POLITECNICO DI TORINO

Master's Degree in Mathematical Engineering

Master's Degree Thesis

ROF MODEL FOR IMAGE DENOISING
analysis and implementation

Supervisor

Prof. Luca LUSSARDI

Candidate

Edoardo VOGLINO

3 October 2023

Acknowledgements

Dedicated to my grandfather *Sergio Gennari*, who sadly passed away in December 2022, for all the support he provided on my studies and during my whole life. I am so grateful that I had you in my life, my university achievement are also yours.

At first, I have to thank my family, especially my father and my mother for all the care they provided and the effort they put to carry me in my education.

I would like also to thank the professors who guided me in these years at Politecnico di Torino. In particular *L.Lussardi*, the supervisor for this work, and *M.Morandotti*, two of the most effective teachers I have ever had the luck to meet and who were able to convey me the interest for mathematical analysis and support my preparation.

I would like to reserve my last and most sincere thanks for my classmates and friends, who with their mathematical passion, initiative and ambitions pulled me forward, allowing me to make many growth experiences with an amazing company.

"The man and society are completely different ideals now, but the true road to victory requires embracing a man's values... and you should be able to see it now. Continue the race and find out for yourself the path of light. That is what I will pray for. [...] Welcome to the true man's world. "
Hirohiko Araki

Summary

The purpose of this thesis is presenting the Rudin-Osher-Fatemi filter for image analysis. We will consider mostly the application for denoising, however there are other relevant application, such as restoring damaged inscriptions.

This filter involves solving a minimization problem with the objective function built summing an L^2-norm squared, which measure the distance from the reference image, and the total variation functional. Therefore, one of the main concerns of this work is to understand the behaviour of this problem under a variational point of view, this involves studying the total variation functional and the main properties of the bounded variation functions, on which it is defined.

In the first chapter the analytical preliminaries will be presented, in particular we will focus on the BV function, the Caccioppoli sets and the coarea formula, which links the total variation with the perimeter of the function level sets (which are Caccioppoli sets). Then, the idea of reduced boundary will be presented and studied, in order to provide a better understanding of the perimeter of Caccioppoli sets and a glimpse on the differentiation of BV functions. This preliminary section will end with the definition and some basic properties of Γ-convergence.

The second chapter is mostly dedicated to the study of the study of the Rudin-Osher-Fatemi problem, in particular: the existence and uniqueness of the solution, finding the associated Euler-Lagrange equation and finding an analytical solution for a simple test case.

The last chapter will open with an idea for the discretization of the considered problem and the proof that it actually converges to it (in the sense of Γ-convergence). Then a modified Arrow-Hurwitz algorithm will be presented with the purpose of implementing a solver for this problem. Finally we will comment some of the images produced in this way and compare the analytical solution with the computational one.

Table of Contents

Chapter 1

Preliminaries

In this chapter we will introduce some mathematical fundamentals for the analytical formulation and study of the thesis core problem. In particular, there will be presented the space of bounded variation functions and the class of Caccioppoli sets, with some of their fundamental properties such as the "*compactness*" in L^1, the approximation through regular functions, the coarea formula and the isoperimetric inequality. In the second section we will focus more on some further results on Cacciopoli set for the definition of inner normal vector and a better understanding of their perimeter. The last section will regard the Γ - convergence, which will be used to prove the validity of a continuum optimization problem discretization.

1.1 Functions of Bounded Variations and Caccioppoli sets

Definitions and basic properties

Let's begin with the definition of total variation.

Definition 1.1.1 (Total variation)**.** Let $\Omega \subset \mathbb{R}^N$ open and $f \in L^1(\Omega)$. We define the *total variation* of f on Ω as:

$$V(f, \Omega) \coloneqq \sup \left\{ \int_\Omega f \operatorname{div}(g) \, \mathrm{d}x \mid g \in C_c^1 \left(\Omega; \mathbb{R}^N\right), \, \|g\|_\infty \leq 1 \right\}.$$

We can extend this definition to a generic Ω (not open) as follows:

$$V(f, \Omega) = \inf \left\{ V(f, A) : \ A \supset \Omega \ \text{ open} \right\}.$$

We can notice that this is a generalization, by duality, of $\int_\Omega |\mathrm{grad}(f)| \, \mathrm{d}x$, that for $N = 1$ represent indeed the distance spanned on the image of f.

Example 1.1.2. *If $f \in C^1(\Omega)$, we have:*

$$\forall g \in C_c^1\left(\Omega; \mathbb{R}^N\right), \quad \int_\Omega f\, div(g)\,\mathrm{d}x = -\int_\Omega grad(f) \cdot g\,\mathrm{d}x.$$

Therefore, extracting the sup over g we get:

$$V(f,\Omega) = \int_\Omega |grad(f)|\,\mathrm{d}x.$$

More generally, this result can be achieved also for $g \in W^{1,1}(\Omega)$.

From this example we also verify that for f constant on Ω, then $V(f,\Omega) = 0$.

Definition 1.1.3 (BV space)**.** We define the space of *functions of bounded variations* on Ω as follows:

$$BV(\Omega) = \{f \in L^1(\Omega) \,|\, V(f,\Omega) < \infty\}.$$

This definition well explain the space's name, however it may be useful to look at it under another, yet equivalent, definition: the functions whose distributional derivative is a Radon measure.

Definition 1.1.4. A function $f \in BV(\Omega)$ if and only if $f \in L^1(\Omega)$ and exists a vector valued finite Radon measure μ such that, $\forall \phi \in C_c^\infty(\Omega)$:

$$\int_\Omega f \partial_{x_i} \phi\,\mathrm{d}x = -\int_\Omega \phi\, d\,\mu_i; \qquad i = 1, \dots, N.$$

We call $\mu = Df$.

This two definitions turn out to be equivalent (see the appendix B).

Observation 1.1.5. We can indeed notice that the following properties are true:

- $V(f,\Omega) \geq 0$, because $g \equiv 0$ is always a competitor for the supremum in 1.1.1.
- $V(c,\Omega) = 0$ for c constant, because for Gauss-Green theorem $\int_\Omega \mathrm{div}(g)\,\mathrm{d}x = 0$ for every $g \in [C_c^1(\Omega)]^n$.
- $\Omega_1 \cap \Omega_2 = \varnothing \implies V(f, \Omega_1 \cup \Omega_2) = V(f,\Omega_1) + V(f,\Omega_2)$.
- $\Omega_1 \subset \Omega_2 \implies V(f,\Omega_1) \leq V(f,\Omega_2)$.

This suggests that $V(f,\cdot)$ is a positive measure. Indeed, generalizing the example 1.1.2, it turned out that $V(f,\Omega) = |Df|(\Omega)$ (see B.0.1 for the definition of $|Df|$).

We can also show the following scaling property:

$$V(f_\lambda, \Omega) = \frac{1}{\lambda^{n-1}} V(f, \lambda\Omega), \tag{1.1}$$

where $f_\lambda(x) = f(\lambda x)$. Indeed, given g as in definition 1.1.3

$$\int_\Omega f(\lambda x) \mathrm{div} g(x)\, \mathrm{d}x = \int_{\lambda\Omega} f(x) \mathrm{div} g\left(\frac{1}{\lambda} x\right) \frac{1}{\lambda^n}\, \mathrm{d}x,$$

but since $\mathrm{div} g_{\frac{1}{\lambda}}(x) = \mathrm{div} g\left(\frac{1}{\lambda}x\right)\frac{1}{\lambda}$, we can conclude

$$\int_\Omega f(\lambda x) \mathrm{div} g(x)\, \mathrm{d}x = \frac{1}{\lambda^{n-1}} \int_{\lambda\Omega} f(x) \mathrm{div} g_{\frac{1}{\lambda}}(x)\, \mathrm{d}x,$$

which leads to the equation (1.1).

From the example 1.1.2 we have that $W^{1,1} \subset BV$, however we can show that the equality doesn't hold.

Example 1.1.6. *Take a relatively compact set $E \subset\subset \Omega$ (with closure included in Ω) and, most importantly, with C^2 boundary. Let χ_E the characteristic function on E. Then*

$$\int_\Omega \chi_E\, \mathrm{d}x = |E|;$$

therefore $\chi_E \in L^1(\Omega)$. However, $D\chi_E$ is non zero and supported only on ∂E, thus it can not be an L^1 function. That is $\chi_E \in L^1(\Omega) \setminus W^{1,1}(\Omega)$. Still we can verify that $\chi_E \in BV(\Omega)$. Take $g \in C_0^1(\Omega, \mathbb{R}^N)$ with $\|g\|_\infty \leq 1$. Applying the Gauss-Green theorem we have:

$$\int_\Omega \chi_E\, div(g)\, \mathrm{d}x = \int_E div(g)\, \mathrm{d}x = \int_{\partial E} g \cdot \nu\, d\mathcal{H}_{n-1}, \tag{1.2}$$

where ν is the outward normal unit vector to ∂E and $\mathcal{H}_{n-1}$ the $(n-1)$-dimensional Hausdorff measure. Since $|g \cdot \nu| \leq 1$,

$$\int_{\partial E} g \cdot \nu\, d\mathcal{H}_{n-1} \leq \mathcal{H}_{n-1}(\partial E).$$

Thus, $V(\chi_E, \Omega) \leq \mathcal{H}_{n-1}(\partial E) < \infty$, that is $\chi_E \in BV(\Omega)$. Furthermore, since the boundary is C^2, we can extend ν to the whole Ω as a $C_0^1(\Omega)$ with $\|\nu\|_\infty \leq 1$.

$$V(\chi_E, \Omega) \geq \int_\Omega \chi_E\, div(\nu)\, \mathrm{d}x = \mathcal{H}_{n-1}(\partial E)$$

$$\implies V(\chi_E, \Omega) = \mathcal{H}_{n-1}(\partial E).$$

More in general for any open A: $V(\chi_E, A) = \mathcal{H}_{n-1}(\partial E \cap A)$.

This example suggests a way to give a more general definition of perimeter, using the total variation.

Definition 1.1.7. Let E be a Borel set and Ω an open set, both in $\mathbb{R}^N$. We define *perimeter* of E in Ω as:

$$P(E, \Omega) \coloneqq V(\chi_E, \Omega).$$

For semplicity we denote $P(E) = P(E, \mathbb{R}^N)$.

We say that E is a *Caccioppoli set* or *set of locally finite perimeter* if, for every bounded Ω, $P(E, \Omega) < +\infty$.

This definition of perimeter, as a measure $V(\chi_E, \Omega) = |D\chi_E|(\Omega)$, is localized on ∂E, as χ_E is constant elsewhere. Clearly, the properties in 1.1.5 still hold for the perimeter, but we also have the following two:

Proposition 1.1.8. *E, F Caccioppoli sets and Ω open, then:*

1. $P(E, \Omega) + P(F, \Omega) \geq P(E \cap F, \Omega) + P(E \cup F, \Omega)$.
2. $|E| = 0 \implies P(E, \Omega) = 0$.
3. $|E \setminus F \cup F \setminus E| = 0 \implies P(E, \Omega) = P(F, \Omega)$.
4. $P(E, \Omega) = P(\mathbb{R}^N \setminus E, \Omega)$.
5. $P(tE, \Omega) = t^{n-1} P\left(E, \frac{1}{t}\Omega\right)$.

Proof. (1): For this part we need a density result we will prove later, see 1.1.13 and 1.1.10. There exist two sequences of smooth functions $\{u_j\}$ and $\{v_j\}$ such that $0 \leq u_j, v_j \leq 1$ and :

$$\begin{cases} u_j \longrightarrow \chi_E & \text{in } L^1(\Omega) \\ \lim\limits_{j \to +\infty} V(u_j, \Omega) = P(E, \Omega) \end{cases} \quad \bigwedge \quad \begin{cases} v_j \longrightarrow \chi_F & \text{in } L^1(\Omega) \\ \lim\limits_{j \to +\infty} V(v_j, \Omega) = P(F, \Omega) \end{cases}.$$

Now $u_j v_j \longrightarrow \chi_{E \cap F}$ in L^1, indeed

$$\|\chi_{E\cap F} - u_j v_j\|_{L^1} = \|\chi_E \chi_F - u_j v_j\|_{L^1} \leq \|\chi_E(\chi_F - v_j)\|_{L^1} + \|(\chi_E - u_j)v_j\|_{L^1} \leq$$

$$\leq \|\chi_F - v_j\|_{L^1} + \|\chi_E - u_j\|_{L^1} \longrightarrow 0.$$

Similarly, $u_j + v_j - u_j v_j \longrightarrow \chi_{E \cup F}$:

$$|\mathrm{grad}(u_j v_j)| + |\mathrm{grad}(u_j + v_j - u_j v_j)| =$$

$$= |u_j \mathrm{grad}(v_j) + v_j \mathrm{grad}(u_j)| + |(1 - v_j)\mathrm{grad}(u_j) + (1 - u_j)\mathrm{grad}(v_j)| \leq$$

$$\leq u_j |\mathrm{grad}(v_j)| + v_j |\mathrm{grad}(u_j)| + (1 - v_j)|\mathrm{grad}(u_j)| + (1 - u_j)|\mathrm{grad}(v_j)| =$$

$$= |\mathrm{grad}(u_j)| + |\mathrm{grad}(v_j)|.$$

In conclusion, by 1.1.10 and that for smooth functions $V(u,\Omega) = \int_\Omega |\mathrm{grad}(u)|\,\mathrm{d}x$:

$$P(E\cup F,\Omega) + P(E\cap F,\Omega) \leq$$

$$\leq \liminf_j \int_\Omega |\mathrm{grad}(u_j + v_j - u_j v_j)|\,\mathrm{d}x + \liminf_j \int_\Omega |\mathrm{grad}(u_j v_j)|\,\mathrm{d}x \leq$$

$$\leq \liminf_j \left\{ \int_\Omega |\mathrm{grad}(u_j + v_j - u_j v_j)|\,\mathrm{d}x + \int_\Omega |\mathrm{grad}(u_j v_j)|\,\mathrm{d}x \right\} \leq$$

$$\leq \liminf_j \left\{ \int_\Omega |\mathrm{grad}(u_j)|\,\mathrm{d}x + \int_\Omega |\mathrm{grad}(v_j)|\,\mathrm{d}x \right\} = P(E,\Omega) + P(F,\Omega).$$

(2): straightforward.
(3): Without loss of generality suppose $\Omega = \mathbb{R}^N$.

$$|E\setminus F| \cup |F\setminus E| = 0 \implies |E\setminus F| = 0 \bigwedge |F\setminus E| = 0$$

$$\implies \int_{E\cup F} \mathrm{div}(g)\,\mathrm{d}x = \int_E \mathrm{div}(g)\,\mathrm{d}x + \int_{E\setminus F} \mathrm{div}(g)\,\mathrm{d}x = \int_E \mathrm{div}(g)\,\mathrm{d}x$$

$$\implies P(F\cup E) = P(E).$$

But similarly $P(E\cup F) = P(F)$.
(4): $\chi_{\mathbb{R}^N\setminus E} = 1 - \chi_E$

$$\implies \int_\Omega \chi_{\mathbb{R}^N\setminus E}\,\mathrm{div}(g)\,\mathrm{d}x = \int_\Omega \mathrm{div}(g)\,\mathrm{d}x - \int_\Omega \chi_E\,\mathrm{div}(g)\,\mathrm{d}x.$$

But since $g \in C^1_c(\Omega)$, from Gauss-Green we have:

$$\int_\Omega \mathrm{div}(g)\,\mathrm{d}x = 0.$$

$$\implies \int_\Omega \chi_{\mathbb{R}^N\setminus E}\,\mathrm{div}(g)\,\mathrm{d}x = -\int_\Omega \chi_E\,\mathrm{div}(g)\,\mathrm{d}x.$$

Thus, we conclude extracting the supremum.
(5): Since $\chi_{tE}(x) = \chi_E\left(\frac{1}{t}x\right)$, from (1.1) we conclude

$$P(tE,\Omega) = V(\chi_{tE},\Omega) = t^{n-1} V\left(\chi_E, \frac{1}{t}\Omega\right) = t^{n-1} P\left(E, \frac{1}{t}\Omega\right).$$

□

Observation 1.1.9. From definition 1.1.4, given a Caccioppoli set E, $D\chi_E$ is a vector valued finite Radon measure. Thus, for any C^1 function g we have:

$$\int_E \operatorname{div}(g)\,\mathrm{d}x = -\int_{\mathbb{R}^n} g \cdot \mathrm{d}D\chi_E.$$

As $D\chi_E$ has support only in ∂E, we can see this equality as a generalized Gauss-Green formula.

