf{GY}^p(B(0,1);\mathbb{R}^{m\times d})$ with

$$\left| \int_{B(0,1)} \varphi_k(x) \int h_l(A) \,\mathrm{d}[(\nu_j)_x - \nu_x](A) \,\mathrm{d}x \right| \le \frac{1}{j} \qquad \text{for all } k,l \le j.$$

For every ν_j we may find $\nabla u_j \in \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ such that

$$\left| \int_{B(0,1)} \varphi_k(x) h_l(\nabla u_j(x)) \,\mathrm{d}x - \int_{B(0,1)} \varphi_k(x) \int h_l \,\mathrm{d}\nu_x \,\mathrm{d}x \right| \le \frac{1}{j} \qquad \text{for all } k,l \le j.$$

Moreover, we may assume, adding $(\varphi_0 \otimes h_0)(x,A) := |A|^p$ to our collection $\{\varphi_k \otimes h_l\}_{k,l}$ and using the Poincaré–Friedrichs inequality, that the sequence (u_j) is uniformly $\mathrm{W}^{1,p}$-bounded. Hence it generates a gradient Young measure $\tilde{\nu} \in \mathbf{GY}^p(B(0,1);\mathbb{R}^{m\times d})$. By Lemma 3.19 we must have $\nu = \tilde{\nu}$ and the assertion follows. □

Comment...

Proof of Theorem 5.12. By Lemma 3.25 only the sufficiency of (5.17) remains to be proved.

Step 1. We first show the result for homogeneous Young measures, so assume that $\mu = (\mu_x) \in \mathbf{Y}^p(B(0,1);\mathbb{R}^{m\times d}) \subset \mathscr{M}(\mathbb{R}^{m\times d})$ satisfies

$$h(A_0) \leq \int h \,\mathrm{d}\mu, \qquad A_0 := [\mu]. \tag{5.18}$$

Assume that μ is not a gradient Young measure. From the preceding lemma we know that $\mathbf{GY}^p(B(0,1);\mathbb{R}^{m\times d})$ is convex and weakly* closed in the topology of $\mathrm{C}_0(\mathbb{R}^{m\times d})^*$. Then, applying the Hahn–Banach separation theorem in the version of Theorem A.19 ($X := \mathrm{C}_0(\mathbb{R}^{m\times d})$, there is $h \in \mathrm{C}_0(\mathbb{R}^{m\times d})$ such that

$$\int h \,\mathrm{d}\mu < \inf_{\nu\in\mathbf{GY}^p(B(0,1);\mathbb{R}^{m\times d})} \int h \,\mathrm{d}\nu.$$

In particular, we may test with all $\nu := \overline{\delta_{A_0+\nabla\psi}}$ for $\psi \in \mathrm{W}_0^{1,\infty}(B(0,1))$ to see via (5.1) (also cf. Lemma 5.8)

$$\int h \,\mathrm{d}\mu < Qh(A_0) \leq h(A_0).$$

However, this contradicts (5.18).

Step 2. We now treat the inhomogeneous case for $1 < p < \infty$ and assuming that $u = 0$ almost everywhere. So let $\nu = (\nu_x) \in \mathbf{Y}^p(\Omega;\mathbb{R}^{m\times d})$ with (5.17). Take a countable collection $\{\varphi_k \otimes h_l\}_{k,l}$ as in Lemma 3.19. By the first step we have that $\nu_x \in \mathbf{GY}^p(B(0,1);\mathbb{R}^{m\times d})$ for almost every $x \in \Omega$. Moreover, almost every $x \in \Omega$ is a Lebesgue point of the functions

$$x \mapsto \langle h_l, \nu_x\rangle \qquad \text{and} \qquad x \mapsto \langle |\cdot|^p, \nu_x\rangle,$$

see Theorem A.9.

Fix $\varepsilon > 0$ and cover Ω (removing the points x at which (5.17) does not hold) with a countable collection of disjoint balls $B(a_k, r_k)$, where $a_k \in \Omega$, $r_k > 0$ ($k \in \mathbb{N}$) such that

$$\Omega = Z \cup \bigcup_{k=1}^{N} B(a_k, r_k), \qquad |Z| = 0,$$

and

$$\left| \fint_{B(a_k,r_k)} \langle h_l, \nu_y\rangle - \langle h_l, \nu_{a_k}\rangle \right| + \left| \fint_{B(a_k,r_k)} \langle |\cdot|^p, \nu_y\rangle \,\mathrm{d}y - \langle |\cdot|^p, \nu_{a_k}\rangle \right| \leq \varepsilon. \tag{5.19}$$

The last condition can be achieved by the Lebesgue point property and the fact that in the Vitali cover we may choose every radius r_k arbitrarily small (how small may depend on x).

For each $k \in \mathbb{N}$ take a generating sequence $(v_j^{(k)}) \subset \mathrm{W}_0^{1,p}(B(0,1);\mathbb{R}^m)$ with $\nabla v_j^{(k)} \xrightarrow{\mathrm{Y}} \nu_{a_k}$, cf. Lemma 3.23 and Lemma 3.22. Define for $j \in \mathbb{N}$,

$$w_j^{\varepsilon}(x) := r_k v_j^{(k)}\left(\frac{x-a_k}{r_k}\right) \quad \text{if } x \in B(a_k, r_k), \qquad x \in \Omega.$$

Comment...

We estimate

$$\begin{aligned}\int_\Omega |\nabla w_j^\varepsilon|^p \,\mathrm{d}x &= \sum_{k\in\mathbb{N}} \int_{B(a_k,r_k)} \left|\nabla v_j^{(k)}\left(\frac{x-a_k}{r_k}\right)\right|^p \mathrm{d}x \\ &= \sum_{k\in\mathbb{N}} r_k^d \int_{B(0,1)} |\nabla v_j^{(k)}|^p \,\mathrm{d}y \\ &\le \sum_{k\in\mathbb{N}} |B(a_k,r_k)| \left\langle |\cdot|^p, \nu_{a_k}\right\rangle + |\Omega| \\ &\le \sum_{k\in\mathbb{N}} \int_{B(a_k,r_k)} \left\langle |\cdot|^p, \nu_y\right\rangle \mathrm{d}y + (1+\varepsilon)|\Omega| < \infty\end{aligned}$$

where we discarded some leading elements of $(v_j^{(k)})$ and also used (5.19). By the Poincaré–Friedrich inequality from Theorem A.21, we thus get that (w_j^ε) is uniformly $\mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$-bounded and so, selecting a subsequence, we may assume that

$$\nabla w_j^\varepsilon \xrightarrow{\mathrm{Y}} \nu^\varepsilon \in \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d}).$$

By a similar calculation as above and also using the uniform continuity of φ_k, and (5.19), we get

$$\begin{aligned}\left\langle\!\left\langle \varphi_k \otimes h_l, \nu^\varepsilon \right\rangle\!\right\rangle &= \lim_{j\to\infty} \int_\Omega \varphi_k(x) h_l(\nabla w_j^\varepsilon(x)) \,\mathrm{d}x \\ &= \lim_{j\to\infty} \sum_{k\in\mathbb{N}} r_k^d \left[\int_{B(0,1)} \varphi_k(a_k) h_j(\nabla v_j^{(k)}(y)) \,\mathrm{d}y + E(\varepsilon)\right] \\ &= \sum_{k\in\mathbb{N}} |B(a_k,r_d)| \cdot \varphi_k(a_k) \cdot \left\langle h_j, \nu_{a_k}\right\rangle + E(\varepsilon) \\ &= \sum_{k\in\mathbb{N}} \varphi_k(a_k) \int_{B(a_k,r_k)} \left\langle h_j, \nu_y\right\rangle \mathrm{d}y + E(\varepsilon) \\ &= \int_\Omega \varphi(y) \left\langle h_j, \nu_y\right\rangle \mathrm{d}y + E(\varepsilon)\end{aligned}$$

Here, $E(\varepsilon)$ is an error term that may change from line to line and vanishes as $\varepsilon \downarrow 0$. Thus, as $\varepsilon \downarrow 0$ we get that

$$\nu^\varepsilon \to \nu \quad \text{in } \mathbf{Y}^p(\Omega;\mathbb{R}^{m\times d}).$$

As all the ν^ε are gradient Young measures, a diagonal argument yields that also ν is a gradient Young measure.

Step 3. Let now $1 < p < \infty$ and u not equal to zero. Then define the shifted Young measure $\hat{\nu} = (\hat{\nu}_x)_x \in \mathbf{Y}^p(\Omega;\mathbb{R}^{m\times d})$ through $\hat{\nu}_x := \nu_x \star \delta_{-\nabla u(x)}$, that is

$$\int h \,\mathrm{d}\hat{\nu}_x = \int h(A - \nabla u(x)) \,\mathrm{d}\nu_x, \qquad \text{for all } h \in \mathrm{C}_0(\mathbb{R}^{m\times d}).$$

We have $[\hat{\nu}] = 0$ and the Jensen-inequalities still hold for $\hat{\nu}$:

$$h([\hat{\nu}]) = h(0) = h([\nu] - \nabla u(x)) \le \int h(A - \nabla u(x)) \,\mathrm{d}\nu_x(A) = \int h \,\mathrm{d}\hat{\nu}_x.$$

Comment...

Thus, the previous step applies. Consequently, $\hat{\nu} \in \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$ and there is $(w_j) \subset \mathrm{W}^{1,p}(\Omega;\mathbb{R}^m)$ such that $\nabla w_j \xrightarrow{\mathrm{Y}} \hat{\nu}$. Then,

$$u_j(x) := w_j(x) + u(x)$$

generates ν and so $\nu \in \mathbf{GY}^p(\Omega;\mathbb{R}^{m\times d})$. This finishes the proof for $1 < p < \infty$.

Step 4. For the sufficiency part of Theorem 5.12 in the case $p = \infty$, we simply apply the result for any exponent $1 < q < \infty$. This yields that $\nu \in \mathbf{GY}^q(\Omega;\mathbb{R}^{m\times d})$. On the other hand, since $\nu \in \mathbf{Y}^\infty(\Omega;\mathbb{R}^{m\times d})$, there exists a compact set $K \subset \mathbb{R}^{m\times d}$ with $\operatorname{supp} \nu_x \subset K$ for almost every $x \in \Omega$. Thus, Zhang's Lemma 5.9 yields that also $\nu \in \mathbf{GY}^\infty(\Omega;\mathbb{R}^{m\times d})$. □

The Kinderlehrer–Pedregal theorem is conceptually very important. It entails that if we could understand the class of quasiconvex functions, then we also could understand gradient Young measures and thus sequences of gradients. Unfortunately, our knowledge of quasiconvex functions (and hence of gradient Young measures) is very limited at present. Most results are for specific situations only and much of it has an "ad-hoc" flavor.

