Measure and Image

Lecture notes

Tuomo Valkonen
tuomov@iki.fi

University of Cambridge, Michaelmas 2014.

Last updated December 23, 2014

Contents

Chapter 1

Introduction

1.1. About the course

Photographs and other natural images are usually not smooth maps, they contain edges (discontinuities) and other non-smooth geometric features that should be preserved by image enhancement techniques. The correct mathematical modelling of these features involves the space of functions of bounded variation and, in consequence, aspects of geometric measure theory. The aim of this course is to provide an introduction to functions of bounded variation and their applications in image processing.

Although we review measure theory in the beginning of this course, it is strongly recommended that the reader be well acquainted with the topic. Introductions may be found in [28, 19] among other books. It is also advisable, although not required, that the student be acquainted with the basics of Sobolev spaces and weak topologies in function spaces. An introduction to the former may be found in [14], and to the latter in [23, 4]

This course is mainly based on the treatment of functions of bounded variation in [2] and [15]. The books [10, 3] on mathematical imaging also provides good additional reading. Scientific articles detailing certain aspects will be pointed to as encountered. Hausdorff measures and rectifiable sets will be important throughout the course. They also play an important role in the fascinating topic of fractal geometry. The interested student may read more about Hausdorff measures and fractals in [25] and, in a more introductory style, in [16].

1.2. Photographs

Take a look at the photograph in Figure 1.1. What do you see? A flower, yes! But as a mathematician? A function! Yes, photographs are functions, they map points on a rectangular domain $\Omega \subset \mathbb{R}^2$ into colour values, often in an RGB (Red-Green-Blue) space, embedded into $\mathbb{R}^3$. In the print industry CMYK (Cyan-Magenta-Yellow-Key(blacK)) is more popular. Mathematicians often still live in the 19th century, and study greyscale photographs $u : \Omega \to \mathbb{R}$.

What properties does the function portrayed in Figure 1.1 have? The green background features *smooth* and buttery *bokeh*; an artistic effect created by the background being out-of-focus in the narrow depth of field of the camera lens caused by a large aperture (small f-number). The flower itself features plenty of *texture* in the form of alternating streaks of red and yellow. But what about points x where there is a transition from flower into background? There is an *edge*! The colour values change abruptly. Mathematically, u has a jump; the point $x \in J_u$, the jump set of u.

A large part of this course will concern J_u. It is important for image processing algorithms, such as denoising, deblurring, and more advanced techniques in medical imaging, to preserve and to restore edges. They represent depth information, and separate regions consisting of different materials. It is also equally important to preserve and to restore smooth areas and texture, but that will be of less concern to us in this course. The question then is, what kind of function spaces allow for the jump set J_u in a controlled manner? Such that we can expect more regularity from the image than the spaces $L^p(\Omega)$ provide?

Figure 1.1: A colourful flower

Of course, our example image is of high quality, and does not need to be be processed. Even in photography, with advanced technology, there is however, still need for image processing. Take low-light photography, for example. In low light conditions, there is noise. The good photographs obtained with high ISO numbers by top-of-the line DSLR (digital single lens reflex) cameras are the result of image processing! In magnetic resonance imaging, even physical limitations restrict the resolution of images, as does subject movement (patient comfort).

1.3. Regularisation of inverse problems

Many image processing problems can be seen as instances of solving an operator equation

$$Ku + v = f,$$

for our known data f, noise v, and unknown image u. This is an *inverse problem*. In general, such problems are ill-posed, and we cannot expect to have a unique solution, or a solution at all. In order to impose well-posedness, we introduce a regulariser R that models our prior assumptions on a good solution u, as well as a fidelity functional F that models the noise v. The choice of R is specific to the problem at hand; a prototypical choice in image processing is the *total variation*, which we will also discuss in this course. More recent research has focused on higher-order [5, 9, 24] and curvature-based [11, 29] extensions, as well as non-convex regularisers [22].

If we know a noise level σ, we may then try to solve the problem

$$\min_u R(u) \quad \text{subject to} \quad F(Ku - f) \leq \sigma. \tag{1.1}$$

Often the noise level is not known. Moreover, this problem nformulation can be numerically very difficult. It is therefore more common to solve the *Tikhonov regularised* problem

$$\min_u F(Ku - F) + \alpha R(u), \tag{1.2}$$

for a suitable *regularisation parameter* α. We refer to [13] the student interested in reading on more about inverse problems theory, and the role α and σ play especially in their limit.

In image processing, for *denoising* $K = I$ is the identity. For *deblurring*, K can be a convolution operation, $Ku = \rho * u$ for a suitable blur kernel ρ. For sub-sampled reconstruction from Fourier samples, as is the case with magnetic resonance imaging (MRI) reconstructions, $K = S\mathcal{F}$ for S a sub-sampling operator, and $\mathcal{F}$ the Fourier transform. If simply $K = S$ for a sub-sampling operator, then we are talking about *inpainting*. This might be used, for example, to hide hairs and scratches in old photographs or films. For a detailed treament of various image processing tasks, see, for example [10, 3].

The questions now are, when do (1.1) and (1.2) have solutions, and what are their properties? We will discuss some of these questions at the end of the course. First we however need to find the right space for the image u.

1.4. Sobolev functions

Sobolev spaces are common in various areas of mathematics. Given a domain $\Omega \subset \mathbb{R}^n$, the space $W^{1,p}(\Omega)$, for an exponent $p \in [1, \infty)$ may be defined as

$$W^{1,p}(\Omega) := \{u \in L^p(\Omega) \mid \nabla u \in L^p(\Omega; \mathbb{R}^n)\}.$$

Here $L^p(\Omega; \mathbb{R}^k)$ is the space of functions $v : \Omega \to \mathbb{R}^k$, $v = (v_1, \ldots, v_k)$, such that

$$\begin{aligned} \|v\|_{L^p(\Omega;\mathbb{R}^k)} &:= \left(\sum_{i=1}^k \int_\Omega |v_i(x)|^p \, dx\right)^{1/p} < \infty, \quad (1 \le p < \infty), \quad \text{and} \\ \|v\|_{L^\infty(\Omega;\mathbb{R}^k)} &:= \operatorname{ess\,sup}_{x\in\Omega} |u(x)| < \infty. \end{aligned} \tag{1.3}$$

The integral is to be understood in the Lebesgue sense, and the supremum is the measure-theoretic essential supremum that ignores the value of u on negligible sets. (We will get back to these later!) In other words,

$$W^{1,p}(\Omega) = \{u \in L^1(\Omega) \mid \|u\|_{L^p(\Omega)} + \|\nabla u\|_{L^p(\Omega;\mathbb{R}^k)} < \infty\}.$$

The space can be normed with

$$\|u\|_{W^{1,p}(\Omega)} := \|u\|_{L^p(\Omega)} + \|\nabla u\|_{L^p(\Omega;\mathbb{R}^k)},$$

and can be formulated as the completion

$$W^{1,p}(\Omega) = \overline{C^\infty(\Omega)}$$

with respect to the norm $\|\cdot\|_{W^{1,p}(\Omega)}$, $(1 \le p < \infty)$.

To be precise the gradient ∇u is to be understood in the distributional sense: We require that the distributional gradient Du is actually a function, and denote this by ∇u. For $u : \Omega \to \mathbb{R}$, the distributional gradient is defined as the linear functional

$$Du(\varphi) := -\int_\Omega u(x) \operatorname{div} \varphi(x) \, dx, \quad (\varphi \in C_c^\infty(\Omega; \mathbb{R}^n)).$$

For Du to be a function, we require that we can actually write

$$Du(\varphi) = \int_\Omega \langle \varphi(x), g(x) \rangle \, dx \tag{1.4}$$

for some $g \in L^1(\Omega; \mathbb{R}^n)$. If u is smooth, the expression (1.4) holds for $g = \nabla u$ by the Gauss-Green theorem. For non-smooth u, we therefore also denote $\nabla u := g$ when (1.4) holds.

The nice thing about Sobolev functions is that the distributional differential Du is a function much more often than u is differentiable in the classical sense. Moreover, Sobolev functions are practical for formulations of partial differential equations in a weak form, that can be solved and manipulated more easily than a classical form; we point the interested reader to [14].

With $p = 1$, there is however a problem. For $p \in (1, \infty)$, the unit ball of $W^{1,p}(\Omega)$ is *weakly compact*. For $p = \infty$ it is *weakly* compact*. Hence, given a bounded sequence $\{u^i\}_{i=1}^{\infty} \subset W^{1,p}(\Omega)$, $(1 < p \leq \infty)$, we can extract a weakly(*) convergent subsequence $\{u^{i_j}\}_{j=1}^{\infty}$. This means in both cases that there exist $u \in W^{1,p}(\Omega)$ such that

$$u^{i_j} \to u \quad \text{strongly in } L^p(\Omega) \quad \text{and} \quad \nabla u^{i_j} \to \nabla u \quad \text{weakly(*) in } L^p(\Omega; \mathbb{R}^n).$$

For $p = 1$ this is not the case! The unit ball is neither weakly nor weakly* compact. There however exists a slightly larger space, the space $\mathrm{BV}(\Omega) \supset W^{1,1}(\Omega)$ of *functions of bounded variation* that has a weakly* compact unit ball. This space will be the main topic of this course.

Why do we want to use and study this space? Let us consider a simple example, and define for $\epsilon > 0$ on $\Omega := [-1, 1]$ the function

$$u_\epsilon(x) := \begin{cases} -1, & x < \epsilon, \\ x/\epsilon, & -\epsilon \leq x \leq \epsilon, \\ 1, & x > \epsilon. \end{cases} \tag{1.5}$$

This function is classically differentiable, and

$$\nabla u_\epsilon(x) = \begin{cases} 1/\epsilon, & -\epsilon \leq x \leq \epsilon, \\ 0, & \text{otherwise.} \end{cases}$$

It can easily be seen that $u_\epsilon \in W^{1,p}(\Omega)$ for every $p \geq 1$.

The pointwise limit of u_ϵ is

$$u_0(x) := \begin{cases} -1, & x < 0, \\ 0, & x = 0, \\ 1, & x > 0. \end{cases} \tag{1.6}$$

This function is not differentiable. Indeed the distributional gradient Du_0 is given by

$$Du_0(\varphi) = 2\varphi(0), \tag{1.7}$$

which does not have the integral representation of (1.4). So clearly $u_0 \notin W^{1,p}(\Omega)$ for any $p \geq 1$. Oh, wait…Sobolev spaces cannot model edges?! Although useful in other contexts, they are not the right space for image processing.

We may also calculate $\|\nabla u_\epsilon\|_{L^p(\Omega)} = 2/\epsilon^{p-1}$, so

$$\lim_{\epsilon \searrow 0} \|\nabla u_\epsilon\|_{L^p(\Omega)} = \infty, \quad \text{for} \quad p \in (1, \infty).$$

Thus u_0 cannot even be approximated in the corresponding Sobolev norm by elements of $W^{1,p}(\Omega)$ for $p > 1$. But how about $p = 1$. We have

$$\|\nabla u_\epsilon\|_{L^1(\Omega)} = 2, \quad \text{for} \quad \epsilon > 0,$$

so $\{u_\epsilon\}_{\epsilon > 0}$ forms a bounded set in $W^{1,1}(\Omega)$. It therefore has a weakly* convergent subsequence in $\mathrm{BV}(\Omega)$. In fact, the whole sequence must converge weakly* in $\mathrm{BV}(\Omega)$ to u as $\epsilon \searrow 0$, because it converges in $L^1(\Omega)$. Thus we have found a function $u \in \mathrm{BV}(\Omega) \setminus W^{1,1}(\Omega)$. We have also discovered that the space $\mathrm{BV}(\Omega)$ can model edges in a controlled manner. That sounds very very good and interesting for further study!

1.5. Functions of bounded variation in $\mathbb{R}^1$

We now briefly look at the case $\Omega = (a, b) \subset \mathbb{R}^1$, in order to see what kind of bizarre creatures can inhabit $\mathrm{BV}(\Omega)$.

Definition 1.1. Let $\Omega = (a, b)$ and $u \in L^1(\Omega)$. We define the *pointwise variation*

$$\mathrm{pV}(u, \Omega) := \sup\Big\{ \sum_{i=1}^{n} |u(t_{i+1}) - u(t_i)| \;\Big|\; n \geq 2,\, a < t_1 < \cdots < t_n < b \Big\}.$$

Roughly speaking $\mathrm{pV}(u, \Omega) < \infty$ defines functions in $\mathrm{BV}(\Omega)$, but we have to be careful: functions in $L^1(\Omega)$ are only defined almost everywhere, so we have to set

$$\mathrm{eV}(u, \Omega) := \inf\{\mathrm{pV}(v, \Omega) \mid v = u \text{ almost everywhere in } \Omega\}.$$

Then

$$\mathrm{BV}(\Omega) = \{u \in L^1(\Omega) \mid \mathrm{eV}(u, \Omega) < \infty\}.$$

We will get back to the rigorous definition of "almost everywhere"; for the rest of this section, we will only work with such representatives u that $\mathrm{pV}(u, \Omega) = \mathrm{eV}(u, \Omega)$.

Example 1.1. We now demonstrate that rather weird creatures satisfy $\mathrm{pV}(u, \Omega) < \infty$. We begin by constructing the Cantor middle-third set. We set $C_0 = [0, 1]$, and assuming C_k has been constructed, we construct C_{k+1} as

$$C_{k+1} := \psi_1(C_k) + \psi_2(C_k)$$

where

$$\psi_1(x) = x/3, \quad \text{and} \quad \psi_2(x) = 2/3 + x/3.$$

In other words, we create C_{k+1} by removing the middle-third of $(0, 1)$ altogether, and replacing the ends by scaled-down copies of C_k. The Cantor set is then

$$C := \bigcap_{k=0}^{\infty} C_k.$$

By construction, it is a *self-similar fractal*;

$$C = \psi_1^{-1}(C \cap (0, 1/3)) = \psi_2^{-1}(C \cap (2/3, 1)).$$

Let us then define

$$f_k(x) := \int_0^x (3/2)^k \chi_{C^k}(t)\, dt, \quad (x \in (0, 1)).$$

It can be seen that $\{f_k\}_{k=1}^{\infty}$ is a Cauchy sequence in $C(0, 1)$, and hence convergent to some $f \in C(0, 1)$. This is called the *Cantor-Vitali* function. Observe that $f_k' = 0$ on $[0, 1] \setminus C_k$, so that also $f' = 0$ on $[0, 1] \setminus \bigcup_{i=1}^{k} C_k$ for any $k \geq 1$. It follows that in a sense (that we will make strict later), $f' = 0$ almost everywhere. Yet f is continuous with $f(0) = 0$ and $f(1) = 1$. Indeed, $\mathrm{pV}(f, (0, 1)) = 1$!

It turns out that Df is zero almost everywhere with respect to the Lebesgue measure, but is concentrated on the set C, which has Hausdorff dimension $\log_3 2 < 1$. Thus the differential of C is "too small to be detected" by classical tools. But now we really have to be rigorous and start defining the measure-theoretic notions we've been waving our hands about.

Chapter 2

Sets and measures

2.1. Basic measure theory

A *measure* is a way to formalise the intuitive idea of the area of a set in $\mathbb{R}^2$, as well as the length of a curve, and the number of elements in a discrete set, under one concept. Measures are defined on sets that belong to a *σ-algebra*. The σ-algebra defines sets that are *measurable*. We have to place restrictions on sets that are measurable, otherwise we may reach conclusions like the Banach-Tarski paradox, which allows reconstructing a solid three-dimensional ball into two exact copies of itself!

Definition 2.1. Let Ω be a set, and denote by $\mathcal{P}(\Omega)$ the collection of all subsets of Ω. Then a family $\Sigma \subset \mathcal{P}(\Omega)$ is a *σ-algebra*, if the following properties are satisfied:

(i) Non-empty: $\Sigma \neq \emptyset$.

(ii) Closed under complement: If $A \in \Sigma$, then also $A^c \in \Sigma$.

(iii) Closed under countable unions: If $\{A_i\}_{i=1}^{\infty} \subset \Sigma$, then also $\bigcup_{i=1}^{\infty} A_i \in \Sigma$.

A set $A \in \Sigma$ is called a *(Σ-)measurable set*, and we call the pair (Ω, Σ) a *measure space*.

The idea is that if we can measure, let's say, the area of a set, and take something measurable out of or add to it, then the result should be measurable. Typically, in vector spaces, also any translation of a measurable set is measurable. Indeed, in the setting of metric spaces, it usually suffices to work with the smallest σ-algebra containing all open (and hence all closed) sets.

Definition 2.2. Let Ω be a metric space. Then we denote by $\mathcal{B}(\Omega)$ the smallest σ-algebra containing all open sets of Ω. We call it the *Borel* σ-algebra of Ω. A set $A \in \mathcal{B}(\Omega)$ is called *Borel-measurable*.

Example 2.1. Another example of a σ-algebra is the *discrete $\sigma-$algebra* consisting of all subsets of Ω. It is however rarely practical.

Definition 2.3. Let (Ω, Σ) be a measure space. Then a mapping $\mu : \Sigma \to [0, \infty]$ is a *(positive) measure* if the following properties are satisfied:

(i) Non-negativity: $\mu(A) \geq 0$ for all $A \subset \Sigma$.

(ii) Null empty set: $\mu(\emptyset) = 0$.

(iii) Countable additivity: If the sets $\{A_i\}_{i=1}^{\infty} \subset \Sigma$ are pairwise disjoint, $A_i \cap A_j = \emptyset$ for $i \neq j$, then $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.

Example 2.2. Let us define for every $A \subset \Omega$ the set function

$$\#(A) := \begin{cases} n, & A = \{x_1, \ldots, x_n\} \text{ is a finite set}, \\ \infty, & \text{otherwise}. \end{cases}$$

Then $\#$ is a measure on the discrete σ-algebra of Example 2.1.

Example 2.3. Let us pick $x \in \Omega$ and define for every $A \subset \Omega$ the set function

$$\delta_x(A) := \begin{cases} 1, & x \in A, \\ 0, & \text{otherwise.} \end{cases}$$

Then δ_x is a measure on the discrete σ-algebra of Example 2.1, called the *Dirac measure* (at x). It is the measure-theoretic realisation of what is sometimes incorrectly referred to as the *Dirac δ-function.*

Definition 2.4. We call a positive measure μ *finite* if $\mu(\Omega) < \infty$. If $\Omega = \bigcup_{i=1}^{\infty} A_i$ with $\mu(A_i) < \infty$, we call μ *σ-finite.* A positive measure μ on $\mathcal{B}(\Omega)$ is a *Borel measure.* If, moreover, $|\mu(K)| < \infty$ on compact sets $K \subset \Omega$, we call μ a *positive Radon measure*, and denote $\mu \in \mathcal{M}(\Omega)$.

We can also define signed and general vector-valued measures. These will be useful later on, when we want to differentiate functions of bounded variation.

Definition 2.5. Let (Ω, Σ) be a measure space. Then a mapping $\mu : \Sigma \to \mathbb{R}^m$ is a *(vector-valued) measure* if the following properties are satisfied:

(i) Null empty set: $\mu(\emptyset) = 0$.

(ii) Countable additivity: If $\{A_i\}_{i=1}^{\infty} \subset \Sigma$ with A_i pairwise disjoint, then $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.

If $m = 1$, we also call μ a signed measure. We define the *total variation of μ* for every $A \in \Sigma$ by

$$|\mu|(A) := \sup\{\sum_{k=1}^{\infty} \|\mu(A_i)\| \mid A = \bigcup_{k=1}^{\infty} A_i,\, A_i \in \Sigma \text{ pairwise disjoint}\}.$$

If $\mu : \Sigma \to \mathbb{R}$ is a signed real measure, we define the positive and negative parts by

$$\mu^+ := \frac{1}{2}(|\mu| + \mu), \quad \mu^- := \frac{1}{2}(|\mu| - \mu).$$

We immediately observe that if μ is an $\mathbb{R}^m$-valued measure, then μ_i, defined by

$$\mu_i(A) := \langle e_i, \mu(A) \rangle$$

is a signed measure. Here e_i is the standard unit vector with the i:th entry 1 and other entries zero.

Definition 2.6. We call a $\mathbb{R}^m$-valued measure μ on $\mathcal{B}(\Omega)$ a *finite Radon measure*, and denote $\mu \in \mathcal{M}(\Omega; \mathbb{R}^m)$.