We will now present now some relevant results about functions in BV and Caccioppoli sets.

Theorem 1.1.10 (Semicontinuity)**.** *Let $\Omega \subset \mathbb{R}^N$ open and $\{f_j\}_{j\in\mathbb{N}} \subset BV(\Omega)$ a sequence converging weakly $\star$ to f in $L^1_{LOC}(\Omega)$. Then*

$$V(f,\Omega) \leq \liminf_{j\to+\infty} V(f_j,\Omega).$$

Proof. Let $g \in C^1_c(\Omega)$ s.t. $\|g\|_\infty \leq 1$:

$$\int_\Omega f \operatorname{div}(g)\mathrm{d}x = \lim_{j\to+\infty} \int_\Omega f_j \operatorname{div}(g)\mathrm{d}x = \liminf_{j\to+\infty} \int_\Omega f_j \operatorname{div}(g)\mathrm{d}x \leq \liminf_{j\to+\infty} V(f,\Omega).$$

Then extracting the supremum over g we get the thesis.

□

We can observe that this is a quite general property, telling us that the total variation is lower semicontinuous with respect to any L^p convergence, with $p \in [1,\infty)$, strong or weak, since all of them imply the weak $\star$ one in L^1_{LOC}.

We can set on BV the norm

$$\|f\|_{BV} = \|f\|_{L^1} + V(f,\Omega).$$

Proposition 1.1.11. *$(BV, \|\cdot\|_{BV})$ is a Banach space.*

Proof. Let's verify $\|\cdot\|_{BV}$ is a norm:

- $\|f\|_{BV} \geq 0 \wedge \|f\|_{BV} = 0 \iff f = 0$,
- $V(\lambda f,\Omega) = \sup_g \lambda \int_\Omega f \operatorname{div}(g)\,\mathrm{d}x = |\lambda| \sup_g \int_\Omega f \operatorname{div}(g)\,\mathrm{d}x = |\lambda| V(f,\Omega)$. This leads to: $\|\lambda f\|_{BV} = |\lambda| \|f\|_{BV}$.
- $V(f_1+f_2,\Omega) = \sup_g \{\int_\Omega (f_1+f_2)\operatorname{div}(g)\,\mathrm{d}x\} \leq$

$$\leq \sup_g \left\{\int_\Omega f_1 \operatorname{div}(g)\,\mathrm{d}x\right\} + \sup_g \left\{\int_\Omega f_2 \operatorname{div}(g)\,\mathrm{d}x\right\} = V(f_1,\Omega) + V(f_2,\Omega).$$

This gives the triangular inequality.

About the completeness, a Cauchy sequence on this norm is also Cauchy in L^1, that is $f_j \to f$ in L^1. But the sequence $\{f_j\}$ is bounded in BV norm, so their total variations are, therefore (from 1.1.10) $V(f,\Omega) < +\infty$, that is $f \in BV(\Omega)$. f_j is a Cauchy sequence, then

$$\forall \epsilon > 0, \exists N \in \mathbb{N}\, s.t.\, \forall j,k > N \quad \|f_j - f_k\|_{BV} < \epsilon.$$

$f_j - f_k \to f_j - f$ in L^1, then for $j > N$ by 1.1.10:

$$V(f_j - f, \Omega) \leq \liminf_k V(f_j - f_k, \Omega) \leq \liminf_k \|f_j - f_k\|_{BV} \leq \epsilon.$$

$$\implies f_j \longrightarrow f \text{ in } BV(\Omega).$$

□

Observation 1.1.12. In the last two points of the proof that $\|\cdot\|_{BV}$ is a norm, we have also shown that $V(\cdot,\Omega)$ is a convex functional.

We can notice, from example 1.1.2, that for functions in $W^{1,1}$ the BV norm is equivalent to the one in $W^{1,1}$. This mean that we have no hopes to approximate a generic BV function with a sequence in $W^{1,1}$, since this Sobolev space is already closed, and much less with smooth functions. Nevertheless, we have the following "density" result.

Theorem 1.1.13. *For every $f \in BV(\Omega)$ there is a sequence $\{f_j\} \subset C^\infty(\Omega)$ such that:*

$$\begin{cases} f_j \longrightarrow f \quad in\, L^1(\Omega); \\ \lim_{j\to+\infty} V(f_j,\Omega) = V(f,\Omega). \end{cases}$$

Proof. For $\varepsilon > 0$, we define

$$\Omega_k = \left\{ x \in \Omega : \ \text{dist}(x,\Omega) < \frac{1}{k+m} \right\},$$

for $k = 0,1,2,\dots$ and $m \in \mathbb{N}$ is big enough to have $V(f, \Omega \setminus \Omega_0) < \varepsilon$. We define the sets A_i such that:

$$A_1 = \Omega_2 \bigwedge A_i = \Omega_{i+1} \setminus \bar{\Omega}_{i-1}, \text{ for } i > 1.$$

We then define ϕ_i a partition of the unity related to the covering A_i, that is:

- $\forall i \in \mathbb{N},\ \phi_i \in C_c^\infty(A_i)$;
- $\forall i \in \mathbb{N},\ 0 \leq \phi \leq 1$;
- $\sum_{i=1}^\infty \phi_i = 1$;

Given η_ε a family of mollifiers, for every i we can find an ε_i that satisfy:

1. $\operatorname{supp}\{\eta_{\varepsilon_i} * (f\phi_i)\} \subset \Omega_{i+2} \setminus \bar{\Omega}_{i-2}$;
2. $\int |\eta_{\varepsilon_i} * (f\phi_i) - f\phi_i| \, \mathrm{d}x < \varepsilon 2^{-i}$;
3. $\int |\eta_{\varepsilon_i} * (fD\phi_i) - fD\phi_i| \, \mathrm{d}x < \varepsilon 2^{-i}$;

We can finally define the approximating function:

$$f_\varepsilon = \sum_{i=1}^{\infty} \eta_{\varepsilon_i} * (f\phi_i).$$

From 1 we have that this sum is locally finite, therefore $f_\varepsilon \in C^\infty(\Omega)$ and also

$$\int_\Omega |f_\varepsilon - f| \, \mathrm{d}x = \int_\Omega \left| \sum_{i=1}^{\infty} \eta_{\varepsilon_i} * (f\phi_i) - f\phi_i \right| \mathrm{d}x \leq \sum_{i=1}^{\infty} \int_\Omega |\eta_{\varepsilon_i} * (f\phi_i) - f\phi_i| \, \mathrm{d}x < \varepsilon.$$

$$\Longrightarrow \lim_{\varepsilon \to 0} f_\varepsilon = f \quad \text{in } L^1.$$

It remains to show that the variation is converging. From 1.1.10 we already have $V(f,\Omega) \leq \liminf_{\varepsilon \to 0} V(f_\varepsilon, \Omega)$. For the limsup inequality consider a function g as in the definition of total variation, given that η_{ε_i} has radial symmetry, we get:

$$\int_\Omega f_\varepsilon \operatorname{div}(g) \, \mathrm{d}x = \sum_{i=1}^{\infty} \int_\Omega (f\phi_i) * \eta_{\varepsilon_i} \operatorname{div}(g) \, \mathrm{d}x = \sum_{i=1}^{\infty} \int_\Omega f\phi_i \operatorname{div}(\eta_{\varepsilon_i} * g) \, \mathrm{d}x;$$

$$\phi_i \operatorname{div}(\eta_{\varepsilon_i} * g) = \operatorname{div}(\phi_i(\eta_{\varepsilon_i} * g)) - \operatorname{grad}(\phi_i) \cdot (\eta_{\varepsilon_i} * g);$$

$$\Longrightarrow \int_\Omega f_\varepsilon \operatorname{div}(g) \, \mathrm{d}x = \sum_{i=1}^{\infty} \int_\Omega f \operatorname{div}(\phi_i(\eta_{\varepsilon_i} * g)) \, \mathrm{d}x - \sum_{i=1}^{\infty} \int_\Omega (f \operatorname{grad}(\phi_i)) * \eta_{\varepsilon_i} \cdot g \, \mathrm{d}x.$$

To deal with the first integral we notice that $|\phi_i(\eta_{\varepsilon_i} * g)| \leq 1$, therefore

$$\int_\Omega f \operatorname{div}(\phi_1(\eta_{\varepsilon_1} * g)) \, \mathrm{d}x \leq V(f, \Omega),$$

while for $i > 1$, since the sum locally involves at most 3 addends, $|\sum_{i=2}^{\infty} \phi_i(\eta_{\varepsilon_i} * g)| \leq 3$. Furthermore it is supported just in $\Omega - \Omega_0$, then we have:

$$\sum_{i=2}^{\infty} \int_\Omega f \operatorname{div}(\phi_i(\eta_{\varepsilon_i} * g)) \, \mathrm{d}x \leq 3V(f, \Omega - \Omega_0) < 3\varepsilon.$$

For the second integral we can notice that $\sum_{i=1}^{\infty} \operatorname{grad}(\phi_i) = 0$, then we can rewrite it as:

$$\sum_{i=1}^{\infty} \langle g, (f \operatorname{grad}(\phi_i)) * \eta_{\varepsilon_i} - f \operatorname{grad}(\phi_i) \rangle < \sum_{i=1}^{\infty} \varepsilon 2^{-i} = \varepsilon.$$

$$\implies \int_\Omega f_\varepsilon \mathrm{div}(g)\,\mathrm{d}x < V(f,\Omega) + 4\varepsilon$$
$$\implies \limsup_{\varepsilon\to 0} V(f_\varepsilon,\Omega) \le V(f,\Omega).$$

□

This result result allow us to prove also that the embedding of BV in L^1 is compact, under some constraints on the domain Ω.

Theorem 1.1.14 (Compactness)**.** *If Ω is an open set in $\mathbb{R}^N$, bounded and with Lipschitz boundary, then $BV(\Omega)$ is compactly embedded in $L^1(\Omega)$.*

Proof. Let $\{f_j\}$ a bounded sequence in $BV(\Omega)$, then, by 1.1.13, for each j I can find a smooth function $\tilde{f}_j$ s.t.:

$$\begin{cases} \left\| f_j - \tilde{f}_j \right\|_{L^1(\Omega)} < \dfrac{1}{j}; \\ \left\| \tilde{f}_j \right\|_{BV} < \| f_j \|_{BV} + \epsilon. \end{cases}$$

This means that $\{\tilde{f}_j\}$ is a bounded sequence in the BV norm, but for smooth functions this coincide with the one in $W^{1,1}$. On this domain, $W^{1,1}$ is compactly embedded in $L^1(\Omega)$, therefore there exists a subsequence $\{\tilde{f}_{j_k}\}$ converging in L^1 at a function f. But $\{f_{j_k}\}$ has the same L^1-limit as $\{\tilde{f}_{j_k}\}$ and from 1.1.10 $f \in BV(\Omega)$. That is, there exists a subsequence converging in L^1 to a function in BV. □

Leibniz rule and relevant inequalities

Concerning the differentiation, we have a sort of Leibniz rule.

Proposition 1.1.15. *Let $f \in BV(\Omega)$ and ψ a locally Lipschitz function, then $\psi f \in BV(\Omega)$ and*
$$D(\psi f) = \psi Df + \nabla\psi f \mathcal{L}_n.$$

Proof. If $\psi \in C^\infty$, $D(\psi f) = \psi Df + \nabla \psi f$ as distribution. Furthermore, since $f \in L^1(\Omega)$, f as a distribution is $f\mathcal{L}_n$ as a Radon measure. Hence, in this particular case the differentiation rule is true. Now, for $\psi \in Lip_{LOC}$, take a test function ϕ with support on a compact K. Then, there is a sequence $\{\psi_j^K\}$ equi-Lipschitz converging to ψ uniformly on K. As $\{\psi_j^K\}$ are equi-Lipschitz, $\{\nabla\psi_j^K\}$ are equibounded (almost everywhere). Furthermore, this sequence is given by the convolution of ψ with mollifiers, then $\nabla\psi_j^K = \nabla\psi * \rho_j \to \nabla\psi$ in L^1. Thus, we can chose a subswquence for which we have $\nabla\psi_j^K \to \nabla\psi$ almost everywhere. We want to prove

$$\langle D(\psi f), \phi\rangle = -\langle \psi f, \nabla\phi\rangle = -\lim_{j\to\infty} \langle \psi_j^K f, \nabla\phi\rangle =$$

$$= \lim_{j\to\infty} \left[\langle \nabla\psi_j^K f, \phi\rangle + \langle \psi_j^K Df, \phi\rangle \right] = \langle \nabla\psi f, \phi\rangle + \langle \psi Df, \phi\rangle,$$

that is, we just need to prove the limits. As $\{\psi_j^K\}$ is uniformly bounded, for dominated convergence we have the limits:

$$\lim_{j\to\infty} \langle \psi_j^K f, \nabla\phi\rangle = \langle \psi f, \nabla\phi\rangle,$$

$$\lim_{j\to\infty} \langle \psi_j^K Df, \phi\rangle = \lim_{j\to\infty} \int_K \psi_j^K \phi \, \mathrm{d}Df = \int_K \psi\phi \, \mathrm{d}Df = \langle \psi Df, \phi\rangle.$$

Now, as $\{\nabla\psi_j^K\}$ are equibounded and converge almost everywhere to $\nabla\psi$, we can apply again the dominated convergence:

$$\lim_{j\to\infty} \langle \nabla\psi_j^K f, \phi\rangle = \langle \nabla\psi f, \phi\rangle.$$

In this way we showed that

$$D(\psi f) = \psi Df + \nabla\psi f$$

as distributions, but since $\nabla\psi f$ behaves as $\nabla\psi f \mathcal{L}_n$, we can say that $D(\psi f)$ is a Radon measure and the derivation rule holds □

Corollary 1.1.16. *Let $u \in BV(\Omega)$ with Ω open, $\forall x \in \Omega$ and $\forall \rho \in (0, \rho_0)$, with $\rho_0 = dist(x, \partial\Omega)$:*

$$V(u\chi_{B_\rho}, \mathbb{R}^n) \le V(u, \overline{B}_\rho) + \frac{d}{d\rho_+}\left(\int_{B_\rho} |u(x)| \, \mathrm{d}x \right),$$

where $B_\rho = B(x, \rho)$ and

$$\frac{d}{d\rho_+}\phi(\rho) = \liminf_{\sigma\to 0^+} \frac{\phi(\rho+\sigma) - \phi(\rho)}{\sigma}.$$

Proof. Consider a Lipschitz-regularized version of χ_{B_ρ}:

$$\gamma_\sigma(x) = \begin{cases} 1 & |x| \in [0, \rho) \\ 1 - \dfrac{|x| - \rho}{\sigma} & |x| \in [\rho, \rho+\sigma) \\ 0 & |x| \in [\rho+\sigma, +\infty). \end{cases}$$

Then by the Leibniz rule:

$$D(u\gamma_\sigma)(A) = \int_A \gamma_\sigma(x) \, \mathrm{d}Du + \int_{A\cap B_{\rho+\sigma}\setminus B_\rho} -\frac{1}{\sigma}\frac{x}{|x|} u(x) \, \mathrm{d}x,$$

for a given set A. Then by definition B.0.1 of total variation of a measure and B.0.3 we have

$$|D(u\gamma_\sigma)|(\mathbb{R}^n) \leq |\gamma_\sigma||Du|(\mathbb{R}^N) + \frac{1}{\sigma}\int_{B_{\rho+\sigma}\setminus B_\rho}|u(x)|\,\mathrm{d}x \leq$$

$$\leq |Du|(B_{\rho+\sigma}) + \frac{1}{\sigma}\int_{B_{\rho+\sigma}\setminus B_\rho}|u(x)|\,\mathrm{d}x.$$

$u\gamma_\sigma$ converge $\mathcal{L}_n$-a.e. to $u\chi_{B_\rho}$, then, by dominated convergence, also in L^1. Hence, for semicontinuity and $|Du|$ continuity from above, we get

$$\left|D(u\chi_{B_\rho})\right|(\mathbb{R}^n) \leq \liminf_{\sigma\to 0^+}|D(u\gamma_\sigma)|(\mathbb{R}^n) \leq |Du|(\overline{B}_\rho) + \liminf_{\sigma\to 0^+}\frac{1}{\sigma}\int_{B_{\rho+\sigma}\setminus B_\rho}|u(x)|\,\mathrm{d}x.$$

□

Observation 1.1.17. If we take $u = \chi_E$, this inequality become

$$P(E\cap B_\rho) \leq P(E,\overline{B}_\rho) + \frac{d}{d\rho_+}|E\cap B_\rho|.$$

In BV we can find a generalization of some Sobolev inequalities, here we present the Poincaré and Poincaré-Wirtinger inequalities.