5.6 Quasiconvexity versus rank-one convexity

We close this chapter with a result showing that quasiconvex functions are not easy to understand. In Proposition 3.3 we proved that quasiconvexity implies rank-one convexity and Proposition 4.1 showed that polyconvexity implies quasiconvexity. Hence, in a sense, rank-one convexity is a lower bound and polyconvexity an upper bound on quasiconvexity. Moreover, the Alibert–Dacorogna–Marcellini function (see Examples 3.5, 4.2)

$$h_\gamma(A) := |A|^2\big(|A|^2 - 2\gamma \det A\big), \qquad A \in \mathbb{R}^{2\times 2},\ \gamma \in \mathbb{R}.$$

has the following convexity properties:

- h_γ is convex if and only if $|\gamma| \le \dfrac{2\sqrt{2}}{3} \approx 0.94$,
- h_γ is rank-one convex if and only if $|\gamma| \le \dfrac{2}{\sqrt{3}} \approx 1.15$,
- h_γ is quasiconvex if and only if $|\gamma| \le \gamma_{\mathrm{QC}}$ for some $\gamma_{\mathrm{QC}} \in \left(1, \dfrac{2}{\sqrt{3}}\right]$,
- h_γ is polyconvex if and only if $|\gamma| \le 1$.

However, it is open whether $\gamma_{\mathrm{QC}} = \frac{2}{\sqrt{3}}$ and so it could be that quasiconvexity and rank-one convexity are actually equivalent. This is one of the central questions of the field:

Conjecture 5.14 (Morrey 1952). *Rank-one convexity does not imply quasiconvexity.*

Comment...

This question was completely open for a long time until Švérak's 1992 counterexample [103], which settled the question in the negative, at least for $d \geq 2$, $m \geq 3$, see below. The case $d = m = 2$ is still open and considered to be very important, also because of its connection to other branches of mathematics (a partial result for diagonal matrices is in [83]).

Example 5.15 (Švérak). We will show the non-equivalence of quasiconvexity and rank-one convexity for $m = 3$, $d = 2$ only. Higher dimensions can be treated using an embedding of $\mathbb{R}^{3\times 2}$ into $\mathbb{R}^{m\times d}$. To accomplish this, we construct $h\colon \mathbb{R}^{3\times 2} \to \mathbb{R}$ that is quasiconvex, but not rank-one convex. Define a linear subspace L of $\mathbb{R}^{3\times 2}$ as follows:

$$L := \left\{ \begin{pmatrix} x & 0 \\ 0 & y \\ z & z \end{pmatrix} \;:\; x,y,z \in \mathbb{R} \right\}.$$

We denote by $P\colon \mathbb{R}^{3\times 2} \to L$ a linear projection onto L given as

$$P(A) := \begin{pmatrix} a & 0 \\ 0 & d \\ (e+f)/2 & (e+f)/2 \end{pmatrix} \qquad \text{for} \qquad A = \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix} \in \mathbb{R}^{3\times 2}.$$

Also define $g\colon L \to \mathbb{R}$ to be

$$g \begin{pmatrix} x & 0 \\ 0 & y \\ z & z \end{pmatrix} := -xyz.$$

For $\alpha, \beta > 0$, we let $h_{\alpha,\beta}\colon \mathbb{R}^{3\times 2} \to \mathbb{R}$ be given as

$$h_{\alpha,\beta}(A) := g(P(A)) + \alpha\big(|A|^2 + |A|^4\big) + \beta |A - P(A)|^2.$$

We show the following two properties of $h_{\alpha,\beta}$:

(i) For every $\alpha > 0$ sufficiently small and all $\beta > 0$, the function $h_{\alpha,\beta}$ is *not* quasiconvex.

(ii) For every $\alpha > 0$, there exists $\beta = \beta(\alpha) > 0$ such that $h_{\alpha,\beta}$ is rank-one convex.

Ad (i): For $A = 0$, $D = (0,1)^2$, and a periodic $\psi \in \mathrm{W}^{1,\infty}(D;\mathbb{R}^3)$ given as

$$\psi(x_1,x_2) := \frac{1}{2\pi} \begin{pmatrix} \sin 2\pi x_1 \\ \sin 2\pi x_2 \\ \sin 2\pi(x_1 + x_2) \end{pmatrix},$$

we have $\nabla\psi \in L$ and so $P(\nabla\psi) = \nabla\psi$. Thus, we may compute (using the identity for the cosine of a sum and the fact that the sine is an odd function)

$$\int_D g(\nabla\psi)\,\mathrm{d}x = -\int_0^1 \int_0^1 (\cos 2\pi x_1)^2 (\cos 2\pi x_2)^2 \,\mathrm{d}x_1\,\mathrm{d}x_2 = -\frac{1}{4} < 0.$$

Comment...

Then, for $\alpha > 0$ sufficiently small and all $\beta > 0$,

$$\int_D h_{\alpha,\beta}(\nabla\psi)\,\mathrm{d}x < 0 = h_{\alpha,\beta}(0). \tag{5.20}$$

It turns out that in the definition of quasiconvexity we may alternatively test with functions that have merely periodic boundary values on D (see Proposition 5.13 in [29]). Hence, (i) follows.

Ad (ii): Rank-one convexity is equivalent to the **Legendre–Hadamard condition**

$$\mathrm{D}^2 h_{\alpha,\beta}(A)[B,B] := \frac{\mathrm{d}^2}{\mathrm{d}t^2} h_{\alpha,\beta}(A+tB)\bigg|_{t=0} \geq 0 \qquad \text{for all } A,B \in \mathbb{R}^{3\times 2},\ \operatorname{rank} B \leq 1.$$

Here, $\mathrm{D}^2 h_{\alpha,\beta}(A)[B,B]$ is the second (Gâteaux-)derivative of $h_{\alpha,\beta}$ at A in direction B.

The function g is a homogeneous polynomial of degree three, whereby we can find $c > 0$ such that for all $A,B \in \mathbb{R}^{3\times 2}$ with $\operatorname{rank} B \leq 1$,

$$G(A,B) := \mathrm{D}^2(g\circ P)(A)[B,B] := \frac{\mathrm{d}^2}{\mathrm{d}t^2} g(P(A+tB))\bigg|_{t=0} \geq -c|A||B|^2.$$

A computation then shows that

$$\begin{aligned}\mathrm{D}^2 h_{\alpha,\beta}(A)[B,B] &= G(A,B) + 2\alpha|B|^2 + 4\alpha|A|^2|B|^2 + 8\alpha(A:B)^2 + 2\beta|B-P(B)|^2 \\ &\geq (-c + 4\alpha|A|)|A||B|^2.\end{aligned}$$

where we used the Frobenius product $A : B := \sum_{i,j} A^i_j B^i_j$. Thus, for $|A| \geq c/(4\alpha)$, the Legendre–Hadamard condition (and hence rank-one convexity) holds.

We still need to prove the Legendre–Hadamard condition for $|A| < c/(4\alpha)$ and any fixed $\alpha > 0$. Since $\mathrm{D}^2 h_{\alpha,\beta}(A)[B,B]$ is homogeneous of degree two in B, we only need to consider A,B from the compact set

$$K := \left\{ (A,B) \in \mathbb{R}^{3\times 2} \times \mathbb{R}^{3\times 2} \ :\ |A| \leq \frac{c}{4\alpha},\ |B| = 1,\ \operatorname{rank} B = 1 \right\}.$$

From the estimate

$$\mathrm{D}^2 h_{\alpha,\beta}(A)[B,B] \geq G(A,B) + 2\alpha|B|^2 + 2\beta|B-P(B)|^2 =: \kappa(A,B,\beta)$$

it suffices to show that there exists $\beta = \beta(\alpha) > 0$ such that $\kappa(A,B,\beta) \geq 0$ for all $(A,B) \in K$. Assume that this is not the case. Then there exists $\beta_j \to \infty$ and $(A_j,B_j) \in K$ with

$$0 > \kappa(A_j,B_j,\beta_j) = G(A_j,B_j) + 2\alpha + 2\beta|B_j - P(B_j)|^2.$$

As K is compact, we may assume $(A_j,B_j) \to (A,B) \in K$, for which it holds that

$$G(A,B) + 2\alpha \leq 0, \qquad P(B) = B, \qquad \text{and} \qquad \operatorname{rank} B = 1.$$

However, since $\operatorname{rank} B = 1$, we have $g(P(A+tB)) = 0$ for all $t \in \mathbb{R}$. This implies in particular $G(A,B) = \mathrm{D}^2(g\circ P)(A)[B,B] = 0$, yielding $2\alpha \leq 0$, which contradicts our assumption $\alpha > 0$. Thus, we have shown that for a suitable choice of $\alpha,\beta > 0$ it indeed holds that $h_{\alpha,\beta}$ is rank-one convex.

Comment...

This also shows that the theory of elliptic systems of PDEs, for which traditionally the Legendre–Hadamard condition is the expression of "ellipticity", is incomplete. In particular, non-variational systems (i.e. systems of PDEs which do not arise as the Euler–Lagrange equations to a variational problem) are only poorly understood.

Based on Švérak's example, it has been shown that quasiconvexity is not a local condition. Let $\mathscr{Q}\colon \mathrm{C}^\infty(\mathbb{R}^{m\times d}) \to \mathscr{X}^{m\times d}$ be a nonlinear operator, where $X^{m\times d}$ is the space of functions from $\mathbb{R}^{m\times d}$ to $[-\infty,+\infty]$. Call $\mathscr{Q}$ **local** if $h = g$ in a neighborhood of $A \in \mathbb{R}^{m\times d}$ implies that also $\mathscr{Q}(h) = \mathscr{Q}(g)$ in a neighborhood of A.

Theorem 5.16 (Kristensen 1997). *Let $d \geq 2$ and $m \geq 3$. Then, there exists no local nonlinear operator $\mathscr{Q}\colon \mathrm{C}^\infty(\mathbb{R}^{m\times d}) \to X^{m\times d}$ such that*

$$\mathscr{Q}(h) = 0 \quad \Longleftrightarrow \quad h \text{ is quasiconvex} \qquad (h \in \mathrm{C}^\infty(\mathbb{R}^{m\times d})).$$

This is of course in contrast to rank-one convexity, which is characterized by the local operator

$$\mathscr{R}(h)(A) := \inf\Big\{ \mathrm{D}^2 h(A)[a\otimes b, a\otimes b] \ : \ a \in \mathbb{R}^m,\ b \in \mathbb{R}^d \Big\},$$

where $h \in \mathrm{C}^\infty(\mathbb{R}^{m\times d})$ and $A \in \mathbb{R}^{m\times d}$.