The finiteness in this definition is justified by the exercise in example sheet 1, showing that $|\mu|$, μ^+, and μ^- are positive finite measures if μ is signed or vector-valued measure.

We will often talk about a property holding "almost everywhere" with respect to a measure. Let us make this precise.

Definition 2.7. Let μ be a measure on a measure space (Ω, Σ), and $A \in \Sigma$. We say that a property P holds for *μ-almost every $x \in A$*, in short *μ-a.e.*, if there exists $E \subset A$ such that P holds for every $x \in E$, and $\mu(A \setminus E) = 0$.

Functions can also be measurable. This is important for defining integration by measures.

Definition 2.8. Let (Ω, Σ) be a measure space, and $f : \Omega \to \mathbb{R}$. Then we call f *Σ-measurable* if $f^{-1}(A) \in \Sigma$ for every open set $A \subset \mathbb{R}$. If μ is a measure on (Ω, Σ), we call f *μ-measurable* if f is Σ-measurable. If $\Sigma = \mathcal{B}(\Omega)$, we call f *Borel-measurable.*

2.2. Integration

Definition 2.9. Let (Ω, Σ) be a measure space, $f : \Omega \to [0, \infty)$ a measurable function, and μ a positive measure on Σ. Let $E \in \Sigma$. We define the integral

$$\int_E f(x)\, d\mu(x) := \sup_s \sum_{i=1}^n \alpha_i \mu(A_i \cap E)$$

with the supremum taken over all *simple functions*

$$s = \sum_{i=1}^n \alpha_i \chi_{A_i} \le f, \quad (\alpha_i > 0, A_i \in \Sigma;\ i = 1, \ldots, n;\ n \in \mathbb{N}).$$

If $f : \Omega \to \mathbb{R}$, we write $f = f^+ - f^-$ for $f^\pm \ge 0$, and set

$$\int_E f(x)\, d\mu(x) = \int_E f^+(x)\, d\mu(x) - \int_E f^-(x)\, d\mu(x).$$

We call f *integrable*, and denote $f \in L^1(\Omega; \mu)$, if

$$\int_\Omega |f|(x)\, d\mu(x) < \infty.$$

Definition 2.10. Let (Ω, Σ) be a measure space. If μ is a signed measure on Σ, and $f \in L^1(\Omega; |\mu|)$, we define

$$\int_E f\, d\mu = \int_E f\, d\mu^+ - \int_E f\, d\mu^-, \quad (E \in \Sigma).$$

If μ is an $\mathbb{R}^m$-valued measure on Σ, and $f = (f_1, \ldots, f_m) : \Omega \to \mathbb{R}^m$ with each $f_i \in L^1(\Omega; |\mu_i|)$, we define

$$\int_E \langle f(x), d\mu(x)\rangle := \sum_{i=1}^m \int_E f_i(x)\, d\mu_i(x), \quad (E \in \Sigma).$$

Theorem 2.1 (Fatou's lemma). *Suppose $\{u^i : \Omega \to [0, \infty]\}_{i=1}^\infty$ are μ-measurable with μ a positive measure. Then*

$$\int_\Omega \liminf_{i\to\infty} u^i(x)\, d\mu(x) \le \liminf_{i\to\infty} \int_\Omega u^i(x)\, d\mu(x)$$

Definition 2.11 (Product σ-algebras). Let (Ω_1, Σ_1) and (Ω_2, Σ_2) be measure spaces. Then we define the *product σ-algebra* $\Sigma_1 \times \Sigma_2$ on $\Omega_1 \times \Omega_2$ as the smallest σ-algebra that contains all sets of the form $A_1 \times A_2$ for $A_1 \in \Sigma_1$ and $A_2 \in \Sigma_2$.

Theorem 2.2 (Fubini). *Let (Ω_1, Σ_1) and (Ω_2, Σ_2) be measure spaces and μ_1 and μ_2 be corresponding positive σ-finite measures. Then there exists a unique positive σ-finite measure μ on $(\Omega_1 \times \Omega_2, \Sigma_1 \times \Sigma_2)$ such that*

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2).$$

If $u : \Omega_1 \times \Omega_2 \to [0, \infty]$ is μ-measurable, then the marginal mappings

$$x \mapsto \int_{\Omega_2} u(x, y)\, d\mu_2(y) \quad \text{and} \quad y \mapsto \int_{\Omega_1} u(x, y)\, d\mu_1(x)$$

are, respectively, Σ_1-measurable and Σ_2-measurable. Moreover, we can change the order of integration by the formula

$$\begin{aligned}\int_{\Omega_1\times\Omega_2} u(x, y)\, d\mu(x, y) &= \int_{\Omega_1} \int_{\Omega_2} u(x, y)\, d\mu_2(y)\, d\mu_1(x) \\ &= \int_{\Omega_2} \int_{\Omega_1} u(x, y)\, d\mu_1(x)\, d\mu_2(y).\end{aligned}$$

2.3. Lebesgue measure

Enough with toys like the counting measure! Let's define something really useful: a way to measure the volume of general sets in $\mathbb{R}^n$.

To begin with, if $Q = [a_1, b_1] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$ is an n-dimensional cube, we define the volume

$$v(Q) = \prod_{j=1}^{n} (b_j - a_j).$$

We desire to extend this definition to general sets.

Definition 2.12. Let $A \subset \mathbb{R}^n$, and define *the Lebesgue outer measure* by

$$\widetilde{\mathcal{L}}^n(A) = \inf \sum_{i=1}^{\infty} v(Q_i)$$

where the infimum is taken over all countable collections $\{Q_i\}_{i=1}^{\infty}$ of cubes in $\mathbb{R}^n$ such that

$$A \subset \bigcup_{i=1}^{\infty} Q_i.$$

Theorem 2.3. *Let $\Omega \in \mathcal{B}(\mathbb{R}^n)$. The mapping $\widetilde{\mathcal{L}}^n$ restricted to the Borel σ-algebra $\mathcal{B}(\Omega)$ is a positive measure. We call it the* Lebesgue measure*, and denote it by $\mathcal{L}^n$. We have*

$$\mathcal{L}^n(Q) = v(Q).$$

Proof. Let us show that $\widetilde{\mathcal{L}}^n(Q) = v(Q)$ for a cube Q. Clearly $\widetilde{\mathcal{L}}^n(Q) \leq v(Q)$. To show the other inequality, we take an ϵ-cover $\{Q_j\}_{j=1}^{\infty}$ of Q, i.e.,

$$\sum_{j=1}^{\infty} v(Q_j) \leq \widetilde{\mathcal{L}}^n(Q) + \epsilon.$$

We may subdivide each Q_j into a family of cubes $Q_j^1, \ldots, Q_j^{2^n}$ such that $Q_j^1 = Q_j \cap Q$ and $v(Q_j) = \sum_{i=1}^{2^n} v(Q_j^i)$. But then $v(Q) \leq \sum_{j=1}^{\infty} v(Q_j^1)$, so that

$$v(Q) \leq v(Q) + \sum_{j=1}^{\infty} \sum_{i=2}^{2^n} v(Q_j^i) = \sum_{j=1}^{\infty} v(Q_j) \leq \widetilde{\mathcal{L}}^n(Q) + \epsilon.$$

Since $\epsilon > 0$ was arbitrary, this concludes the proof.

In order to show that $\widetilde{\mathcal{L}}^n$ is indeed a measure on $\mathcal{B}(\Omega)$, we have to verify the axioms in Definition 2.3. That $\mathcal{L}^n(A) \geq 0$ for $A \in \mathcal{B}(\Omega)$ is immediate because $v \geq 0$. Picking $Q_i = [0, \epsilon/2^{i/n}]$, we have $\sum_{i=1}^{\infty} v(Q_i) = \epsilon$, so letting $\epsilon \searrow 0$, we see that clearly $\mathcal{L}^n(\emptyset) = 0$. We therefore just have to verify countable additivity. So suppose $A_i \in \mathcal{B}(\Omega)$, $(i = 1, 2, 3, \ldots)$ are pairwise disjoint, and let $A := \bigcup_{i=1}^{\infty} A_i$. We may assume that $\widetilde{\mathcal{L}}^n(A_i) < \infty$, $(i = 1, 2, \ldots)$, because otherwise clearly $\widetilde{\mathcal{L}}^n(A) = \infty$. Pick arbitrary $\epsilon > 0$. By the definition of $\widetilde{\mathcal{L}}^n$, we may find collections $\{Q_i^j\}_{j=1}^{\infty}$ of n-dimensional cubes such that

$$A_i \subset \bigcup_{j=1}^{\infty} Q_i^j,$$

and

$$\sum_{j=1}^{\infty} v(Q_i^j) \leq \widetilde{\mathcal{L}}^n(A_i) + \epsilon/2^i.$$

Then

$$A \subset \bigcup_{i=1}^{\infty} \bigcup_{j=1}^{\infty} Q_i^j,$$

so

$$\widetilde{\mathcal{L}}^n(A) \le \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} v(Q_i^j) \le \sum_{i=1}^{\infty} \widetilde{\mathcal{L}}^n(A_i) + \epsilon.$$

Since $\epsilon > 0$ was arbitrary, we conclude that

$$\widetilde{\mathcal{L}}^n(A) \le \sum_{i=1}^{\infty} \widetilde{\mathcal{L}}^n(A_i).$$

We now have to prove the opposite inequality. Let us, first of all, verify that it holds for countably many sets, if it holds for two disjoint sets, that is

$$\widetilde{\mathcal{L}}^n(A) + \widetilde{\mathcal{L}}^n(B) \le \widetilde{\mathcal{L}}^n(A \cup B), \quad (A, B \in \mathcal{B}(\Omega),\ A \cap B = \emptyset). \tag{2.1}$$

Indeed, then for any $k \ge 2$, we have

$$\sum_{i=1}^{k} \widetilde{\mathcal{L}}^n(A_i) \le \widetilde{\mathcal{L}}^n(\bigcup_{i=1}^{k} A_i) \le \widetilde{\mathcal{L}}^n(A), \tag{2.2}$$

where the last inequality holds because any cover of A by cubes also covers $\bigcup_{i=1}^{\infty} A_i \subset A$. Letting $k \to \infty$ proves the claim under (2.1). Hence countable additivity follows from Lemma 2.3 below, where we prove (2.1). It follows that $\mathcal{L}^n = \widetilde{\mathcal{L}}^n$ is a measure on $\mathcal{B}(\Omega)$. □

Definition 2.13. For $A, B \subset \mathbb{R}^n$, we define the distance

$$d(A, B) := \inf_{x \in A, y \in B} \|x - y\|.$$

Lemma 2.1. *Let $A, B \subset \Omega$. Then the inequality (2.1) holds in the following cases.*

(i) $d(A, B) = r > 0$.

(ii) *A and B are closed and disjoint.*

Proof. (i) Pick $\epsilon > 0$, and let $\{Q_j\}_{j=1}^{\infty}$ be a covering of $A \cup B$ by cubes such that

$$\sum_{j=1}^{\infty} v(Q_j) \le \widetilde{\mathcal{L}}^n(A \cup B) + \epsilon.$$

Subdividing each Q_j into smaller sub-cubes, we may assume that there are two subsets of indices $\mathcal{I}, \mathcal{J} \subset \mathbb{N}$ such that

$$A \subset \bigcup_{j \in \mathcal{I}} Q_j, \quad B \subset \bigcup_{j \in \mathcal{J}} Q_j,$$

and

$$\sum_{j \in \mathcal{I} \cap \mathcal{J}} v(Q_j) \le \epsilon. \tag{2.3}$$

Indeed, let k be such that

$$\sum_{j=k+1}^{\infty} v(Q_j) \le \epsilon,$$

and define

$$\widetilde{A} := A \cap \bigcup_{j=1}^{k} Q_j, \quad \text{and} \quad \widetilde{B} := B \cap \bigcup_{j=1}^{k} Q_j.$$

Because $d(\widetilde{A},\widetilde{B}) \geq d(A,B) = r > 0$, by possibly subdividing each Q_j for $j = 1,\dots,k$ into smaller cubes, we can find $\widetilde{\mathcal{I}},\widetilde{\mathcal{J}} \subset \{1,\dots,k\}$, such that $\widetilde{\mathcal{I}} \cap \widetilde{\mathcal{J}} = \emptyset$ and

$$\widetilde{A} \subset \bigcup_{j\in\widetilde{\mathcal{I}}} Q_j, \quad \widetilde{B} \subset \bigcup_{j\in\widetilde{\mathcal{J}}} Q_j.$$

We now simply let $\mathcal{I} = \widetilde{\mathcal{I}} \cup \{k, k+1, \dots\}$ and $\mathcal{J} = \widetilde{\mathcal{J}} \cup \{k, k+1, \dots\}$ to construct covers of A and B. The estimate (2.3) follows. With that at hand, we may calculate

$$\widetilde{\mathcal{L}}^n(A) + \widetilde{\mathcal{L}}^n(B) \leq \sum_{j\in\mathcal{I}} v(Q_j) + \sum_{j\in\mathcal{J}} v(Q_j) + \epsilon \leq \sum_{j=1}^{\infty} v(Q_j) + \epsilon \leq \widetilde{\mathcal{L}}^n(A\cup B) + 2\epsilon.$$

Since $\epsilon > 0$ was arbitrary, this concludes the proof of (2.1).

(ii) Finally, if A and B are closed and disjoint, then $\widetilde{A}$ and $\widetilde{B}$ are closed, bounded, and disjoint. Therefore $d(\widetilde{A},\widetilde{B}) > 0$. The previous argument applies. □

Lemma 2.2. *Let $A \in \mathcal{B}(\mathbb{R}^n)$ and $\epsilon > 0$. Suppose $\mathcal{L}^n(A) < \infty$. Then there exist closed $F \subset A$ and open $G \supset A$ such that $\widetilde{\mathcal{L}}^n(A \setminus F) \leq \epsilon$ and $\widetilde{\mathcal{L}}^n(G \setminus A) \leq \epsilon$.*

Proof. Suppose first that A is open. We may then choose $G = A$. Regarding F, we set

$$F_i := \{x \in A \mid d(\partial A, x) \geq 1/i\}.$$

Then F_i is closed, and

$$A = \bigcup_{i=1}^{\infty} F_i.$$

Let $T_{i+1} := F_{i+1} \setminus F_i$, and observe that $d(T_j,T_i) > 0$ whenever $|j-i| \geq 2$. Therefore by Lemma 2.1 and (2.2), we have

$$\sum_{j=1}^{\infty} \widetilde{\mathcal{L}}^n(T_{2j}) \leq \widetilde{\mathcal{L}}^n(\bigcup_{j=1}^{\infty} T_{2j}) \leq \widetilde{\mathcal{L}}^n(A) < \infty,$$

and likewise

$$\sum_{j=1}^{\infty} \widetilde{\mathcal{L}}^n(T_{1+2j}) \leq \widetilde{\mathcal{L}}^n(\bigcup_{j=1}^{\infty} T_{1+2j}) \leq \widetilde{\mathcal{L}}^n(A) < \infty.$$

Thus

$$\sum_{j=N+1}^{\infty} \widetilde{\mathcal{L}}^n(T_j) < \epsilon.$$

for large enough N. But $A \setminus F_N = \bigcup_{j=N+1}^{\infty} T_j$, so that

$$\widetilde{\mathcal{L}}^n(A \setminus F_N) \leq \epsilon.$$

If A is closed, we may choose $F = A$ and G is obtained by a construction analogous to the above using

$$G_i := \{x \in \Omega \mid d(A,x) < 1/i\}.$$

Let $\mathcal{D}$ be the class of sets A on which the claim of the lemma holds for any $\epsilon > 0$. Clearly, by the above, $\mathcal{D}$ contains all open and closed sets A with $\mathcal{L}^n(A) < \infty$. Here we recall that we have assumed the atter from our set of interest as well. Moreover, if $A, B \in \mathcal{D}$, then $A \setminus B \in \mathcal{D}$. Indeed, if $F \subset A$ and $G \supset B$, then

$$(A \setminus B) \setminus (F \setminus G) \subset (A \setminus F) \cup (G \setminus B).$$

Thus

$$\widetilde{\mathcal{L}}^n\Big((A \setminus B) \setminus (F \setminus G)\Big) \leq \epsilon.$$

whenever $\widetilde{\mathcal{L}}^n(A \setminus F) \leq \epsilon/2$ and $\widetilde{\mathcal{L}}^n(G \setminus B) \leq \epsilon/2$. As $F \setminus G$ is closed for F closed and G open, one side of the claim follows. The other is analogous.

Let then $\{A_i\}_{i=1}^\infty \in \mathcal{D}$, and pick closed $F_i \subset A_i$ and open $G_i \supset A_i$ such that

$$\widetilde{\mathcal{L}}^n(A_i \setminus F_i) \leq \epsilon/2^i, \quad \text{and} \quad \widetilde{\mathcal{L}}^n(G_i \setminus A_i) \leq \epsilon/2^i.$$

If we set

$$A := \bigcap_{i=1}^\infty A_i \quad \text{and} \quad F := \bigcap_{i=1}^\infty F_i,$$

Then

$$A \setminus F \subset \bigcup_{i=1}^\infty (A_i \setminus F_i),$$

so that

$$\widetilde{\mathcal{L}}^n(A \setminus F) \leq \sum_{i=1}^\infty \widetilde{\mathcal{L}}^n(A_i \setminus F_i) \leq \sum_{i=1}^\infty \epsilon/2^i = \epsilon.$$

Moreover, if we pick $N \in \mathbb{N}$, and set

$$A^N := \bigcap_{i=1}^N A_i, \quad \text{and} \quad G^N := \bigcap_{i=1}^N G_i,$$

then

$$G^N \setminus A^N \subset \bigcup_{i=1}^N (G_i \setminus A_i),$$

so that

$$\widetilde{\mathcal{L}}^n(G^N \setminus A^N) \leq \sum_{i=1}^N \widetilde{\mathcal{L}}^n(G_i \setminus A_i) \leq \epsilon.$$

It follows that $\mathcal{D}$ contains finite intersections. Analogously we show that $\mathcal{D}$ contains finite unions, and that if

$$A := \bigcup_{i=1}^\infty A_i \quad \text{and} \quad G := \bigcup_{i=1}^\infty G_i,$$

then

$$\widetilde{\mathcal{L}}^n(G \setminus A) \leq \sum_{i=1}^\infty \widetilde{\mathcal{L}}^n(G_i \setminus A_i) \leq \epsilon.$$

Let us then consider countable unions. We set

$$A := \bigcup_{i=1}^\infty A_i \quad \text{and} \quad F^k := \bigcup_{i=1}^k F_i,$$

By the above, by possibly replacing A_i by $A_i \setminus \bigcup_{j=1}^{i-1} A_j$, we may assume that the sets $\{A_i\}_{i=1}^\infty$ are disjoint. Hence also $\{F_i\}_{i=1}^\infty$ are disjoint. Now

$$\widetilde{\mathcal{L}}^n(A \setminus \bigcup_{i=1}^\infty F_i) \leq \sum_{i=1}^\infty \mathcal{L}^n(A_i \setminus F_i) \leq \epsilon.$$

It remains to show show that for some k

$$\widetilde{\mathcal{L}}^n(\bigcup_{i=1}^\infty F_i \setminus F^k) \leq \epsilon, \tag{2.4}$$

because then

$$\widetilde{\mathcal{L}}^n(A \setminus F^k) \le \widetilde{\mathcal{L}}^n(A \setminus \bigcup_{i=1}^{\infty} F_i) + \widetilde{\mathcal{L}}^n(\bigcup_{i=1}^{\infty} F_i \setminus F_k) \le 2\epsilon.$$

But the sets F_i are closed and disjoint, so by Lemma 2.1 and (2.2), and

$$\infty > \mathcal{L}^n(A) \ge \widetilde{\mathcal{L}}^n(\bigcup_{i=1}^{\infty} F_i) = \sum_{i=1}^{\infty} \widetilde{\mathcal{L}}^n(F_i).$$

In particular $\mathcal{L}^n(F_i) \to 0$. But

$$\widetilde{\mathcal{L}}^n(\bigcup_{i=1}^{\infty} F_i \setminus F^k) \le \sum_{i=k+1}^{\infty} \widetilde{\mathcal{L}}^n(F_i),$$

Thus (2.4) holds for large enough k, and in consequence $\mathcal{D}$ contains countable unions.