Theorem 1.1.18.
- *(Poincaré) Let $f \in BV(\mathbb{R}^n)$, with support bounded in at least one direction, then there exists $c_n > 0$ such that*

$$\|f\|_{L^1(\mathbb{R}^n)} \leq c_n V(f,\mathbb{R}^n).$$

- *(Poincaré-Wirtinger) Let Ω bounded and $f \in BV(\Omega)$. Call $\overline{f}$ the integral mean of f over Ω. Then there exists $c_n > 0$ such that*

$$\left\|f - \overline{f}\right\|_{L^1(\Omega)} \leq c_n V(f,\Omega).$$

Proof. These inequalities for sure are true in $W^{1,1}(\Omega)$ (in the first point $\Omega = \mathbb{R}^n$). But by 1.1.13 there exists a sequence $\{f_j\}$ of smooth functions converging in L^1 to f and $V(f_j,\Omega)$ is converging to $V(f,\Omega)$. Therefore the first inequality comes immediately:

$$\|f\|_{L^1(\mathbb{R}^n)} \leq \liminf_{j\to\infty}\|f_j\|_{L^1(\mathbb{R}^n)} \leq \liminf_{j\to\infty} c_n V(f_j,\mathbb{R}^n) = c_n V(f,\mathbb{R}^n).$$

For the second inequality we can argue at the same way, just notice that $\overline{f}_j \to \overline{f}$. □

Observation 1.1.19. We have to notice that as $L^{\frac{n}{n-1}}$ is embedded in $W^{1,1}$, also BV is and the generalized Poincaré inequalities can be similarly proven with the norm in $L^{\frac{n}{n-1}}$ in place of the L^1. Indeed, $\{f_j\}$ is bounded in $W^{1,1}$ and consequently also $L^{\frac{n}{n-1}}$, then there is a subsequence converging weakly to f in $L^{\frac{n}{n-1}}$. As the norms are lower semicontinous with respect to the weak topology, the inequalities in the proof works as well. Then we also have:

- $\|f\|_{L^{\frac{n}{n-1}}(\Omega)} \leq c_n \|f\|_{BV(\Omega)}$.
- $\|f\|_{L^{\frac{n}{n-1}}(\mathbb{R}^n)} \leq c_n V(f, \mathbb{R}^n)$, for f compactly supported.
- $\left\|f - \overline{f}\right\|_{L^{\frac{n}{n-1}}(\Omega)} \leq c_n V(f, \Omega)$, with Ω bounded.

Corollary 1.1.20 (Isoperimetric inequalities)**.** *Let E be a Caccioppoli set. Then:*

1. *E bounded* $\implies |E|^{\frac{n-1}{n}} \leq c_n P(E)$.
2. *B bounded* $\implies \min\{|E \cap A|, |(\mathbb{R}^n \setminus E) \cap B|\}^{\frac{n-1}{n}} \leq c_n P(E, A)$.

Proof. Both this results comes plugging $f = \chi_E$ in the inequalities from the observation above. The former is straightforward. For the latter,

$$\overline{\chi}_E = \frac{1}{|A|} \int_A \chi_E \,\mathrm{d}x = \frac{|A \cap E|}{|A|}$$

$$\int_A |\chi_E - \overline{\chi}_E|^{\frac{n}{n-1}} \,\mathrm{d}x = \left(1 - \frac{|A \cap E|}{|A|}\right)^{\frac{n}{n-1}} |A \cap E| + \left(\frac{|A \cap E|}{|A|}\right)^{\frac{n}{n-1}} |A \cap (\mathbb{R}^n \setminus E)|,$$

but, since $|A \cap (\mathbb{R}^n \setminus E)| = |A| - |A \cap E|$, we have:

$$\int_A |\chi_E - \overline{\chi}_E|^{\frac{n}{n-1}} \,\mathrm{d}x \geq \min\{|E \cap A|, |(\mathbb{R}^n \setminus E) \cap B|\} \frac{|E \cap A|^{\frac{n}{n-1}} + |(\mathbb{R}^n \setminus E)|^{\frac{n}{n-1}}}{|A|^{\frac{n}{n-1}}}$$

We recall that $\forall a, b > 0 \land p \geq 1,\ a^p + b^b \geq 2^{1-p}(a+b)^p$, which implies

$$\frac{|E \cap A|^{\frac{n}{n-1}} + |(\mathbb{R}^n \setminus E)|^{\frac{n}{n-1}}}{|A|^{\frac{n}{n-1}}} \geq 2^{-\frac{1}{n-1}}.$$

Now extracting the $\frac{n-1}{n}$-th root and applying the Poincaré-Wirtinger inequality we conclude. □

This last result is very important as provide a way to bound the area of a figure with its perimeter. We also point out that the inequalities are true also without the $\frac{n-1}{n}$ exponent, however in the applications the exponent is often necessary.

Coarea formula and some consequences

Another relevant result in BV is the so called *coarea formula*, which relates the total variation with the perimeter of sublevel sets.

Theorem 1.1.21 (Coarea formula)**.** *Let $f \in BV(\Omega)$, with Ω open, and define*

$$F_t = \{x \in \Omega : \ f(x) < t\}.$$

Then, for each $E \subset\subset \Omega$

$$V(f, E) = \int_{-\infty}^{+\infty} P(F_t, E)\,\mathrm{d}t.$$

Proof. Consider as domain the whole Ω. Let's first split f in positive and negative part f_+ and f_-. We notice that:

$$f_+(x) = \int_0^{f_+(x)} \mathrm{d}t = \int_0^{+\infty} \chi_{\{t \le f(x)\}}(t)\,\mathrm{d}t = \int_0^{+\infty} 1 - \chi_{F_t}(x)\,\mathrm{d}t;$$

$$\implies \int_\Omega f_+ \operatorname{div}(g)\,\mathrm{d}x = \int_0^{+\infty} \int_\Omega (1 - \chi_{F_t}(x)) \operatorname{div}(g)\,\mathrm{d}x\mathrm{d}t \le \int_0^{+\infty} P(F_t, \Omega)\,\mathrm{d}t.$$

And similarly for f_-:

$$\int_\Omega -(f_-) \operatorname{div}(g)\,\mathrm{d}x \le \int_{-\infty}^0 P(F_t, \Omega)\,\mathrm{d}t.$$

Then adding these inequalities we conclude:

$$\int_\Omega f \operatorname{div}(g)\,\mathrm{d}x \le \int_{-\infty}^{+\infty} P(F_t, \Omega)\,\mathrm{d}t; \qquad \implies \qquad V(f, \Omega) \le \int_{-\infty}^{+\infty} P(F_t, \Omega)\,\mathrm{d}t.$$

Let's at first show that the equality is true for $f \in BV(\Omega)$ continuous and piecewise linear. This means that there exist a finite partition $\{\Omega_i\}$ of Ω s.t.

$$f(x) = c_i \cdot x + b_i, \qquad \text{for } x \in \Omega_i.$$

therefore, f is in $W^{1,1}$, which implies:

$$V(f, \Omega) = \int_\Omega |\mathrm{grad}(f)|\,\mathrm{d}x = \sum_i |c_i||\Omega_i|.$$

On the other hand, $P(F_t, \Omega_i) = \mathcal{H}_{n-1}(\{x \in \Omega_i : \ c_i \cdot x + b_i = t\})$,

$$\implies \int_{-\infty}^{+\infty} P(F_t, \Omega_i)\,\mathrm{d}t = \int_{-\infty}^{+\infty} \mathcal{H}_{n-1}(\{x \in \Omega_i : \ c_i \cdot x = t\})\,\mathrm{d}t,$$

where we applied the shift $t - b_i$ in the integral. We now do a change of basis $(\xi_1, \dots, \xi_N) = R(e_1, \dots, e_N)$, where (e_j) is the standard base, so that $\xi_1 = \frac{c_i}{|c_i|}$ while $c_i \cdot \xi_j = 0$ for $j > 1$. We call $y = Rx$, then $c_i \cdot x = |c_i| y_1$. Hence, changing variable $|c_i|s = t$, we finally get:

$$\int_{-\infty}^{+\infty} \mathcal{H}_{n-1}(\{x \in \Omega_i :\ c_i \cdot x = t\})\,\mathrm{d}t = |c_i| \int_{-\infty}^{+\infty} \mathcal{H}_{n-1}(\{y \in R\Omega_i :\ y_1 = s\})\,ds =$$

$$= |c_i| \int_{\mathbb{R}^N} \chi_{\Omega_i}\,dy = |c_i||\Omega|.$$

$$\implies V(f;\Omega) = \sum_i \int_{-\infty}^{+\infty} P(F_t, \Omega_i)\,\mathrm{d}t = \int_{-\infty}^{+\infty} P(F_t, \Omega)\,\mathrm{d}t.$$

Now we would like to extend this result at first to C^∞ functions then to the whole BV. We will show that if we have a sequence $\{f_j\} \subset BV(\Omega)$ for which the coarea formula is true and this is sequence satisfy:

$$\begin{cases} f_j \longrightarrow f & \text{in } L^1(\Omega), \\ \lim\limits_{j \to +\infty} V(f_j, \Omega) = V(f, \Omega), \end{cases} \tag{1.3}$$

then the coarea formula is true also for f. Indeed, let's call $F_{tj} \coloneqq \{x \in \Omega :\ f_j(x) < t\}$, then we have:

$$|f_j - f| = \int_0^{|f_j - f|} \mathrm{d}t =$$

$$= \begin{cases} \displaystyle\int_0^{f_j - f} \mathrm{d}t & \text{for } x :\ f_j(x) > f(x) \\ \displaystyle\int_0^{f - f_j} \mathrm{d}t & \text{for } x :\ f(x) > f_j(x) \end{cases} = \begin{cases} \displaystyle\int_f^{f_j} \mathrm{d}t & \text{for } x :\ f_j(x) > f(x) \\ \displaystyle\int_{f_j}^{f} \mathrm{d}t & \text{for } x :\ f(x) > f_j(x) \end{cases} =$$

$$= \int_{-\infty}^{+\infty} \chi_{F_t \setminus F_{tj}} + \chi_{F_{tj} \setminus F_t}\,\mathrm{d}t = \int_{-\infty}^{+\infty} \left|\chi_{F_t} - \chi_{F_{tj}}\right| \mathrm{d}t.$$

$$\implies \|f_j - f\|_{L^1} = \int_{-\infty}^{+\infty} \left\|\chi_{F_t} - \chi_{F_{tj}}\right\|_{L1} \mathrm{d}t.$$

Thus, we can extract a subsequence $\chi_{F_{tj}}$ converging in $L^1(\Omega)$ to χ_{F_t} for almost every t. Now using the coarea formula for f_j, the Fatou's lemma and the theorem 1.1.10, we have:

$$V(f, \Omega) = \lim_{j \to +\infty} V(f_j, \Omega) = \lim_{j \to +\infty} \int_{-\infty}^{+\infty} P(F_{tj}, \Omega)\,\mathrm{d}t \geq$$

$$\geq \int_{-\infty}^{+\infty} \liminf_{j \to +\infty} P(F_{tj}, \Omega)\,\mathrm{d}t \geq \int_{-\infty}^{+\infty} P(F_t, \Omega)\,\mathrm{d}t$$

$$\Longrightarrow V(f,\Omega) = \int_{-\infty}^{+\infty} P(F_t,\Omega)\,\mathrm{d}t.$$

Now we can approximate in $W^{1,1}(\Omega)$ compactly supported smooth functions with piecewise continuous linear functions (with linear interpolation on progressively finer simplex meshes), therefore 1.3 are satisfied, and so is the coarea formula. But then we can generalize by density to every function in $W^{1,1}(\bar{\Omega})$. Hence, the coarea formula is true for smooth functions in $BV(\Omega)$. At the same way, from 1.1.13, we can conclude that the coarea formula is true on the whole $BV(\Omega)$.

All this just to show the formula for $E = \Omega$, but then we can quickly see that it holds for every open $E \subset \Omega$. Consider now a closed E, then

$$V(f,E) = \inf_{A \supset E\, open} V(f,A) = \lim_j V(f,A_j),$$

for a given sequence of open sets A_j, such that $E \subset A_j \subsetneq \Omega$. We can take this sequence to be such that $A_{j+1} \subsetneq A_j$ and $E = \bigcap_j A_j$. Indeed, if the first property is not true, we can just take the sequence defined as $A'_1 = A_1$ and $A'_{j+1} = A_{j+1} \cap A'_j$. Instead, if the second property does not holds, then we must have $A \coloneqq \bigcap_j A_j \supset E$. We can take a sequence of points $\{x_j\}$ that is filling densely $A \setminus E$ and we can define a new sequence of open sets A'_j such that $E \subset A'_j \subset A$ and $x_j \notin A_j$, in this way $\bigcap_j A'_j = E$. Now, let's apply the coarea formula for the open A_j

$$V(f,E) = \lim_j \int_{-\infty}^{+\infty} P(F_t,A_j)\,\mathrm{d}t.$$

Because $P(F_t,A_j) \leq P(F_t,\Omega)$ for almost every t, we can take the limit inside for dominated convergence. Then, using the measures' continuity from above we have:

$$V(f,E) = \int_{-\infty}^{+\infty} \lim_j P(F_t,A_j)\,\mathrm{d}t = \int_{-\infty}^{+\infty} P(F_t,E)\,\mathrm{d}t.$$

□

The coarea formula can be used to relate volumes with perimeters. An example of this follows.

Example 1.1.22. *Take the function $f(x) = |x - x_0|$, for some fixed x_0, and an open bounded domain Ω. Therefore, $F_t = \{f < t\} = B(x_0,t)$ and $f \in W^{1,1}(\Omega)$ with $\nabla f(x) = \frac{x}{|x|}$:*

$$\int_0^{+\infty} P(B(x_0,t),\Omega)\,\mathrm{d}t = V(f,\Omega) = \int_\Omega |\nabla f|\,\mathrm{d}x = |\Omega|.$$

Using the results in 1.1.6 and that Ω is bounded, we further get:

$$|\Omega| = \int_0^R \mathcal{H}_{n-1}(\partial B(x_0,t) \cap \Omega)\,\mathrm{d}t.$$

Observation 1.1.23. From the previous example we notice that if we take a set $E \subset\subset \Omega$, we have:

$$\int_{\mathbb{R}} P(B(x_0,t),\partial E)\,\mathrm{d}t = |\partial E| = 0$$
$$\implies P(B(x_0,t),\partial E) = 0 \quad a.e.\, t.$$

From the Radon measure point of view, the coarea formula is written as:

$$|Df|(\Omega) = \int_{-\infty}^{+\infty} |D\chi_{F_t}|(\Omega)\,\mathrm{d}t.$$

This suggest that there may be a way to move an integral from $\mathrm{d}|Df|$ to $\mathrm{d}|D\chi_{F_t}|\mathrm{d}t$.