Notes and historical remarks

Classically, one often defines the quasiconvex envelope through (5.2) and not as we have done through (5.1). In this case, (5.1) is often called the **Dacorogna formula**, see Section 6.3 in [29] for further references.

The construction of Lemma 5.2 is based on a construction of Šverák [102] and some ellipticity arguments similar to those in Lemma 2.7 of [85].

Further relaxation formulas can be found in Chapter 11 of the recent text [2]; historically, Dacorogna's lecture notes [28] were also influential.

The Kinderlehrer–Pedregal Theorem 5.12 also holds for $p = \infty$, which was in fact the first case to be established. The original works are [57, 59], also see [94].

A different proof of Proposition 5.2 can be found in Section 5.3.9 of [29].

The conditions (i)–(iii) at the beginning of Section 5.3 are modeled on the concept of Γ-convergence (introduced by De Giorgi) and are also the basis of the usual treatment of parameter-dependent integral functionals. See [18] for an introduction and [30] for a thorough treatise.

Comment...

Appendix A

Prerequisites

This appendix recalls some facts that are needed throughout the course.

A.1 Linear algebra

First, recall the following version of the fundamental **Young inequality**

$$ab \leq \frac{\delta}{2}a^2 + \frac{1}{2\delta}b^2,$$

which holds for all $a, b \geq 0$ and $\delta > 0$.

In this text, the matrix space $\mathbb{R}^{m\times d}$ of $(m \times d)$-matrices always comes equipped with the the **Frobenius matrix (inner) product**

$$A : B := \mathrm{tr}(A^T B) = \mathrm{tr}(AB^T) = \sum_{j,k} A^j_k B^j_k, \qquad A, B \in \mathbb{R}^{m\times d},$$

which is just the Euclidean product if we identify such matrices with vectors in $\mathbb{R}^{md}$. This inner product induces the **Frobenius matrix norm**

$$|A| := \sqrt{\sum_{j,k} (A^j_k)^2}, \qquad A \in \mathbb{R}^{m\times d}.$$

Very often we will use the **tensor product** of vectors $a \in \mathbb{R}^m$, $b \in \mathbb{R}^d$,

$$a \otimes b := ab^T \quad \in \mathbb{R}^{m\times d}.$$

The tensor product interacts well with the Frobenius norm,

$$|a \otimes b| = |a| \cdot |b|, \qquad a \in \mathbb{R}^m,\ b \in \mathbb{R}^d.$$

While all norms on the finite-dimensional space $\mathbb{R}^{m\times d}$ are equivalent, some "finer" arguments require to specify a matrix norm; if nothing else is stated, we always use the Frobenius norm.

We recall the following elementary lemma:

Comment...

Lemma A.1. *Let $M \in \mathbb{R}^{m\times d}$ be a matrix. Then,* $\operatorname{rank} M \leq 1$ *if and only if there exist vectors $a \in \mathbb{R}^m$, $b \in \mathbb{R}^d$ such that $M = a \otimes b$.*

It is proved in texts on matrix analysis, see for instance [56], that this Frobenius norm can also be expressed as

$$|A| = \sqrt{\sum_i \sigma_i(A)^2}, \qquad A \in \mathbb{R}^{m\times d},$$

where $\sigma_i(A) \geq 0$ is the i'th singular value of A, where $i = 1, \ldots, \min(d,m)$. Recall that every matrix $A \in \mathbb{R}^{m\times d}$ has a **(real) singular value decomposition**

$$A = P\Sigma Q^T$$

for orthogonal $P \in \mathbb{R}^{m\times m}$, $Q \in \mathbb{R}^{d\times d}$, and a *diagonal* $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_{\min(d,m)}) \in \mathbb{R}^{m\times d}$ with only positive entries

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0 = \sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_{\min(d,m)},$$

called the **singular values** of A, where r is the rank of A.

For $A \in \mathbb{R}^{d\times d}$ the **cofactor matrix** $\operatorname{cof} A \in \mathbb{R}^{d\times d}$ is the matrix whose (j,k)'th entry is $(-1)^{j+k} M_{\neg k}^{\neg j}(A)$ with $M_{\neg k}^{\neg j}(A)$ being the (j,k)-minor of A, i.e. the determinant of the matrix that originates from A by deleting the j'th row and k'th column. For

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

we have

$$\operatorname{cof} A = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}.$$

One important formula (and in fact another way to define the cofactor matrix) is

$$\partial_A[\det A] = \operatorname{cof} A.$$

Furthermore, **Cramer's rule** entails that

$$A(\operatorname{cof} A)^T = (\operatorname{cof} A)^T A = \det A \cdot \mathrm{Id} \qquad \text{for all } A \in \mathbb{R}^{d\times d},$$

where Id denotes the identity matrix as usual. In particular, if A is invertible,

$$A^{-1} = \frac{(\operatorname{cof} A)^T}{\det A}. \tag{A.1}$$

Sometimes, the matrix $(\operatorname{cof} A)^T$ is called the **adjugate matrix** to A in the literature (we will not use this terminology here, however).

Comment...

One particular consequence of this is **Jacobi's formula**, which says that for a function $A(t)\colon \mathbb{R} \to \mathbb{R}^{d\times d}$ it holds that

$$\frac{\mathrm{d}}{\mathrm{d}t}\det A(t) = \operatorname{cof} A(t) : \frac{\mathrm{d}A(t)}{\mathrm{d}t} = \operatorname{tr}\left[(\operatorname{cof} A(t))^T \frac{\mathrm{d}A(t)}{\mathrm{d}t}\right].$$

In particular, if $A(t) = A_0 + tB$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\det[A_0+tB] = \operatorname{tr}\big[(\operatorname{cof} A_0)^T B + t(\operatorname{cof} B)^T B\big] = (\operatorname{cof} A_0) : B + td\det B.$$

As a consequence, we derive that if $\frac{\mathrm{d}}{\mathrm{d}t}\det[A_0+tB]$ is constant, then necessarily $\det B = 0$.

The **special orthogonal group** $\mathrm{SO}(d)$ is defined as

$$\mathrm{SO}(d) := \big\{\, Q \in \mathbb{R}^{d\times d} \;:\; Q \text{ invertible}, Q^{-1} = Q^T, \det Q = 1 \,\big\}.$$

It has the following useful property, which can be verified via (A.1):

$$\operatorname{cof} Q = Q \qquad \text{for all } Q \in \mathrm{SO}(d).$$

Any $Q \in \mathrm{SO}(2) \subset \mathbb{R}^{2\times 2}$ (a rotation) has the following form

$$Q = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

for some $\theta \in [0, 2\pi)$. For the convex hull $\mathrm{SO}(2)^{**}$ of $\mathrm{SO}(2)$ we get

$$\mathrm{SO}(2)^{**} = \left\{ A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \;:\; a^2 + b^2 \le 1 \right\}.$$

Moreover, one can see from a Taylor expansion and the fact that the Lie-algebra of the Lie group $\mathrm{SO}(2)$ is the vector space of all skew-symmetric matrices that

$$\operatorname{dist}(\mathrm{Id} + A, \mathrm{SO}(2)) \le \frac{1}{2}|A + A^T| + C|A|^2 \tag{A.2}$$

for some $C > 0$.

Any square matrix $A \in \mathbb{R}^{d\times d}$ has a real **polar decomposition**

$$A = QS, \qquad \text{where} \qquad Q \in \mathrm{SO}(d),\ S \text{ symmetric and positive definite.}$$

A fundamental inequality involving the determinant is the **Hadamard inequality**: Let $A \in \mathbb{R}^{d\times d}$ with columns $A_j \in \mathbb{R}^n$ $(j = 1, \dots, d)$. Then,

$$|\det A| \le \prod_{j=1}^{d} |A_j| \le |A|^d.$$

Of course, an analogous formula holds with the rows of A.

We recall a special case of the Jordan normal form theorem for real (2×2)-matrices: Let $A \in \mathbb{R}^{2\times 2}$. Then, there exists an invertible matrix $S \in \mathbb{R}^{2\times 2}$ such that

$$S^{-1}AS = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \qquad \text{or} \qquad S^{-1}AS = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

with $a, b, c \in \mathbb{R}$.

The reader is referred to [56] for advanced matrix analysis.

Comment...

A.2 Measure theory

We assume that the reader is familiar with the notion of Lebesgue- and Borel-measurability, nullsets, and L^p-spaces; a good introduction is [97]. For the d-dimensional Lebesgue measure we write $\mathscr{L}^d$ (or $\mathscr{L}^d_x$ if we want to stress the integration variable), but the Lebesgue measure of a Borel- or Lebesgue-measurable set $A \subset \mathbb{R}^d$ is simply $|A|$. The **characteristic function** of such an A is

$$\mathbb{1}_A(x) := \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise,} \end{cases} \qquad x \in \mathbb{R}^d.$$

In the following, we only recall some results that are needed elsewhere.

Lemma A.2 (Fatou). *Let $f_j \colon \mathbb{R}^N \to [0,+\infty]$, $j \in \mathbb{N}$, be a sequence of Lebesgue-measurable functions. Then,*

$$\liminf_{j\to\infty} \int f_j(x)\,\mathrm{d}x \geq \int \liminf_{j\to\infty} f_j(x)\,\mathrm{d}x.$$

Lemma A.3 (Monotone convergence). *Let $f_j \colon \mathbb{R}^N \to [0,+\infty]$, $j \in \mathbb{N}$, be a sequence of Lebesgue-measurable functions with $f_j(x) \uparrow f(x)$ for almost every $x \in \Omega$. Then, $f \colon \mathbb{R}^N \to [0,+\infty]$ is measurable and*

$$\lim_{j\to\infty} \int f_j(x)\,\mathrm{d}x = \int f(x)\,\mathrm{d}x.$$

Lemma A.4. *If $f_j \to f$ **strongly** in $\mathrm{L}^1(\mathbb{R}^d)$, $1 \leq p \leq \infty$, that is,*

$$\|f_j - f\|_{\mathrm{L}^p} \to 0 \quad \text{as} \quad j \to \infty,$$

then there exists a subsequence (not relabeled[1]) such that $f_j \to f$ almost everywhere.

Theorem A.5 (Lebesgue dominated convergence theorem). *Let $f_j \colon \mathbb{R}^N \to \mathbb{R}^n$, $j \in \mathbb{N}$, be a sequence of Lebesgue-measurable functions such that there exists an L^p-integrable majorant $g \in \mathrm{L}^p(\mathbb{R}^N)$, $1 \leq p < \infty$, that is,*

$$|f_j| \leq g \qquad \text{for all } j \in \mathbb{N}.$$

If $f_j \to f$ pointwise almost everywhere, then also $f_j \to f$ strongly in L^p.