Finally, $\mathcal{D}$ contains countable intersections: If $A = \bigcap_{i=1}^{\infty} A_i$, we may assume $A_1 \supset A_2 \supset \ldots$. Then $A_1 \setminus A = \bigcup_{i=1}^{\infty}(A_1 \setminus A_i)$. But then $A_1 \setminus A \in \mathcal{D}$, and so also $A = A_1 \setminus (A_1 \setminus A) \in \mathcal{D}$.

It follows that $\mathcal{D}$ contains all Borel sets A with $\widetilde{\mathcal{L}}^n(A) < \infty$, constructed out of open and closed sets A_i satisfying this assumption. To finish the proof, we observe that we do not in fact need sets A_i with $\widetilde{\mathcal{L}}^n(A_i) = \infty$ in the construction of Borel sets, as we can always write $A_i = \bigcup_{j=1}^{\infty} A_i \cap B(0,j)$, where $\widetilde{\mathcal{L}}^n(A_i \cap B(0,j)) < \infty$. □

Lemma 2.3. *Let $A, B \in \mathcal{B}(\Omega)$ with $A \cap B = \emptyset$. Then (2.1) holds.*

Proof. Let $\epsilon > 0$. By Lemma 2.2, we may find closed sets $F_A \subset A$ and $F_B \subset B$ such that $\widetilde{\mathcal{L}}^n(A \setminus F_A) \le \epsilon$ and $\widetilde{\mathcal{L}}^n(B \setminus F_B) \le \epsilon$. Then

$$\begin{aligned}\widetilde{\mathcal{L}}^n(A) + \widetilde{\mathcal{L}}^n(B) &\le \widetilde{\mathcal{L}}^n(F_A) + \widetilde{\mathcal{L}}^n(A \setminus F_A) + \widetilde{\mathcal{L}}^n(F_B) + \widetilde{\mathcal{L}}^n(B \setminus F_B) \\ &\le \widetilde{\mathcal{L}}^n(F_A) + \widetilde{\mathcal{L}}^n(F_B) + 2\epsilon\end{aligned}$$

It is an easy exercise to show that $\widetilde{\mathcal{L}}^n(A_1) \le \widetilde{\mathcal{L}}^n(A_2)$ if $A_1 \subset A_2$. Therefore, because (2.1) holds for disjoint closed sets by Lemma 2.1, and $F_A \cup F_A \subset A \cup B$, we further obtain

$$\begin{aligned}\widetilde{\mathcal{L}}^n(A) + \widetilde{\mathcal{L}}^n(B) &\le \widetilde{\mathcal{L}}^n(F_A \cup F_B) + 2\epsilon \\ &\le \widetilde{\mathcal{L}}^n(A \cup B) + 2\epsilon.\end{aligned}$$

As $\epsilon > 0$ was arbitrary, (2.1) follows. □

Remark 2.1. The Lebesgue measure can be extended to a larger σ-algebra than $\mathcal{B}(\Omega)$, consisting of so-called *Lebesgue measurable* sets. For our purposes, it suffices to limit the attention to Borel-measurable sets.

Definition 2.14. We will denote integration by the Lebesgue measure by the usual notation for Riemann integration, as these agree on continuous functions. That is

$$\int_A f(x)\,dx := \int_A f(x)\,d\mathcal{L}^n(x).$$

We also denote $L^1(\Omega) := L^1(\Omega; \mathcal{L}^n)$.

2.4. Hausdorff measure

In the definition of the Lebesgue measure, we could calculate the the volume $v(Q_j)$ of the cube Q_j through the diameter $\operatorname{diam} Q_j$ as $v(Q_j) = \omega(n)2^{-n}(\operatorname{diam} Q_j)^n$ for a suitable dimensional constant $\omega(n)$, which gives the volume of the unit ball. We could further place an upper bound on the diameter, $\operatorname{diam} Q_j < \epsilon$, because we can always decompose a cube into smaller subcubes without altering the total volume. In fact, we do not even have to use cubes, since we can approximate an arbitrary set by cubes. This leads us to the definition of the Hausdorff measure, where we also allow instead of the dimension n of $\Omega \subset \mathbb{R}^n$ an arbitrary possibly non-integer dimension $k \in [0, \infty)$.

Definition 2.15. Let $A \in \mathcal{B}(\Omega)$, $\Omega \subset \mathbb{R}^n$, and $k \in [0, \infty)$. Then the *k-dimensional Hausdorff measure* of A is defined by

$$\mathcal{H}^k(A) := \lim_{\epsilon \searrow 0} \mathcal{H}^k_\epsilon(A),$$

where

$$\mathcal{H}^k_\epsilon(A) := \inf\{\sum_{j=1}^\infty \omega(k) 2^{-k} (\operatorname{diam} E_j)^k \mid A \subset \bigcup_{j=1}^\infty E_j, \operatorname{diam} E_j < \epsilon\}.$$

The normalisation constant is defined as

$$\omega(k) := \frac{\pi^{k/2}}{\Gamma(1 + k/2)}, \quad \Gamma(s) = \int_0^\infty e^{-x} x^{s-1} \, dt.$$

The point of forcing $\epsilon \searrow 0$ is to measure complicated sets accurately. For example, the set

$$A := \{(x, \sin(1/x)) \mid 0 < x \leq 1\} \subset \mathbb{R}^2$$

has $\mathcal{H}^1_\epsilon(A) < \infty$ for every $\epsilon > 0$, but $\mathcal{H}^1(A) = \infty$. We will get back to how to arrive at these, when we talk about rectifiable sets.

The idea behind using the 1-dimensional measure $\mathcal{H}^1$ to measure sets $A \subset \mathbb{R}^2$, is that $\mathcal{H}^1$ measures the length of curves. Likewise $\mathcal{H}^0$ measures the number of points; it just the the counting measure. In $\Omega \subset \mathbb{R}^2$, the measure $\mathcal{H}^2$ coincides with the Lebesgue measure $\mathcal{L}^2$ (when both are restricted to the Borel measurable sets $\mathcal{B}(\Omega)$). But we may also use $\mathcal{H}^2$ on $\Omega \subset \mathbb{R}^3$ to measure the area of more complicated surfaces, such as the unit sphere $\mathbb{S}^2$, where we generally define

$$\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n \mid \|x\| = 1\}.$$

How about $\mathcal{H}^s$ non-integral s? These measure fractals! The Cantor middle-thirds set C defined in Example 1.1 has $\mathcal{H}^k(C) = 1$ for $k = k^* := \log 2 / \log 3$, but $\mathcal{H}^k(C) = \infty$ for $k < k^*$ and $\mathcal{H}^k(C) = 0$ for $k > k^*$. This motivates the following.

Definition 2.16. Let $A \in \mathcal{B}(\Omega)$. The *Hausdorff dimension* of A is

$$\dim_{\mathcal{H}} A := \inf\{s \geq 0 \mid \mathcal{H}^s(A) = 0\}.$$

The enthusiastic student may read more about fractals and Hausdorff dimensions in the literature listed in the Bibliography. In the rest of this course, we will primarily concentrate on a particular type of sets of dimension $n - 1$ in $\mathbb{R}^n$, as well as a "remainder set" of arbitrary dimension $k \in (n - 1, n)$, arising from the differentiation of functions of bounded variation.

2.5. Densities and derivation of measures

Definition 2.17. Let μ be a measure on a measure space (Ω, Σ), and $f : \Omega \to \mathbb{R}$ measurable. We then define the measure $f\mu$ by

$$(f\mu)(A) = \int_A f(x) \, d\mu(x), \quad (A \in \Sigma).$$

Example 2.4. Let $f(x) := e^{-x^2/2} / \sqrt{2\pi}$ and $\mathcal{N} := f\mathcal{L}^1$. Then $\mathcal{N}$ is a *probability measure* for the Gaussian distribution. A probability measure is a positive measure with $\mathcal{N}(\Omega) = 1$.

Definition 2.18. Let ν be a (signed or vector) measure and μ a positive measure. If

$$\mu(A) = 0 \implies |\nu|(A) = 0,$$

then we say that ν *is absolutely continuous with respect to* μ, denoted $\nu \ll \mu$. If, on the other hand, μ and ν are positive measures, and there exists $A \in \Sigma$ such that $\mu(E) = 0$ and $\nu(\Omega \setminus E) = 0$, we say that μ and ν are *mutually singular*, denoted $\mu \perp \nu$. If μ and ν are vector-valued, we say the same if this holds for $|\mu|$ and $|\nu|$.

Example 2.5. We have $\delta_x \perp \mathcal{L}^n$ for $x \in \mathbb{R}^n$, and $\mathcal{N} \ll \mathcal{L}^1$.

Obviously $\mu \ll |\mu|$. In fact, we have the following.

Theorem 2.4 (Polar decomposition)**.** *Let μ be a (signed or vector) measure on a measure space (Ω, Σ). Then $\mu \ll |\mu|$. In this case we write*

$$\mu = \left(\frac{d\mu}{d|\mu|}\right)|\mu|$$

and call this the polar decomposition *of μ.*

This result is a corollary of the Radon-Nikodým theorem.

Theorem 2.5 (Radon-Nikodým)**.** *Let $\nu : \Sigma \to \mathbb{R}^m$ be a (signed or vector) measure and μ a σ-finite measure on a measure space (Ω, Σ). Then there exist unique measures ν^a and ν^s such that $\nu^a \ll \mu$ and $\nu^s \perp \mu$ with $\nu = \nu^a + \nu^s$. Moreover, there exists a unique function $f \in [L^1(\Omega; \mu)]^m$ such that $\nu^a = f\mu$.*

Theorem 2.6 (Besicovitch derivation theorem)**.** *In Theorem 2.5, suppose ν and μ are Radon measures on $\mathcal{B}(\Omega)$. Then $\nu^a = f\mu$ for*

$$f = \lim_{\rho \searrow 0} \frac{\nu(B(x,\rho))}{\mu(B(x,\rho))},$$

where the limit exists μ-a.e. If $\nu \ll \mu$, we thus denote

$$\frac{d\nu}{d\mu} := f.$$

Definition 2.19. We introduce the average integral notation

$$\fint_A f(x)\, d\mu(x) := \frac{1}{\mu(A)} \int_A f(x)\, d\mu(x).$$

Corollary 2.1 (Lebesgue points)**.** *Let $f \in L^1(\Omega)$ with $\Omega \subset \mathbb{R}^n$ open. Then for $\mathcal{L}^n$-a.e. $x \in \Omega$ we have*

$$\lim_{\rho \searrow 0} \fint_{B(x,\rho)} |f(y) - f(x)|\, dy = 0.$$

Such a point x is called a Lebesgue *point of f.*

Proof. Apply Theorem 2.6 to $\nu = |f - q|\mathcal{L}^n$ and $\mu = \mathcal{L}^n$ with $q \in \mathbb{Q}$ to get

$$|f(x) - q| = \lim_{\rho \searrow 0} \fint_{B(x,\rho)} |f(y) - q|\, dy$$

for $x \in A_q \subset \Omega$, where $\mathcal{L}^n(\Omega \setminus A_q) = 0$. Let $A = \bigcap_{q \in \mathbb{Q}} A_q$, and observe that $\mathcal{L}^n(\Omega \setminus A) = 0$ because $\mathbb{Q}$ is countable. For $x \in A$, pick $\{q_i\}_{i=1}^\infty \subset \mathbb{Q}$ with $q_i \to f(x)$. Then

$$\lim_{\rho \searrow 0} \fint_{B(x,\rho)} |f(y) - f(x)|\, dy \le \lim_{\rho \searrow 0} \fint_{B(x,\rho)} |f(y) - q_i|\, dy + |f(x) - q_i| = 2|f(x) - q_i|,$$

where the right hand side tends to zero as $i \to \infty$. □

Definition 2.20. Let $\Omega \subset \mathbb{R}^n$ be open, and μ a positive Radon measure on Ω. For $k \ge 0$, we define the upper and lower k–dimensional densities of μ at $x \in \Omega$ by

$$\Theta^*_k(\mu, x) := \limsup_{\rho \searrow 0} \frac{\mu(B(x,\rho))}{\omega(k)\rho^k}, \quad \text{and} \quad \Theta_{*k}(\mu, x) := \liminf_{\rho \searrow 0} \frac{\mu(B(x,\rho))}{\omega(k)\rho^k}.$$

If these densities agree, we denote the common value by $\Theta_k(\mu, x)$.

In order to define densities of sets, we introduce the following notation.

Definition 2.21. Let μ be measure on a measure space (Ω, Σ), and $A \in \Sigma$. Then we define the restriction of μ to A, denoted $\mu \llcorner A$, by

$$(\mu \llcorner A)(E) := \mu(A \cap E), \quad (E \in \Sigma).$$

Definition 2.22. Let $A \in \mathcal{B}(\Omega)$ We then define the upper and lower k-dimensional densities of A at $x \in \Omega$, by

$$\Theta_k^*(A,x) := \Theta_k^*(\mathcal{H}^k \llcorner A, x), \quad \text{and} \quad \Theta_{*k}(A,x) := \Theta_{*k}(\mathcal{H}^k \llcorner A, x).$$

Again, the common value, if it exists, is denoted $\Theta_k(A,x)$.

Remark 2.2. Observe in Theorem 2.6 that if $\mu = \mathcal{L}^n$, then $f(x) = \Theta_n(\mu,x)$. Thus $\Theta_n(\mu,x)$ gives the density of μ with respect to the Lebesgue measure. Next we study densities with respect to the Hausdorff measure.

2.6. Rectifiable sets

Definition 2.23. We call a set $A \in \mathcal{B}(\mathbb{R}^n)$ *countably $\mathcal{H}^k$-rectifiable* if there exist Lipschitz functions $f_j : \mathbb{R}^k \to \mathbb{R}^n$, $(j = 1, 2, \ldots)$ such that

$$\mathcal{H}^k\left(A \setminus \bigcup_{j=1}^{\infty} f_j(\mathbb{R}^k)\right) = 0.$$

If also $\mathcal{H}^k(A) < \infty$, we say that A is $\mathcal{H}^k$-rectifiable.

Theorem 2.7. *Let μ be a positive Radon measure on an open set $\Omega \subset \mathbb{R}^n$. If $\mu = \theta \mathcal{H}^k \llcorner S$ and S is $\mathcal{H}^k$-rectifiable, then $\theta(x) = \Theta_k(\mu,x)$ for $\mathcal{H}^k$-a.e. $x \in S$. Consequently*

$$\Theta_k(\mu,x) = \frac{d\mu}{d\mathcal{H}^k \llcorner S}(x).$$

Theorem 2.8 (Area formula)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz with $m \geq n$, and $E \subset \mathbb{R}^n$ be $\mathcal{L}^n$-measurable. Then the function $y \mapsto \mathcal{H}^0(E \cap f^{-1}(y)) = \sum_{x \in f^{-1}(y)} \chi_E(x)$ is measurable and*

$$\int_{\mathbb{R}^m} \mathcal{H}^0(E \cap f^{-1}(y))\, d\mathcal{H}^n(y) = \int_E \mathcal{J}_n[\nabla f(x)]\, dx$$

where the n-dimensional Jacobian of a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ is defined as

$$\mathcal{J}_n L := \sqrt{\det(L^* L)}.$$

Remark 2.3. By the area formula, if f is one-to-one on E, then

$$\mathcal{H}^n(f(E)) = \int_E \mathcal{J}_n[\nabla f(x)]\, dx.$$

Example 2.6. Consider the set

$$A := \{(x, \sin(1/x)) \mid 0 < x \leq 1\} \subset \mathbb{R}^2$$

from Section 2.4. Let us set $I_0 = [1, \pi)$ and $I_i = i\pi + [0, \pi)$ for $i = 1, 2, 3, \ldots$. Write $[a_i, b_i) = I_i$ and

$$f_i(x) = \begin{cases} (1/x, \sin(x)), & x \in I_i, \\ (1/a_i, \sin(a_i)), & x < a_i, \\ (1/b_i, \sin(b_i)), & x \geq b_i. \end{cases}$$

Then each f_i Lipschitz, and $A \subset \bigcup f_i(I_i)$. Thus A is countably $\mathcal{H}^1$-rectifiable. Let us show that it is not $\mathcal{H}^1$-rectifiable. Indeed, we have

$$\nabla f_i(x) = (-1/x^2, \cos(x)) \quad \text{on } I_i,$$

so that given $c \in (0,1)$, we can find $\epsilon > 0$ such that

$$\mathcal{J}_1[\nabla f_i(x)] \geq c \quad \text{on } (a_i, a_i + \epsilon) \cup (b_i - \epsilon, b_i), \quad (i \geq 1).$$

Thus by the area formula, Theorem 2.8, we have for $i \geq 1$ that

$$\mathcal{H}^1(A \cap f_i(I_i)) = \int_{\mathbb{R}^2} \mathcal{H}^0(I_i \cap f_i^{-1}(y))\, d\mathcal{H}^1(y) = \int_{I_i} \mathcal{J}_1[\nabla f_i(x)]\, dx \geq 2\epsilon c.$$

Thus

$$\mathcal{H}^1(A) \geq \sum_{i=1}^{\infty} \mathcal{H}^1(A \cap f_i(I_i)) = \infty.$$

2.7. Interlude: Convolution

Mollification by convolution is often used to approximate non-smooth functions and even measures by smooth functions. Generally, convolution is defined as follows.

Definition 2.24 (Convolution)**.** Let $f \in L^1(\mathbb{R}^n)$ and $g \in L^1(\mathbb{R}^n; \mathbb{R}^m)$. We then define the *convolution* $f * g$ by

$$(f * g)(x) := \int_{\mathbb{R}^n} f(x-y) g(y)\, dy.$$

Analogously, if $\mu \in \mathcal{M}(\mathbb{R}^n; \mathbb{R}^m)$, we define the convolution $f * \mu$ by

$$(f * \mu)(x) := \int_{\mathbb{R}^n} f(x-y)\, d\mu(y).$$

Definition 2.25 (Family of mollifiers)**.** We require a function $\rho \in C_c^\infty(\Omega)$ satisfying $\rho \geq 0$, $\int \rho\, dx = 1$ and $\operatorname{supp} \rho \subset B(0,1)$. We then set $\rho_\epsilon(x) := \epsilon^{-n} \rho(x/\epsilon)$ and call $\{\rho_\epsilon\}_{\epsilon > 0}$ a *family of mollifiers.*

Definition 2.26 (Standard mollifier)**.** The *standard mollifier* that can be used in Definition 2.25 is

$$\rho(x) := \begin{cases} e^{-1/(1-\|x\|^2)}, & \|x\| < 1, \\ 0, & \|x\| \geq 1. \end{cases}$$

Theorem 2.9. *Let $f \in L^p(\mathbb{R}^n; \mathbb{R}^m)$, ($p \in [1, \infty)$), and $\{\rho_\epsilon\}_{\epsilon>0}$ be a family of mollifiers. Define $f_\epsilon := f * \rho_\epsilon$. Then*

1. *$f_\epsilon(x) \to f(x)$ for every Lebesgue point x of f, hence hence almost everywhere in $\mathbb{R}^n$.*
2. *$f_\epsilon|K \to f|K$ in $L^p(K; \mathbb{R}^m)$ for every compact set K.*

2.8. Weak* convergence

If $\{\mu^i\}_{i=1}^\infty$ is a sequence of measures on a measure space (Ω, Σ), they converge to μ *strongly* if $|\mu - \mu^i|(\Omega) \to 0$. This kind of convergence is often difficult to achieve. For example, let us define on $\Omega = \mathbb{R}^n$ the Borel measure

$$\mu(A) := \delta_0(A) = \chi_A(0).$$

concentrated at the point $0 \in \mathbb{R}^n$. Let then $\{\rho_\epsilon\}_{\epsilon>0}$ be a family of mollifiers. We define $\mu^i := (\rho_{1/i} * \mu)\mathcal{L}^n$, that is

$$\mu^i(A) := \int_A \int_{\mathbb{R}^n} \rho_{1/i}(x-y)\, d\mu(y)\, dx = \int_A i^n \rho(ix)\, dx.$$

Thus $\mu^i \ll \mathcal{L}^n$, but $\mu \perp \mathcal{L}^n$. By mutual singularity

$$|\mu^i - \mu|(\mathbb{R}^n) = |\mu^i|(\mathbb{R}^n) + |\mu|(\mathbb{R}^n) = 2,$$

so μ^i does not converge to μ strongly. However

$$\int_{\mathbb{R}^n} \varphi(x)\, d\mu^i(x) = \int_{\mathbb{R}^n} \varphi(x) i^n \rho(ix)\, dx \to \varphi(0) = \int_{\mathbb{R}^n} \varphi(x)\, d\mu(x)$$

for every $\varphi \in C_0(\mathbb{R}^n)$. This motivates the following mode of convergence.