Proposition 1.1.24. *Let f be a measurable function and $u \in BV(\Omega)$, then for every borel set $A \subseteq \Omega$*

$$f|Du|(A) = \int_{-\infty}^{+\infty} f\Big|D\chi_{\{u<t\}}\Big|(A)\,\mathrm{d}t,$$

where $\{u < t\} = \{x \in \Omega :\ u(x) < t\}$.

Proof. Let's prove it for simple function, then positive functions, and finally generalize. Take $f = \sum_{i=1}^m c_i\chi_{A_i}$, for some real coefficients c_i and borel sets A_i.

$$f|Du|(A) = \sum_i c_i|Du|(A_i) = \sum_i c_i \int_{-\infty}^{+\infty} \Big|D\chi_{\{u<t\}}\Big|(A_i)\,\mathrm{d}t =$$
$$= \int_{-\infty}^{+\infty} \sum_i c_i\Big|D\chi_{\{u<t\}}\Big|(A_i)\,\mathrm{d}t = \int_{-\infty}^{+\infty} f\Big|D\chi_{\{u<t\}}\Big|(A)\,\mathrm{d}t.$$

For $f \geq 0$, we can take a sequence of simple functions $\{\phi_k\}$ monotonically converging to f, therefore by monotone convergence we have:

$$f|Du|(A) = \lim_k \int_A \phi_k d|Du| = \lim_k \int_{-\infty}^{+\infty} \int_A \phi_k d\Big|D\chi_{\{u<t\}}\Big|\,\mathrm{d}t =$$
$$= \int_{-\infty}^{+\infty} f\Big|D\chi_{\{u<t\}}\Big|(A)\,\mathrm{d}t.$$

From this the last generalization is straightforward. □

With the coarea formula we can show that the inequality of 1.1.8(1) can be extended to the total variation on open sets.

Proposition 1.1.25. *Let Ω be an open set , then for any $u, v \in BV(\Omega)$*

$$V(\max\{u,v\},\Omega) + V(\min\{u,v\},\Omega) \leq V(u,\Omega) + V(v,\Omega).$$

Proof. Let's call $E_t = \{u < t\}$ and $F_t = \{v < t\}$. Then we notice that

$$E_t \cap F_t = \{u < t \,\wedge\, v < t\} = \{\max\{u, v\} < t\},$$

$$E_t \cup F_t = \{u < t \,\vee\, v < t\} = \{\min\{u, v\} < t\}.$$

Then from the coarea formula and 1.1.8(1) we conclude:

$$V(\max\{u, v\}, \Omega) + V(\min\{u, v\}, \Omega) =$$

$$= \int_{\mathbb{R}} P(\{\max\{u, v\} < t\}, \Omega) + P(\{\min\{u, v\} < t\}, \Omega)\,\mathrm{d}t =$$

$$= \int_{\mathbb{R}} P(E_t \cap F_t, \Omega) + P(E_t \cup F_t, \Omega)\,\mathrm{d}t \leq$$

$$\leq \int_{\mathbb{R}} P(E_t, \Omega) + P(F_t, \Omega)\,\mathrm{d}t = V(u, \Omega) + V(v, \Omega).$$

□

This inequality allows a relevant property of the total variation, that it decrease by truncation.

Corollary 1.1.26. *Let m and M real numbers and $u \in BV(\Omega)$. We define $u \wedge M \coloneqq \min\{u, M\}$ and $u \vee m \coloneqq \max\{u, m\}$, then*

$$V(m \vee (u \wedge M), \Omega) \leq V(u, \Omega).$$

Proof. Given the inequality (1.1.25) and that V is 0 on constants, then:

$$V(m \vee (u \wedge M), \Omega) \leq V(m, \Omega) + V(u \wedge M, \Omega) \leq V(u, \Omega) + V(M, \Omega) = V(u, \Omega).$$

□

Another consequence of the coarea formula is the approximation of Caccioppoli sets with smooth sets. At first we need the following lemma.

Lemma 1.1.27. *Let E be a Caccioppoli set and f_ϵ a mollification of χ_E. For $0 < t < 1$ we define:*

$$E_\epsilon \coloneqq \left\{x \in \mathbb{R}^N : \, f_\epsilon(x) > t\right\}.$$

Then

$$\|\chi_{E_\epsilon} - \chi_E\|_{L^1} \leq \frac{1}{\min(1-t, t)} \|f_\epsilon - \chi_E\|_{L^1}.$$

Proof.

$$\int |f_\epsilon - \chi_E| \,\mathrm{d}x \geq \int_{E\setminus E_\epsilon} |f_\epsilon - \chi_E| \,\mathrm{d}x + \int_{E_\epsilon \setminus E} |f_\epsilon - \chi_E| \,\mathrm{d}x \geq$$
$$\geq t|E_\epsilon \setminus E| + (1-t)|E \setminus E_\epsilon| \geq \min(t,1-t)(|E_\epsilon \setminus E| + |E \setminus E_\epsilon|) =$$
$$= \min(t,1-t) \int |\chi_{E_\epsilon} - \chi_E| \,\mathrm{d}x.$$

□

Theorem 1.1.28. *A bounded Caccioppoli set E can be aproximated by a sequence of sets $\{E_j\}$ with C^∞ boundary as follows:*

$$\begin{cases} \chi_{E_j} \longrightarrow \chi_E \quad in\, L^1(\Omega); \\ \lim\limits_{j\to+\infty} P(E_j,\Omega) = P(E,\Omega). \end{cases}$$

Proof. From 1.1.13 χ_E can be approximated by the mollification of χ_E, we call them f_ϵ, s.t. $0 \leq f_\epsilon \leq 1$ and

$$P(E) = \lim_{\epsilon\to 0} V(f_\epsilon). \tag{1.4}$$

If we define $E_{\epsilon,t} \coloneqq \left\{x \in \mathbb{R}^N :\, f_\epsilon(x) > t\right\}$, then by lemma 1.1.27

$$\forall t \in (0,1), \quad \lim_{\epsilon\to 0} \chi_{E_{\epsilon,t}} = \chi_E \quad \text{in}\, L^1,$$

as f_ϵ goes to χ_E in L^1. Thus, by 1.1.10:

$$\liminf_{\epsilon\to 0} P(E_{\epsilon,t}) \geq P(E), \quad \forall t \in (0,1). \tag{1.5}$$

Now applying the coarea formula to 1.4 and the Fatou's lemma we have:

$$P(E) = \lim_{\epsilon\to 0} \int_0^1 P(\mathbb{R}^N \setminus E_{\epsilon,t}) \,\mathrm{d}t = \lim_{\epsilon\to 0} \int_0^1 P(E_{\epsilon,t}) \,\mathrm{d}t \geq \int_0^1 \liminf_{\epsilon\to 0} P(E_{\epsilon,t}) \,\mathrm{d}t \geq P(E)$$
$$\implies \int_0^1 \liminf_{\epsilon\to 0} P(E_{\epsilon,t}) \,\mathrm{d}t = P(E).$$

But for 1.5 we must have:

$$\liminf_{\epsilon\to 0} P(E_{\epsilon,t}) = P(E),$$

for almost every $t \in (0,1)$. Finally, for Sard's lemma $\partial E_{\epsilon,t}$ is smooth for almost every T. This means we can find a t for which we can extract a sequence E_j with smooth boundary s.t.:

$$\begin{cases} \lim\limits_{j\to+\infty} \chi_{E_j} = \chi_E \quad \text{in}\, L^1; \\ \lim\limits_{j\to+\infty} P(E_j) = P(E). \end{cases}$$

□

Combining this result and the coarea formula we can prove a lemma that will be useful later on.

Lemma 1.1.29. *Let E be a bounded Caccioppoli set. For almost every $\rho > 0$ we have*

$$P(E, \partial B_\rho) = 0.$$

Proof. From 1.1.28 we can find a sequence of regular sets $\{E_h\}$ converging in measure to E and such that $\lim_h P(E_h, \partial B_\rho) = P(E, \partial B_\rho)$. We notice that $P(E, B_\rho)$ is a measurable function in ρ as it is monotone, then also $P(E, \overline{B}_\rho) = \inf_{t>\rho} P(E, B_t)$. this means that also $P(E, \partial B_\rho)$ is measurable and then it can be integrated, then by Fatou's lemma:

$$\int_{-\infty}^{+\infty} P(E, \partial B_t)\,\mathrm{d}t = \int_{-\infty}^{+\infty} \lim_h P(E_h, \partial B_t)\,\mathrm{d}t \leq \liminf_h \int_{-\infty}^{+\infty} P(E_h, \partial B_t)\,\mathrm{d}t.$$

But as both B_t and E_h have smooth boundary, from 1.1.6 and 1.1.23,

$$P(E_h, \partial B_t) = \mathcal{H}_{n-1}(\partial E_h \cap \partial B_t) = P(B_t, \partial E) = 0, \quad a.e.$$

Finally,

$$\int_{-\infty}^{+\infty} P(E, \partial B_t)\,\mathrm{d}t = 0 \implies P(E, \partial B_\rho) = 0 \quad a.e.$$

□

Then we present a standard result of existence of minimal perimeter.

Theorem 1.1.30. *Let Ω be a bounded open set in $\mathbb{R}^N$ and L be a Caccioppoli set. We define:*

$$\mathcal{C} := \{F \textit{ Caccioppoli set s.t } F \setminus \Omega = L \setminus \Omega\}.$$

Then the problem

$$\min_{F\in\mathcal{C}} P(F)$$

has a solution.

Observation 1.1.31. We notice that the role of L is to capture the constrained boundary. Indeed, we can read the problem as: "find the subset of Ω with minimal perimeter, given that $\partial\Omega \cap L$ has to be part of its boundary".

Proof. From 1.1.5 we can write

$$P(F) = P(F, \bar{\Omega}) + P(F, \mathbb{R}^N \setminus \bar{\Omega}).$$

But $P(F, \mathbb{R}^N \setminus \bar{\Omega})$ is fixed by constraint, therefore we want to minimize just $P(F, \bar{\Omega}) = V(\chi_F, \bar{\Omega})$. The perimeter is always positive, this means that the problem is bounded from below and so any minimizing sequence E_j has uniformly bounded perimeters. Furthermore, $\left\|\chi_{E_j}\right\|_{L^1(\Omega)} = |E_j \cap \Omega| \leq |\Omega|$. Therefore, $\{\chi_{E_j}\}$ is bounded in BV. By the compact embedding, there exists $f \in L^1(\Omega)$ to which a subsequence of χ_{E_j} is converging in L^1. We can choose this subsequence to converge also a.e. to f, this implies that $f(x) \in \{0,1\}$ a.e. If we call $E := \{x \in \Omega : \ f(x) = 1\}$, we can say:

$$f = \chi_E \quad a.e.$$

But by 1.1.10, we have:

$$P(E,\Omega) = V(\chi_E,\Omega) \leq \liminf_j V(\chi_{E_j},\Omega) = \inf_{F \in \mathcal{C}} P(F,\Omega).$$

This means that E is the minimum. □

We can point out that similar proofs can be used for optimization problem with functionals of the kind:

$$\mathcal{G}(F) = P(F,\Omega) + \int_F H(x)\,\mathrm{d}x.$$

1.2 Reduced boundary

Definition

In this section the idea of reduced boundary will be present with some relevant properties. In particular, on the reduced boundary is possible to define in some sense the the normal vector and s tangent space. Furthermore, we can write the perimeter of a Caccioppoli set as the Hausdorff measure on the reduced boundary.

We will develop further theory on the Caccioppoli sets considering them as a class of equivalence, this comes naturally as we identify the set with $\chi_E \in L^1_{LOC}$.

Definition 1.2.1. We say that two sets E and F are *equivalent* if

$$|(E \setminus F) \cup (F \setminus E)| = 0.$$

This equivalence relation is consistent with the perimeter from 1.1.8 (3). From now on we will always take for any Caccioppoli set E its representative satisfying the following property:

$$0 < |E \cap B(x,\rho)| < \omega_n \rho^n \quad \forall x \in \partial E, \forall \rho > 0; \tag{1.6}$$

where $B(x,\rho)$ is the ball centered in x of radius ρ, while $\omega_n = |B(0,1)|$. We can now prove that always exists an equivalent set with such a property.

Proposition 1.2.2. *For any Borel set E exists an equivalent set $\tilde{E}$ satisfying* (1.6)*.*

Proof. We define two sets:

$$E_0 = \{x \in \mathbb{R}^n : \exists \rho > 0, |B(x,\rho) \cap E| = 0\},$$

$$E_1 = \{x \in \mathbb{R}^n : \exists \rho > 0, |B(x,\rho) \cap E| = \omega_n \rho^n\}.$$

We first notice that $E_0 \cap E_1 = \emptyset$. Indeed, if we suppose $x \in E_0 \cap E_1$, then there must exist $\rho_1 > \rho_0 > 0$ s.t.

$$|B(x,\rho_0) \cap E| = 0 \bigwedge |B(x,\rho_1) \cap E| = \omega_n \rho_1^n$$

$$\implies |(B(x,\rho_1) \setminus B(x,\rho_0)) \cap E| = |B(x,\rho_1) \cap E| - |B(x,\rho_0) \cap E| = \omega_n \rho_1^n.$$

But at the same time $|(B(x,\rho_1) \setminus B(x,\rho_0)) \cap E| \leq |B(x,\rho_1) \setminus B(x,\rho_0)| < \omega_n \rho_1^n$, contradiction.

Then we can show that those sets are open. Take $x \in E_0$, then there exist $\rho > 0$ s.t. $|B(x,\rho) \cap E| = 0$. Let $y \in B(x,\rho)$, then we can find $\rho_y > 0$ s.t $B(y,\rho_y) \subset B(x,\rho)$ and consequently $|B(y,\rho_y) \cap E| = 0$. This show that $B(x,\rho) \subseteq E_0$. Similarly, for $x \in E_1$ we can find $\rho > 0$ s.t. $|B(x,\rho) \cap E| = \omega_n \rho^n$ and for every $y \in B(x,\rho)$ there is a ball $B(y,\rho_y) \subset B(x,\rho)$.

$$\omega_n \rho^n = |(B(x,\rho) \setminus B(t,\rho_y)) \cap E| + |B(y,\rho_y) \cap E| \leq \omega_n \rho^n - \omega_n \rho_y^n + |B(y,\rho_y) \cap E|$$

$$\implies |B(y,\rho_y) \cap E| \geq \omega_n \rho_y^n.$$

Then, $B(x,\rho) \subseteq E_1$.