The following strengthening of Lebesgue's theorem is very convenient:

Theorem A.6 (Pratt). *If $f_j \to f$ pointwise almost everywhere and there exists a sequence $(g_j) \subset \mathrm{L}^1(\mathbb{R}^N)$ with $g_j \to g$ in L^1 such that $|f_j| \leq g_j$, then $f_j \to f$ in L^1.*

The following convergence theorem is of fundamental significance:

[1] We do not choose different indices for subsequences if the original sequence is discarded.

Comment...

Theorem A.7 (Vitali's convergence theorem). *Let $\Omega \subset \mathbb{R}^d$ be bounded and let $(f_j) \subset \mathrm{L}^p(\Omega;\mathbb{R}^m)$, $1 \le p < \infty$. Assume furthermore that*

(i) ***No oscillations:*** *$f_j \to f$ in measure, that is, for all $\varepsilon > 0$,*

$$\mathscr{L}^d\big(\{x \in \Omega \,:\, |f_j - f| > \varepsilon\}\big) \to 0 \quad as \quad j \to \infty.$$

(ii) ***No concentrations:*** *$\{f_j\}_j$ is p-equiintegrable, i.e.*

$$\lim_{K\uparrow\infty} \sup_j \int_{\{|f_j|>K\}} |f_j|^p \,\mathrm{d}x = 0.$$

Then, $f_j \to f$ strongly in L^p.

Theorem A.8 (Radon–Riesz). *Let $1 < p < \infty$ and assume that for $(u_j) \subset \mathrm{L}^p(\Omega)$ it holds that $u_j \rightharpoonup u$ and $\|u_j\|_{\mathrm{L}^p} \to \|u\|_{\mathrm{L}^p}$. Then, $u_j \to u$ in L^p.*

Lebesgue-integrable functions also have a very useful "continuity" property:

Theorem A.9. *Let $f \in \mathrm{L}^1(\mathbb{R}^d,\mu)$, where $\mu \in \mathscr{M}_+(\mathbb{R}^d)$ is a positive (local) Radon measure. Then, μ-almost every $x_0 \in \mathbb{R}^d$ is a* ***Lebesgue point*** *of f with respect to μ, i.e.*

$$\lim_{r\downarrow 0} \fint_{B(x_0,r)} |f(x) - f(x_0)| \,\mathrm{d}\mu(x) = 0,$$

where $B(x_0,r) \subset \mathbb{R}^d$ is the ball with center x_0 and radius $r > 0$ and $\fint_{B(x_0,r)} := \frac{1}{|B(x_0,r)|}\int_{B(x_0,r)}$.

The following *covering theorem* is a handy tool for several constructions, see Theorems 2.19 and 5.51 in [1] for a proof:

Theorem A.10 (Vitali covering theorem). *Let $B \subset \mathbb{R}^N$ be a bounded Borel set and let $K \subset \mathbb{R}^d$ be compact. Then, there exist $a_k \in B$, $r_k > 0$, where $k \in \mathbb{N}$, such that we may write B as the disjoint union*

$$B = Z \cup \bigcup_{k=1}^{N} K(a_k,r_k), \qquad K(a_k,r_k) = a_k + r_k K,$$

with $Z \subset B$ a Lebesgue-nullset. Moreover, if we are given for almost every $x \in \Omega$ a real number $R(x) > 0$, then we may additionally require of the cover that $r_k < R(a_k)$.

We also need other measures than Lebesgue measure on subsets of $\mathbb{R}^N$ (this could be either $\mathbb{R}^d$ or a matrix space $\mathbb{R}^{m\times d}$, identified with $\mathbb{R}^{md}$). All of these abstract measures will be **Borel measures**, that is, they are defined on the **Borel σ-algebra** $\mathfrak{B}(\mathbb{R}^N)$ of $\mathbb{R}^N$. In this context, recall that the Borel-σ-algebra is the smallest σ-algebra that contains the open sets. All Borel measures defined on $\mathbb{R}^N$ that do not take the value $+\infty$ are collected in the set $\mathscr{M}(\mathbb{R}^N)$ of **(finite) Radon measures**, its subclass of **probability measures** is $\mathscr{M}_1(\mathbb{R}^N)$. We also use $\mathscr{M}(\Omega), \mathscr{M}_1(\Omega)$ for the subset of measures that only charge $\Omega \subset \mathbb{R}^N$. Besides

Comment...

the d-dimensional Lebesgue measure $\mathscr{L}^d$, we denote by $\mathscr{H}^s$ the s-dimensional Hausdorff measure, $0 \le s < \infty$. We denote by ω_d the volume of the d-dimensional unit ball.

We define the **duality pairing**

$$\langle h, \mu \rangle := \int h(A) \, \mathrm{d}\mu(A)$$

for $h\colon \mathbb{R}^N \to \mathbb{R}$ and $\mu \in \mathscr{M}(\mathbb{R}^N)$, whenever this integral makes sense (we need at least h μ-measurable). The following notation is convenient for the **barycenter** of a finite Borel measure μ:

$$[\mu] := \langle \mathrm{id}, \mu \rangle = \int A \, \mathrm{d}\mu(A).$$

The **restriction** of a Borel measure $\mu \in \mathscr{M}(\mathbb{R}^d)$ a Borel set A is denoted by $\mu \llcorner A$ and defined via $(\mu \llcorner A)(B) := \mu(A \cap B)$ for any Borel set B.

Probability measures and convex functions interact well:

Lemma A.11 (Jensen inequality). *For all probability measures $\mu \in \mathscr{M}_1(\mathbb{R}^N)$ and all convex $h\colon \mathbb{R}^N \to \mathbb{R}$ it holds that*

$$h([\mu]) \le \int h(A) \, \mathrm{d}\mu(A).$$

There is an alternative, "dual", view on finite measures, expressed in the following important theorem:

Theorem A.12 (Riesz representation theorem). *The space of Radon measures $\mathscr{M}(\mathbb{R}^N)$ is isometrically isomorphic to the dual space $\mathrm{C}_0(\mathbb{R}^N)^*$ via the duality pairing*

$$\langle h, \mu \rangle = \int h \, \mathrm{d}\mu.$$

As an easy, often convenient, consequence, we can *define* measures through their **action** on $\mathrm{C}_0(\mathbb{R}^N)$. Note, however, that we then need to check the boundedness $|\langle h, \mu \rangle| \le \|h\|_\infty$.

If the element $\mu \in \mathrm{C}_0(\mathbb{R}^N)^*$ is additionally positive, that is, $\langle h, \mu \rangle \ge 0$ for $h \ge 0$, and normalized, that is, $\langle \mathbb{1}, \mu \rangle = 1$ (here, $\mathbb{1} = 1$ on the whole space), then μ from the Riesz representation theorem is in fact a *probability measure*, $\mu \in \mathscr{M}_1(\mathbb{R}^N)$.

The following is an extension of Theorem A.10 to general measures:

Theorem A.13 (Vitali–Besicovitch covering theorem). *Let $\mu \in \mathscr{M}^+(\mathbb{R}^d)$, $B \subset \mathbb{R}^N$ be a bounded Borel set, and let $K \subset \mathbb{R}^d$ be compact. Then, there exist $a_k \in B$, $r_k > 0$, where $k \in \mathbb{N}$, such that we may write B as the disjoint union*

$$B = Z \cup \bigcup_{k=1}^{N} K(a_k, r_k), \qquad K(a_k, r_k) = a_k + r_k K,$$

with $Z \subset B$ a μ-nullset. Moreover, if we are given for μ-almost every $x \in \Omega$ a real number $R(x) > 0$, then we may additionally require of the cover that $r_k < R(a_k)$.

Comment...

We finally exhibit a few aspects of the theory of **(Borel) vector measures** (often just called "measures" in this text), which are σ-additive set function $\mu\colon \mathfrak{B}(\mathbb{R}^d) \to \mathbb{R}^m$, where $\mathfrak{B}(\mathbb{R}^d)$ denotes the Borel σ-algebra on $\mathbb{R}^d$. All such measures μ are collected in the space $\mathscr{M}(\mathbb{R}^d;\mathbb{R}^N)$; likewise define $\mathscr{M}(\Omega;\mathbb{R}^N)$ and $\mathscr{M}(\overline{\Omega};\mathbb{R}^N)$ for an open set $\Omega \subset \mathbb{R}^d$. We denote the **support** of μ by

$$\operatorname{supp}\mu := \overline{\left\{ x \in \mathbb{R}^d \ : \ \mu(B(x,r)) \neq 0 \text{ for all } r > 0 \right\}}.$$

The **total variation measure** of a $\mu \in \mathscr{M}(\mathbb{R}^d;\mathbb{R}^N)$ is the positive measure $|\mu| \in \mathscr{M}_+(\mathbb{R}^d)$ defined as

$$|\mu|(B) := \left\{ \sum_{j=1}^{\infty} |\mu(B_k)| \ : \ B = \bigcup_{j=1}^{\infty} B_k \text{ as a disjoint union of Borel sets} \right\}.$$

Of fundamental importance is the following theorem:

Theorem A.14 (Besicovitch differentiation theorem). *Given a $\mu \in \mathscr{M}(\mathbb{R}^d;\mathbb{R}^N)$ and $\nu \in \mathscr{M}_+(\mathbb{R}^d)$, for ν-almost every $x_0 \in \mathbb{R}^d$ in the support of ν, the limit*

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x_0) := \lim_{r\downarrow 0} \frac{\mu(B(x_0,r))}{\nu(B(x_0,r))}$$

*exists in $\mathbb{R}^N$ and is called the **Radon–Nikodým derivative** of μ with respect to ν, Moreover, the **Lebesgue–Radon–Nikodým decomposition** of μ is given as*

$$\mu = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\nu + \mu^s,$$

where $\mu^s = \mu \llcorner E$ and

$$E = (\mathbb{R}^d \setminus \operatorname{supp}\mu) \cup \left\{ x \in \operatorname{supp}\mu \ : \ \lim_{r\downarrow 0} \frac{|\mu|(B(x,r))}{\nu(B(x,r))} = \infty \right\}.$$

For a Borel measure $\mu \in \mathscr{M}(\Omega)$ and a Borel map $\varphi\colon \Omega \to \Omega' \subset \mathbb{R}^n$, we define the **push-forward measure** of μ under φ via

$$\varphi_{\#}\mu := \mu \circ \varphi^{-1}.$$

We have the following transformation formula for any $g\colon \Omega' \to \mathbb{R}^m$:

$$\int_{\Omega'} g \,\mathrm{d}(\varphi_{\#}\mu) = \int_{\Omega} g \circ \varphi \,\mathrm{d}\mu,$$

provided these integrals are defined.