Definition 2.27. Let $\{\mu^i\}_{i=1}^\infty \subset \mathcal{M}(\Omega;\mathbb{R}^m)$. If there exists $\mu \in \mathcal{M}(\Omega;\mathbb{R}^m)$ such that

$$\int_\Omega \varphi(x)\, d\mu^i(x) \to \int_\Omega \varphi(x)\, d\mu(x) \quad \text{for every } \varphi \in C_0(\Omega;\mathbb{R}^m),$$

we say that *the measures μ^i converge to μ weakly**, denoted $\mu^i \overset{*}{\rightharpoonup} \mu$.

Theorem 2.10 (Weak* compactness)**.** *Let $\{\mu^i\}_{i=1}^\infty \subset \mathcal{M}(\Omega;\mathbb{R}^m)$, and suppose*

$$\sup_i |\mu^i|(\Omega) < \infty.$$

Then there exists $\mu^* \in \mathcal{M}(\Omega;\mathbb{R}^m)$ and a subsequence $\{\mu^{i_j}\}_{j=1}^\infty$ such that $\mu^{i_j} \overset{*}{\rightharpoonup} \mu^*$. Moreover, the map $\mu \mapsto |\mu|(\Omega)$ is lower semicontinuous with respect to weak convergence.*

Proof. For convenience, we use the notation

$$\mu(\varphi) := \int_\Omega \langle \varphi(x), d\mu(x)\rangle.$$

The proof is a simple diagonal argument combined with the Riesz representation theorem, which we state below in Theorem 2.11. Indeed, we may pick a countable set $G = \{\varphi_j\}_{j=1}^\infty$ with $\|\varphi_j\|_\infty \le 1$ such that G has dense linear span in $C_0(\Omega;\mathbb{R}^m)$. By a diagonal argument, we can find a subsequence $\{\mu^{i_h}\}_{h=1}^\infty$ such that there exist limits

$$\alpha_j := \lim_{h\to\infty} \int_\Omega \langle \varphi_j(x), d\mu^{i_h}(x)\rangle, \quad (j = 1, 2, \ldots).$$

Then $|\alpha_j| \le M$ for $M := \sup_i |\mu^i|(\Omega)$.

If $\varphi = \sum_{\ell=1}^k \beta_\ell \varphi_\ell \in \operatorname{span} G$, let us define

$$L(\varphi) := \sum_{\ell=1}^k \beta_\ell \alpha_\ell.$$

Then L is linear on $\operatorname{span} G$, and given $\epsilon > 0$, for large enough $h \ge h_0$, we have

$$\left\| L(\varphi) - \int_\Omega \varphi(x)\, d\mu^{i_h}(x) \right\| \le \epsilon. \tag{2.5}$$

It follows that

$$\|L(\varphi)\| \le M\|\varphi\| + \epsilon,$$

so that by the arbitrariness of ϵ, we have $\|L\| \le M$.

Picking arbitrary $\varphi \in C_0(\Omega;\mathbb{R}^m)$, we may for any $\epsilon > 0$ find k, β_ℓ, ($\ell = 1, \ldots, k$) such that

$$\widetilde{\varphi} = \sum_{\ell=1}^k \beta_\ell \varphi_\ell$$

satisfies

$$\|\varphi - \widetilde{\varphi}\|_\infty \le \epsilon/(2M). \tag{2.6}$$

Then

$$L(\widetilde{\varphi}) = \sum_{\ell=1}^k \beta_\ell \alpha_\ell,$$

and

$$\|\mu^{i_h}(\varphi) - L(\widetilde{\varphi})\| \le \|\mu^{i_h}(\varphi) - \mu^{i_h}(\widetilde{\varphi})\| + \|\mu^{i_h}(\widetilde{\varphi}) - L(\widetilde{\varphi})\| \le \epsilon/2 + \|\mu^{i_h}(\widetilde{\varphi}) - L(\widetilde{\varphi})\|.$$

Using (2.5), we thus observe the existence of $h_0 \in \mathbb{N}^+$ such that

$$\|\mu^{i_h}(\varphi) - L(\widetilde{\varphi})\| \le \epsilon, \quad (h \ge h_0).$$

It can easily be seen that approximating φ this way, L extends to a bounded linear functional $\overline{L}$ on $C_0(\Omega; \mathbb{R}^m)$. Thus Theorem 2.11 below shows that there exists $\mu^* \in \mathcal{M}(\Omega; \mathbb{R}^m)$ such that

$$\overline{L}(\varphi) = \int_\Omega \langle \varphi(x), d\mu^*(x) \rangle, \quad (\varphi \in C_0(\Omega; \mathbb{R}^m)),$$

and

$$|\mu^*|(\Omega) = \|\overline{L}\| = M = \sup_i |\mu^i|(\Omega).$$

We have to show that $\mu^{i_h} \overset{*}{\rightharpoonup} \mu^*$. Clearly $\mu^{i_h}(\varphi) \to \mu(\varphi)$ for $\varphi \in \operatorname{span} G$. By approximating general $\varphi \in C_0(\Omega; \mathbb{R}^m)$ as in (2.6), we get

$$|\nu(\varphi - \widetilde{\varphi})| \le |\nu|(\epsilon \chi_\Omega) \le \epsilon M, \quad (\nu = \mu^*, \mu^1, \mu^2, \ldots). \tag{2.7}$$

Letting $\epsilon \searrow 0$, we deduce from $\mu^{i_h}(\widetilde{\varphi}) \to \mu^*(\widetilde{\varphi})$ that $\mu^{i_h}(\varphi) \to \mu^*(\varphi)$.

Finally, the claimed lower semicontinuity of $\mu \mapsto |\mu|(\Omega)$ follows easily from the Definition (2.27) of weak* converge □

2.9. The Riesz representation theorem

We required the following Riesz representation theorem in the proof of weak compactness. It will be important for us in the next section as well, as we introduce functions of bounded variation. The content is: *bounded linear functionals are measures.*

Theorem 2.11 (Riesz representation theorem). *Let $\Omega \subset \mathbb{R}^n$, and suppose $L : C_0(\Omega; \mathbb{R}^n) \to \mathbb{R}$ is linear and bounded, i.e.,*

$$\|L\| := \sup\{L(\varphi) \mid u \in C_0(\Omega; \mathbb{R}^n), \sup_{x \in \Omega} \|\varphi(x)\| \le 1\} < \infty.$$

Then there exists a unique $\mu \in \mathcal{M}(\Omega; \mathbb{R}^n)$ such that

$$L(\varphi) = \int_\Omega \langle \varphi(x), d\mu(x) \rangle.$$

Moreover

$$|\mu|(\Omega) = \|L\|.$$

Chapter 3

Functions of bounded variation

We are finally ready to start a proper treatment of the main subject of the course.

3.1. Definition and basic properties

Definition 3.1. Let $u \in L^1(\Omega)$ for an open set $\Omega \subset \mathbb{R}^n$. We say that u is of *bounded variation*, denoted $u \in \mathrm{BV}(\Omega)$, if

$$\mathrm{TV}(u) := \sup\Big\{\int_\Omega \operatorname{div}\varphi(x)u(x)\,dx \;\Big|\; \varphi \in C_c^\infty(\Omega;\mathbb{R}^n),\ \sup_{x\in\Omega}\|\varphi(x)\| \le 1\Big\} < \infty.$$

Remark 3.1. Observe that we have not defined the finite-dimensional norm used in the constraint

$$\|\varphi(x)\| \le 1.$$

For the basic theory, this makes no difference, since all finite-dimensional norms are *topologically equivalent* in the sense that any two norms $\|\cdot\|$ and $\|\cdot\|'$ on $\mathbb{R}^n$ satisfy for some $c, C > 0$ the inequalities

$$c\|x\| \le \|x\|' \le C\|x\|.$$

Geometrically the two norms however can be very different, as the unit balls $\|x\| \le 1$ and $\|x\|' \le 1$ can differ. Consequently, the choice of the finite-dimensional norm will play a role in image processing applications. Typically we choose the 2-norm to get *isotropic*, direction-invariant, behaviour, but sometimes the ∞-norm makes sense to get *anisotropic* behaviour that enhances vertical and horizontal lines, for example.

Theorem 3.1 (Structure theorem)**.** *Let $u \in \mathrm{BV}(\Omega)$. Then there exists a measure $Du \in \mathcal{M}(\Omega;\mathbb{R}^n)$ such that*

$$\mathrm{TV}(u) = |Du|(\Omega)$$

and the following generalised Green's identity *holds,*

$$\int_\Omega \operatorname{div}\varphi(x)u(x)\,dx + \int_\Omega \langle \varphi(x), dDu(x)\rangle = 0, \quad (\varphi \in C_c^\infty(\Omega;\mathbb{R}^n)).$$

Moreover, if $u \in C^1(\overline{\Omega})$ and $\partial\Omega$ is of class C^1, then $Du = \nabla u \mathcal{L}^n$.

Proof. Let us set

$$L(\varphi) := -\int_\Omega \operatorname{div}\varphi(x)u(x)\,dx.$$

Then L is a linear functional on $C_c^\infty(\Omega;\mathbb{R}^n)$, and

$$|L(\varphi)| \le \mathrm{TV}(u)\|\varphi\|_{L^\infty(\Omega;\mathbb{R}^n)}, \quad (\varphi \in C_c^\infty(\Omega;\mathbb{R}^n)). \tag{3.1}$$

We want to extend L to $C_0(\Omega;\mathbb{R}^n)$ and apply the Riesz representation theorem. As $C_0(\Omega;\mathbb{R}^n)$ is the closure of $C_c(\Omega;\mathbb{R}^n)$ in the infinity norm, and $C_c(\Omega;\mathbb{R}^n)$ can approximated by elements of $C_c^\infty(\Omega)$, given $\varphi \in C_0(\Omega;\mathbb{R}^n)$, we can indeed find $\varphi^i \in C_c^\infty(\Omega)$, $(i = 1, 2, \ldots)$ with $\varphi^i \to \varphi$. Using (3.1), we see that

$$|L(\varphi^i) - L(\varphi^j)| \le \mathrm{TV}(u)\|\varphi^i - \varphi^j\|, \quad (i, j = 1, 2, \ldots).$$

Therefore $\{L(\varphi^i)\}_{i=1}^{\infty}$ forms a Cauchy sequence, so that the following limit exists

$$\bar{L}(\varphi) := \lim_{i\to\infty} L(\varphi^i).$$

An analogous argument shows that limit is independent of the approximating sequence.

We also easily see that $\bar{L}$ is linear and $\|\bar{L}\| = \mathrm{TV}(u)$. An application of the Riesz representation theorem, Theorem 2.11, therefore shows that

$$\bar{L}(\varphi) = \int_\Omega \langle \varphi(x), dDu(x)\rangle$$

for some measure $Du \in \mathcal{M}(\Omega;\mathbb{R}^n)$. Moreover

$$|Du|(\Omega) = \|\bar{L}\| = \mathrm{TV}(u).$$

Finally, if $u \in C^1(\overline{\Omega})$, using the fact that $\varphi|\partial\Omega = 0$, Green's identity shows that

$$L(\varphi) = -\int_\Omega \operatorname{div}\varphi(x)u(x)\,dx = \int_\Omega \langle \varphi(x), \nabla u(x)\rangle\,dx.$$

Since φ was arbitrary, it follows that $\nabla u\mathcal{L}^n = Du$. □

Theorem 3.2. *The space* $\mathrm{BV}(\Omega)$ *is a Banach space when equipped with the norm*

$$\|u\|_{\mathrm{BV}(\Omega)} := \|u\|_{L^1(\Omega)} + |Du|(\Omega).$$

Moreover, $u \mapsto |Du|(\Omega)$ *is lower semicontinuous with respect to convergence in* $L^1(\Omega)$.

Proof. We have to show that $\mathrm{BV}(\Omega)$ is complete with respect to the the *BV-norm* $\|\cdot\|_{\mathrm{BV}(\Omega)}$. Let $\{u^i\}_{i=1}^{\infty}$ be a Cauchy sequence in $\mathrm{BV}(\Omega)$ with respect to the BV-norm. Then $\{u^i\}_{i=1}^{\infty}$ is also a Cauchy sequence in $L^1(\Omega)$, and hence converges to some $u \in L^1(\Omega)$ strongly in $L^1(\Omega)$.

Given $\varphi \in C_c^\infty(\Omega)$, by Hölder's inequality we have

$$\left|\int_\Omega \operatorname{div}\varphi(x)(u - u^i(x))\,dx\right| \le \|\operatorname{div}\varphi\|_{L^\infty(\Omega;\mathbb{R}^n)}\|u-u^i\|_{L^1(\Omega)}.$$

It therefore follows that

$$\int_\Omega \operatorname{div}\varphi(x)u(x)\,dx \le \liminf_{i\to\infty}\int_\Omega \operatorname{div}\varphi(x)u^i(x)\,dx \le \liminf_{i\to\infty}|Du^i|(\Omega).$$

Thus $u \in \mathrm{BV}(\Omega)$, and by Theorem 3.1, Du exists. This also shows the claimed lower semicontinuity.

We want to show that $|Du^i - Du| \to 0$, as that would establish the convergence $u^i \to u$ in the BV-norm, and show completeness. Minding that $\{Du^i\}_{i=1}^{\infty}$ is a Cauchy sequence in $\mathcal{M}(\Omega;\mathbb{R}^m)$, let us pick $\epsilon > 0$, and choose i_0 large enough that $|Du^i - Du^j|(\Omega) \le \epsilon$ for $i,j \ge i_0$. Then for $\varphi \in C_c^\infty(\Omega;\mathbb{R}^n)$ with $\|\varphi\|_{L^\infty(\Omega;\mathbb{R}^n)} \le 1$, we have

$$\begin{aligned}\int_\Omega \langle\varphi(x), d(Du - Du^i)(x)\rangle &= \int_\Omega \langle\varphi(x), d(Du - Du^j)(x)\rangle + \int_\Omega \langle\varphi(x), d(Du^j - Du^i)(x)\rangle \\ &= -\int_\Omega \operatorname{div}\varphi(x)(u-u^j)(x)\,dx + \int_\Omega \langle\varphi(x), d(Du^j - Du^i)(x)\rangle \\ &\le \|\operatorname{div}\varphi\|_{L^\infty(\Omega)}\|u-u^j\|_{L^1(\Omega)} + \epsilon.\end{aligned}$$

Choosing j large enough, we conclude that

$$\int_\Omega \langle\varphi(x), d(Du - Du^i)(x)\rangle \le 2\epsilon.$$

Thus

$$|Du - Du^i|(\Omega) = \sup_\varphi \int_\Omega \langle\varphi(x), d(Du - Du^i)(x)\rangle \le 2\epsilon$$

with the supremum over $\varphi \in C_c^\infty(\Omega;\mathbb{R}^n)$ with $\|\varphi\|_{L^\infty(\Omega;\mathbb{R}^n)} \le 1$. Since ϵ was arbitrary, this concludes the proof. □

3.2. Smooth approximation

Theorem 3.3 (Smooth approximation). *Suppose $\Omega \subset \mathbb{R}^n$ is open and let $u \in \mathrm{BV}(\Omega)$. Then there exists a sequence $\{u^i\}_{i=1}^\infty \in C^\infty(\Omega)$ with $u^i \to u$ in $L^1(\Omega)$ and $|Du^i|(\Omega) \to |Du|(\Omega)$.*

Proof. Given a positive integer m, we set $\Omega_0 = \emptyset$ as well as

$$\Omega_k := B(0, k+m) \cap \{x \in \Omega \mid \inf_{y \in \partial\Omega} \|x - y\| \geq 1/(m+k))\}.$$

We pick m large enough that

$$|Du|(\Omega \setminus \Omega_1) < 1/i. \tag{3.2}$$

With

$$V_k := \Omega_{k+1} \setminus \overline{\Omega}_{k-1},$$

each $x \in \Omega$ belongs to at most four sets V_k. We may then find a partition of unity $\{\zeta_k\}_{k=1}^\infty$ with $\zeta_k \in C_c^\infty(V_k)$, $0 \leq \zeta_k \leq 1$ and $\sum_{k=1}^\infty \zeta_k \equiv 1$ on Ω.

With $\{\rho_\epsilon\}_{\epsilon>0}$ a family of mollifiers, and $\epsilon_k > 0$, we let

$$u_k := \rho_{\epsilon_k} * (u\zeta_k).$$

We select $\epsilon_k > 0$ small enough that $\operatorname{supp} u_k \subset V_k$ (doable because $\zeta_k \in C_c^\infty(V_k)$), and

$$\|u_k - u\zeta_k\| \leq 1/(2^k i), \quad \text{and} \quad \|\rho_{\epsilon_k} * (u\nabla\zeta_k) - u\nabla\zeta_k\| \leq 1/(2^k i). \tag{3.3}$$

We then let

$$u^i := \sum_{k=1}^\infty u_k.$$

By the construction of the partition of unity, for every $x \in \Omega$ there is a neighbourhood of x such that there are only finitely many non-zero terms in this sum. Hence $u^i \in C^\infty(\Omega)$. Moreover, as $u = \sum_{k=1}^\infty \zeta_k u$, (3.3) gives

$$\|u - u^i\| \leq \sum_{k=1}^\infty \|u_k - u\zeta_k\| < 1/i.$$

Thus $u^i \to u$ in $L^1(\Omega)$ as $i \to \infty$.