Now we define $\tilde{E} = (E \cup E_1) \setminus E_0 = (E \setminus E_0) \cup E_1$. W now show that $\tilde{E}$ is equivalent to E.

$$E \setminus \tilde{E} = E \cap ((E \cup E_1) \cap E_0^c)^c = E \cap ((E \cup E_1)^c \cup E_0) =$$

$$= (E \cap E_0) \cup (E \cap (E \cup E_1)^c) = E \cap E_0.$$

$$\tilde{E} \setminus E = (E \cup E_1) \cap E_0^c \cap E^c = (E_1 \cap E^c) \cap E_0^c = E_1 \setminus E$$

$$\implies \left|(E \setminus \tilde{E}) \cup (\tilde{E} \setminus E)\right| \leq |E_1 \setminus E| + |E \cap E_0|.$$

Now, since $\mathbb{R}^n$ is separable, we can find a sequence $\{x_i\} \subset E_0$ dense in E_0 and consequently such that $E_0 \subseteq \bigcup_{i \in \mathbb{N}} B(x_i,\rho_i)$ and $|E \cap B(x_i,\rho_i)| = 0$.

$$\implies |E \cap E_0| \leq \sum_{i \in \mathbb{N}} |E \cap B(x_i,\rho_i)| = 0.$$

At the same way we can find a numerable ball covering for E_1 s.t. $|E \cap B(x_i,\rho_i)| = \omega_n \rho_i^n$, then

$$|E_1 \setminus E| \leq \sum_{i \in \mathbb{N}} |B(x_i,\rho_i) \setminus E| = \sum_{i \in \mathbb{N}} |B(x_i,\rho_i)| - |B(x_i,\rho_i) \cap E| = 0.$$

$$\implies \left|(E \setminus \tilde{E}) \cup (\tilde{E} \setminus E)\right| = 0.$$

Now, to conclude, we have to prove that $\partial\tilde{E} \cap (E_1 \cup E_0) = \emptyset$. Suppose $x \in E_1 \cup E_0$, since those sets are open we can find a ball $B(x,\rho)$ totally contained either in E_0 or E_1.

- $B(x,\rho) \subset E_0$: this implies $B(x,\rho) \cap \tilde{E} \subset E_0 \cap E_0^c = \emptyset$, and so $x \notin \partial\tilde{E}$.
- $B(x,\rho) \subset E_1$: that is $B(x,\rho) \cap \tilde{E}^c = B(x,\rho) \cap ((E^c \cup E_0) \cap E_1^c) = \emptyset$, which again implies $x \notin \partial\tilde{E}$.

□

From now on we will always refer to the representative set satisfying (1.6).

Observation 1.2.3. A first relevant consequence of this property is that the support of $|D\chi_E|$ is exactly ∂E. Indeed, $\mathrm{supp}|D\chi_E| \subseteq \partial E$, while given an A open such that

$$|D\chi_E|(A) = 0,$$

then for each $x \in \partial E$ we can find a ball $B = B(x,\rho)$ that is $|D\chi_E|$-negligible. But from isoperimetric inequality 1.1.20 this implies:

$$\min\{|B \cap E|, |B \cap (\mathbb{R}^n \setminus E)|\} = 0.$$

Although, this contradicts the property (1.6).

We go on defining the *reduced boundary*.

Definition 1.2.4 (Reduced Boundary)**.** Given a Caccioppoli set E, we define the *reduced boundary* $\partial^* E$ as the set of x s.t.

- $|D\chi_E|(B(x,\rho)) > 0$, $\forall \rho > 0$;
- the limit $\nu(x) = \lim_{\rho\to 0} \frac{D\chi_E(B(x,\rho))}{|D\chi_E|(B(x,\rho))}$ exists;
- $|\nu(x)| = 1$.

Observation 1.2.5. Since $|D\chi_E|$ is supported on ∂E and the property 1.6 is assumed, then the first requirement cannot be satisfied for $x \notin \partial E$. Therefore, $\partial^* E \subseteq \partial E$.

ν actually is the integrable function of the polar decomposition of $D\chi_E$ (see (B.0.4)), that is $|\nu(x)| = 1$, $|D\chi_E|$-a.e., and ν is the Radon-Nikodym derivative (see appendix A) between $D\chi_E$ and $|D\chi_E|$:

$$D\chi_E = \nu|D\chi_E|.$$

Observation 1.2.6. This means that ν is defined as above for $|D\chi_E|$-almost every point on ∂E. This combined with the fact that $|D\chi_E|$ is supported in ∂E, let us conclude:

$$P(E,\Omega) = |D\chi_E|(\Omega) = |D\chi_E|(\Omega \cap \partial E) = |D\chi_E|(\Omega \cap \partial^* E). \tag{1.7}$$

We can look at this ν as a generalization of the inward unitary normal vector and the reduced boundary as the portion of ∂E where this normal is well defined. The following example will show it.

Example 1.2.7. *Take E set with C^2 boundary. From example 1.1.6 we have*

$$|D\chi_E|(A) = P(E,A) = \mathcal{H}_{n-1}(A \cap \partial E)$$

for any measurable set A. But we also know that in this case the distributional derivative of χ_E is given by:

$$D\chi_E = \nu \mathcal{H}_{n-1}, \ \textit{on}\, \partial E,$$

where ν is the normal vector on ∂E pointing inside E.

$$\implies \frac{D\chi_E(B(x,\rho))}{|D\chi_E|(B(x,\rho))} = \frac{1}{\mathcal{H}_{n-1}(\partial E \cap B(x,\rho))} \int_{\partial E \cap B(x,\rho)} \nu(x) d\mathcal{H}_{n-1}.$$

This is the integral average of ν on a ball, thus by continuity it converges to $\nu(x)$ itself. Then

$$\partial^* E = \partial E.$$

If we take E to be instead a polygon, then the corners are not in the reduced boundary. Indeed, there the integral average will be the average between the normal vectors of two sides $\frac{\nu_1+\nu_2}{2}$, which is not a unitary vector.

Approximation by tangent spaces

We can see that with this definition of normal vector we can define a tangent space, which locally approximate the reduced boundary. In particular for any $z \in \partial^* E$ we can define the tangent space and the approximating half-space:

$$T(z) = \{x \in \mathbb{R}^n : \ \langle \nu(z), x - z \rangle = 0\}, \tag{1.8}$$

$$T^+(z) = \{x \in \mathbb{R}^n : \ \langle \nu(z), x - z \rangle > 0\}, \tag{1.9}$$

where in T^+ we take ">0" because ν is the inner normal vector. In order to prove this approximation we need the some inequalities.

Proposition 1.2.8. *Let E be a Caccioppoli set and take $x \in \partial^* E$, then $\forall \varepsilon > 0$ exists $\rho_0 > 0$ and a constant $c = c(n, \varepsilon) > 0$ such that $\forall \rho < \rho_0$:*

$$P(E, B(x, \rho)) \leq c\rho^{n-1}.$$

Proof. Without loss of generality consider $x = 0$ and E bounded (as we want to verify a local property) and denote $B_\rho = B(0, \rho)$. From 1.1.9, if we take any constant $g \in \mathbb{R}^n$, then

$$g \cdot D\chi_{E\cap B_\rho}(\mathbb{R}^n) = 0 \implies D\chi_{E\cap B_\rho} = 0.$$

But as well

$$\begin{aligned} &D\chi_{E\cap B_\rho}(\mathbb{R}^n) = D\chi_{E\cap B_\rho}(B_\rho) + D\chi_{E\cap B_\rho}(\mathbb{R}^n \setminus B_\rho) \\ \implies &\left|D\chi_{E\cap B_\rho}(B_\rho)\right| = \left|D\chi_{E\cap B_\rho}(\mathbb{R}^n \setminus B_\rho)\right| \leq P(E \cap B_\rho, \mathbb{R}^n \setminus B_\rho). \end{aligned}$$

Then, from 1.1.17 and the fact that for any A open $P(E, A) = P(E \cap A, A)$:

$$P(E\cap B_\rho, \mathbb{R}^n \setminus B_\rho) = P(E\cap B_\rho, \mathbb{R}^n) - P(E\cap B_\rho, B_\rho) \leq P(E, \partial B_\rho) + \frac{d}{d\rho_+}|E \cap B_\rho|.$$

However, from 1.1.29 $P(E, \partial B_\rho) = 0$ for almost every ρ, and in addiction from 1.1.22 we have

$$|E \cap B_\rho| = \int_0^\rho \mathcal{H}_{n-1}(\partial B_t \cap (E \cap B_\rho))\,\mathrm{d}t = \int_0^\rho \mathcal{H}_{n-1}(\partial B_t \cap E)\,\mathrm{d}t.$$

Thus $|E \cap B_\rho|$ is actually almost everywhere differentiable and consequently

$$\frac{d}{d\rho_+}|E \cap B_\rho| = \frac{d}{d\rho}|E \cap B_\rho| = \mathcal{H}_{n-1}(\partial B_\rho \cap E) \leq n\omega_n \rho^{n-1}.$$

On the other hand, $D\chi_{E\cap B_\rho}(B_\rho) = D\chi_E(B_\rho)$, since for any $\phi \in \mathcal{D}(B_\rho)$

$$\langle D_j\chi_{E\cap B_\rho}(B_\rho), \phi\rangle = -\langle \chi_{E\cap B_\rho}, D_j\phi\rangle = -\langle \chi_E, D_j\phi\rangle.$$

To summarize we have

$$|D\chi_E(B_\rho)| \leq n\omega_n\rho^{n-1}.$$

Now we assume $0 \in \partial^* E$, therefore

$$\frac{|D\chi_E(B_\rho)|}{|D\chi_E|(B_\rho)} \longrightarrow |\nu(0)| = 1,$$

then for every ε there are ρ small enough such that: $|D\chi_E(B_\rho)| \geq (1-\varepsilon)|D\chi_E|(B_\rho)$,

$$\implies |D\chi_E|(B_\rho) \leq \frac{n\omega_n}{1-\varepsilon}\rho^{n-1}, \qquad a.e.\rho < \rho_0.$$

However, using the continuity from below of $|D\chi_E|$ we can extend this inequality to every ρ.

$\square$

Proposition 1.2.9. *Let E be a Caccioppoli set and take $x \in \partial^* E$, then $\forall \varepsilon > 0$ exists $\rho_0 > 0$ and a constant $c = c(n, \varepsilon) > 0$ such that $\forall \rho < \rho_0$:*

$$\min\{\, |E \cap B(x,\rho)|,\ |(\mathbb{R}^n \setminus E) \cap B(x,\rho)|\,\} \geq c\rho^n.$$

Proof. Let's call B_ρ the ball $B(x,\rho)$. In seek of simplicity during the proof I will always use the notation c for a generic constant in the inequality, even though its actual value is changing. Recalling some steps of the proof in 1.2.8, we can say that

$$P(E \cap B_\rho, B_\rho) = P(E, B_\rho) \leq \frac{1}{1-\varepsilon}|D\chi_E(B_\rho)| \leq \frac{1}{1-\varepsilon} P(E \cap B_\rho, \mathbb{R}^n \setminus B_\rho),$$

$$\implies P(E \cap B_\rho) \leq c\, P(E \cap B_\rho, \mathbb{R}^n \setminus B_\rho).$$

But we also have

$$P(E \cap B_\rho, \mathbb{R}^n \setminus B_\rho) \leq \frac{d}{d\rho_+}|E \cap B_\rho| = \frac{d}{d\rho}|E \cap B_\rho|$$

while with the isoperimetric inequality 1.1.20 we have

$$|E \cap B_\rho|^{\frac{n-1}{n}} \leq c\, P(E \cap B_\rho)$$

$$\implies |E \cap B_\rho|^{\frac{n-1}{n}} \leq c \frac{d}{d\rho}|E \cap B_\rho|.$$

Pairing this differential inequality with the initial condition $|E \cap B_0| = 0$, we can solve:

$$|E \cap B_\rho| \geq c\rho^n.$$

We can just put $\mathbb{R}^n \setminus E$ in place of E to get the other inequality. □

Theorem 1.2.10. *Take $z \in \partial^* E$ and $t > 0$, define the set:*

$$E_t = \{x \in R^n :\ t(x-z) + z \in E\}.$$

Then, E_t converges locally in measure to $T^+(z)$ for t going to 0 and in particular for every bounded set A s.t. $\mathcal{H}_{n-1}(\partial A \cap T(z)) = 0$ we have:

$$\lim_{t \to 0} P(E_t, A) = P(T^+(z), A) = \mathcal{H}_{n-1}\left(A \cap T(z)\right).$$

Proof. Without loss of generality we assume $z = 0$ and $\nu(0) = -e_1$, so that

- $E_t = \frac{1}{t}E$,
- $T^+ = T^+(0) = \{x :\ x_i < 0\}$.

Moreover, from 1.2.8 and 1.1.8 (5) we have:

$$P(E_t, B_\rho) = \frac{1}{t^{n-1}} P(E, B_{t\rho}) \leq c\rho^{n-1},$$

that is, $P(E_t, B_\rho)$ is bounded as a function of t. In order to prove that E_t converges to $T^+(z)$, we show that for every sequence E_{t_j}, such that $t_j \to 0^+$, there is a subsequence, we call it E_j, that converges to $T^+(z)$. Beacause $P(E_t, B_\rho)$ is bounded, from the compactness theorem 1.1.14, we can extract a subsequence E_j such that χ_{E_j} converges in L^1_{LOC}, but we can chose it to converge also almost everywhere, therefore the limit function must be a characteristic of a set:

$$\chi_{E_j} \xrightarrow[j\to\infty]{L^1_{LOC}} \chi_F.$$

But the L^1_{LOC} convergence implies the convergence in distribution and therefore also $D\chi_{E_j}$ converge in distribution to $D\chi_F$. Finally, by density of C_c^∞ in C_0, we have

$$D\chi_{E_j} \overset{\star}{\rightharpoonup} D\chi_F,$$

for the meaning of this convergence see B.0.7. The scaling property 1.1.8 (5) is true also for the $D\chi_E$ measure for the same argument as in (1.1), using it again we have:

$$\frac{D\chi_{E,t_j}}{t_j^{n-1}} \overset{\star}{\rightharpoonup} D\chi_F,$$

where $D\chi_{E,t_j}(A) = D\chi_E(t_jA)$. Now we, taking into account 1.2.8 and B.0.12, we can apply B.0.14 to conclude:

$$\left|D\chi_{E_j}\right| = \frac{\left|D\chi_{E,t_j}\right|}{t_j^{n-1}} \overset{\star}{\rightharpoonup} |D\chi_F|, \quad \bigwedge \quad D\chi_F = \nu(0)|D\chi_F| = -e_1|D\chi_F|.$$

But this means that χ_F depends actually only on x_1. Indeed, take the mollified version $\chi_F * \rho_\varepsilon$, where ε is a standard mollifier. Then for $i \geq 2$

$$\frac{\partial}{\partial x_i}(\chi_F * \rho_\varepsilon) = D_i\chi_F * \rho_\varepsilon = (|D\chi_F| * \rho_\varepsilon)\nu_i(0) = 0,$$

that is $\chi_F * \rho_\varepsilon$ depends only on x_1 for all ε. Thus, the same is true moving to the limit $\varepsilon \to 0^+$. This means that there exists a set $J \subset \mathbb{R}$ such that:

$$\chi_F(x) = \chi_J(x_1),$$

and consequently $F = J \times \mathbb{R}^{n-1}$. Since, for any $a \in \mathbb{R}$

$$D\chi_J((a,x)) = D_1\chi_F((a,x) \times (0{,}1)^{n-1}) = -|D\chi_F|((a,x) \times (0{,}1)^{n-1}) \leq 0,$$

we must have that χ_J is a decreasing function, although this is possible only if $J = (-\infty, \alpha)$ for some $\alpha \in \mathbb{R}$, that is F is a halfspace. Now suppose $\alpha < 0$, then

$$\lim_{j\to\infty} \frac{\left|E \cap B_{\alpha t_j}\right|}{t_j^n} = \lim_{j\to\infty} |E_j \cap B_\alpha| = |F \cap B_\alpha| = 0.$$

However this contradicts 1.2.9, for which $\frac{\left|E\cap B_{\alpha t_j}\right|}{t_j^n} \geq c > 0$ for every j. Similarly, if $\alpha > 0$, we can use the same argument (adding that $\mathbb{R}^n \setminus E_j$ converges in measure to $\mathbb{R}^n \setminus F$) and we find a contradiction with $\frac{\left|(\mathbb{R}^n\setminus E)\cap B_{\alpha t_j}\right|}{t_j^n} \geq c > 0$. In conclusion $F = T^+$.