Finally, the weak* convergence of vector measures is defined exactly as for positive measures. We also recall that if $\mu_j \overset{*}{\rightharpoonup} \mu$ in $\mathscr{M}(\mathbb{R}^d;\mathbb{R}^N)$ and $|\mu_j| \overset{*}{\rightharpoonup} \Lambda$, then $\mu_j(B) \to \mu(B)$ for all Borel sets $B \subset \mathbb{R}^N$ with $\Lambda(\partial B) = 0$.

Comment...

A.3 Functional analysis

We assume that the reader has a solid foundation in the basic notions of functional analysis such as Banach spaces and their duals, weak/weak* convergence (and topology[2]), reflexivity, and weak/weak* compactness. The application of $x^* \in U^*$ to $u \in U$ is often written via the **duality pairing** $\langle x, x^* \rangle = \langle x^*, x \rangle := x^*(x)$. We write $x_j \rightharpoonup x$ for weak convergence and $x_j \overset{*}{\rightharpoonup} x$ for weak* convergence. The weak* topology is metrizable on norm-bounded sets in the dual to a separable Banach space. Likewise, in reflexive and separable Banach spaces also the weak topology is metrizable on norm-bounded sets. In this context recall that in Banach spaces topological weak compactness is equivalent to sequential weak compactness, this is the Eberlein–Šmulian Theorem. Also recall that the norm in a Banach space is lower semicontinuous with respect to weak convergence: If $u_j \rightharpoonup u$ in U, then $\|x\| \leq \liminf_{j\to\infty} \|x_j\|$. One very thorough reference for most of this material is [27]. The following are just reminders:

Theorem A.15 (Weak compactness). *Let U be a reflexive Banach space. Then, norm-bounded sets in U are sequentially weakly relatively compact.*

Theorem A.16 (Banach–Alaoglu). *Let U be a separable Banach space. Then, norm-bounded sets in the dual space U^* are weakly* sequentially relatively compact.*

Theorem A.17 (Dunford–Pettis). *Let $\Omega \subset \mathbb{R}^d$ be bounded. A bounded family $\{f_j\} \subset \mathrm{L}^1(\Omega)$ ($j \in \mathbb{N}$) is equiintegrable if and only if it is weakly relatively (sequentially) compact in $\mathrm{L}^1(\Omega)$.*

Weak convergence can be "improved" to strong convergence in the following way:

Lemma A.18 (Mazur). *Let $x_j \rightharpoonup x$ in a Banach space U. Then, there exists a sequence $(y_j) \subset U$ of convex combinations,*

$$y_j = \sum_{n=j}^{N(j)} \theta_{j,n} x_n, \qquad \theta_{j,n} \in [0,1], \qquad \sum_{n=j}^{N(j)} \theta_{j,n} = 1.$$

such that $y_j \to x$ in U.

We also need the following version of the Hahn–Banach separation theorem:

Theorem A.19 (Hahn–Banach). *Let U be a Banach space and let $K \subset U^*$ be convex and compact. Then, for every $x_0^* \in U^* \setminus K$ there is $x_0 \in U$ such that*

$$\langle x_0, x_0^* \rangle < \inf_{x^* \in K} \langle x_0, x^* \rangle.$$

[2]However, we here really only need convergence of sequences.

Comment...

A.4 Sobolev and other function spaces

We give a brief introduction to Sobolev spaces, see [69] or [42] for more detailed accounts and proofs.

Let $\Omega \subset \mathbb{R}^d$. As usual we denote by $\mathrm{C}(\Omega) = \mathrm{C}^0(\Omega), \mathrm{C}^k(\Omega)$, $k = 1, 2, \ldots$, the spaces of continuous and k-times continuously differentiable functions. As norms in these spaces we have

$$\|u\|_{\mathrm{C}^k} := \sum_{|\alpha| \le k} \|\partial^\alpha u\|_\infty, \qquad u \in \mathrm{C}^k(\Omega), \qquad k = 0, 1, 2, \ldots,$$

where $\|\cdot\|_\infty$ is the supremum norm. Here, the sum is over all **multi-indices** $\alpha \in (\mathbb{N} \cup \{0\})^d$ with $|\alpha| := \alpha_1 + \cdots + \alpha_d \le k$ and

$$\partial^\alpha := \partial_1^{\alpha_1} \partial_2^{\alpha_2} \cdots \partial_d^{\alpha_d}$$

is the α-derivative operator. Similarly, we define $\mathrm{C}^\infty(\Omega)$ of infinitely-often differentiable functions (but this is not a Banach space anymore). A subscript "$_c$" indicates that all functions u in the respective function space (e.g. $\mathrm{C}_c^\infty(\Omega)$) must have their **support**

$$\operatorname{supp} u := \overline{\{x \in \Omega \ : \ u(x) \neq 0\}}$$

compactly contained in Ω (so, $\operatorname{supp} u \subset \Omega$ and $\operatorname{supp} u$ compact). For the compact containment of a A in an open set B we write $A \Subset B$, which means that $\overline{A} \subset B$. In this way, the previous condition could be written $\operatorname{supp} u \Subset \Omega$.

For $k \in \mathbb{N}$ a positive integer and $1 \le p \le \infty$, the **Sobolev space** $\mathrm{W}^{k,p}(\Omega)$ is defined to contain all functions $u \in \mathrm{L}^p(\Omega)$ such that the **weak derivative** $\partial^\alpha u$ exists in $\mathrm{L}^p(\Omega)$ for all multi-indices $\alpha \in (\mathbb{N} \cup \{0\})^d$ with $|\alpha| \le k$. This means that for each such α, there is a (unique) function $v_\alpha \in \mathrm{L}^p(\Omega)$ satisfying

$$\int v_\alpha \cdot \psi \, \mathrm{d}x = (-1)^{|\alpha|} \int u \cdot \partial^\alpha \psi \, \mathrm{d}x \qquad \text{for all } \psi \in \mathrm{C}_c^\infty(\Omega),$$

and we write $\partial^\alpha u$ for this v_α. The uniqueness follows from the Fundamental Lemma of the calculus of variations 2.21. Clearly, if $u \in \mathrm{C}^k(\Omega)$, then the weak derivative is the classical derivative. For $u \in \mathrm{W}^{1,p}(\Omega)$ we further define the **weak gradient** and **weak divergence**,

$$\nabla u := (\partial_1 u, \partial_2 u, \ldots, \partial_d u), \qquad \operatorname{div} u := \partial_1 u + \partial_2 u + \cdots + \partial_d u.$$

As norm in $\mathrm{W}^{k,p}(\Omega)$, $1 \le p < \infty$, we use

$$\|u\|_{\mathrm{W}^{k,p}} := \left(\sum_{|\alpha| \le k} \|\partial^\alpha u\|_{\mathrm{L}^p}^p \right)^{1/p}, \qquad u \in \mathrm{W}^{k,p}(\Omega).$$

For $p = \infty$, we set

$$\|u\|_{\mathrm{W}^{k,\infty}} := \max_{|\alpha| \le k} \|\partial^\alpha u\|_{\mathrm{L}^\infty}, \qquad u \in \mathrm{W}^{k,\infty}(\Omega).$$

Under these norms, the $\mathrm{W}^{k,p}(\Omega)$ become Banach spaces.

Concerning the boundary values of Sobolev functions, we have:

Comment...

Theorem A.20 (Trace theorem). *For every domain $\Omega \subset \mathbb{R}^d$ with Lipschitz boundary and every $1 \le p \le \infty$, there exists a bounded linear* ***trace operator*** $\operatorname{tr}_\Omega \colon \mathrm{W}^{1,p}(\Omega) \to \mathrm{L}^p(\partial\Omega)$ *such that*

$$\operatorname{tr}\varphi = \varphi|_{\partial\Omega} \qquad \text{if } \varphi \in \mathrm{C}(\overline{\Omega}).$$

We write $\operatorname{tr}_\Omega u$ simply as $u|_{\partial\Omega}$. Furthermore, tr_Ω is bounded and weakly continuous between $\mathrm{W}^{1,p}(\Omega)$ and $\mathrm{L}^p(\partial\Omega)$.

For $1 < p < \infty$ denote the image of $\mathrm{W}^{1,p}(\Omega)$ under tr_Ω by $\mathrm{W}^{1-1/p,p}(\partial\Omega)$, called the **trace space** of $\mathrm{W}^{1,p}(\Omega)$. As a makeshift norm on $\mathrm{W}^{1-1/p,p}(\partial\Omega)$ one can use the $\mathrm{L}^p(\partial\Omega)$-norm, but $\mathrm{W}^{1-1/p,p}(\partial\Omega)$ is not closed under this norm. This can be remedied by introducing a more complicated norm involving "fractional" derivatives (hence the notation for $\mathrm{W}^{1-1/p,p}(\partial\Omega)$), under which this set becomes a Banach space.

For $p = 1$ the trace space of $\mathrm{W}^{1,1}(\Omega)$ is $\mathrm{L}^1(\partial\Omega)$.

We write $\mathrm{W}^{1,p}_0(\Omega)$ for the linear subspace of $\mathrm{W}^{1,p}(\Omega)$ consisting of all $\mathrm{W}^{1,p}$-functions with zero boundary values (in the sense of trace). More generally, we use $\mathrm{W}^{1,p}_g(\Omega)$ with $g \in \mathrm{W}^{1-1/p,p}(\partial\Omega)$, for the affine subspace of all $\mathrm{W}^{1,p}$-functions with boundary trace g.

The following are some properties of Sobolev spaces, stated for simplicity only for the first-order space $\mathrm{W}^{1,p}(\Omega)$ with $\Omega \subset \mathbb{R}^d$ a bounded Lipschitz domain.

Theorem A.21 (Poincaré–Friedrichs inequality). *Let $u \in \mathrm{W}^{1,p}(\Omega)$.*

(i) If $u|_{\partial\Omega} = 0$, then

$$\|u\|_{\mathrm{L}^p} \le C\|\nabla u\|_{\mathrm{L}^p},$$

where $C = C(\Omega) > 0$ is a constant. If $1 < p < \infty$ and $g \in \mathrm{W}^{1-1/p,p}(\partial\Omega)$, then

$$\|u\|_{\mathrm{L}^p} \le C\big(\|\nabla u\|_{\mathrm{L}^p} + \|g\|_{\mathrm{W}^{1-1/p,p}}\big).$$

For $p = 1$ and $g \in \mathrm{L}^1(\partial\Omega)$, it holds that

$$\|u\|_{\mathrm{L}^1} \le C\big(\|\nabla u\|_{\mathrm{L}^1} + \|g\|_{\mathrm{L}^1}\big).$$

(ii) Setting $[u]_\Omega := \fint_\Omega u \,\mathrm{d}x$, it furthermore holds that

$$\|u - [u]_\Omega\|_{\mathrm{L}^p} \le C\|\nabla u\|_{\mathrm{L}^p},$$

where $C = C(\Omega) > 0$ is a constant.