By Theorem 3.2, we have

$$|Du|(\Omega) \leq \liminf_{i\to\infty} |Du^i|(\Omega).$$

It therefore only remains to prove the opposite inequality. Let $\varphi \in C_c^1(\Omega; \mathbb{R}^n)$ with $\sup_{x\in\Omega} |\varphi(x)| \leq 1$. We have

$$\begin{aligned}
\int_\Omega \operatorname{div} \varphi(x) u_k(x) &= \int_\Omega \operatorname{div} \varphi(x)(\rho_{\epsilon_k} * \zeta_k u)(x)\, dx \\
&= \int_\Omega \operatorname{div}(\rho_{\epsilon_k} * \varphi)(x)\zeta_k(x)u(x)\, dx \\
&= \int_\Omega \operatorname{div}[\zeta_k(\rho_{\epsilon_k} * \varphi)](x)u(x)\, dx - \int_\Omega \langle \nabla\zeta_k(x), (\rho_{\epsilon_k} * \varphi)(x)\rangle u(x)\, dx \\
&= \int_\Omega \operatorname{div}[\zeta_k(\rho_{\epsilon_k} * \varphi)](x)u(x)\, dx \\
&\quad - \int_\Omega \langle \varphi(x), (\rho_{\epsilon_k} * (u\nabla\zeta_k))(x) - (u\nabla\zeta_k)(x)\rangle\, dx - \int_\Omega \langle \varphi(x), (u\nabla\zeta_k)(x)\rangle\, dx.
\end{aligned}$$

Since $\sum_{k=1}^\infty \nabla\zeta_k = 0$, we have

$$\sum_{k=1}^\infty \int_\Omega \langle \varphi(x), (u\nabla\zeta_k)(x)\rangle\, dx = 0.$$

Thus using (3.3), we get

$$\begin{aligned}\int_\Omega \operatorname{div}\varphi(x)u^i(x) &= \sum_{k=1}^\infty \int_\Omega \operatorname{div}\varphi(x)u_k(x)\\ &= \sum_{k=1}^\infty \int_\Omega \operatorname{div}[\zeta_k(\rho_{\epsilon_k}*\varphi)](x)u(x)\,dx\\ &\quad - \sum_{k=1}^\infty \left(\int_\Omega \langle \varphi(x), (\rho_{\epsilon_k}*(u\nabla\zeta_k))(x) - (u\nabla\zeta_k)(x)\rangle\,dx\right)\\ &\leq \sum_{k=1}^\infty \int_\Omega \operatorname{div}[\zeta_k(\rho_{\epsilon_k}*\varphi)](x)u(x)\,dx + 1/i\end{aligned}$$

Observing that $\zeta_k(\rho_{\epsilon_k}*\varphi) \leq 1$, and using the fact that $\sum_{k=1}^\infty \chi_{V_k} \leq 4$, we further get

$$\begin{aligned}\int_\Omega \operatorname{div}\varphi(x)u^i(x) &\leq \int_\Omega \operatorname{div}[\zeta_1(\rho_{\epsilon_1}*\varphi)](x)u(x)\,dx + \sum_{k=2}^\infty \int_\Omega \operatorname{div}[\zeta_k(\rho_{\epsilon_k}*\varphi)](x)u(x)\,dx + 1/i\\ &\leq |Du|(\Omega) + \sum_{k=2}^\infty |Du|(V_k) + 1/i\\ &\leq |Du|(\Omega) + 4|Du|(\Omega\setminus\Omega_1) + 1/i\\ &\leq |Du|(\Omega) + 5/i.\end{aligned}$$

In the final step we have used (3.2). This concludes the proof. □

3.3. Traces and extensions

Theorem 3.4. *Suppose $\Omega \subset \mathbb{R}^n$ is open and bounded with a Lipschitz boundary. Then there exists a bounded linear mapping*

$$T : \mathrm{BV}(\Omega) \to L^1(\partial\Omega; \mathcal{H}^{n-1}),$$

such that

$$\int_\Omega \langle\varphi(x), dDu(x)\rangle = -\int_\Omega \operatorname{div}\varphi(x)u(x)\,dx + \int_{\partial\Omega} \langle\varphi(x), \nu(x)\rangle Tu(x)\,d\mathcal{H}^{n-1}, \quad (u\in \mathrm{BV}(\Omega),\ \varphi\in C_c^\infty(\mathbb{R}^n;\mathbb{R}^n)), \tag{3.4}$$

with ν the unit outer normal field to $\partial\Omega$. We call Tu the trace *of u on $\partial\Omega$.*

Proof. Since $\partial\Omega$ is Lipschitz, at every $x \in \partial\Omega$, we can find a neighbourhood Q of x such that $\partial\Omega\cap Q$ is the graph of a Lipschitz map ψ. By rotation and translation if necessary, we may assume that $x = 0$ and

$$Q = U \times (-\rho, \rho)$$

for some ball $U \subset \mathbb{R}^{n-1}$, as well as

$$\partial\Omega \cap Q = \{(v, \psi(v)) \mid v \in U\}.$$

and

$$\Omega\cap Q = Q \cap \{(v,t) \mid v\in U,\ \rho > t > \psi(v)\}.$$

Suppose first that $u \in C^\infty(\Omega)$, and define the slice

$$u_\epsilon(y) := u(v, \psi(v)+\epsilon), \quad (y = (v,t) \in \partial\Omega\cap Q)$$

whenever $\epsilon \in (0, \rho/2)$. Then for $0 < \delta < \epsilon < \rho/2$ we have

$$|u_\delta(y) - u_\epsilon(y)| \leq \int_\delta^\epsilon \left|\frac{\partial u}{\partial x_n}(v, \psi(v)+t)\right| dt \leq \int_\delta^\epsilon |\nabla u(v,\psi(v)+t)|\,dt.$$

Since ψ is Lipschitz, the area formula shows for some constant $C > 0$ that

$$\int_{\partial\Omega\cap Q} |u_\delta(y) - u_\epsilon(y)|\, d\mathcal{H}^{n-1}(y) \le C \int_{A_{\delta,\epsilon}} |\nabla u(y)|\, dy = C|Du|(A_{\delta,\epsilon}),$$

where

$$A_{\delta,\epsilon} := \{(v,t) \mid v \in U, t \in \psi(v) + (\delta,\epsilon)\}.$$

It follows that $\{u_\epsilon\}_{\epsilon>0}$ is Cauchy in $L^1(\partial\Omega\cap Q;\mathcal{H}^{n-1})$, and hence convergent to some $Tu \in L^1(\partial\Omega\cap Q;\mathcal{H}^{n-1})$. Moreover, the above arguments yield

$$\int_{\partial\Omega\cap Q} |Tu(y) - u_\epsilon(y)|\, d\mathcal{H}^{n-1}(y) \le C|Du|(A_{0,\epsilon}). \tag{3.5}$$

Let then $\varphi \in C^1_c(Q;\mathbb{R}^n)$. By the Gauss-Green theorem, we have

$$\int_{\Omega\cap Q\setminus A_{0,\epsilon}} \operatorname{div}\varphi(y)u(y)\, dy = -\int_{\Omega\cap Q\setminus A_{0,\epsilon}} \langle\varphi(y),\nabla u(y)\rangle\, dy + \int_{\partial\Omega\cap Q} u_\epsilon(y)\langle\varphi(y),\nu_{\partial\Omega}(y)\rangle\, d\mathcal{H}^{n-1}(y).$$

Here we have translated the boundary $\{(v,\psi(v)+\epsilon) \mid v \in U\}$ to $\partial\Omega$. Letting $\epsilon \searrow 0$, we thus get

$$\int_{\Omega\cap Q} \operatorname{div}\varphi(y)u(y)\, dy = -\int_{\Omega\cap Q} \langle\varphi(y),\, dDu(y)\rangle + \int_{\partial\Omega\cap Q} Tu(y)\langle\varphi(y),\nu_{\partial\Omega}(y)\rangle\, d\mathcal{H}^{n-1}(y) \tag{3.6}$$

Let us then finally consider the general case $u \in \mathrm{BV}(\Omega)$. We construct using Theorem 3.3 a sequence $\{u^i\}_{i=1}^\infty \in C^\infty(\Omega)$ with $u^i \to u$ in $L^1(\Omega)$ and $|Du^i|(\Omega) \to |Du|(\Omega)$. By Theorem 2.10, we may assume that $Du^i \overset{*}{\rightharpoonup} Du$ weakly* in $\mathcal{M}(\Omega;\mathbb{R}^n)$. We claim that Tu^i is Cauchy in $L^1(\partial\Omega\cap Q;\mathcal{H}^{n-1})$. Indeed, let

$$u^{i,\epsilon}(y) := \frac{1}{\epsilon}\int_0^\epsilon u^i(v,\psi(v)+t)\, dt = \frac{1}{\epsilon}\int_0^\epsilon u^i_\epsilon(y)\, dt. \tag{3.7}$$

Then using (3.5) we get

$$\int_{\partial\Omega\cap Q} |Tu^i(y) - u^{i,\epsilon}(y)|\, d\mathcal{H}^{n-1}(y) \le \frac{1}{\epsilon}\int_0^\epsilon \int_{\partial\Omega\cap Q} |Tu^i(y) - u^i_\epsilon(y)|\, d\mathcal{H}^{n-1}(y)\, dt \le C|Du^i|(A_{0,\epsilon}).$$

It follows

$$\begin{aligned}\int_{\partial\Omega\cap Q} |Tu^i(y) - Tu^j(y)|\, d\mathcal{H}^{n-1}(y) &\le \int_{\partial\Omega\cap Q} |Tu^i(y) - u^{i,\epsilon}(y)|\, d\mathcal{H}^{n-1}(y)\\ &\quad+ \int_{\partial\Omega\cap Q} |Tu^j(y) - u^{j,\epsilon}(y)|\, d\mathcal{H}^{n-1}(y)\\ &\quad+ \int_{\partial\Omega\cap Q} |u^{i,\epsilon}(y) - u^{j,\epsilon}(y)|\, d\mathcal{H}^{n-1}(y)\\ &\le C|Du^i|(A_{0,\epsilon}) + C|Du^j|(A_{0,\epsilon}) + \frac{C}{\epsilon}\int_{A_{0,\epsilon}} |u^i(y) - u^j(y)|\, dy.\end{aligned} \tag{3.8}$$

In the final step we have used the definition (3.7). Using $u^i \to u$ in $L^1(\Omega)$ for the final term, thus

$$\limsup_{i,j\to\infty} \int_{\partial\Omega\cap Q} |Tu^i(y) - Tu^j(y)|\, d\mathcal{H}^{n-1}(y) \le C'|Du|(A_{0,\epsilon}\cap Q).$$

Since $\epsilon > 0$ was arbitrary, and $|Du|(A_{0,\epsilon}\cap Q) \to 0$ as $\epsilon \searrow 0$, the claim is proved. Thus we may again define

$$Tu := \lim_{i\to\infty} Tu^i$$

in the sense of strong convergence in $L^1(\partial\Omega\cap Q;\mathcal{H}^{n-1})$. Reasoning as in (3.8), it is not difficult to see that this limit is independent of the selection of u^i in the sense of L^1 equivalence classes. Further, the reasoning in (3.8) shows that T is bounded on $C^\infty(\Omega\cap Q)$, and consequently on $\mathrm{BV}(\Omega\cap Q)$. Taking the limit in (3.6) for u^i, we moreover see that it holds for u as well.

Finally, since $\partial\Omega$ is compact, we can cover it by finitely many sets $\{U_i\}_{i=1}^N$ as above. Then we may form a partition of unity $\{\zeta_i\}_{0=1}^N$ with $0\le\zeta_i\le 1$, $\operatorname{supp}\zeta_i\subset U_i$ for $i=1,\dots,N$, and $\operatorname{supp}\zeta_0\subset\Omega$. Moreover $\sum_{i=1}^N\zeta_i(x)=1$ for $x\in\partial\Omega$ and $\sum_{i=0}^N\zeta_i(x)=1$ for $x\in\Omega$. Then we define $T: C^\infty(\Omega)\to L^1(\partial\Omega;\mathcal{H}^{n-1})$ by

$$Tu(x) := \sum_{i=1}^N \zeta_i(x)[T_i u](x),$$

where T_i is the extension operator on U_i as constructed above. Observe from the construction above that, moreover

$$\zeta_i T_i u = T_i(\zeta_i u).$$

Thus for $\varphi\in C_c^1(\mathbb{R}^n;\mathbb{R}^n)$, using (3.6), we get

$$\begin{aligned}
\int_\Omega \operatorname{div}\varphi(y)u(y)\,dy &= \sum_{i=0}^N\int_\Omega \operatorname{div}\varphi(y)\zeta_i(y)u(y)\,dy\\
&= -\sum_{i=0}^N\int_\Omega\langle\varphi(y),\,dD(\zeta_i u)(y)\rangle + \sum_{i=1}^N\int_{\partial\Omega}\zeta_i(y)T_iu(y)\langle\varphi(y),\nu_{\partial\Omega}(y)\rangle\,d\mathcal{H}^{n-1}(y)\\
&= -\int_\Omega\langle\varphi(y),\,dDu(y)\rangle + \int_{\partial\Omega}Tu(y)\langle\varphi(y),\nu_{\partial\Omega}(y)\rangle\,d\mathcal{H}^{n-1}(y).
\end{aligned}$$

This establishes (3.4). The boundedness and linearity of T follows from the boundedness and linearity of T_i □

Theorem 3.5. *Suppose $\Omega\subset\mathbb{R}^n$ is open and bounded with a Lipschitz boundary, and $u\in\mathrm{BV}(\Omega)$. Let*

$$w(x) := \begin{cases} u(x), & x\in\Omega,\\ 0, & \text{otherwise},\end{cases}$$

Then $w\in\mathrm{BV}(\mathbb{R}^n)$ with $\|w\|_{\mathrm{BV}(\mathbb{R}^n)}\le C\|u\|_{\mathrm{BV}(\Omega)}$ for some constant $C=C(\Omega)$.

Proof. Given $\varphi\in C_c^\infty(\mathbb{R}^n)$, we have

$$\int_{\mathbb{R}^n}\operatorname{div}\varphi(x)w(x)\,dx = \int_\Omega\operatorname{div}\varphi(x)u(x)\,dx$$

Using Theorem 3.4, we obtain

$$\int_\Omega\operatorname{div}\varphi(x)u(x)\,dx = -\int_\Omega\langle\varphi(x),\,dDu(x)\rangle + \int_{\partial\Omega}\langle\varphi(x),\nu(x)\rangle Tu(x)\,d\mathcal{H}^{n-1},$$

with the trace $Tu\in L^1(\partial\Omega;\mathcal{H}^{n-1})$, and ν the unit outer normal field to $\partial\Omega$. Thus

$$\int_{\mathbb{R}^n}\operatorname{div}\varphi(x)w(x)\,dx \le \|\varphi\|_{L^\infty(\Omega;\mathbb{R}^n)}\left(|Du|(\Omega)+\|Tu\|_{L^1(\partial\Omega;\mathcal{H}^{n-1})}\right).$$

It follows that $w\in\mathrm{BV}(\mathbb{R}^n)$ and

$$|Dw|(\Omega)\le|Du|(\Omega)+\|Tu\|_{L^1(\partial\Omega;\mathcal{H}^{n-1})}\le C\|u\|_{\mathrm{BV}(\Omega)}.$$

Since $\|w\|_{L^1(\mathbb{R}^n)}=\|u\|_{L^1(\Omega)}$, this concludes the proof. □

Sometimes the above extension introduces difficulties because $|Dw|(\partial\Omega)\ne 0$. We can also make this kind of extension.

Theorem 3.6. *Suppose $\Omega \subset \mathbb{R}^n$ is open and bounded with a Lipschitz boundary, and $u \in \mathrm{BV}(\Omega)$. Then there exists $w \in \mathrm{BV}(\mathbb{R}^n)$ with $w|\Omega = u$, satisfying $|Dw|(\partial\Omega) = 0$ and $\|w\|_{\mathrm{BV}(\mathbb{R}^n)} \leq C\|u\|_{\mathrm{BV}(\Omega)}$ for some constant $C = C(\Omega)$.*

Proof. We skip the details of the proof. It may be found in [2], and is based on locally "mirroring" u over Ω. Namely if Q and ψ are is Theorem 3.4, we define

$$w(y) = \begin{cases} u(v,t), & y = (v,t) \in Q, t > \psi(v), \\ Tu(v,t), & y = (v,t) \in Q, t = \psi(v), \\ u(v, 2\psi(v) - t), & y = (v,t) \in Q, t < \psi(v). \end{cases}$$

Then we glue the extensions together with a partition of unity. The most work is in showing that $|Dw|(Q \setminus \overline{\Omega})$ is bounded by $C|Du|(Q \cap \Omega)$. This depends on results showing that $u \circ g$ is of bounded variation for Lipschitz g. □

3.4. Weak modes of convergence

Definition 3.2. Let $\{u, u^1, u^2, \ldots\} \subset \mathrm{BV}(\Omega)$. We say that $\{u^i\}_{i=1}^\infty$ converge to u

(i) *strongly*, denoted $u^i \to u$, if $\|u^i - u\|_{\mathrm{BV}(\Omega)} \to 0$.

(ii) *strictly*, if $u^i \to u$ strongly in $L^1(\Omega)$, and $|Du^i|(\Omega) \to |Du|(\Omega)$.

(iii) *weakly**, if $u^i \to u$ strongly in $L^1(\Omega)$, and $Du^i \overset{*}{\rightharpoonup} Du$ weakly* in $\mathcal{M}(\Omega; \mathbb{R}^n)$.

Clearly strong convergence implies strict convergence, but the converse is not true. Exercise 2.24 shows that weak* convergence does not imply strict convergence. A consequence of the next proposition is that strict convergence implies weak* convergence.

Proposition 3.1. *Let $\{u^i\}_{i=1}^\infty \subset \mathrm{BV}(\Omega)$. Then $u^i \overset{*}{\rightharpoonup} u$ weakly* in* $\mathrm{BV}(\Omega)$ *if and only if* $\sup_i |Du^i|(\Omega) < \infty$ *and $u^i \to u$ strongly in $L^1(\Omega)$.*

Proof. Suppose $\sup_i |Du^i|(\Omega) < \infty$ and $u^i \to u$ strongly in $L^1(\Omega)$. By Theorem 2.10, we can find a subsequence $\{u^{i_k}\}_{k=1}^\infty$ such that $Du^{i_k} \overset{*}{\rightharpoonup} \mu$ for some $\mu \in \mathcal{M}(\Omega; \mathbb{R}^n)$. If we show that $\mu = Du$, the weak* convergence $u^i \overset{*}{\rightharpoonup} u$ follows. Indeed, by the convergence of u^{i_k} to u in L^1, we have

$$\int_\Omega \langle \varphi(x), dDu^{i_k}(x)\rangle = -\int_\Omega \operatorname{div}\varphi(x) u^{i_k}(x) \to -\int_\Omega \operatorname{div}\varphi(x) u(x), \quad (\varphi \in C_c^\infty(\Omega; \mathbb{R}^n)).$$

But also by $Du^{i_k} \overset{*}{\rightharpoonup} \mu$ we have

$$\int_\Omega \langle \varphi(x), dDu^{i_k}(x)\rangle \to \int_\Omega \langle \varphi(x), d\mu(x)\rangle, \quad (\varphi \in C_c^\infty(\Omega; \mathbb{R}^n)).$$

By the definition of Du, this shows that $Du = \mu$.

For the opposite direction, one may apply the Banach-Steinhaus theorem, also known as the *principle of uniform boundedness*. Applied to our present situation, it says that *if*

$$\sup_{i=1,2,3,\ldots} \|Du^i(\varphi)\| < \infty \textit{ for all } \varphi \in C_c^\infty(\Omega; \mathbb{R}^n),$$

then

$$\sup_{i=1,2,3,\ldots} |Du^i|(\Omega) < \infty.$$

This is exactly what we need. □

Lemma 3.1. *Let $\{\rho_\epsilon\}_{\epsilon>0}$ be a family of mollifiers. We have*

$$\int_{\mathbb{R}^n} |(w * \rho_\epsilon)(x) - w(x)|\, dx \le \epsilon |Dw|(\Omega), \quad (w \in \mathrm{BV}(\mathbb{R}^n) \text{ with compact support}).$$

Proof. Indeed, by Theorem 3.3 we may assume that $w \in C_c^\infty(\mathbb{R}^n)$. Thus with $y \in B(0,\epsilon)$ and $x \in \mathbb{R}^n$, by the fundamental theorem of calculus

$$w(x-y) - w(x) = -\int_0^1 \langle \nabla w(x-ty), y\rangle\, dt.$$

Taking norms, integrating, applying Fubini's theorem and $\|y\| \le \epsilon$, gives

$$\int |w(x-y) - w(x)|\, dx \le \int_0^1 \int |\nabla w(x-ty)| \|y\|\, dx\, dt \le \epsilon |Dw|(\mathbb{R}^n).$$

Multiplying by $\rho_\epsilon(y)$ and integrating over y, we have

$$\int\int |w(x-y) - w(x)| \rho_\epsilon(y)\, dx\, dy \le \epsilon |Dw|(\mathbb{R}^n).$$

But

$$\begin{aligned}\int |(w * \rho_\epsilon)(x) - w(x)|\, dx &= \int \left| \int w(x-y)\rho_\epsilon(y)\, dy - w(x) \right| dx \\ &= \int \left| \int w(x-y) - w(x))\rho_\epsilon(y)\, dy \right| dx \\ &\le \int \int |w(x-y) - w(x)| \rho_\epsilon(y)\, dx\, dy,\end{aligned}$$

which shows the claim. □

Theorem 3.7 (Weak* compactness)**.** *Suppose $\Omega \subset \mathbb{R}^n$ is open and bounded with Lipschitz boundary. The space $\mathrm{BV}(\Omega)$ has a weak* compact unit ball. That is, any bounded sequence $\{u^i\}_{i=1}^\infty \subset \mathrm{BV}(\Omega)$ has a weak* convergent subsequence.*

Proof. We extend each u^i by Theorem 3.5 to $w^i \in \mathrm{BV}(\mathbb{R}^n)$. Then $\|w^i\|_{\mathrm{BV}(\mathbb{R}^n)} \le C\|u^i\|_{\mathrm{BV}(\Omega)}$ for some constant $C = C(\Omega)$. If we can show that $w^{i_k} \overset{*}{\rightharpoonup} w$ weakly* in $\mathrm{BV}(\mathbb{R}^n)$ for some subsequence, then $u^{i_k} \overset{*}{\rightharpoonup} u$ for $u := w|\Omega$ weakly* in $\mathrm{BV}(\Omega)$. Indeed, we immediately see that $u^{i_k} \to u$ strongly in $L^1(\Omega)$. Therefore, as $\sup_{i_k} |Du^{i_k}|(\Omega) < \infty$, Proposition 3.1 proves the claim.