The second statement comes from the more general fact that if a sequence of positive Radon measures $\mu_h \overset{*}{\rightharpoonup} \mu$ locally and $\mu(\partial A) = 0$, where A is a relatively compact set, then $\lim_h \mu_h(A) = \mu(A)$. Indeed, we can take an approximation with Lipschitz functions of χ_A from above (a_ε) and from below (b_ε), that is:

$$a_\varepsilon = \sup\{\chi_A(y) - \varepsilon|x-y|,\, y \in \mathbb{R}^n\}, \qquad b_\varepsilon = \inf\{\chi_A(y) + \varepsilon|x-y|,\, y \in \mathbb{R}^n\}.$$

For $\varepsilon \to 0^+$, a_ε converge pointwise to $\chi_{\bar{A}}$, while b_ε to $\chi_{\mathring{A}}$. Nevertheless, a_ε and b_ε are bounded functions on a compact domain, that is they can be dominated, therefore those convergences are also in $L^1(K,\mu)$ and $L^1(K,\mu_h)$, where K is a wide enough compact. Therefore we have

$$\lim_{\varepsilon\to 0^+} \int_K b_\varepsilon \,\mathrm{d}\mu = \int_K \chi_{\mathring{A}} \,\mathrm{d}\mu = \mu(\mathring{A})$$

$$\lim_{\varepsilon\to 0^+} \int_K a_\varepsilon \,\mathrm{d}\mu = \int_K \chi_{\bar{A}} \,\mathrm{d}\mu = \mu(\bar{A}).$$

Taking into account that $\mu(\partial A) = 0$, we conclude

$$\lim_{\varepsilon\to 0^+} \int_K b_\varepsilon \,\mathrm{d}\mu = \lim_{\varepsilon\to 0^+} \int_K a_\varepsilon \,\mathrm{d}\mu = \mu(A).$$

However for each ε we have:

$$\int_K b_\varepsilon \,\mathrm{d}\mu = \lim_h \int_K b_\varepsilon \,\mathrm{d}\mu_h \leq \liminf_h \int_K \chi_A \,\mathrm{d}\mu_h \leq$$

$$\leq \limsup_h \int_K \chi_A \,\mathrm{d}\mu_h \leq \lim_h \int_K a_\varepsilon \,\mathrm{d}\mu_h = \int_K a_\varepsilon \,\mathrm{d}\mu.$$

Then moving to the limit for $\varepsilon \to 0^+$ we have the thesis. □

Corollary 1.2.11. *For each $x \in \partial^* E$*

$$\lim_{\rho\to 0} \frac{P(E, B(x,\rho))}{\omega_{n-1}\rho^{n-1}} = 1.$$

Proof. Because the perimeter is invariant on translations, we can take $x = 0$. From 1.1.8 property (5), $P\left(\frac{1}{\rho}E, B(0,1)\right) = \frac{1}{\rho^{n-1}}P(E, B(0,\rho))$. But from the previous theorem we have:

$$\lim_{\rho\to 0} P\left(\frac{1}{\rho}E, B(0,1)\right) = \mathcal{H}_{n-1}(B(0,1)\cap T(0)) = \omega_{n-1}.$$

Then we can easily conclude.

□

This corollary is a generalization of a known result for Hausdorff measures on regular surfaces. Indeed, let V be a set with regular boundary, then $\partial V = \partial^* V$ and $|D\chi_V| = \mathcal{H}_{n-1}\lfloor_{\partial V}$ (the Hausdorff measure restricted on ∂V). Thus, we can verify

$$\lim_{\rho\to 0} \frac{\mathcal{H}_{n-1}(B(x,\rho)\cap\partial V)}{\omega_{n-1}\rho^{n-1}} = 1, \quad \forall x \in \partial V.$$

Perimeter and Hausdorff measures

The most relevant result, for which we need to introduce the reduced boundary, is that:

$$P(E,\Omega) = \mathcal{H}_{n-1}(\partial^* E\cap\Omega).$$

In order to show it we need an useful decomposition of the reduced boundary, that we are not going to prove, for more insight see [1] chapter 4.

Theorem 1.2.12. *If E is a Caccippoli set, then*

$$\partial^* E = \bigcup_{i=1}^{\infty} C_i \cup N;$$

where N is $|D\chi_E|$-negligible, while for all i C_i is such that there exist a set $A_i \supset C_i$ and a C^1 function $f_i : A_i \longrightarrow \mathbb{R}$ satisfying for all $x \in \bar{C}_i$:

$$f_i(x) = 0, \quad \bigwedge \quad Df_i(x) \neq 0.$$

Furthermore, C_i can be chosen so that $|D\chi_E|(\partial^* E \setminus C_i) < \frac{1}{i}$.

We also need the following inequality.

Lemma 1.2.13. *Let E a Caccioppoli set and $C \subseteq \partial^* E$, then $\exists\, b_n$ such that*

$$\mathcal{H}_{n-1}(C) \le 2b_n|D\chi_E|(C).$$

Observation 1.2.14. On $\partial^* E$ we have $\mathcal{H}_{n-1} << |D\chi_E|$.

Proof. For outer regularity of Radon measures (see A.0.3), for any $\eta > 0$ we can find an open set $A \supset C$ such that:

$$|D\chi_E|(A) \leq |D\chi_E|(C) + \eta.$$

Take an arbitrary $\varepsilon > 0$, then for each $x \in C$ we can find (from 1.2.11) a $\varepsilon > \bar{\rho} > 0$ such that for each $\rho \leq \bar{\rho}$ $B(x,\rho) \subset A$ and

$$|D\chi_E|(B(x,\rho)) \geq \frac{1}{2}\omega_{n-1}\rho^{n-1}.$$

This is fine ball covering so by Besicovitch covering theorem (A.0.8) we can find a countable family of such balls $B_i = B(x_i, \rho_i)$ which still cover C and each point is in at most b_n balls. Therefore,

$$\sum_i \omega_{n-1}\rho_i^{n-1} \leq 2\sum_i |D\chi_E|(B_i) \leq 2b_n|D\chi_E|(A) \leq 2b_n\left(|D\chi_E|(C) + \eta\right).$$

Recalling that

$$\mathcal{H}_{n-1}(C) = \lim_{\varepsilon\to 0}\ \inf\left\{\sum_i \omega_{n-1}\left(\frac{\operatorname{diam}(U_i)}{2}\right)^{n-1}\middle|\ C \subset \bigcup_i U_i,\ \operatorname{diam}(U_i) < \varepsilon\right\},$$

then for arbitrariness of ε and η we conclude that

$$\mathcal{H}_{n-1}(C) \leq 2b_n|D\chi_E|(C).$$

$\square$

Theorem 1.2.15. *Let E be a Caccioppoli set, then for every $B \subseteq \partial^*E$*

$$|D\chi_E|(B) = \mathcal{H}_{n-1}(B),$$

and in particular, for any open Ω:

$$P(E,\Omega) = \mathcal{H}_{n-1}(\Omega \cap \partial^*E).$$

Proof. Take the decomposition given in 1.2.12, then by the previous lemma:

$$|D\chi_E|(B \setminus C_i) < \frac{1}{i};\ \Longrightarrow\ \mathcal{H}_{n-1}(B \setminus C_i) < \frac{2b_n}{i}.$$

For now suppose the statement is true for $B \cap C_i$, for any i, then

$$|D\chi_E|(B) = |D\chi_E|(B \setminus C_i) + |D\chi_E|(B \cap C_i) < \frac{1}{i} + \mathcal{H}_{n-1}(B \cap C_i) \leq \frac{1}{i} + \mathcal{H}_{n-1}(B),$$

$$\implies |D\chi_E|(B) \leq \mathcal{H}_{n-1}(B).$$

But, similarly using $\mathcal{H}_{n-1}(B \setminus C_i) < \frac{2b_n}{i}$, we can get also $\mathcal{H}_{n-1}(B) \leq |D\chi_E|(B)$. Therefore, we just need to show that the statement is true for $B \cap C_i$, which means we have to verify the equality for a generic set C for which there exist a set $A \supset C$ and a C^1 function $f : A \to \mathbb{R}$ such that for all $x \in \bar{C}$:

$$f(x) = 0, \quad \bigwedge \quad \nabla f(x) \neq 0.$$

As $f \in C^1$, we can take A so that $\nabla f(x) \neq 0$ on A. Therefore, $\Gamma = \{x \in A : f(x) = 0\}$ defines a regular hypersurface. Then, on each $x \in \Gamma$ we have

$$\lim_{\rho \to 0} \frac{\mathcal{H}_{n-1}(B(x, \rho \cap \Gamma))}{\omega_{n-1}\rho^{n-1}} = 1.$$

However, we also know from 1.2.11 that

$$\lim_{\rho \to 0} \frac{|D\chi_E|(B(x,\rho))}{\omega_{n-1}\rho^{n-1}} = 1,$$

$$\implies \lim_{\rho \to 0} \frac{\mathcal{H}_{n-1}(B(x, \rho \cap V))}{|D\chi_E|(B(x,\rho))} = 1.$$

Then from the measure differentiation theorem A.0.11 we can conclude that $\mathcal{H}_{n-1} = |D\chi_E|$ on Γ, and in particular

$$\mathcal{H}_{n-1}(C) = |D\chi_E|(C).$$

Finally from (1.7), we can conclude:

$$\mathcal{H}_{n-1}(\Omega \cap \partial^* E) = |D\chi_E|(\Omega \cap \partial^* E) = P(E, \Omega).$$

□

Corollary 1.2.16. *$\partial^* E$ is dense in ∂E*

Proof. From observation 1.2.5 and the fact that ∂E is a close set, we already have that $\overline{\partial^* E} \subseteq \partial E$. For the reverse inclusion we can notice that if we take an open set A outside of $\partial^* E$, that is $A \cap \partial^* E = \emptyset$, then

$$|D\chi_E|(A) = P(E, A) = \mathcal{H}_{n-1}(A \cap \partial^* E) = 0.$$

But this implies that $A \cap \operatorname{supp}|D\chi_E| = A \cap \partial E = \emptyset$. This means that the interior of $\mathbb{R}^n \setminus \partial^* E$ is contained in the interior of $\mathbb{R}^n \setminus \partial E$, therefore

$$\partial E \subseteq \overline{\partial^* E}.$$

□

We can conclude observing that from the generalize Gauss-Green formula in observation 1.1.9 we can generalize the equation (1.2):

$$\int_E \operatorname{div}(g)\,\mathrm{d}x = -\int_E g\,\mathrm{d}D\chi_E = -\int_E g\cdot\nu\,\mathrm{d}|D\chi_E| = -\int_{\partial E} g\cdot\nu\, d\mathcal{H}_{n-1},$$

which is consistent with the Gauss-Green formula as ν here is the inward normal.

We can extend the idea of reduced boundary to a general function in BV via the so called *jump set*.

Definition 1.2.17. Given $u\in BV(\Omega)$, we say that $x\in J_u$, the jump set of u, if there exist $u_+(x)$, $u_-(x)\in\mathbb{R}$ and $\nu_u(x)\in\mathbb{R}^n$ such that $u_+(x)\neq u_-(x)$ and

$$\lim_{\varepsilon\to 0} u(x+\varepsilon y) = \begin{cases} u_+(x), & \text{if } y\cdot\nu_u \geq 0 \\ u_-(x), & \text{if } y\cdot\nu_u < 0 \end{cases},$$

in $L^1_y(B(0{,}1))$.

The quantities u_+, u_- and ν_u are not uniquely defined, because the same limit function can be described switching u_+ and u_- while changing the sign to ν_u. Conventionally, we take $u_+ > u_-$ and therefore ν_u is directed toward the increasing direction of the jump.

Observation 1.2.18. From theorem 1.2.10 we notice that the jump set of $u=\chi_E$ is the reduced boundary $\partial^* E$. Furthermore, from theorem 1.2.15 and the polar decomposition we can write

$$Du = \nu_u \mathcal{H}_{n-1}\restriction_{J_u}$$

For a general function in BV we have the Federer - Volpert decomposition

$$Du = \nabla u\mathcal{L}_n + (u_+ - u_-)\nu_u\mathcal{H}_{n-1}\restriction_{J_u} + Cu, \tag{1.10}$$

where ∇u is the weak gradient, while Cu is the so called *Cantor part*, which is a residual measure in between the Lebesgue and the Hausdorff measure. More precisely Cu is supported on a $\mathcal{L}_n$-negligible set, but $\mathcal{H}_{n-1}(supp\{Cu\})=\infty$.

The classical example to observe the Cantor part is the Cantor-Vitali function, a.k.a. the devil's staircase (figure 1.1). This is a monotonically increasing function on $[0{,}1]$, this means it belongs to $BV([0{,}1])$, and it is defined as the uniform limit of continuous functions, therefore it is continuous itself ($\implies J_u=\emptyset$). Furthermore, its most interesting feature is that it is constant almost everywhere, even though it is continuous and increasing, in particular it is not constant only on the Cantor set. This implies that $\nabla u = 0$ $\mathcal{L}_n$-a.e. and consequently

$$Du = Cu.$$

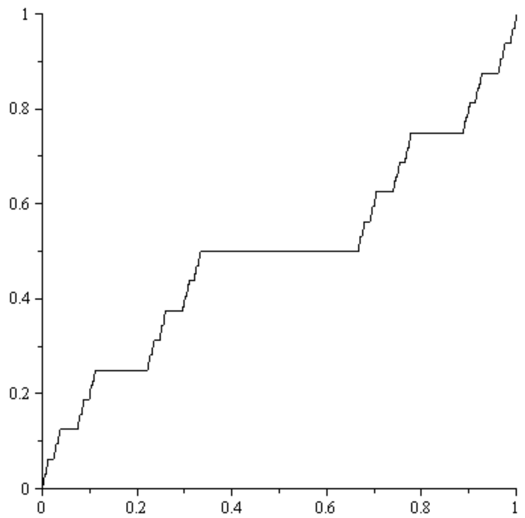

Figure 1.1: The "devil's staircase" or Cantor-Vitali function.

For more insights see [2]. The Cantor part is usually very hard to deal with and refers to some pathological cases, hence sometimes we rather simplify the computations taking into account only the BV functions with no Cantor part, that is the set:

$$SBV(\Omega) := \{u \in BV(\Omega) : \ Cu = 0\}.$$

1.3 Gamma convergence

In this section we will present a convergence criteria for functionals with a nice behaviour with respect to the minima (the minima converge to minima). To deepen the subject you can refer to [3]. This notion is useful in image analysis to jump between the discrete formulation through pixels and a continuous formulation, reached when the density of pixels tends toward infinity.

Definition 1.3.1 (Γ-convergence)**.** Let (X, d) be a complete metric space, $\{F_j\}$ a sequence of functionals $X \longrightarrow \mathbb{R}$. We say that F_j converge to a functional F, and denote it

$$F_j \xrightarrow{\Gamma} F,$$

if for each $x \in X$ we have:

1. for any sequence $x_j \longrightarrow x, \qquad F(x) \leq \liminf_j F_j(x_j);$
2. there exists a sequence $x_j \longrightarrow x, \qquad F(x) \geq \limsup_j F_j(x_j);$

The sequence in (2) is called *recovery sequence.*

Observation 1.3.2. We can notice that the condition (2), given (1), is equivalent to:

$$F(x) = \lim_j F_j(x_j);$$

This definition of limit is well posed, that is if the Γ-limit exists, it is unique. Indeed, let F' and F'' two limit of the sequence F_n, for every $x \in X$ take the recovery sequence x'_n for F', then

$$F'(x) \geq \limsup_n F_n(x'_n) \geq \liminf_n F_n(x'_n) \geq F''(x).$$

Similarly, we also have the reverse inequality.

Observation 1.3.3. In general it is not true that if $F_j = F$ constant sequence, then $F_j \xrightarrow{\Gamma} F$. Indeed, condition (1) implies:

$$\forall x \in X, \quad x_j \longrightarrow x \implies F(x) \leq \liminf_j F(x_j),$$

that is possible only if F is lower semicontionuous.