Sobolev functions have higher integrability than originally assumed:

Theorem A.22 (Sobolev embedding). *Let $u \in \mathrm{W}^{1,p}(\Omega)$.*

(i) If $p < d$, then $u \in \mathrm{L}^{p^*}(\Omega)$, where

$$p^* = \frac{dp}{d-p}$$

and there is a constant $C = C(\Omega) > 0$ such that $\|u\|_{\mathrm{L}^{p^*}} \le C\|u\|_{\mathrm{W}^{1,p}}$.

Comment...

(ii) If $p = d$, then $u \in \mathrm{L}^q(\Omega)$ for all $1 \le q < \infty$ and $\|u\|_{\mathrm{L}^q} \le C_q \|u\|_{\mathrm{W}^{1,p}}$.

(iii) If $p > d$, then $u \in \mathrm{C}(\Omega)$ and $\|u\|_\infty \le C\|u\|_{\mathrm{W}^{1,p}}$.

The second part can in fact be made more precise by considering embeddings into Hölder spaces, see Section 5.6.3 in [42] for details.

Theorem A.23 (Rellich–Kondrachov). *Let $(u_j) \subset \mathrm{W}^{1,p}(\Omega)$ with $u_j \rightharpoonup u$ in $\mathrm{W}^{1,p}$.*

(i) If $p < d$, then $u_j \to u$ in $\mathrm{L}^q(\Omega;\mathbb{R}^N)$ (strongly) for any $q < p^* = dp/(d-p)$.

(ii) If $p = d$, then $u_j \to u$ in $\mathrm{L}^q(\Omega;\mathbb{R}^N)$ (strongly) for any $q < \infty$.

(iii) If $p > d$, then $u_j \to u$ uniformly.

The following is a consequence of the *Stone–Weierstrass Theorem*:

Theorem A.24 (Density). *For every $1 \le p < \infty$, $u \in \mathrm{W}^{1,p}(\Omega)$, and all $\varepsilon > 0$ there exists $v \in (\mathrm{W}^{1,p} \cap \mathrm{C}^\infty)(\Omega)$ with $u - v \in \mathrm{W}^{1,p}_0(\Omega)$ and $\|u-v\|_{\mathrm{W}^{1,p}} < \varepsilon$.*

Finally, we note that for $0 < \gamma \le 1$ a function $u\colon \Omega \to \mathbb{R}$ is **γ-Hölder-continuous function**, in symbols $u \in \mathrm{C}^{0,\gamma}(\Omega)$, if

$$\|u\|_{\mathrm{C}^{0,\gamma}} := \|u\|_\infty + \sup_{\substack{x,y\in\Omega \\ x\neq y}} \frac{|u(x)-u(y)|}{|x-y|^\gamma} < \infty.$$

Functions in $\mathrm{C}^{0,1}(\Omega)$ are called **Lipschitz-continuous**. We construct higher-order spaces $\mathrm{C}^{k,\gamma}(\Omega)$ analogously.

Next, we define a **family of mollifiers** as follows: Let $\eta \in \mathrm{C}^\infty_c(\mathbb{R}^d)$ be radially symmetric and positive. Then, the family $(\eta_\delta)_{\delta>0}$ consists of the functions

$$\eta_\delta(x) := \frac{1}{\delta^d}\eta\left(\frac{x}{\delta^d}\right), \qquad x \in \mathbb{R}^d.$$

For $u \in \mathrm{W}^{k,p}(\mathbb{R}^d)$, where $k \in \mathbb{N} \cup \{0\}$ and $1 \le p < \infty$, we define the **mollification** $u_\delta \in \mathrm{W}^{k,p}(\mathbb{R}^d)$ as the **convolution** between u and η_δ, i.e.

$$u_\delta(x) := (\eta_\delta \star u)(x) := \int \eta_\delta(x-y)u(y)\,\mathrm{d}y, \qquad x \in \mathbb{R}^d.$$

Lemma A.25. *If $u \in \mathrm{W}^{k,p}(\mathbb{R}^d)$, then $u_\delta \to u$ strongly in $\mathrm{W}^{k,p}$ as $\delta \downarrow 0$.*

Analogous results also hold for continuously differentiable functions.

Lemma A.26 (Young's inequality for convolutions). *Let $u \in \mathrm{L}^p(\mathbb{R}^d)$, $v \in \mathrm{L}^q(\mathbb{R}^d)$ and $1 \le p,q,r \le \infty$ such that*

$$\frac{1}{r} + 1 = \frac{1}{p} + \frac{1}{q}.$$

Comment...

Then,

$$\|u \star v\|_{\mathrm{L}^r} \leq \|u\|_{\mathrm{L}^p} \cdot \|v\|_{\mathrm{L}^q}.$$

Finally, all the above notions and theorems continue to hold for vector-valued functions $u = (u^1, \dots, u^m)^T \colon \Omega \to \mathbb{R}^m$ and in this case we set

$$\nabla u := \begin{pmatrix} \partial_1 u^1 & \partial_2 u^1 & \cdots & \partial_d u^1 \\ \partial_1 u^2 & \partial_2 u^2 & \cdots & \partial_d u^2 \\ \vdots & \vdots & & \vdots \\ \partial_1 u^m & \partial_2 u^m & \cdots & \partial_d u^m \end{pmatrix} \quad \in \mathbb{R}^{m \times d}.$$

We use the spaces $\mathrm{C}(\Omega;\mathbb{R}^m), \mathrm{C}^k(\Omega;\mathbb{R}^m), \mathrm{W}^{k,p}(\Omega;\mathbb{R}^m), \mathrm{C}^{k,\gamma}(\Omega;\mathbb{R}^m)$ with analogous meanings.

We also use *local* versions of all spaces, namely $\mathrm{C}_{\mathrm{loc}}(\Omega), \mathrm{C}^k_{\mathrm{loc}}(\Omega), \mathrm{W}^{k,p}_{\mathrm{loc}}(\Omega), \mathrm{C}^{k,\gamma}_{\mathrm{loc}}(\Omega)$, where the defining norm is only finite on every compact subset of Ω.

Let us also quote the following classical result about extensions of Lipschitz functions:

Theorem A.27 (Kirszbraun). *Let $f \colon \Omega \to \mathbb{R}^m$ be a Lipschitz continuous map. Then, f can be extended to $\bar{f} \colon \mathbb{R}^d \to \mathbb{R}^m$ with the same Lipschitz constant as f.*

A.5 Harmonic analysis

The Fourier transform is an indispensable tool for every analyst, we here only need a few basics and a multiplier theorem. A thorough introduction can be found in [50, 51].

Define for $u \in \mathrm{L}^1(\mathbb{R}^d)$ (or vector-valued) the **Fourier transform** $\hat{u} = \mathscr{F} u \in \mathrm{L}^\infty(\mathbb{R}^d)$ as follows:

$$\hat{u}(\xi) := \int_{\mathbb{R}^d} u(x) \mathrm{e}^{-2\pi \mathrm{i} \xi \cdot x} \, \mathrm{d}x, \qquad \xi \in \mathbb{R}^d.$$

We also define the **inverse Fourier transform** $\check{v} = \mathscr{F}^{-1} v$ for $v \in \mathrm{L}^1(\mathbb{R}^d)$ to be

$$\check{u}(x) := \int_{\mathbb{R}^d} v(\xi) \mathrm{e}^{2\pi \mathrm{i} \xi \cdot x} \, \mathrm{d}x, \qquad x \in \mathbb{R}^d.$$

One can also define $\mathscr{F}, \mathscr{F}^{-1}$ on $\mathrm{L}^2(\mathbb{R}^d)$ and it turns out that they are isometries from $\mathrm{L}^2(\mathbb{R}^d)$ to itself, this is known as **Plancherel's relation**,

$$\|\hat{u}\|_{\mathrm{L}^2} = \|u\|_{\mathrm{L}^2}. \tag{A.3}$$

Equivalently, we have **Parseval's relation**

$$\int u \cdot \bar{v} \, \mathrm{d}x = \int \hat{u} \cdot \bar{\hat{v}} \, \mathrm{d}\xi \tag{A.4}$$

for all $u, v \in \mathrm{L}^2(\mathbb{R}^d)$; equivalent relations hold for $\mathbb{C}^N$-valued functions (with the Hermitian transpose in the second function).

The following theorem is the classical result concerning the $(\mathrm{L}^p \to \mathrm{L}^p)$-boundedness of Fourier multiplier operators, see for instance [50] and [15] (Theorem 6.1.6) for a proof.

Comment...

Theorem A.28 (Mihlin multiplier theorem). *Let $m \in \mathrm{C}^{\lfloor d/2 \rfloor + 1}(\mathbb{R}^d \setminus \{0\}; \mathbb{C})$ satisfy*

$$|\partial^\alpha m(\xi)| \le K|\xi|^{-|\alpha|}, \qquad \xi \in \mathbb{R}^d \setminus \{0\},$$

for all multi-indices $\alpha \in \mathbb{N}_0^d$ with $|\alpha| := |\alpha_1| + \cdots + |\alpha_d| \le \lfloor d/2 \rfloor + 1$ and some $K > 0$ ($\partial^\alpha := \partial_1^{\alpha_1} \cdots \partial_d^{\alpha_d}$). Then,

$$Tu := \mathscr{F}^{-1}[m(\xi)\hat{u}(\xi)],$$

which for $u \in \mathrm{L}^2(\mathbb{R}^d)$ is well-defined via Plancherel's relation (A.3)*, extends to a bounded operator $T\colon \mathrm{L}^p(\mathbb{R}^d) \to \mathrm{L}^p(\mathbb{R}^d)$ for all $1 < p < \infty$, which satisfies the estimate*

$$\|T\|_{\mathrm{L}^p \to \mathrm{L}^p} \le C \max(p, (p-1)^{-1})K,$$

where $C = C(d) > 0$ is a dimensional constant. Furthermore, for $p = 1$ the weak-type estimate

$$\mathscr{L}^d\big(\{\, x \in \mathbb{R}^d \;:\; |(Tu)(x)| \ge \lambda \,\}\big) \le \frac{CM}{\lambda}\|u\|_{\mathrm{L}^1}$$

holds for all $\lambda > 0$ and a dimensional constant $C = C(d) > 0$.