So we have to show the weak* convergence of a subsequence $\{w^{i_k}\}_{k=1}^\infty$. Let us pick a family $\{\rho_\epsilon\}_{\epsilon>0}$ of mollifiers, and define $w_\epsilon^i := \rho_\epsilon * w^i$. Using Lemma 3.1 we have

$$\begin{aligned}\int_{\mathbb{R}^n} |w^i(x) - w^j(x)|\, dx &\le \int_{\mathbb{R}^n} |w_\epsilon^i(x) - w_\epsilon^j(x)|\, dx + \int_{\mathbb{R}^n} |w_\epsilon^i - w^i(x)| + |w^j(x) - w_\epsilon^j(x)|\, dx \\ &\le \int_{\mathbb{R}^n} |w_\epsilon^i(x) - w_\epsilon^j(x)|\, dx + 2\epsilon \sup_k |Dw^k|(\Omega).\end{aligned}$$

If we can find a subsequence $\{w^{i_k}\}_{k=1}^\infty$ such that $\{w_{1/\ell}^{i_k}\}_{k=1}^\infty$ convergences for every $\ell = 1,2,3,\ldots$, this shows that $\{w^{i_k}\}_{k=1}^\infty$ is a Cauchy sequence in $L^1(K)$ for any compact set $K \supset \Omega + B(0,1)$. By completeness w^{i_k} converges to some $w \in L^1(K)$. Since

$$\sup_k |Dw^k|(\Omega) \le C \sup_i \|u^i\|_{\mathrm{BV}(\Omega)},$$

Proposition 3.1 shows the required weak* convergence.

We still have to find $\{w^{i_k}\}_{k=1}^\infty$. We have

$$\|w^i_\epsilon\|_{C(\mathbb{R}^n)} \le \|w^i\|_{L^1(\Omega)}\|\rho_\epsilon\|_{C(\mathbb{R}^n)} \quad \text{and} \quad \|\nabla w^i_\epsilon\|_{C(\mathbb{R}^n;\mathbb{R}^n)} \le \|w^i\|_{L^1(\Omega)}\|\nabla\rho_\epsilon\|_{C(\mathbb{R}^n;\mathbb{R}^n)}.$$

It follows that for fixed $\epsilon > 0$, the sequence $\{w^i_\epsilon\}_{i=1}^\infty$ is uniformly bounded and equicontinuous. By the Arzelà-Ascoli theorem, we can therefore find a subsequence $\{w^{i_k}_\epsilon\}_{k=1}^\infty$ convergent in $C(\mathbb{R}^n)$. By diagonalising, we can thus find $\{w^{i_k}\}_{k=1}^\infty$ such that $\{w^{i_k}_{1/\ell}\}_{k=1}^\infty$ is convergent in $C(\mathbb{R}^n)$ for every $\ell = 1, 2, 3, \ldots$. This concludes the proof. □

3.5. The Poincaré inequality

Theorem 3.8. *Let $\Omega \subset \mathbb{R}^n$ be a* connected *bounded open set with Lipschitz boundary, and define*

$$u_\Omega := \fint_\Omega u(x)\,dx := \frac{1}{\mathcal{L}^n(\Omega)}\int_\Omega u(x)\,dx.$$

Then there exists a constant $C = C(\Omega)$ such that

$$\|u - u_\Omega\|_{L^1(\Omega)} \le C|Du|(\Omega), \quad (u \in \mathrm{BV}(\Omega)). \tag{3.9}$$

Proof. Suppose the inequality (3.9) does not hold for any $C > 0$. Then there exist a sequence $\{u^i\}_{i=1}^\infty \subset \mathrm{BV}(\Omega)$ with

$$\|u^i - u^i_\Omega\|_{L^1(\Omega)} \ge i|Du^i|(\Omega), \quad (i = 1, 2, 3, \ldots).$$

Since $D(u^i - u^i_\Omega) = Du^i$, and this inequality is homogeneous on both sides, we may assume that $\|u^i\|_{L^1(\Omega)} = 1$ and $u^i_\Omega = 0$. Thus we have

$$1/i \ge |Du^i|(\Omega), \quad \int_\Omega |u^i(x)|\,dx = 1, \quad \text{and} \quad \int_\Omega u^i(x)\,dx = 0, \quad (i = 1, 2, 3, \ldots). \tag{3.10}$$

Hence by Theorem 3.7, we may assume that $u^i \overset{*}{\rightharpoonup} u$ for some $u \in \mathrm{BV}(\Omega)$. But $|Du^i|(\Omega) \le 1/i$, so by lower semicontinuity $|Du|(\Omega) = 0$. Therefore $u = c$ is a constant by Exercise 3.4. But by (3.10) we also have

$$|c|\mathcal{L}^n(\Omega) = \int_\Omega |u(x)|\,dx = 1, \quad \text{and} \quad |c|\mathcal{L}^n(\Omega) = \left|\int_\Omega u(x)\,dx\right| = 0.$$

This is a contradiction. Therefore (3.9) must hold for some $C = C(\Omega)$. □

Example 3.1. In recent years, research in variational image processing techniques has attempted many higher-order generalisation of total variation. This is due to the stair-casing effect that the latter exhibits: due to noise and other imperfections ind ata, processed images may exhibit large flat areas where there originally was a smooth gradient. This is avoided by higher-order approaches. One of them is *total generalised variation* [5] or TGV. For two parameters $\beta, \alpha > 0$, in the second order case, we may write it as the differentiation cascade

$$\mathrm{TGV}^2_{(\beta,\alpha)}(u) := \min_{w \in \mathrm{BV}(\Omega;\mathbb{R}^n)} \alpha\|Du - w\|_{\mathcal{M}(\Omega;\mathbb{R}^n)} + \beta\|Ew\|_{\mathcal{M}(\Omega;\mathbb{R}^{n\times n})}. \tag{3.11}$$

Here Ew is the symmetrised gradient, which for $w \in C^1(\Omega;\mathbb{R}^n)$ may be written as $Ew = \mathcal{E}w\mathcal{L}^n$ for

$$\mathcal{E}w(x) := \frac{1}{2}\Big(\nabla w(x) + [\nabla w(x)]^T\Big).$$

(Observe that by smooth approximation, we may always take an infimum over $w \in C^1(\Omega;\mathbb{R}^n)$ in (3.11).)

Using Poincaré and more general Sobolev-Korn inequalities – the counterpart of the Poincaré inequality for the symmetrised gradient – it can be shown [7, 6] that the norm

$$\|u\|_{\mathrm{BGV}(\Omega)} := \|u\|_{L^1(\Omega)} + \mathrm{TGV}^2_{(\beta,\alpha)}(u),$$

is equivalent to the standard BV-norm $\|u\|_{\mathrm{BV}(\Omega)}$. This allows us to use a large part of the BV theory, including existing of weak* converging subsequence and solutions to variational problems, to image processing with TGV^2 regularisation.

Corollary 3.1. *Let $x \in \Omega \subset \mathbb{R}^n$ and $\rho > 0$ be such that $B(x,\rho) \subset \Omega$. Then*

$$\|u - u_{B(x,\rho)}\|_{L^1(\Omega)} \le C\rho|Du|(B(x,\rho)), \quad (u \in \mathrm{BV}(\Omega)),$$

for some constant $C = C(n)$.

The proof is an easy exercise employing the area formula.

3.6. Fine properties

We now study the fine stucture of u through the decomposition of Du into different parts. Of particular interest to use are the *jump part* and the *absolutely continuous part*, which, roughly, correspond to image edges and smooth parts. However, the *Cantor part* causes some extra headaches!

Definition 3.3. Let $u \in L^1(\Omega)$, and $x \in \Omega$. Then u has an *approximate limit* at x if there exists $z \in \mathbb{R}$ such that

$$\lim_{\rho \searrow 0} \fint_{B(x,\rho)} |u(y) - z| \, dx = 0.$$

We then set $\widetilde{u}(x) := z$. We denote the set of points where the $\widetilde{u}(x)$ does not exist by S_u, and call it the *approximate discontinuity set.*

Remark 3.2. The definition of S_u is independent of the representative of u in the L^1 equivalence class. Approximate continuity depends on the representative, but we can always find a representative that is absolutely continuous in $\Omega \setminus S_u$: we just set $u(x) = \widetilde{u}(x)$ outside S_u.

Remark 3.3. By Corollary 2.1, we have $\mathcal{L}^n(S_u)$. The next result shows that $x \notin S_u$ when both u and Du are, in a sense, non-singular.

Proposition 3.2. *Let $u \in \mathrm{BV}(\Omega)$, and $x \in \Omega$. Suppose $\Theta^*_n(Du,x) < \infty$ and $\Theta^*_n(u\mathcal{L}^n,x) < \infty$. Then the approximate limit $\widetilde{u}(x)$ exists.*

Proof. Let

$$z_\rho := \fint_{B(x,\rho)} u(y) \, dy.$$

Then by Corollary 3.1, for small enough ρ that $B(0,\rho) \subset \Omega$, we have

$$\int_{B(x,\rho)} |u(y) - z_\rho| \, dy \le C\rho|Du|(B(x,\rho)).$$

Thus

$$\limsup_{\rho \searrow 0} \fint_{B(x,\rho)} |u(y) - z_\rho| \, dy \le \limsup_{\rho \searrow 0} C\rho \frac{|Du|(B(x,\rho))}{\omega(n)\rho^n} = 0, \tag{3.12}$$

where the latter inequality follows from

$$\limsup_{\rho \searrow 0} \frac{|Du|(B(x,\rho))}{\omega(n)\rho^n} = \Theta^*_n(Du,x) < \infty.$$

If we can show for equence $\rho_i \searrow 0$ that $z_{\rho_i} \to z$ for some $z \in \mathbb{R}$, then $\widetilde{u}(x) = z$. Indeed z satisfies

$$\lim_{i \to \infty} \fint_{B(x,\rho_i)} |u(y) - z| \, dy \le \lim_{i \to \infty} \left(\fint_{B(x,\rho_i)} |u(y) - z_{\rho_i}| \, dy + |z - z_{\rho_i}| \right) = 0.$$

To see that only a subsequence suffices, suppose we had two subsequences $z_{\rho_i^1} \to \bar{z}_1$ and $z_{\rho_i^2} \to \bar{z}_2$. Then

$$\begin{aligned}
|\bar{z}_1 - \bar{z}_2| &\le \liminf_{\rho \searrow 0} \left(\fint_{B(y,\rho)} |u(y) - \bar{z}_1|\, dy + \fint_{B(y,\rho)} |u(y) - \bar{z}_2|\, dy \right) \\
&\le \limsup_{i\to\infty} \left(\fint_{B(y,\rho_i^2)} |u(y) - \bar{z}_1|\, dy + \fint_{B(y,\rho_i^1)} |u(y) - \bar{z}_2|\, dy \right) \\
&\le \limsup_{i\to\infty} \left(\fint_{B(y,\rho_i^2)} |u(y) - z_{\rho_1}|\, dy + \fint_{B(y,\rho_i^1)} |u(y) - z_{\rho_2}|\, dy + |\bar{z}_1 - z_{\rho_1}| + |\bar{z}_2 - z_{\rho_2}| \right) = 0.
\end{aligned}$$

We still have to produce $z_{\rho_i} \to z$. Given ϵ, by (3.12) for small enough $\rho > 0$, we have

$$|z_\rho| \le \fint_{B(x,\rho)} |u(y)|\, dy + \epsilon. \le \Theta_n^*(u\mathcal{L}^n, x) + 2\epsilon.$$

It follows that $\{z_\rho\}_{\rho>0}$ is bounded, and we may find a subsequence convergent to some z. □

In order to shed more light on S_u, we next look at the jumps of u.

Definition 3.4. Let $u \in L^1(\Omega)$, and $x \in \Omega$. Then x is an *approximate jump point* of u if there exist $a^+, a^- \in \mathbb{R}$ and $\nu \in \mathbb{S}^{n-1}$ such that $a^+ \neq a^-$ and

$$\lim_{\rho \searrow 0} \fint_{B_\nu^\pm(x,\rho)} |u(y) - a^\pm|\, dy = 0,$$

where the half-ball

$$B_\nu^\pm(x,\rho) = \{y \in B(x,\rho) \mid \pm\langle y - x, \nu\rangle \ge 0\}.$$

We then set $u^\pm(x) := a^\pm$ and $\nu_u(x) := \nu$. We denote the set of points x where $(u^+(x), u^-(x), \nu(x))$ exists J_u, and call it the *(approximate) jump set.*

Obviously $J_u \subset S_u$. The next result details the relationship for $u \in \mathrm{BV}(\Omega)$.

Theorem 3.9 (Federer–Vol'pert). *Let $u \in \mathrm{BV}(\Omega)$. Then the approximate discontinuity set S_u is countably $\mathcal{H}^{n-1}$-rectifiable and $\mathcal{H}^{n-1}(S_u \setminus J_u) = 0$. Moreover*

$$Du \llcorner J_u = (u^+ - u^-)\nu_u \mathcal{H}^{n-1} \llcorner J_u.$$

How about Du outside J_u? Just as approximate limits, we may define approximate differentials.

Definition 3.5. Let $u \in L^1(\Omega)$, and $x \in \Omega \setminus S_u$. Then u is *approximately differentiable* at x if there exists $z \in \mathbb{R}^n$ such that

$$\lim_{\rho \searrow 0} \fint_{B(x,\rho)} \frac{|u(y) - \widetilde{u}(x) - \langle z, y - x\rangle|}{\rho}\, dy = 0. \tag{3.13}$$

We then denote $\nabla u(x) := z$, and say that $\nabla u(x)$ is the *approximate differential* of u at x.

Theorem 3.10 (Calderón–Zygmund). *Let $u \in \mathrm{BV}(\Omega)$. Then u is approximately differentiable at $\mathcal{L}^n$-almost every $x \in \Omega$, and*

$$\frac{dD^a u}{d\mathcal{L}^n}(x) = \nabla u(x), \quad (\mathcal{L}^n\text{-a.e. } x \in \Omega). \tag{3.14}$$

Proof. Let $v = dD^a u/d\mathcal{L}^n$. We show that $z = \widetilde{v}(x)$ satisfies (3.13) for every $x \in A$ with

$$A := \{x \in \Omega \setminus (S_u \cup S_v) \mid \Theta_n(D^s u; x) = 0\}$$

By the Besicovitch derivation Theorem 2.6 and Remark 2.2, $\mathcal{L}^n$-almost every $x \in \Omega \setminus (S_u \cup S_v)$ satisfies $\Theta_n(D^s u; x) = 0$. Since $\mathcal{L}^n(S_u \cup S_v) = 0$, we have $\mathcal{L}^n(\Omega \setminus A) = 0$, so proving (3.13) and (3.14) for $x \in A$ will show our claims.

Let us do that. We pick $x \in A$, and set

$$w(y) := u(y) - \widetilde{u}(x) - \langle \widetilde{v}(x), y - x \rangle.$$

Then

$$Dw = \big(v - \widetilde{v}(x)\big)\mathcal{L}^n + D^s u,$$

so that

$$\lim_{\rho \searrow 0} \frac{|Dw|(B(x,\rho))}{\mathcal{L}^n(B(x,\rho))} = \lim_{\rho \searrow 0} \left(\fint_{B(x,\rho)} |v(y) - \widetilde{v}(x)| \, dy + \frac{|D^s u|(B(x,\rho))}{\mathcal{L}^n(B(x,\rho))} \right) = 0 + \Theta_n(D^s u; x) = 0.$$

Using $\widetilde{w}(x) = 0$ then

$$\begin{aligned}
\lim_{\rho \searrow 0} \fint_{B(x,\rho)} \frac{|u(y) - \widetilde{u}(x) - \langle \widetilde{v}(x), y - x \rangle|}{|y - x|} \, dy &= \lim_{\rho \searrow 0} \fint_{B(x,\rho)} \frac{|w(y) - \widetilde{w}(x)|}{|y - x|} \, dy \\
&\le \lim_{\rho \searrow 0} \sup_{t \in (0,\rho)} \frac{|Dw|(B(x,t))}{\mathcal{L}^n(B(x,t))} = 0.
\end{aligned}$$

In the final step, we have applied Lemma 3.2 below. Thus $\nabla u(x) = \widetilde{v}(x)$. □

Lemma 3.2. *Suppose $u \in \mathrm{BV}(B(x,r))$ and that the approximate limit $\widetilde{u}(x)$ exists. Then*

$$\fint_{B(x,r)} \frac{|u(y) - \widetilde{u}(x)|}{|y - x|} \, dy \le \fint_0^r \frac{|Du|(B(x,t))}{\mathcal{L}^n(B(x,t))} \, dt \le \sup_{t \in (0,r)} \frac{|Du|(B(x,t))}{\mathcal{L}^n(B(x,t))} \, dt.$$

Proof. We me assume without loss of generality that $x = 0$. Then, if $u \in C^\infty(B(0,r))$ and $\rho \in (0,1)$, the fundamental theorem of calculus gives

$$\frac{|u(y) - u(\rho y)|}{|y|} \le \int_\rho^1 |\nabla u|(ty) \, dt.$$

Thus application of Fubini's theorem and the area formula give

$$\begin{aligned}
\int_{B(0,r)} \frac{|u(y) - u(\rho y)|}{|y|} \, dy &\le \int_\rho^1 \int_{B(0,r)} |\nabla u|(ty) \, dy \, dt \\
&= \int_\rho^1 \int_{B(0,tr)} t^{-n} |\nabla u|(y) \, dy \, dt \qquad (3.15) \\
&= \int_\rho^1 t^{-n} |Du|(B(0,tr)) \, dt.
\end{aligned}$$

Smoothing $u \in \mathrm{BV}(B(0,r))$ using Theorem 3.3, we see that this inequality holds generally. (Fatou's inequality on the left, dominated convergence theorem on the right.) By assumption $0 \notin S_u$. Therefore

$$\lim_{\rho \searrow 0} \int_{B(0,r)} |u(\rho y) - \widetilde{u}(0)| \, dy = \lim_{\rho \searrow 0} \rho^{-n} \int_{B(0,\rho r)} |u(y) - \widetilde{u}(0)| \, dy = 0.$$

Thus the blow-up mappings $v_\rho(y) := u(\rho y)$ converge to $v(y) := \widetilde{u}(0)$ in $L^1(B(0,r))$. Consequently, we can find $\rho_i \searrow 0$ such that $v_{\rho_i}(y) \to \widetilde{u}(0)$ for $\mathcal{L}^n$-a.e. $y \in B(0,r)$. Fatou's inequality and (3.15) now give

$$\begin{aligned}
\fint_{B(0,r)} \frac{|u(y)-\widetilde{u}(0)|}{|y|}\,dy &= \fint_{B(0,r)} \liminf_{i\to\infty} \frac{|u(y)-u(\rho_i y)|}{|y|}\,dy \\
&\le \liminf_{i\to\infty} \fint_{B(0,r)} \frac{|u(y)-u(\rho_i y)|}{|y|}\,dy \\
&\le \liminf_{i\to\infty} \frac{1}{\mathcal{L}^n(B(0,r))} \int_{\rho_i}^1 t^{-n}|Du|(B(0,tr))\,dt \\
&= \int_0^1 \frac{|Du|(B(0,tr))}{\mathcal{L}^n(B(0,tr))}\,dt = \fint_0^r \frac{|Du|(B(0,t))}{\mathcal{L}^n(B(0,t))}\,dt.
\end{aligned}$$

□

Lemma 3.3. *Let $u \in \mathrm{BV}(\Omega)$, and $A \subset \Omega$ be a Borel set. Then we have the following.*

(i) If $\mathcal{H}^{n-1}(A) = 0$, then $|Du|(A) = 0$.

(ii) If $\mathcal{H}^{n-1}(A) < \infty$ and $S_u \cap A = 0$, then $|Du|(A) = 0$.

Remark 3.4. The first part of the lemma in particular shows that Du has no features of dimension less than $n-1$, which turn out to correspond principally to the jump set J_u, as we see in the next theorem.

Motivated by the above results, we make the following definition.

Definition 3.6. Let $u \in \mathrm{BV}(\Omega)$, and let $D^a u$ and $D^s u$, respectively, be the *absolutely continuous* and *singular parts* of Du with respect to $\mathcal{L}^n$, as given by the Radon-Nikodým Theorem 2.5. Then we call

$$D^j u := D^s u \llcorner J_u$$

the *jump part* of Du, and

$$D^c u := D^s u \llcorner (\Omega \setminus S_u).$$

the *Cantor part* of Du.