Indeed, it actually turns out that the Γ-limit is the lower semicontinuous envelope of F.

We will present some relevant results about Gamma convergence.

Proposition 1.3.4. *Let $U \subset X$ open and $F_j \longrightarrow F$ uniformly on U,*

$$F \textit{ lower semicontinuous} \implies F_j \xrightarrow{\Gamma} F \textit{ on} U.$$

Proof. 1.) Let $u_j \longrightarrow u \in U$, then for j big enough also $u_j \in U$.

$$\implies |F_j(u_j) - F(u_j)| \leq \sup_{v \in U} |F_j(v) - F(v)| \longrightarrow 0.$$

This and the semicontinuity of F let us conclude:

$$\liminf_j F_j(u_j) \leq \liminf_j F_j(u_j) - F(u_j) + \liminf_j F(u_j) \leq F(u).$$

2.) We can take the constant sequence $u_j = u$, then:

$$\lim_j F_j(u_j) = F(u).$$

□

The most important property is the relation between the minima of a the sequence of functionals and the ones of their Γ limit, expressed in the following theorem.

Theorem 1.3.5 (Fundamental theorem of Γ convergence). *Let X be a complete metric space and $F_j, F : X \longrightarrow \mathbb{R} \cup \{+\infty\}$ such that $F_j \xrightarrow{\Gamma} F$ and F_j equi mildly coercive, that is:*

$$\exists K \subset X\, compact : \inf_{u \in X} F_j(u) = \inf_{u \in K} F_j(u) \quad \forall j \in \mathbb{N}.$$

Then the following statements are true:

1. $\exists \min_X F$
2. $\lim_{j\to+\infty}(\inf_X F_J) = \inf_X F$
3. $\exists\{u_j\}$ *relatively compact sequence s.t.* $\lim_{j\to+\infty} F_j(u_j) = \inf_X F$.

Proof. Because the infima of all F_j are located in the same compact K, then they admit a minimizer in K, that is

$$\exists\{\tilde{u}_j\} \subset K \text{ s.t.} \quad \inf_X F_j = F_j(\tilde{u}_j).$$

But since K is compact, I can extract a subsequence (I still call it $\tilde{u}_j$) converging to a point $\bar{u} \in K$. Let's call instead $\{\bar{u}_j\}$ the recovery sequence for $\bar{u}$, then:

$$F(\bar{u}) \leq \liminf_j F_j(\tilde{u}_j) = \liminf_j \inf_X F_j \leq \limsup_j \inf_X F_j \leq \limsup_j F_j(\bar{u}_j) \leq F(\bar{u})$$

$$\implies F(\bar{u}) = \lim_j (\inf_X F_j).$$

Consider now $u \in X$ with $\{u_j\}$ its recovery sequence, then

$$F(u) \geq \limsup_j F_j(u_j) \geq \limsup_j \inf_X F_j = F(\bar{u})$$

$$\implies F(\bar{u}) = \min_X F.$$

$\square$

Observation 1.3.6. We can notice that in this proof we also showed that a sequence $\tilde{u}_j$ of minimizers for F_j is converging, up to a subsequence, to a minimizer of F $\bar{u}$.

With this kind of convergence the property

$$\begin{cases} F_j \xrightarrow{\Gamma} F \\ G_j \xrightarrow{\Gamma} G \end{cases} \implies F_j + G_j \xrightarrow{\Gamma} F + G$$

is <u>not</u> true in general. This is understandable as the minima displacement changes with the sum, therefore the property that minima converge to minima can be lost.

Example 1.3.7. *Take $f_j(x) = \sin^2(jx)$ and $g_j(x) = \cos^2(jx)$. We can rapidly verify that:*

$$f_j, g_j \xrightarrow{\Gamma} 0.$$

1. *$\liminf_j f_j(x_j) \geq 0$ independently on $\{x_j\}$.*
2. *Take as recovery sequence $x_j = \frac{\pi}{j}\lfloor \frac{j}{\pi}x \rfloor \xrightarrow[j\to\infty]{} x$, but $f_j(x_j) = 0$ for each j.*

and similarly for g_j. However, $f_j + g_j = 1 \xrightarrow{\Gamma} 1$.

But the convergence is preserved by adding continuous functionals.

Theorem 1.3.8 (Stability under continuous perturbations)**.** *Let $G : X \longrightarrow \mathbb{R}$ be a continuous functional,*

$$F_j \xrightarrow{\Gamma} F \implies F_j + G \xrightarrow{\Gamma} F + G.$$

Proof. 1. Take $u_j \longrightarrow u$, then

$$\liminf_j (F_j(u_j) + G(u_j)) \geq \liminf_j F_j(u_j) + \liminf_j G(u_j) \geq F(u) + G(u)$$

2. Take $\{u_j\}$ the recovery sequence for u over F_j, then

$$F_j(u_j) + G(u_j) \longrightarrow F(u) + G(u).$$

□

Often to show a Γ-limit it is convenient to show that the $\Gamma - \limsup$ and the $\Gamma - \liminf$ are equal.

Definition 1.3.9. Given a sequence of functionals F_j on a metric space X we define

1. $(\Gamma - \liminf_j F_j)(u) = \inf \{\liminf_j F(u_j) : u_j \to u\}$,
2. $\left(\Gamma - \limsup_j F_j\right)(u) = \inf \left\{\limsup_j F(u_j) : u_j \to u\right\}$.

Usually we call $F' \coloneqq \Gamma - \liminf_j F_j$ and $F'' \coloneqq \Gamma - \limsup_j F_j$.

It turned out that these quantities exist and are never $-\infty$, furthermore those inf are actually min. Let's verify that

$$\Gamma - \liminf_j F_j = \Gamma - \limsup_j F_j = F \iff \exists \Gamma - \lim_j F_j = F. \tag{1.11}$$

Indeed, from (1)

$$\forall u,\, \forall u_j \to u,\; F(u) \le \liminf_j F_j(u_j),$$

while, since those are minimum, from (2), there exists a sequence such that $F(u) = \limsup_j F_j(u_j)$. Conversely, if $F = \Gamma - \lim_j F_j$, then for the liminf property, $F \le \Gamma - \liminf_j F_j$, while the limsup property lead us to conclude that $F \ge \Gamma - \limsup_j F_j$. But, since clearly $\Gamma - \liminf_j F_j \le \Gamma - \limsup_j F_j$, we conclude that those are both equal to F.

With these definition we can finally show the relation between Γ-convergence and lower semicontinuity.

Proposition 1.3.10. *Given a sequence F_j, then the $\Gamma - \liminf_j F_j$ and $\Gamma - \limsup_j F_j$ are lower semicontinuous.*

Proof. Take a sequence $\{u_k\}$ converging to u. From the definition 1.3.9 (1)

$$\forall k,\; \exists \{u_k^j\}\; s.t.\, u_k^j \underset{j}{\to} u_k, \quad F'(u_k) = \liminf_j F_j(u_k^j).$$

Therefore, for all k we can find a j_k such that $j_k \ge j_{k-1}$, $d(u_k^{j_k}, u_k) < \frac{1}{k}$ and $F_{j_k}(u_k^{j_k}) < F'(u_k) + \frac{1}{k}$. Then we define

$$v_j := \begin{cases} u_k^{j_k}, & \text{if for some } k\ j = j_k \\ u, & \text{otherwise.} \end{cases}$$

We can see that $v_j \underset{j}{\to} u$, because $u_k^{j_k} \underset{k}{\to} u$. Then by the definition of $\Gamma - \liminf$:

$$F'(u) \le \liminf_j F_j(v_j) \le \liminf_k F_{j_k}(u_k^{j_k}) \le \liminf_k F'(u_k) + \frac{1}{k} = \liminf_k F'(u_k).$$

This shows that the $\Gamma - \liminf$ is lower semicontinuous. The same can be proven for the $\Gamma - \limsup$ with a similar argument. $\square$

Corollary 1.3.11. *If the Γ-limit exists then it is lower semicontinuous.*

Furthermore, we can use this semicontinuity for this extension result for the limsup inequality.

Proposition 1.3.12. *Let $\{F_n\}$ and F functionals over a metric space X and let $Y \subset X$ such that for any $x \in X$ exists $\{y_n\} \subset Y$ satisfying:*

$$y_n \to x \quad \bigwedge \quad F(y_n) \to F(x).$$

Then, calling $F'' = \Gamma - \limsup_n F_n$,

$$F''(y) \le F(y), \quad \forall y \in Y \implies F''(x) \le F(x), \quad \forall x \in X.$$

Proof. From the convergence of the restrictions we have:

$$F''(u) \leq F(u), \qquad \forall u \in Y.$$

For every $x \in X$ we take the sequence $\{y_n\}$ as in the hypothesis. Then, using the semicontinuity of F'' we conclude

$$F''(x) \leq \liminf_n F''(y_n) \leq \liminf_n F(y_n) = F(x).$$

□

38

Chapter 2

The Rudin-Osher-Fatemi problem

In this chapter will be presented the problem core of this thesis, the Rudin-Osher-Fatemi model, that is an application of calculus of variation in image analysis. The setting is denoising, even though the approach we will develop has applications also in other image analysis problems.

This model consists in the minimization of a functional involving the total variation, with the desired effect of reducing the oscillations given by the noise without introducing a blurring effect.

The study of this problem will proceed through the proof of existence and uniqueness of the solution, finding the associated Euler-Lagrange equation and in conclusion an example of analytic solution, which will be achieved finding the optimal level sets.

2.1 Variational continuous models

We model a image as a function $g : \Omega \longrightarrow \mathbb{R}$ so that at each point $x \in \Omega$ associate the intensity of grey $g(x)$ (you can simply consider a triplet of functions for the RGB representation). For analytic purpose, we set the model on a continuous domain Ω, but since the aim is working with digital pictures we take $\Omega \subset \mathbb{R}^2$, bounded and with at least Lipschitz boundaries, we will often consider $\Omega = [0,1]^2$. For the same reason we assume $g \in L^\infty(\Omega)$ and non negative, therefore it is not restrictive to take $\mathrm{Im}(g) \subseteq [0,1]$.

Getting into the problem, given a noisy image $g = \bar{g} + \sigma \cdot \varepsilon$, where $\bar{g}$ is the original image and $\sigma \cdot \varepsilon$ is a white noise of scale σ, we want to reconstruct as accurately as possible the original $\bar{g}$. This can be translated in a variational setting with an optimization problem of this kind:

$$\min_{u\in X\cap L^2(\Omega)} \lambda F(u) + \frac{1}{2}\|u-g\|^2_{L^2(\Omega)}, \tag{2.1}$$

whose minimizer is the reconstructed image. The objective function consists of two terms, the second $\frac{1}{2}\|u-g\|^2_{L^2(\Omega)}$ pushes the minimizer to be close to the reference image, indeed we expect the noise to change the original image not drastically. The first term, represented by the functional $F : X \longrightarrow \mathbb{R}$, has to behave, instead, in a "*regularizing*" way, such as to mitigate the vibrating effect of the noise. $\lambda > 0$ is a tunable parameter representing the ratio of relevance between the two terms.

With this idea of regularization, we can take F as the Dirichlet functional, that is:

$$F(u) = \int_\Omega |\nabla u|^2 \, dx,$$

with the domain $X = H^1(\Omega)$. This choice leads to the Euler-Lagrange equation:

$$-\Delta u + u - g = 0, \tag{2.2}$$

which is known to have an extremely regularizing effect. This is actually not ideal for images, because on the depicted objects' contours the gray intensity will present a discontinuity. Therefore we would like to preserve this kind of irregularities, whereas the previous proposal will produce a blur effect, making the edges unrecognizable (see figure 2.1).

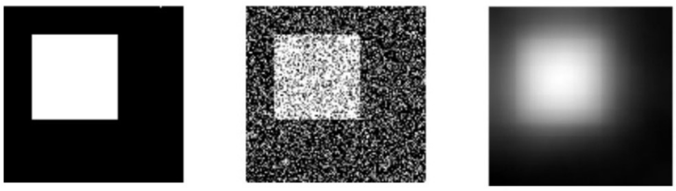

Figure 2.1: In order: original image, noisy reference, reconstruction with the Dirichlet term.

In order to preserve the contours, D. Mumford and J. Shah [4] thought of introducing a 1-dimensional curve $\Gamma \subset \Omega$, representing the edges, so that an equation like (2.2) would be valid only on $\Omega \setminus \Gamma$. Then, we minimize on all the possible edges, punishing its length, so as to avoid confusing the noise for a contour. In formulas the problem become:

$$\min_{\Gamma} \min_{u\in H^1(\Omega\setminus\Gamma)} \frac{1}{2}\|u-g\|^2_{L^2(\Omega)} + \lambda \int_{\Omega\setminus\Gamma} |\nabla u|^2 \, dx + \mu\, \mathcal{H}^1(\Gamma). \tag{MS}$$

This problem can be very challenging to approach not only computationally, but also analytically for the lack of convexity in general. Therefore, usually is better to find a sequence of convex functionals Γ - converging to the Mumford - Shah functional. Some approximations are presented in [5], as well as a deeper analysis on this problem. A noteworthy approximation is the one given by L. Ambrosio and V. M. Tortorelli:

$$AT_\varepsilon(z,u) = \frac{1}{2}\|z(u-g)\|^2_{L^2(\Omega)} + \lambda \int_\Omega z|\nabla u|^2 \,\mathrm{d}x + \frac{\beta}{2}\int_\Omega \left\{\varepsilon|\nabla z|^2 + \varepsilon^{-1}(1-z)^2\right\} \mathrm{d}x,$$

where z is a function such that $0 \le z(x) \le 1$, with the purpose of relaxing the contour positioning: z is close to 0 in proximity of an edge. For $\varepsilon \longrightarrow 0^+$, AT_ε does not Γ-converge exactly to Mumford - Shah functional in (MS), but to a slight variation where the double minimization is changed considering $u \in SBV(\Omega)$ and $\Gamma = J_u$ (the jump set of u). In other word we have:

$$AT_\varepsilon(z,u) \xrightarrow{\Gamma} F(z,u),$$

with $F(z,u) = \frac{1}{2}\|u-g\|^2_{L^2(\Omega)} + \lambda \int_{\Omega\setminus\Gamma} |\nabla u|^2 \, dx + \mu\, \mathcal{H}^1(J_u)$, for $u \in SBV(\Omega)$ and $z \equiv 1$, $F(z,u) = \infty$ otherwise. See [6] for the proof. This approach has produced remarkable results, also in edge detection, but at the price of theoretical and implementation difficulties.

2.2 The ROF model

Another choice of F for (2.1) can be the total variation. Indeed, it will tend to dampen the oscillations, without introducing a diffusive effect. This leads to the Rudin-Osher-Fatemi problem:

$$\min_{u \in BV(\Omega)\cap L^2(\Omega)} \lambda V(u,\Omega) + \frac{1}{2}\|u-g\|^2_{L^2(\Omega)}. \tag{ROF}$$

This problem is defined for $\Omega \in \mathbb{R}^N$, even though in the application on image analysis $N = 2$, for which $BV(\Omega)$ is continuously embedded in $L^2(\Omega)$ (see the observation 1.1.19). However, for the rest of this section we will present the result in the more general setting of free N.

We can immediately see that this problem is convex and, furthermore, the objective function is strictly convex, thanks to the fidelity term $\frac{1}{2}\|u-g\|^2_{L^2(\Omega)}$. Therefore, this minimization problem has at most one solution. To further investigate the existence of a minimizer it is useful to restrict the domain, showing that the solution is bounded.