As a special case, the conclusions of the preceding theorem hold for any positively 0-homogeneous smooth multipliers $m\colon \mathbb{R}^d \setminus \{0\} \to \mathbb{C}$.

Another result we need concerns the $(\mathrm{L}^p \to \mathrm{L}^p)$-boundedness of the (centered) **maximal function** $Mf\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ of $f\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, which is defined as

$$Mf(x_0) := \sup_{r>0} \fint_{B(x_0,r)} |f(x)| \,\mathrm{d}x, \qquad x_0 \in \mathbb{R}^d.$$

We quote the following result about the maximal function whose proofs can be found in [50, 100]:

Theorem A.29. *(i) If $1 < p \le \infty$, then*

$$\|Mf\|_{\mathrm{L}^p} \le C\|Mf\|_{\mathrm{L}^p},$$

where $C = C(d,p) > 0$ is a constant.

(ii) If $1 \le p < \infty$, then the weak-type estimate

$$\big|\{\, x \in \mathbb{R}^d \;:\; |Mu| \ge \lambda \,\}\big| \le \frac{C}{\lambda^p}\int_{\{|u| \ge \lambda/2\}} |u|^p \,\mathrm{d}x \le \frac{C}{\lambda^p}\|u\|_{\mathrm{L}^p}^p.$$

holds for all $\lambda > 0$ and a constant $C = C(d) > 0$.

(iii) For every $K > 0$ and $f \in \mathrm{W}^{1,p}(\mathbb{R}^d; \mathbb{R}^m)$, the maximal function Mf is Lipschitz on the set $\{M(|f| + |\nabla f|) < K\}$ and its Lipschitz-constant is bounded by CK, where $C = C(d,m,p)$ is a dimensional constant.

Comment...

Bibliography

[1] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free-Discontinuity Problems*, Oxford Mathematical Monographs, Oxford University Press, 2000.

[2] H. Attouch, G. Buttazzo, and G. Michaille, *Variational Analysis in Sobolev and BV Spaces – Applications to PDEs and Optimization*, MPS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics, 2006.

[3] E. J. Balder, *A general approach to lower semicontinuity and lower closure in optimal control theory*, SIAM J. Control Optim. **22** (1984), 570–598.

[4] J. M. Ball, *Convexity conditions and existence theorems in nonlinear elasticity*, Arch. Ration. Mech. Anal. **63** (1976/77), 337–403.

[5] J. M. Ball, *Global invertibility of Sobolev functions and the interpenetration of matter*, Proc. Roy. Soc. Edinburgh Sect. A **88** (1981), 315–328.

[6] _____, *A version of the fundamental theorem for Young measures*, PDEs and continuum models of phase transitions (Nice, 1988), Lecture Notes in Physics, vol. 344, Springer, 1989, pp. 207–215.

[7] _____, *Some open problems in elasticity*, Geometry, mechanics, and dynamics, Springer, 2002, pp. 3–59.

[8] J. M. Ball, J. C. Currie, and P. J. Olver, *Null Lagrangians, weak continuity, and variational problems of arbitrary order*, J. Funct. Anal. **41** (1981), 135–174.

[9] J. M. Ball and R. D. James, *Fine phase mixtures as minimizers of energy*, Arch. Ration. Mech. Anal. **100** (1987), 13–52.

[10] _____, *Proposed experimental tests of a theory of fine microstructure and the two-well problem*, Phil. Trans. R. Soc. Lond. A **338** (1992), 389–450.

[11] J. M. Ball, B. Kirchheim, and J. Kristensen, *Regularity of quasiconvex envelopes*, Calc. Var. Partial Differential Equations **11** (2000), 333–359.

Comment...

[12] J. M. Ball and V. J. Mizel, *One-dimensional variational problems whose minimizers do not satisfy the Euler-Lagrange equation*, Arch. Ration. Mech. Anal. **90** (1985), 325–388.

[13] Lisa Beck, *Elliptic regularity theory. A first course*, Lecture Notes of the Unione Matematica Italiana, vol. 19, Springer, 2016.

[14] B. Benešová and M. Kružík, *Weak lower semicontinuity of integral functionals and applications*, arXiv:1601.00390.

[15] J. Bergh and J. Löfström, *Interpolation Spaces*, Grundlehren der mathematischen Wissenschaften, vol. 223, Springer, 1976.

[16] H. Berliocchi and J.-M. Lasry, *Intégrandes normales et mesures paramétrées en calcul des variations*, Bull. Soc. Math. France **101** (1973), 129–184.

[17] B. Bojarski and T. Iwaniec, *Analytical foundations of the theory of quasiconformal mappings in* $\mathbf{R}^n$, Ann. Acad. Sci. Fenn. Ser. A I Math. **8** (1983), 257–324.

[18] A. Braides, Γ-*convergence for Beginners*, Oxford Lecture Series in Mathematics and its Applications, vol. 22, Oxford University Press, 2002.

[19] J. Campos Cordero and J. Kristensen, *Full regularity and uniqueness of minimizers in the quasiconvex case*, forthcoming.

[20] C. Carstensen and T. Roubíček, *Numerical approximation of Young measures in non-convex variational problems*, Numer. Math. **84** (2000), 395–415.

[21] C. Castaing, P. R. de Fitte, and M. Valadier, *Young Measures on Topological Spaces: With Applications in Control Theory and Probability Theory*, Mathematics and its Applications, Kluwer, 2004.

[22] Ph. G. Ciarlet, *Mathematical Elasticity*, vol. 1: Three Dimensional Elasticity, North-Holland, 1988.

[23] Ph. G. Ciarlet and G. Geymonat, *Sur les lois de comportement en élasticité non linéaire compressible*, C. R. Acad. Sci. Paris Sér. II Méc. Phys. Chim. Sci. Univers Sci. Terre **295** (1982), 423–426.

[24] Ph. G. Ciarlet, L. Gratie, and C. Mardare, *Intrinsic methods in elasticity: a mathematical survey*, Discrete Contin. Dyn. Syst. **23** (2009), 133–164.

[25] Ph. G. Ciarlet and J. Nečas, *Injectivity and self-contact in nonlinear elasticity*, Arch. Ration. Mech. Anal. **97** (1987), 171–188.

[26] S. Conti, G. Dolzmann, B. Kirchheim, and S. Müller, *Sufficient conditions for the validity of the Cauchy-Born rule close to* $\mathrm{SO}(n)$, J. Eur. Math. Soc. **8** (2006), 515–530.

Comment...

[27] J. B. Conway, *A Course in Functional Analysis*, 2nd ed., Graduate Texts in Mathematics, vol. 96, Springer, 1990.

[28] B. Dacorogna, *Weak Continuity and Weak Lower Semicontinuity for Nonlinear Functionals*, Lecture Notes in Mathematics, vol. 922, Springer, 1982.

[29] ———, *Direct Methods in the Calculus of Variations*, 2nd ed., Applied Mathematical Sciences, vol. 78, Springer, 2008.

[30] G. Dal Maso, *An Introduction to* Γ*-Convergence*, Progress in Nonlinear Differential Equations and their Applications, vol. 8, Birkhäuser, 1993.

[31] E. De Giorgi, *Sulla differenziabilità e l'analiticità delle estremali degli integrali multipli regolari*, Mem. Accad. Sci. Torino. Cl. Sci. Fis. Mat. Nat. **3** (1957), 25–43.

[32] ———, *Un esempio di estremali discontinue per un problema variazionale di tipo ellittico*, Boll. Un. Mat. Ital. **1** (1968), 135–137.

[33] J. Diestel and J. J. Uhl, Jr., *Vector measures*, Mathematical Surveys, vol. 15, American Mathematical Society, 1977.

[34] N. Dinculeanu, *Vector measures*, International Series of Monographs in Pure and Applied Mathematics, vol. 95, Pergamon Press, 1967.

[35] R. J. DiPerna, *Compensated compactness and general systems of conservation laws*, Trans. Amer. Math. Soc. **292** (1985), 383–420.

[36] G. Dolzmann, *Variational methods for crystalline microstructure—analysis and computation*, Lecture Notes in Mathematics, vol. 1803, Springer, 2003.

[37] G. Dolzmann and J. Kristensen, *Higher integrability of minimizing Young measures*, Calc. Var. Partial Differential Equations **22** (2005), 283–301.

[38] W. E and P. Ming, *Cauchy-Born rule and the stability of crystalline solids: static problems*, Arch. Ration. Mech. Anal. **183** (2007), 241–297.

[39] I. Ekeland and R. Temam, *Convex analysis and variational problems*, North-Holland, 1976.

[40] J. L. Ericksen, *On the Cauchy-Born rule*, Math. Mech. Solids **13** (2008), 199–220.

[41] L. C. Evans, *Weak convergence methods for nonlinear partial differential equations*, CBMS Regional Conference Series in Mathematics, vol. 74, Conference Board of the Mathematical Sciences, 1990.

[42] ———, *Partial differential equations*, 2nd ed., Graduate Studies in Mathematics, vol. 19, American Mathematical Society, 2010.

Comment...

[43] I. Fonseca and G. Leoni, *Modern Methods in the Calculus of Variations: L^p Spaces*, Springer, 2007.

[44] I. Fonseca, S. Müller, and P. Pedregal, *Analysis of concentration and oscillation effects generated by gradients*, SIAM J. Math. Anal. **29** (1998), 736–756.

[45] G. A. Francfort, *An introduction to H-measures and their applications*, Variational problems in materials science (Trieste, 2004), Progress in Nonlinear Differential Equations and their Applications, vol. 68, Birkhäuser, 2006, pp. 85–110.

[46] G. Friesecke and F. Theil, *Validity and failure of the Cauchy-Born hypothesis in a two-dimensional mass-spring lattice*, J. Nonlinear Sci. **12** (2002), 445–478.

[47] P. Gérard, *Microlocal defect measures*, Comm. Partial Differential Equations **16** (1991), 1761–1794.

[48] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Grundlehren der mathematischen Wissenschaften, vol. 224, Springer, 1998.

[49] E. Giusti, *Direct methods in the calculus of variations*, World Scientific, 2003.

[50] L. Grafakos, *Classical Fourier Analysis*, 3rd ed., Graduate Texts in Mathematics, vol. 249, Springer, 2014.

[51] _____, *Modern Fourier Analysis*, 3rd ed., Graduate Texts in Mathematics, vol. 250, Springer, 2014.

[52] R. Gratwick, *Singular sets and the Lavrentiev phenomenon*, Proc. Roy. Soc. Edinburgh Sect. A **145** (2015), 513–533.