We may summarise the various results above in the following

Theorem 3.11. *Let $u \in \mathrm{BV}(\Omega)$. Then*

$$Du = D^a u + D^j u + D^c u,$$

where the absolutely continuous part and jump part satisfy

$$D^a u = \nabla u \mathcal{L}^n \quad \text{and} \quad D^j u = (u^+ - u^-)\nu_u \mathcal{H}^{n-1} \llcorner J_u, \tag{3.16}$$

*respectively. Moreover the Cantor part $D^c u$ vanishes on sets A that are σ-*finite with respect to $\mathcal{H}^{n-1}$: $D^c u(A) = 0$ *if $A = \bigcup_{i=1}^\infty A_i = 0$ with $\mathcal{H}^{n-1}(A_i) < \infty$, $(i = 1, 2, 3, \ldots)$.*

Proof. By the Radon-Nikodým Theorem 2.5, we have

$$Du = D^a u + D^s u,$$

By the Calderón–Zygmund Theorem 3.10, and the Federer–Vol'pert Theorem 3.9, the expressions (3.16) hold. It therefore remains to show that

$$D^s u = D^j u + D^c u, \tag{3.17}$$

and that $D^c u$ is σ-finite. By Lemma 3.3, indeed $Du \llcorner (S_u \setminus J_u) = 0$. Because $D^a u \llcorner S_u = 0$ (since $\mathcal{L}^n(S_u) = 0$), it follows that $D^s u \llcorner (S_u \setminus J_u) = 0$. Thus (3.17) holds. Finally, if $A = \bigcup_{i=1}^\infty A_i = 0$ with $\mathcal{H}^{n-1}(A_i) < \infty$, then in particular $\mathcal{H}^{n-1}(\widetilde{A}_i) < \infty$ for $\widetilde{A}_i := A_i \setminus S_u$. Therefore Lemma 3.3 shows that $D^s u(A_i) = Du(\widetilde{A}_i) = 0$. It follows that $D^s u(A) = 0$. □

3.7. The co-area formula

Definition 3.7. We say that a Borel set $E \subset \Omega$ has *finite perimeter* if $\chi_E \in \mathrm{BV}(\Omega)$. We then denote $\mathrm{Per}(E;\Omega) = |D\chi_E|(\Omega)$.

Theorem 3.12 (Co-area formula). *Suppose $\Omega \subset \mathbb{R}^n$ is open and bounded with Lipschitz boundary. Let $u \in \mathrm{BV}(\Omega)$, and denote by*

$$E_t := E_t(u) := \{x \in \Omega \mid u(x) > t\}, \quad (t \in \mathbb{R}),$$

the level sets of u. Then E_t has finite perimeter for $\mathcal{L}^1$-a.e. $t \in \mathbb{R}$, and

$$|Du|(\Omega) = \int_{-\infty}^{\infty} \mathrm{Per}(E_t;\Omega)\,dt. \tag{3.18}$$

Conversely, $u \in \mathrm{BV}(\Omega)$ if $u \in L^1(\Omega)$ and

$$\int_{-\infty}^{\infty} \mathrm{Per}(E_t;\Omega)\,dt < \infty. \tag{3.19}$$

Proof. We first prove that

$$\int_\Omega \operatorname{div}\varphi(x)u(x)\,dx = \int_{-\infty}^{\infty}\left(\int_\Omega \operatorname{div}\varphi(x)\chi_{E_t}(x)\,dx\right)dt, \quad (\varphi \in C_c^1(\Omega;\mathbb{R}^n)). \tag{3.20}$$

Let us write $u = u_+ - u_-$ where $u_\pm \geq 0$. Then

$$u_+(x) = \int_0^\infty \chi_{E_t}(x)\,dt,$$

so that

$$\begin{aligned}\int_\Omega \operatorname{div}\varphi(x)u_+(x)\,dx &= \int_\Omega \operatorname{div}\varphi(x)\left(\int_0^\infty \chi_{E_t}(x)\,dt\right)dx\\ &= \int_0^\infty \int_\Omega \operatorname{div}\varphi(x)\chi_{E_t}(x)\,dx\,dt\end{aligned}$$

Analogously

$$u_-(x) = \int_{-\infty}^0 \chi_{E_t}(x) - 1\,dt,$$

so that

$$\begin{aligned}\int_\Omega \operatorname{div}\varphi(x)u_-(x)\,dx &= \int_\Omega \operatorname{div}\varphi(x)\left(\int_{-\infty}^0 \chi_{E_t}(x) - 1\,dt\right)dx\\ &= \int_0^\infty \int_\Omega \operatorname{div}\varphi(x)(\chi_{E_t}(x)-1)\,dx\,dt\\ &= \int_0^\infty \int_\Omega \operatorname{div}\varphi(x)\chi_{E_t}(x)\,dx\,dt\end{aligned}$$

Summing the contributions from u_+ and u_-, we deduce (3.20). This also shows that

$$\int_\Omega \operatorname{div}\varphi(x)u(x)\,dx \leq \int_{-\infty}^{\infty} \mathrm{Per}(E_t;\Omega)\,dt, \quad (\varphi \in C_c^1(\Omega;\mathbb{R}^n)).$$

Thus

$$|Du|(\Omega) \leq \int_{-\infty}^{\infty} \mathrm{Per}(E_t;\Omega)\,dt. \tag{3.21}$$

In particular $u \in \mathrm{BV}(\Omega)$ if $u \in L^1(\Omega)$ and (3.19) holds.

To prove the converse and (3.18), we suppose first that $u \in \mathrm{BV}(\Omega) \cap C^\infty(\Omega)$. We let

$$m(t) := \int_\Omega 1 - \chi_{E_t}(x)\, d|Du|(x) = \int_\Omega \chi_{\{u \le t\}}(x)\, d|Du|(x).$$

Lemma 3.4 below shows that

$$m'(t) \ge \mathrm{Per}(E_t; \Omega), \quad (\mathcal{L}^1\text{-a.e. } t \in \mathbb{R})$$

and

$$\int_{-\infty}^{\infty} m'(t) \le |Du|(\Omega).$$

Thus

$$\int_{-\infty}^{\infty} \mathrm{Per}(E_t; \Omega)\, dx \le |Du|(\Omega).$$

This proves (3.18) for $u \in \mathrm{BV}(\Omega) \cap C^\infty(\Omega)$.

To prove the general case, we take a sequence $\{u^i\}_{i=1}^\infty \in \mathrm{BV}(\Omega) \cap C^\infty(\Omega)$ strictly approximating u, as given by Theorem 3.3. Since $u^i \to u$ in $L^1(\Omega)$, we also have

$$\chi_{E_t(u^i)} \to \chi_{E_t} \quad \text{in } L^1(\Omega) \text{ for } \mathcal{L}^1\text{-a.e. } t. \tag{3.22}$$

To see this, one may apply Fatou's inequality on

$$\int_{-\infty}^{\infty} \lim_{i\to\infty} \|\chi_{E_t(u^i)} - \chi_{E_t(u)}\|_{L^1(\Omega)}\, dt.$$

By Theorem 2.10,

$$\mathrm{Per}(E_t; \Omega) \le \liminf_{i\to\infty} \mathrm{Per}(E_t(u^i); \Omega).$$

Thus by another referral to Fatou's inequality

$$\int_{-\infty}^{\infty} \mathrm{Per}(E_t; \Omega)\, dt \le \liminf_{i\to\infty} \int_{-\infty}^{\infty} \mathrm{Per}(E_t(u^i); \Omega) = \lim_{i\to\infty} |Du^i|(\Omega) = |Du|(\Omega).$$

Together with (3.21) this proves the coarea formula (3.18). That $\mathrm{Per}(E_t; \Omega) < \infty$ for $\mathcal{L}^1$-a.e. $t \in \mathbb{R}$ is immediate from (3.18). □

Lemma 3.4. *Let $u \in C^\infty(\Omega) \cap \mathrm{BV}(\Omega)$. Letting*

$$m(t) := \int_\Omega \chi_{\{u \le t\}}(x)\, d|Du|(x),$$

we have

$$m'(t) \ge \mathrm{Per}(E_t; \Omega), \quad (\mathcal{L}^1\text{-a.e. } t \in \mathbb{R}). \tag{3.23}$$

and

$$\int_{-\infty}^{\infty} m'(t) \le |Du|(\Omega). \tag{3.24}$$

Proof. From the construction, we immediately see that the function m is non-decreasing and the derivative m' exists $\mathcal{L}^1$-a.e. . The fundamental theorem of calculus shows (3.24). To show (3.23), it suffices to show that given $\varphi \in C_c^\infty(\Omega; \mathbb{R}^n)$ and $\|\varphi\|_{L^\infty(\Omega;\mathbb{R}^n)} \le 1$, it holds

$$m'(t) \ge -\int_\Omega \mathrm{div}\, \varphi(x) \chi_{E_t}\, dx, \quad (\text{when } m'(t) \text{ exists}). \tag{3.25}$$

Fixing any $t \in \mathbb{R}$ and $r > 0$, we define to χ_{E_t} the approximation $\eta \circ u$ using

$$\eta(s) := \begin{cases} 0, & s \le t, \\ \frac{s-t}{r}, & t \le s \le t+r, \\ 1, & s \ge t+r. \end{cases}$$

Then

$$\eta'(s) = \begin{cases} 1/r, & t < s < t + r, \\ 0, & \text{otherwise.} \end{cases}$$

Hence

$$\begin{aligned} -\int_\Omega \operatorname{div} \varphi(x)\eta(u(x))\,dx &= \int_\Omega \eta'(u(x))\langle \nabla u(x), \varphi(x)\rangle\,dx \\ &= \frac{1}{r}\int_\Omega \left(\chi_{E_t}(x) - \chi_{E_{t+r}}(x)\right)\langle \nabla u(x), \varphi(x)\rangle\,dx. \\ &\le \frac{1}{r}\int_\Omega \left(\chi_{E_t}(x) - \chi_{E_{t+r}}(x)\right)|\nabla u(x)|\,dx \\ &= \frac{m(t+r) - m(t)}{r}. \end{aligned}$$

Letting $r \to 0$, (3.25) follows. ☐

Chapter 4

Total variation denoising

4.1. Problem statement

Let $f \in L^2(\Omega)$ be a noisy image. The Rudin-Osher-Fatemi (ROF) total variation denoising problem is stated

$$\min_{u \in \mathrm{BV}(\Omega)} \frac{1}{2}\|f - u\|^2_{L^2(\Omega)} + \alpha \mathrm{TV}(u). \tag{4.1}$$

Here $\alpha > 0$ is a parameter that balances between restoring f perfectly (not removing noise) as $\alpha \searrow 0$ and simply averaging f as $\alpha \nearrow \infty$.

4.2. Level set formulation

Theorem 4.1. *Let $f \in L^2(\Omega) \cap \mathrm{BV}(\Omega)$, where $\Omega \subset \mathbb{R}^n$ is open and bounded with Lipschitz boundary. Then $\hat{u}$ solves (4.1) if and only if the super/sub-level sets*

$$E^*_t(\hat{u}) := \begin{cases} E_t(\hat{u}), & t \geq 0, \\ E_{-t}(-\hat{u}), & t < 0, \end{cases}$$

and the function

$$g_t(x) := \mathrm{sgn}(t)(t - f(x))$$

solve for corresponding t the minimal surface problem

$$\min_{E \subset \Omega} \int_E g_t(x)\, dx + \alpha \mathrm{Per}(E; \Omega), \quad (\mathcal{L}^1\text{-a.e. } t \in \mathbb{R}), \tag{4.2}$$

Lemma 4.1. *Let $f \in L^2(\Omega) \cap \mathrm{BV}(\Omega)$ and $u \in \mathrm{BV}(\Omega)$. Then*

$$\frac{1}{2}\|f - u\|^2_{L^2(\Omega)} + \alpha \mathrm{TV}(u) = \int_0^\infty \left(\int_{E^*_t(u)} g_t(x)\, dx + \mathrm{Per}(E^*_t(u); \Omega) \right) dt + \frac{1}{2}\|f\|^2_{L^2(\Omega)}.$$

Proof. By the co-area formula

$$\mathrm{TV}(u) = \int_{-\infty}^\infty \mathrm{Per}(E_t(u); \Omega)\, dt = \int_{-\infty}^\infty \mathrm{Per}(E^*_t(u); \Omega)\, dt,$$

so it remains to expand the fidelity term in terms of level sets. We have

$$\frac{1}{2}|f(x) - u(x)|^2 - \frac{1}{2}|f(x)|^2 = \frac{1}{2}|u(x)|^2 - f(x)u(x) = \begin{cases} \int_0^{u(x)} \big(t - f(x)\big)\, dt, & u(x) \geq 0, \\ -\int_{u(x)}^0 \big(t - f(x)\big)\, dt, & u(x) < 0, \end{cases}$$

Therefore

$$\begin{aligned}
\frac{1}{2}\|f-u\|^2_{L^2(\Omega)} - \frac{1}{2}\|f\|^2_{L^2(\Omega)} &= \int_\Omega \left(\frac{1}{2}|f(x)-u(x)|^2 - \frac{1}{2}|f(x)|^2\right) dx \\
&= \int_\Omega \left(\int_0^{u(x)} \chi_{E_0(u)}(x)\big(t-f(x)\big)\,dt - \int_{u(x)}^0 \chi_{E_0(-u)}(x)\big(t-f(x)\big)\,dt\right) dx \\
&= \int_\Omega \left(\int_{-\infty}^{\infty} \chi_{E_0(u)\cap E_t(u)}(x)\big(t-f(x)\big)\,dt - \int_{-\infty}^{\infty} \chi_{E_0(-u)\cap E_{-t}(-u)}(x)\big(t-f(x)\big)\,dt\right) dx \\
&= \int_\Omega \int_{-\infty}^{\infty} \chi_{E_t^*(u)}(x)\,\mathrm{sgn}(t)\big(t-f(x)\big)\,dt\,dx.
\end{aligned}$$

Referral to Fubini's theorem now establishes the claim. □

Lemma 4.2. *Suppose $h, g \in L^1(\Omega)$ with $g(x) < h(x)$ for $\mathcal{L}^n$-almost every $x \in \Omega$. If $\hat{E}$ and $\hat{F}$ solve, respectively*

$$\min_{E\subset\Omega} \mathrm{Per}(E;\Omega) - \int_E g(x)\,dx, \quad \textit{and} \quad \min_{F\subset\Omega} \mathrm{Per}(F;\Omega) - \int_F h(x)\,dx,$$

then $\mathcal{L}^n(\hat{E}\setminus\hat{F}) = 0$.

Proof. Observe that if $A, B \subset \Omega$ are Borel sets, then

$$\mathrm{Per}(A\cup B;\Omega) + \mathrm{Per}(A\cap B;\Omega) \le \mathrm{Per}(A;\Omega) + \mathrm{Per}(B;\Omega). \tag{4.3}$$

If now $\mathcal{L}^n(\hat{E}\setminus\hat{F}) > 0$, we have

$$\begin{aligned}
-\int_{\hat{E}} g(x)\,dx &= -\int_{\hat{E}\cap\hat{F}} g(x)\,dx - \int_{\hat{E}\setminus\hat{F}} g(x)\,dx \\
&> -\int_{\hat{E}\cap\hat{F}} g(x)\,dx + \int_{\hat{E}\setminus\hat{F}} h(x)\,dx.
\end{aligned}$$

Thus

$$\begin{aligned}
&\left(\mathrm{Per}(\hat{E};\Omega) - \int_{\hat{E}} g(x)\,dx\right) + \left(\mathrm{Per}(\hat{F};\Omega) - \int_{\hat{F}} h(x)\,dx\right) \\
&\qquad > \left(\mathrm{Per}(\hat{E}\cap\hat{F};\Omega) - \int_{\hat{E}\cap\hat{F}} g(x)\,dx\right) + \left(\mathrm{Per}(\hat{E}\cup\hat{F};\Omega) - \int_{\hat{E}\cup\hat{F}} h(x)\,dx\right).
\end{aligned}$$

This contradicts the optimality of $\hat{E}$ and $\hat{F}$. □

Proof of Theorem 4.1. Applying Lemma 4.2 on $g = -(t-f)$ and $h = -(s-f)$ for $t > s \ge 0$, we see that solutions A_t and A_s to

$$\min_{A\subset\Omega} \int_A \big(t - f(x)\big)\,dx + \alpha\mathrm{Per}(A;\Omega) \tag{4.4}$$

at corresponding levels t and s satisfy $\mathcal{L}^n(A_t\setminus A_s) = 0$.

Likewise, applying Lemma 4.2 on $g = t - f$ and $h = s - f$ for $t < s \le 0$, we see that the solutions C_t and C_s to

$$\min_{C\subset\Omega} \int_C -\big(t - f(x)\big)\,dx + \alpha\mathrm{Per}(C;\Omega) \tag{4.5}$$

at corresponding levels t and s satisfy $\mathcal{L}^n(C_t\setminus C_s) = 0$.

Most of the work is now done. We just have show that we get no growth of the level sets at $t = 0$, that is

$$\mathcal{L}^n(A_t\cap C_s) = 0, \quad (s < 0 < t). \tag{4.6}$$

Indeed, because $\mathrm{Per}(A_0; \Omega) = \mathrm{Per}(\Omega \setminus A_0; \Omega)$, we have

$$\int_{\Omega \setminus A_0} f(x)\, dx + \alpha \mathrm{Per}(\Omega \setminus A_0; \Omega) = \int_\Omega f(x)\, dx + \left(\int_{A_0} -f(x)\, dx + \alpha \mathrm{Per}(A_0; \Omega) \right).$$

As the term in parentheses achieves the minimum in (4.4) for $t = 0$, we see that $C_0' := \Omega \setminus A_0$ solves (4.5) for $t = 0$. Likewise $A_0' := \Omega \setminus C_0$ solves (4.4) for $t = 0$. It follows that $\mathcal{L}^n(A_t \setminus A_0') = 0$ for $t > 0$ and $\mathcal{L}^n(C_t \setminus C_0') = 0$ for $t < 0$. Consequently (4.6) holds. Therefore, if we define

$$E_t = \begin{cases} A_t, & t > 0, \\ \Omega \setminus C_t, & t < 0, \end{cases}$$

and set

$$u(x) = \sup\{t \in \mathbb{R} \setminus \{0\} \mid x \in E_t\},$$

then $E_t(u) = E_t$ for $\mathcal{L}^1$-a.e. $t \in \mathbb{R}$. Now, since by construction E_t solves (4.2) for t, using Lemma 4.1, we see that u solves (4.1).

Conversely, if u solves (4.1), then $E_t = E_t(u)$ necessarily solves (4.2), because otherwise the argument above would construct more optimal A_t or C_t for t in a set of positive measure. This way v violating the optimality of u could be constructed. □

Remark 4.1. We may extend the statement of Theorem 4.1 to $\Omega = \mathbb{R}^n$; the only point where the the assumptions of open and bounded with Lipschitz boundary were needed, was the smooth approximation in the proof of Theorem 3.12. This can also be done in $\Omega = \mathbb{R}^n$ simply by mollification.

Example 4.1. Let $f = \chi_A$ for a Borel set $A \subset \Omega$. Then

$$g_t(x) = \begin{cases} t - \chi_A(x), & t \geq 0, \\ \chi_A(x) - t, & t < 0, \end{cases}$$

so that

$$g_t(x) \begin{cases} \geq 0, & t \geq 1, \\ \leq 0, & t \leq 0, \\ = t\chi_{A^c}(x) + (t-1)\chi_A(x), & 0 < t < 1. \end{cases}$$

It follows that $E_t(\hat{u}) = \emptyset$ if $t > 1$ and $E_t(\hat{u}) = \Omega$ if $t < 0$.

If $\Omega = \mathbb{R}^2$ and A is closed and convex, it can be further shown that $E_t(\hat{u}) \subset A$ for $0 < t < 1$. In fact, applying the results for L^1 fidelities in [12], it is not difficult to show that in this case

$$E_t(\hat{u}) = \bigcup \{B_x \mid x \in \Omega,\, B_x := B(x, \rho(\alpha, t)) \subset A\}$$

for a suitable radius $\rho(\alpha, t)$,

We may have $E_t(\hat{u}) = \emptyset$ for $t \in (0, 1)$. Also, if we had $\Omega \subsetneq \mathbb{R}^2$, we would not necessarily have $E_t(\hat{u}) \subset A$ for $t \in (0, 1)$. Both of these effects are part of the *contrast loss* that the L^2-TV model exhibits, but L^1-TV doesn't.

4.3. Structure of the jump set

Theorem 4.2. *Let $f \in L^\infty(\Omega) \cap \mathrm{BV}(\Omega)$, where $\Omega \subset \mathbb{R}^n$ is open and bounded with Lipschitz boundary. If $\hat{u}$ solves (4.1), then*

$$\mathcal{H}^{n-1}(J_{\hat{u}} \setminus J_f) = 0. \tag{4.7}$$

For the proof, we follow [8]. We require the following result based on regularity results for minimal surfaces.

Theorem 4.3. *Denote the symmetric difference of two sets A, B by*

$$A \Delta B := (A \setminus B) \cup (B \setminus A).$$

Let $\lambda > 0$, and suppose $A \in \mathcal{B}(\Omega)$ satisfies for every compact $K \subset \Omega$ and $B \in \mathcal{B}(\Omega)$ with

$$\mathcal{L}^n((A \Delta B) \setminus K) = 0,$$

the inequality

$$\operatorname{Per}(A; K) \leq \operatorname{Per}(B; K) + \lambda \mathcal{L}^n(A \Delta B).$$

Then, except for a singular set Σ of Hausdorff dimension at most $n-8$, the boundary ∂A is of class $C^{1,\alpha}$ for any $\alpha \in (0,1)$. That is, for every $x \in \partial E \setminus \Sigma$, there exists $\rho > 0$ such that up to rotation, $B(x,\rho) \cap \partial A$ is the graph of a $C^{1,\alpha}$ function, i.e., a continuously differentiable function ψ with $\nabla\psi$ Hölder continuous of exponent α. Moreover,

$$\mathcal{H}^s(\Sigma) = 0, \quad (s > n-8).$$

Proof. See [1]. The proof therein only directly applies to $2 \leq n \leq 7$ (in which case $\Sigma = \emptyset$). The case $n = 1$ is trivial with regard to regularity. For $n > 8$ we may extend the proof based on regularity results for minimal surfaces in higher dimensions. These can be found in [21, 17]. A very recommendable introductory book to the entire topic is [26]. □

Proof of Theorem 4.2. Let $E_t := E_t(\hat{u})$ and $M := \|f\|_{L^\infty(\Omega)}$. By Theorem 4.1, E_t solves (4.2). Thus, with $g_t(x) := \operatorname{sgn}(t)(t - f(x))$, we have

$$\alpha \operatorname{Per}(E_t; \Omega) + \int_{E_t} g_t(x)\, dx \leq \alpha \operatorname{Per}(B; \Omega) + \int_B g_t(x)\, dx, \quad (B \in \mathcal{B}(\Omega)).$$

In particular with $K \subset \Omega$ compact, we have

$$\alpha \operatorname{Per}(E_t; K) + \int_{E_t \cap K} g_t(x)\, dx \leq \alpha \operatorname{Per}(B; K) + \int_{B \cap K} g_t(x)\, dx,$$

whenever $B \in \mathcal{B}(\Omega)$ with $\mathcal{L}^n((E_t \Delta B) \setminus K) = 0$. If $t \in [-M, M]$, we have

$$\left| \int_{E_t} g_t(x)\, dx - \int_B g_t(x)\, dx \right| \leq 2M \mathcal{L}^n(E_t \Delta B)).$$

It follows from Theorem 4.3 that E_t is of class $C^{1,\alpha}$ for $t \in [-M, M]$. Regarding $t > M$ and $t < -M$, we clearly have, $E_t = \emptyset$ and $E_t = \Omega$, respectively. In both cases $\partial E_t = \emptyset$.

We want to show that if $t > s$, then

$$\mathcal{H}^{n-1}(\partial E_s \cap \partial E_t \setminus J_f) = 0. \tag{4.8}$$

This would imply (4.7). Indeed, any point $x \in J_{\hat{u}}$ satisfies $x \in \partial E_{a-\epsilon} \cap \partial E_{b+\epsilon}$ for $a = \max\{u^+(x), u^-(x)\}$, $b = \min\{u^-(x), u^+(x)\}$, and any $\epsilon \in (0, a-b)$. In particular, by the density of $\mathbb{Q}$ in $\mathbb{R}$, we can pick $s, t \in \mathbb{Q}$, $s \neq t$, such that $x \in \partial E_s \cap \partial E_t$. Thus

$$J_{\hat{u}} \subset \bigcup_{s,t \in \mathbb{Q}; s \neq t} \partial E_s \cap \partial E_t.$$

Since the union is over a countable set, (4.8) implies (4.7). We can also take $\operatorname{sgn} t = \operatorname{sgn} s$.

Let Σ_t be the singular set of E_t, given by 4.3. Assuming that (4.7) does not hold, we can find $x \in J_u \setminus J_f$ with $x \notin \Sigma_t \cap \Sigma_s$. By Theorem 3.9, the set S_f is countably $\mathcal{H}^{n-1}$ rectifiable, and $\mathcal{H}^{n-1}(S_f \setminus J_f) = 0$. It follows that for $\mathcal{H}^{n-1}$–a.e. $y \in J_u \setminus J_f$, the approximate limit $\widetilde{f}(y)$ exists. We may therefore assume that $\widetilde{f}(x)$ exists. Choosing an appropriate representative of the L^1 equivalence class of f, we may further assume that x is a Lebesgue point of f.

By rotation and translation if necessary, we may take $x = 0$ and

$$\nu_{E_t}(x) = e_n, \quad \text{(the } n\text{:th unit vector).} \tag{4.9}$$

Therefore, locally E_t can be written as graph of $\psi_t \in C^{1,\alpha}(U_t)$, with $\psi_t(0) = 0$ and

$$U_t = B(0, \delta_t) \subset \mathbb{R}^{n-1}$$

for some $\delta_t > 0$. This means that within $Q_t := U_t \times (-\delta_t, \delta_t)$ we have

$$E_t \cap Q_t = \{(v, \rho) \mid -\delta_t \leq \rho \leq \psi_t(v), v \in U_t\}$$

and

$$\partial E_t \cap Q_t = \{(v, \psi_t(v)) \mid v \in U_t\} =: \Psi_t(U_t).$$

Since $E_s \supset E_t$, $(s < t)$ and $x \in \partial E_t \cup \partial E_s$, we have that the outer normals agree, $\nu_{E_t}(x) = \nu_{E_s}(x)$. Therefore we may locally parametrise ∂E_s analogously to ∂E_t with $U_s = U_t$ and $Q_s = Q_t$. Then $\psi_s \geq \psi_t$. Henceforth we denote $Q := Q_t$ and $U := U_t$.

With $r = t, s$, we may write

$$[\nabla \Psi_r(v)]^* w = (w, 0) + (0, \langle \nabla \psi_r(v), w \rangle).$$

Thus

$$\langle [\nabla \Psi_r(v)]^* w, [\nabla \Psi_r(v)]^* w \rangle = \begin{cases} \|w\|^2(1 + \|\nabla \psi_r(v)\|^2), & w \propto \nabla \psi_r(v), \\ \|w\|^2, & w \perp \nabla \psi_r(v). \end{cases}$$

It follows that the eigenvalues of $\nabla \Psi_r(v)[\nabla \Psi_r(v)]^*$ are $\lambda_1 = 1 + \|\nabla \psi_r(v)\|^2$ and $\lambda_k = 1$ for $k = 1, \ldots, n-2$. Thus the $(n-1)$-dimensional Jacobian

$$\mathcal{J}_{n-1}([\nabla \Psi_r(v)]^*) = \sqrt{\det\Big(\nabla \Psi_r(v)[\nabla \Psi_r(v)]^*\Big)} = \sqrt{1 + \|\nabla \psi_r(v)\|^2}.$$

By the area formula, Theorem 2.8, we can write

$$\operatorname{Per}(E_r; Q) = \mathcal{H}^{m-1}(\partial E_r \cap Q) = \int_U \sqrt{1 + \|\nabla \psi_r(v)\|^2}\, d\mathcal{H}^{m-1}(v)$$

and

$$\int_{E_r \cap Q} \operatorname{sgn}(t)(t - f(x))\, dx = \int_U \int_{-\delta}^{\psi_r(v)} \operatorname{sgn}(t)(t - f(v, \rho))\, d\rho\, dv.$$

Let us now consider local variations to (4.2), adding $h \in C_0^1(U)$ to ψ_r. Then we have the problem

$$\min_{h \in C_0^1(U)} \left(\int_U \sqrt{1 + |\nabla(\psi_r + h)(v)|^2}\, d\mathcal{H}^{m-1}(v) + \int_U \int_{-\delta}^{(\psi_r + h)(v)} \operatorname{sgn}(r)(r - f(v, \rho))\, d\rho\, dv \right).$$

Since E_r is optimal by Theorem 4.1, $h = 0$ must solve this problem. Differentiating to get the first order optimality conditions, we therefore see that

$$-\alpha \operatorname{div} \frac{\nabla \psi_r(v)}{\sqrt{1 + \|\nabla \psi_r(v)\|^2}} + \operatorname{sgn}(r)(r - f(v, \psi_r(v))) = 0, \quad (v \in U; r = t, s) \tag{4.10}$$

in a weak sense. Using results for higher regularity of solutions to elliptic partial differential equations (see, e.g., [20]), we can show that actually $\psi_r \in C^2(U')$ for any $0 \in U' \Subset U$. Thus (4.10) holds in the depicted classical sense on U'. As $x = 0$ is a Lebesgue point of f, the value $f(0) = f(0, \psi_r(0))$ has a pointwise meaning. Using (4.9), we also see that $\nabla \psi_r(0) = 0$, $(r = t, s)$. With $v = 0$, (4.10) therefore gives

$$-\alpha \Delta \psi_r(0) + \operatorname{sgn}(r)(r - f(0, \psi_r(0))) = 0, \quad (r = t, s). \tag{4.11}$$

Subtracting (4.11) for $r = t, s$ we get when $\operatorname{sgn} t = \operatorname{sgn} s = 1$ that

$$\alpha \Delta(\psi_t(0) - \psi_s(0)) = t - s > 0.$$

This contradicts $\psi_s > \psi_t$ (i.e., $E_s \supset E_t$). The case $\operatorname{sgn} t = \operatorname{sgn} s = -1$ is analogous. □

Remark 4.2. We now know that the L^2-TV denoising model does not introduce artefacts in terms of edges. But does it preserve edges? Generally $\mathcal{H}^{m-1}(J_f \setminus J_{\hat{u}}) = 0$ does not hold. But can we say, let's say, that $\lim_{\alpha \searrow 0} \mathcal{H}^{m-1}(J_f \setminus J_{\hat{u}_\alpha}) = 0$, where $\hat{u}_\alpha$ solves (4.1) for α? In some explicit cases yes, as in the case of convex sets in Example 4.1.

Chapter 5

Special functions of bounded variation

We finish the course with a quick look into special functions of bounded variation, and the Mumford–Shah image segmentation problem.

5.1. Basics

Definition 5.1. We call functions $u \in \mathrm{BV}(\Omega)$ with $D^c u = 0$ *special functions of bounded variation*, and denote the space by $\mathrm{SBV}(\Omega)$.

This kind of functions are important in various *free discontinuity problems*, see [2]. These include in particular the *Mumford–Shah image segmentation problem.* Image segmentation is an important task in computer vision, and attempts to discover different objects in the scene by dividing it into segments. Before looking at this problem in more detail, we have to establish some facts about SBV.

5.2. Compactness

Although SBV gets rid of the sometimes nasty Cantor part $D^c u$, it comes with its own problems. In particular, if we have a sequence $\{u^i\}_{i=1}^\infty \subset \mathrm{SBV}(\Omega)$, converging in any of the standard senses – strongly, strictly, or weakly* – to some $u \in \mathrm{BV}(\Omega)$, we do not necessarily have $u \in \mathrm{SBV}(\Omega)$. This problem is related to similar difficulties with compactness in the space $L^1(\Omega)$, and the associated Dunford–Pettis theorem; see [18].

Fortunately, we do have a following compactness result, for whose statement we introduce a few shorthand notations.

Definition 5.2. If $\varphi : [0,\infty) \to [0,\infty)$, we write

$$\varphi_0 := \lim_{t \searrow 0} \varphi(t)/t, \quad \text{and} \quad \varphi^\infty := \lim_{t \nearrow \infty} \varphi(t)/t,$$

implicitly assuming that the (possibly infinite) limits exist.

Definition 5.3. Let $u \in \mathrm{BV}(\Omega)$. Introducing the function $\theta_u \in L^1(J_u)$, we write $|D^j u| = \theta_u \mathcal{H}^{n-1} \llcorner J_u$.

Theorem 5.1 (SBV compactness)**.** *Let $\Omega \subset \mathbb{R}^n$ be open and bounded. Suppose $\varphi, \psi : [0,\infty) \to [0,\infty)$ are lower semicontinuous and increasing with $\varphi^\infty = \infty$ and $\psi_0 = \infty$. Suppose $\{u^i\}_{i=1}^\infty \subset \mathrm{SBV}(\Omega)$ and $u^i \overset{*}{\rightharpoonup} u \in \mathrm{SBV}(\Omega)$ weakly* in $\mathrm{BV}(\Omega)$. If*

$$\sup_{i=1,2,3,\ldots} \left(\int_\Omega \varphi(\|\nabla u^i(x)\|)\, dx + \int_{J_{u^i}} \psi(\theta_{u^i}(x))\, d\mathcal{H}^{n-1}(x) \right) < \infty,$$

then $u \in \mathrm{SBV}(\Omega)$ and there exists a subsequence of $\{u^i\}_{i=1}^\infty$, unrelabelled, such that

$$u^i \to u \text{ strongly in } L^1(\Omega), \tag{5.1}$$

$$\nabla u^i \rightharpoonup \nabla u \text{ weakly in } L^1(\Omega;\mathbb{R}^n), \tag{5.2}$$

$$D^j u^i \overset{*}{\rightharpoonup} D^j u \text{ weakly* in } \mathcal{M}(\Omega;\mathbb{R}^n). \tag{5.3}$$

If, moreover, ψ is subadditive with $\psi(0) = 0$, then

$$\int_{J_u} \psi(\theta_u(x))\, d\mathcal{H}^{n-1}(x) \le \liminf_{i\to\infty} \int_{J_{u^i}} \psi(\theta_{u^i}(x))\, d\mathcal{H}^{n-1}(x). \tag{5.4}$$

Roughly, the idea with $\varphi^\infty = \infty$ is to prevent the absolutely continuous part from forming higher-density Cantor or jump parts in the limit. Likewise, the idea with $\psi^0 = \infty$ is to prevent jumps from flattening out into lower-density Cantor or absolutely continuous parts in the limit.

5.3. The Mumford–Shah problem

The Mumford–Shah segmentation problem, originally stated in a discrete setting [27], is often stated in the form

$$\inf\{J(K,u) \mid K \subset \overline{\Omega} \text{ closed}, u \in C^1(\Omega \setminus K)\} \tag{5.5}$$

for

$$J(K,u) := \int_{\Omega\setminus K} \left(\|\nabla u(x)\|^2 + \alpha(u(x) - f(x))^2\right) dx + \beta\mathcal{H}^{n-1}(K \setminus \Omega).$$

Here $\alpha, \beta > 0$ are two regularisation parameters, and $f \in L^2(\Omega)$ our source image. The idea is to find a set K modelling the boundaries between different image segments, and then approximate f – which may have Gaussian noise as modelled by the squared L^2 term – by a C^1 function u within each of the segments. To keep the solution reasonably simple, we penalise the complexity and count of the segments by the Hausdorff measure of their boundary K. The term $\|\nabla u\|^2$ likewise penalises the complexity of the image within the segments – it should be smooth, and not cross any natural segment boundaries, where the gradient would blow up and not be an L^2 function, but a measure.

The above formulation of the segmentation problem is mathematically troublesome, because of the dependence on both u and K. On the surface of it, there is no simple natural space where the solution lives. A natural idea is to replace K by the jump set J_u of an SBV function u, relaxing the requirements that K is closed and u smooth. The SBV compactness theorem immediately shows existence of solutions to the relaxed problem

$$\min_{u\in \mathrm{SBV}(\Omega)} \alpha\|u - f\|^2_{L^2(\Omega)} + \left(\|\nabla u\|^2_{L^2(\Omega;\mathbb{R}^n)} + \beta\mathcal{H}^{n-1}(J_u)\right). \tag{5.6}$$

The question is, are the solutions of (5.5) and (5.6) related? Fortunately, yes, if we relax the C^1 requirement to $W^{1,2}_{\mathrm{loc}}$.

Theorem 5.2. *Suppose $f \in L^\infty(\Omega) \cap L^2(\Omega)$. Let $v \in \mathrm{SBV}(\Omega)$ solve (5.6). Then $K = \overline{S}_v$ and $u = v \in W^{1,2}_{\mathrm{loc}}(\Omega \setminus K)$ solve (5.5).*

The proof is very long, and we point the interested reader to [2]. The crucial bit in the proof is the *density lower bound*

$$\mathcal{H}^{n-1}(S_u \cap B(x,\rho)) \ge \theta\rho^{n-1}, \quad (0 < \rho < \min\{d(x,\partial\Omega), \beta\alpha^{-1}\|f\|^{-2}_{L^\infty(\Omega)}\}),$$

for some constant $\theta = \theta(n) > 0$ and any $x \in \overline{S}_u$. This allows to show that S_u is not too much of a point cloud, such that taking its closure will not add too much to it.

Bibliography

1. W. Allard, *Total variation regularization for image denoising, I. Geometric theory*, SIAM Journal on Mathematical Analysis, 39 (2008), pp. 1150–1190.

2. L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford University Press, 2000.

3. G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, Springer, 2nd ed., 2006.

4. B. Bollobás, *Linear Analysis: An Introductory Course*, Cambridge mathematical textbooks, Cambridge University Press, 1999.

5. K. Bredies, K. Kunisch, and T. Pock, *Total generalized variation*, SIAM Journal on Imaging Sciences, 3 (2011), pp. 492–526.

6. K. Bredies, K. Kunisch, and T. Valkonen, *Properties of L^1-TGV^2: The one-dimensional case*, Journal of Mathematical Analysis and Applications, 398 (2013), pp. 438–454.

7. K. Bredies and T. Valkonen, *Inverse problems with second-order total generalized variation constraints*, in Proceedings of the 9th International Conference on Sampling Theory and Applications (SampTA) 2011, Singapore, 2011.

8. V. Caselles, A. Chambolle, and M. Novaga, *The discontinuity set of solutions of the TV denoising problem and some extensions*, Multiscale Modeling and Simulation, 6 (2008), pp. 879–894.

9. A. Chambolle and P.-L. Lions, *Image recovery via total variation minimization and related problems*, Numerische Mathematik, 76 (1997), pp. 167–188.

10. T. Chan and J. Shen, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*, SIAM e-books, Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 2005.

11. T. F. Chan, S. H. Kang, and J. Shen, *Euler's elastica and curvature-based inpainting*, SIAM Journal on Applied Mathematics, (2002), pp. 564–592.

12. V. Duval, J. F. Aujol, and Y. Gousseau, *The TVL1 model: A geometric point of view*, Multiscale Modeling and Simulation, 8 (2009), pp. 154–189.

13. H. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Mathematics and Its Applications, Springer Netherlands, 2000.

14. L. C. Evans, *Partial Differential Equations*, Americal Mathematical Society, 1998.

15. L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, 1992.

16. K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, Wiley, 2004.

17. H. Federer, *Geometric Measure Theory*, Springer, 1969.

18. I. Fonseca and G. Leoni, *Modern methods in the calculus of variations: L^p spaces*, Springer Verlag, 2007.

19. A. Friedman, *Foundations of Modern Analysis*, Dover Books on Mathematics, Dover, 1970.

20. D. Gilbarg and N. Trudinger, *Elliptic partial differential equations of second order*, Classics in mathematics, Springer, 2001.

21. E. Giusti, *Minimal Surfaces and Functions of Bounded Variation*, vol. 80 of Monographs in Mathematics, Birkhäuser, 1984.

22. J. Huang and D. Mumford, *Statistics of natural images and models*, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1999.

23. A. N. Kolmogorov and S. V. Fomin, *Introductory Real Analysis*, Dover Books on Mathematics, Dover, 1975.

24. M. Lysaker, A. Lundervold, and X.-C. Tai, *Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time*, IEEE Transactions on Image Processing, 12 (2003), pp. 1579–1590.

25. P. Mattila, *Geometry of sets and measures in Euclidean spaces: Fractals and rectifiability*, Cambridge University Press, 1999.

26. F. Morgan, *Geometric Measure Theory: A Beginner's Guide*, Academic Press, 1987.

27. D. Mumford and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Communications on Pure and Applied Mathematics, 42 (1989), pp. 577–685.

28. W. Rudin, *Real and Complex Analysis*, McGraw-Hill Book Company, 1966.

29. J. Shen, S. Kang, and T. Chan, *Euler's elastica and curvature-based inpainting*, SIAM Journal on Applied Mathematics, 63 (2003), pp. 564–592.