[53] R. Gratwick and D. Preiss, *A one-dimensional variational problem with continuous Lagrangian and singular minimizer*, Arch. Ration. Mech. Anal. **202** (2011), 177–211.

[54] D. Henao and C. Mora-Corral, *Invertibility and weak continuity of the determinant for the modelling of cavitation and fracture in nonlinear elasticity*, Arch. Ration. Mech. Anal. **197** (2010), 619–655.

[55] D. Hilbert, *Mathematische Probleme – Vortrag, gehalten auf dem internationalen Mathematiker-Kongreß zu Paris 1900*, Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse (1900), 253–297.

[56] R. A. Horn and C. R. Johnson, *Matrix analysis*, 2nd ed., Cambridge University Press, 2013.

[57] D. Kinderlehrer and P. Pedregal, *Characterizations of Young measures generated by gradients*, Arch. Ration. Mech. Anal. **115** (1991), 329–365.

Comment...

[58] D. Kinderlehrer and P. Pedregal, *Weak convergence of integrands and the Young measure representation*, SIAM J. Math. Anal. **23** (1992), 1–19.

[59] _____, *Gradient Young measures generated by sequences in Sobolev spaces*, J. Geom. Anal. **4** (1994), 59–90.

[60] B. Kirchheim, *Rigidity and Geometry of Microstructures*, Lecture notes 16, Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig, 2003.

[61] K. Koumatos, F. Rindler, and E. Wiedemann, *Differential inclusions and Young measures involving prescribed Jacobians*, SIAM J. Math. Anal. **47** (2015), 1169–1195, to appear.

[62] _____, *Orientation-preserving Young measures*, Q. J. Math. (2016).

[63] J. Kristensen, *Lower semicontinuity of quasi-convex integrals in BV*, Calc. Var. Partial Differential Equations **7** (1998), no. 3, 249–261.

[64] _____, *Lower semicontinuity in spaces of weakly differentiable functions*, Math. Ann. **313** (1999), 653–710.

[65] J. Kristensen and C. Melcher, *Regularity in oscillatory nonlinear elliptic systems*, Math. Z. **260** (2008), 813–847.

[66] J. Kristensen and G. Mingione, *The singular set of Lipschitzian minima of multiple integrals*, Arch. Ration. Mech. Anal. **184** (2007), 341–369.

[67] M. Kružík, *On the composition of quasiconvex functions and the transposition*, J. Convex Anal. **6** (1999), 207–213.

[68] M. Lavrentiev, *Sur quelques problèmes du calcul des variations*, Ann Mat Pura Appl **4** (1926), 7–28.

[69] G. Leoni, *A first course in Sobolev spaces*, Graduate Studies in Mathematics, vol. 105, American Mathematical Society, 2009.

[70] B. Manià, *Sopra un essempio di lavrentieff*, Boll. Un. Mat. Ital. **13** (1934), 147–153.

[71] M. Marcus and V. J. Mizel, *Transformations by functions in Sobolev spaces and lower semicontinuity for parametric variational problems*, Bull. Amer. Math. Soc. **79** (1973), 790–795.

[72] P. Mattila, *Geometry of Sets and Measures in Euclidean Spaces*, Cambridge Studies in Advanced Mathematics, vol. 44, Cambridge University Press, 1995.

[73] E. J. McShane, *Generalized curves*, Duke Math. J. **6** (1940), 513–536.

[74] _____, *Necessary conditions in generalized-curve problems of the calculus of variations*, Duke Math. J. **7** (1940), 1–27.

Comment...

[75] _____, *Sufficient conditions for a weak relative minimum in the problem of Bolza*, Trans. Amer. Math. Soc. **52** (1942), 344–379.

[76] N. G. Meyers, *Quasi-convexity and lower semi-continuity of multiple variational integrals of any order*, Trans. Amer. Math. Soc. **119** (1965), 125–149.

[77] Giuseppe Mingione, *Regularity of minima: an invitation to the dark side of the calculus of variations*, Appl. Math. **51** (2006), 355–426.

[78] C. B. Morrey, Jr., *On the solutions of quasi-linear elliptic partial differential equations*, Trans. Amer. Math. Soc. **43** (1938), 126–166.

[79] C. B. Morrey, Jr., *Quasi-convexity and the lower semicontinuity of multiple integrals*, Pacific J. Math. **2** (1952), 25–53.

[80] _____, *Multiple Integrals in the Calculus of Variations*, Grundlehren der mathematischen Wissenschaften, vol. 130, Springer, 1966.

[81] J. Moser, *A new proof of De Giorgi's theorem concerning the regularity problem for elliptic differential equations*, Comm. Pure Appl. Math. **13** (1960), 457–468.

[82] _____, *On Harnack's theorem for elliptic differential equations*, Comm. Pure Appl. Math. **14** (1961), 577–591.

[83] S. Müller, *Rank-one convexity implies quasiconvexity on diagonal matrices*, Internat. Math. Res. Notices (1999), no. 20, 1087–1095.

[84] _____, *A sharp version of Zhang's theorem on truncating sequences of gradients*, Trans. Amer. Math. Soc. **351** (1999), 4585–4597.

[85] S. Müller, *Variational models for microstructure and phase transitions*, Calculus of variations and geometric evolution problems (Cetraro, 1996), Lecture Notes in Mathematics, vol. 1713, Springer, 1999, pp. 85–210.

[86] S. Müller and S. J. Spector, *An existence theory for nonlinear elasticity that allows for cavitation*, Arch. Ration. Mech. Anal. **131** (1995), 1–66.

[87] S. Müller and V. Šverák, *Convex integration for Lipschitz mappings and counterexamples to regularity*, Ann. of Math. **157** (2003), 715–742.

[88] F. Murat, *Compacité par compensation*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. **5** (1978), 489–507.

[89] _____, *Compacité par compensation. II*, Proceedings of the International Meeting on Recent Methods in Nonlinear Analysis (Rome, 1978), Pitagora, 1979, pp. 245–256.

Comment...

[90] ———, *Compacité par compensation: condition nécessaire et suffisante de continuité faible sous une hypothèse de rang constant*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. **8** (1981), 69–102.

[91] J. Nash, *Continuity of solutions of parabolic and elliptic equations*, Amer. J. Math. **80** (1958), 931–954.

[92] J. Nečas, *On regularity of solutions to nonlinear variational inequalities for second-order elliptic systems*, Rend. Mat. **8** (1975), 481–498.

[93] P. J. Olver, *Applications of Lie groups to differential equations*, 2nd ed., Graduate Texts in Mathematics, vol. 107, Springer, 1993.

[94] P. Pedregal, *Parametrized Measures and Variational Principles*, Progress in Nonlinear Differential Equations and their Applications, vol. 30, Birkhäuser, 1997.

[95] Y. G. Reshetnyak, *On the stability of conformal mappings in multidimensional spaces*, Siberian Math. J. **8** (1967), 69–85.

[96] T. Roubíček, *Relaxation in Optimization Theory and Variational Calculus*, De Gruyter Series in Nonlinear Analysis & Applications, vol. 4, De Gruyter, 1997.

[97] W. Rudin, *Real and Complex Analysis*, 3rd ed., McGraw–Hill, 1986.

[98] L. Simon, *Schauder estimates by scaling*, Calc. Var. Partial Differential Equations **5** (1997), 391–407.

[99] E. N. Spadaro, *Non-uniqueness of minimizers for strictly polyconvex functionals*, Arch. Ration. Mech. Anal. **193** (2009), 659–678.

[100] E. M. Stein, *Harmonic Analysis*, Princeton University Press, 1993.

[101] V. Šverák, *Regularity properties of deformations with finite energy*, Arch. Ration. Mech. Anal. **100** (1988), 105–127.

[102] ———, *Quasiconvex functions with subquadratic growth*, Proc. Roy. Soc. London Ser. A **433** (1991), 723–725.

[103] ———, *Rank-one convexity does not imply quasiconvexity*, Proc. Roy. Soc. Edinburgh Sect. A **120** (1992), 185–189.

[104] V. Šverák and X. Yan, *A singular minimizer of a smooth strongly convex functional in three dimensions*, Calc. Var. Partial Differential Equations **10** (2000), 213–221.

[105] ———, *Non-Lipschitz minimizers of smooth uniformly convex functionals*, Proc. Natl. Acad. Sci. USA **99** (2002), 15269–15276.

[106] M. A. Sychev, *Characterization of homogeneous gradient Young measures in case of arbitrary integrands*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. **29** (2000), 531–548.

Comment...

[107] L. Székelyhidi, Jr., *The regularity of critical points of polyconvex functionals*, Arch. Ration. Mech. Anal. **172** (2004), 133–152.

[108] Q. Tang, *Almost-everywhere injectivity in nonlinear elasticity*, Proc. Roy. Soc. Edinburgh Sect. A **109** (1988), 79–95.

[109] L. Tartar, *Compensated compactness and applications to partial differential equations*, Nonlinear analysis and mechanics: Heriot-Watt Symposium, Vol. IV, Res. Notes in Math., vol. 39, Pitman, 1979, pp. 136–212.

[110] _____, *The compensated compactness method applied to systems of conservation laws*, Systems of nonlinear partial differential equations (Oxford, 1982), NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., vol. 111, Reidel, 1983, pp. 263–285.

[111] _____, *H-measures, a new approach for studying homogenisation, oscillations and concentration effects in partial differential equations*, Proc. Roy. Soc. Edinburgh Sect. A **115** (1990), 193–230.

[112] _____, *On mathematical tools for studying partial differential equations of continuum physics: H-measures and Young measures*, Developments in partial differential equations and applications to mathematical physics (Ferrara, 1991), Plenum, 1992, pp. 201–217.

[113] _____, *Beyond Young measures*, Meccanica **30** (1995), 505–526.

[114] L. Tonelli, *Sur un méthode directe du calcul des variations*, Rend. Circ. Mat. Palermo **39** (1915), 233–264.

[115] L. C. Young, *Generalized curves and the existence of an attained absolute minimum in the calculus of variations*, C. R. Soc. Sci. Lett. Varsovie, Cl. III **30** (1937), 212–234.

[116] _____, *Generalized surfaces in the calculus of variations*, Ann. of Math. **43** (1942), 84–103.

[117] _____, *Generalized surfaces in the calculus of variations. II*, Ann. of Math. **43** (1942), 530–544.

[118] _____, *Lectures on the calculus of variations and optimal control theory*, 2nd ed., Chelsea, 1980, (reprinted by AMS Chelsea Publishing 2000).

Comment...

Index

Comment...

Comment...

Comment...

Comment...