Proposition 2.2.1. *Let $g \in L^\infty(\Omega)$ with $m \leq g \leq M$ almost everywhere and define $\mathcal{B} := \{u \in L^\infty(\Omega) : m \leq u \leq M \text{ a.e.}\}$, then*

$$\inf_{u \in BV(\Omega) \cap L^2(\Omega)} \lambda V(u, \Omega) + \frac{1}{2}\|u - g\|^2_{L^2(\Omega)} = \inf_{u \in BV(\Omega) \cap \mathcal{B}} \lambda V(u, \Omega) + \frac{1}{2}\|u - g\|^2_{L^2(\Omega)}.$$

Proof. At first we notice that $\mathcal{B} \subset L^2(\Omega)$, since the domain Ω is bounded, therefore we only have to show that the infimum on $BV(\Omega) \cap \mathcal{B}$ is lower than the one on $BV(\Omega) \cap L^2(\Omega)$. Take $u \in BV(\Omega) \cap L^2(\Omega)$ and define $\bar{u} = m \vee (u \wedge M)$, then clearly $|u - g| \geq |\bar{u} - g|$ and consequently the fidelity term decreases. Instead, for the total variation part we recall the inequality 1.1.26, then:

$$V(\bar{u}, \Omega) \leq V(u, \Omega),$$

showing in conclusion that for every $u \in BV(\Omega) \cap L^2(\Omega)$ there exists a $\bar{u} \in BV(\Omega) \cap \mathcal{B}$ such that:

$$\lambda V(\bar{u}, \Omega) + \frac{1}{2}\|\bar{u} - g\|^2_{L^2(\Omega)} \leq \lambda V(u, \Omega) + \frac{1}{2}\|u - g\|^2_{L^2(\Omega)}.$$

□

This property is relevant because the images are encoded in computers with a bounded range of intensity, [0, 255], and we would like to find a solution that still satisfies this intrinsic boundaries. Therefore, from now on we will consider the minimization of (ROF) to happen on

$$\mathcal{U} = \{u \in BV(\Omega) \cap L^\infty(\Omega) |\ u \geq 0,\ \|u\|_\infty \leq \|g\|_\infty\}, \tag{2.3}$$

which still is a convex set.

Observation 2.2.2. The set $\mathcal{U}$ is also compact in $L^1(\Omega)$. Indeed, $BV(\Omega)$ is compact, while the condition $0 \leq u \leq \|g\|_\infty$ is preserved almost everywhere through the L^1 convergence. Hence, $\mathcal{U}$ is the intersection between a compact and a closed set.

Observation 2.2.3. The advantage of working with essentially bounded functions on a bounded domain is that the L^1 and L^2 convergence are equivalent since:

- $\|u\|_{L^1(\Omega)} \leq \sqrt{|\Omega|}\|u\|_{L^2(\Omega)}$,
- $\|u\|^2_{L^2(\Omega)} \leq \|u\|_{L^\infty(\Omega)}\|u\|_{L^1(\Omega)}$.

This allows us to deal simultaneously with the total variation, which works well with the L^1 convergence, and with the L^2 norm of the fidelity term.

Proposition 2.2.4. *The ROF problem admits a unique solution.*

Proof. For the direct method we need to prove that the objective function is bounded from below (but clearly it is positive) and lower semicontinuous on a compact in the L^1 norm. The total variation is lower semicontinuous while for the previous observation the fidelity term is continuous on $\mathcal{U}$ with respect to the L^1 convergence, then the objective function is l.s.c. Then for compactness of $\mathcal{U}$ we can conclude the existence of a minimum, which has to be unique for strict convexity. □

2.2.1 Euler-Lagrange equation

Going on with the standard variational analysis, we will now find the Euler-Lagrange equation associated with this problem. We could compute the Gâteaux derivative (which is not straightforward for the total variation) and pose it to 0 to get:

$$-\lambda \mathrm{div}\left(\frac{\mathrm{d}Du}{\mathrm{d}|Du|}\right) + u - g = 0, \tag{2.4}$$

where $\frac{\mathrm{d}Du}{\mathrm{d}|Du|}$ is the Radon derivative. However, it is more meaningful to use a different approach, which will lead to a formally different equation. For this purpose we need to recall some convex analysis definitions and results, that will be fundamental through out the rest of the thesis. For more insights on convex analysis see [7].

Definition 2.2.5. Given an Hilbert space X and a convex function $F : X \longrightarrow \mathbb{R} \cup \{+\infty\}$, we define the subgradient of F in the point $x \in X$ as the set

$$\partial F(x) = \{v \in X \mid F(y) \geq F(x) + \langle v, y - x\rangle, \forall y \in X\}.$$

Clearly, if $X = \mathbb{R}^N$ and F is convex and differentiable in x, then

$$\partial F(x) = \{\nabla F(x)\},$$

in these cases we may use the simplified notation $\partial F(x) = \nabla F(x)$. This result can be generalized when F is Gâteaux - differentiable, indeed let's call $F'(x)$ such derivative and take $p \in \partial F(x)$, then

$$F(x + hv) \geq F(x) + \langle p, hv\rangle,$$

$$\frac{F(x + hv) - F(x)}{h} \geq \langle p, v\rangle.$$

Then moving to the limit $h \longrightarrow 0^+$, we have $\langle F'(x), v\rangle \geq \langle p, v\rangle$. But for arbitrariness of $v \in X$ we must have $F'(x) = p$.

As in the previous definition, we will refer to functions on an Hilbert space X and codomain in $(-\infty, +\infty]$, then we recall that the domain is the set

$$\mathrm{dom}F = \{x \in X \mid F(x) < +\infty\}.$$

Definition 2.2.6. A function F is called *proper* if $\mathrm{dom}F \neq \emptyset$.

We will call $\Gamma^0(X)$ the set of proper, convex and lower semicontinuous functions on X.

Definition 2.2.7. We define the Legendre conjugate function F^* of F as

$$F^*(p) = \sup_{x \in X} \langle p, x \rangle - F(x),$$

for every $p \in X' \simeq X$.

If F is proper and convex, then $\langle p, x \rangle - F(x)$ admits a maximum for some p, that is F^* is proper. Furthermore, F^* is also convex (is a supremum of affine functions), and more in general the conjugation maps Γ^0 in itself. It follows a fundamental theorem in convex analysis.

Theorem 2.2.8. $F^{**} = \sup \{ f \mid f \in \Gamma^0,\ f \leq F \}$.

From this we can immediately observe that $F \in \Gamma^0 \implies F^{**} = F$.
We conclude this summary with three relevant propositions.

Proposition 2.2.9. *For F convex*

$$p \in \partial F(x) \iff \langle p, x \rangle - F(x) = F^*(p).$$

In particular, if $F \in \Gamma^0$,

$$p \in \partial F(x) \iff x \in \partial F^*(p).$$

Proof. The fact $p \in \partial F(x)$ means $F(y) \geq F(x) + \langle p, y - x \rangle$, $\forall y \in X$, that is equivalent to

$$\langle p, y \rangle - F(y) \leq \langle p, x \rangle - F(x), \quad \forall y \in X.$$

But then we can extract the supremum over y to get $F^*(p) \leq \langle p, x \rangle - F(x) \leq F^*(p)$, proving the first equality. If further $F^{**} = F$, then we also have $\langle p, x \rangle - F^{**}(x) = F^*(p)$, that is $x \in \partial F^*(p)$.

□

Proposition 2.2.10. *For F convex, x is a minimizer of F if and only if $0 \in \partial F(x)$.*

Proof. It is a direct consequence of

$$F(y) \geq F(x) \iff F(y) \geq F(x) + \langle 0, y - x \rangle.$$

□

Proposition 2.2.11. *Let F and G be convex functions such that*

$$\operatorname{dom} G \cap \operatorname{int}(\operatorname{dom} F) \neq \emptyset,$$

then

$$\partial(F+G) = \partial F \oplus \partial G.$$

Proof. Actually if $p \in \partial F(x)$ and $q \in \partial G(x)$, then clearly $p+q \in \partial(F+G)(x)$. Hence,

$$\partial(F+G) \subset \partial F \oplus \partial G$$

is always verified.

For the reverse inclusion we have to show that $\forall p \in \partial(F+G)(x)$ there exist a $q_F \in \partial F(x)$ and a $q_G \in \partial G(x)$ such that $p = q_F + q_G$. Take $p \in \partial(F+G)(x)$, then for every $y \in X$ we have:

$$F(y) + G(y) \geq F(x) + G(x) + \langle p, y-x \rangle,$$

$$G(x) - G(y) \leq F(y) - F(x) - \langle p, y-x \rangle.$$

$$\frown{H}(y) \leq \smile{H}(y),$$

where $\frown{H}(y) = G(x) - G(y)$ is concave while $\smile{H}(y) = F(y) - F(x) - \langle p, y-x \rangle$ is convex, such that $\frown{H}(x) = \smile{H}(x) = 0$. Let's call $C_1, C_2 \subset X \times \mathbb{R}$ respectively the epigraph of $\smile{H}$ and the hypograph of $\frown{H}$. Those are convex sets intersecting only on the boundary, therefore they can be separated by an hyperplane Π. By the hypothesis $\operatorname{dom} G \cap \operatorname{int}(\operatorname{dom} F) \neq \emptyset$, Π cannot be orthogonal to X, that is it can be represented as the graph of an affine function $\alpha + \langle q, y \rangle$. This means that we must have:

$$\frown{H}(y) \leq \alpha + \langle q, y \rangle \leq \smile{H}(y),$$

and choosing $y = x$ we also have $\alpha = -\langle q, x \rangle$. Then

$$G(x) - G(y) \leq \langle q, y-x \rangle \leq F(y) - F(x) - \langle p, y-x \rangle$$

$$\Longrightarrow \begin{cases} F(y) \geq F(x) + \langle p+q, y-x \rangle \\ G(y) \geq G(x) + \langle -q, y-x \rangle. \end{cases}$$

In conclusion, we found the desired decomposition with $q_F = p+q \in \partial F(x)$ and $q_G = -q \in \partial G(x)$. $\square$

Coming back to the Euler - Lagrange equation of the (ROF) problem, from 2.2.10 we can state

$$\partial(\lambda V(u) + \frac{1}{2}\|u-g\|^2_{L^2}).$$

Note that we are not going to write anymore the dependence from Ω explicitly, if not necessary. For the theory here presented we need to set the problem in an Hilbert space, therefore we consider V as a functional on L^2 but with $\mathrm{dom}V = BV(\Omega)$. Since the domain of the fidelity term is the whole L^2, we can apply 2.2.11:

$$0 \in \lambda \partial V(u) + u - g. \tag{2.5}$$

This can be already consider the Euler - Lagrange equation, but let's go on characterizing the subgradient. From 2.2.9 we have

$$\partial V(u) = \{p \mid V(u) = \langle p, u\rangle - V^*(p)\}.$$

Then we need to compute the conjugate, but this can be simply done noticing that V is itself a conjugate of another functional:

$$V(u) = \sup_{\phi \in \mathcal{K}} \int_\Omega -u \operatorname{div} \phi \, \mathrm{d}x = \sup_{\phi} \langle u, -\operatorname{div}\phi\rangle - \mathbb{I}_{\mathcal{K}}(\phi),$$

where $\mathcal{K} := \{\phi \in C_c^1(\Omega; \mathbb{R}^2) \mid |\phi| \leq 1\}$, while $\mathbb{I}_{\mathcal{K}}$ is called indicator function and it is defined as

$$\mathbb{I}_{\mathcal{K}}(\phi) = \begin{cases} 0 & \text{if } \phi \in \mathcal{K} \\ +\infty & \text{otherwise.} \end{cases}$$

We can now see that V is the conjugate of the indicator function on the set

$$\mathcal{K}_{\mathrm{div}} := \left\{ -\operatorname{div}\phi \mid \phi \in \left[C_c^1(\Omega)\right]^N, \ |\phi| \leq 1 \right\},$$

that is $V(u) = \mathbb{I}^*_{\mathcal{K}_{\mathrm{div}}}(u)$. Consequently, from theorem 2.2.8, V^* is the convex lower semicontinuous envelope of $\mathbb{I}_{\mathcal{K}_{\mathrm{div}}}$, that is the indicator function on the closure of $\mathcal{K}_{\mathrm{div}}$ (because this set is already convex). Such a closure turns out to be

$$K = \left\{ -\operatorname{div} z \in L^2(\Omega) \mid |z| \leq 1 \ a.e.,\ z \cdot \nu = 0 \text{ in } \partial\Omega \right\}, \tag{2.6}$$

where the divergence is meant in distributional sense and $z \cdot \nu$ is the orthogonal component to the boundary, defined in a weak sense like a trace operator.

Proposition 2.2.12. *Take $\{p_n\} \subset \mathcal{K}_{\mathrm{div}}$ s.t. $p_n \xrightarrow{L^2} \xi$, then $\xi \in K$.*

Proof. The fact that $p_n \in \mathcal{K}_{\mathrm{div}}$ implies that there is a $\phi_n \in [C_c^1(\Omega)]^N$ such that $\|\phi\|_{L^\infty} \leq 1$ and $p_n = -\operatorname{div} p_n$. Therefore, for all n, $\|\phi_n\|_{L^2(\Omega)} \leq \sqrt{|\Omega|}$, that is the sequence is bounded in L^2. Then, there exists a sub sequence of ϕ_n and a $z \in L^2(\Omega)$ such that

$$\phi_n \xrightharpoonup{L^2} z \implies \phi_n \longrightarrow z \quad \text{in } \mathcal{D}',$$

but then $p_n \longrightarrow -\mathrm{div} z$, hence $\xi = -\mathrm{div} z$.

Let's now take $\psi \in [C^1(\Omega)]^N$ not necessarily with compact support, for the L^2 convergences we still have:

$$\langle -\mathrm{div}(\phi_n), \psi \rangle = \langle \phi_n, \nabla \psi \rangle \;\longrightarrow\; \langle -\mathrm{div}(z), \psi \rangle = \langle z, \nabla \psi \rangle,$$

even though in general $\langle \mathrm{div}(z), \psi \rangle + \langle z, \nabla \psi \rangle = \int_{\partial\Omega} \psi z \cdot \nu \, \mathrm{d}x$. This implies that $z \cdot \nu = 0$ weakly on $\partial\Omega$.

Define $T_k \coloneqq \{|z| \geq 1 + \frac{1}{k}\}$ and suppose $|T_k| > 0$. Since we have

$$\int_{T_k} |z| \, \mathrm{d}x = \sup \left\{ \int_{T_k} z \cdot \psi \, \mathrm{d}x \mid \psi \in \left[C_c^1(\Omega)\right]^N, \ |\psi| \leq 1 \right\},$$

then for every ε we can find a $\psi_\varepsilon \in [C_c^1(\Omega)]^N$ such that $|\psi_\varepsilon| \leq 1$ and

$$\left| \int_{T_k} z \cdot \psi_\varepsilon \, \mathrm{d}x \right| > \int_{T_k} |z| \, \mathrm{d}x - \varepsilon \geq \left(1 + \frac{1}{k}\right) |T_k| - \varepsilon.$$

However, $\left| \int_{T_k} \phi_n \cdot \psi_\varepsilon \, \mathrm{d}x \right| \leq |T_k|$, then, moving to the limit on $n \to +\infty$, we have

$$|T_n| \geq \left| \int_{T_k} z \cdot \psi_\varepsilon \, \mathrm{d}x \right| > \left(1 + \frac{1}{k}\right) |T_k| - \varepsilon,$$

which is impossible for ε small enough. Hence, by contradiction, $|T_k| = 0$ for every k, that is $|z| \leq 1$ a.e. □

Summing up, we have that $V^*(p) = \mathbb{I}_K(p)$ and consequently we have:

$$\partial V(u) = \{p \in K \mid V(u) = \langle p, u \rangle\}. \tag{2.7}$$

Proposition 2.2.13. *Consider $u \in BV(\Omega) \cap L^2(\Omega)$ and $p = -\mathrm{div}\, z \in K$, then we have*

$$\langle p, u \rangle = z \cdot Du(\Omega).$$

Proof. Notice that $\langle p, u \rangle = \int_\Omega -\mathrm{div}(z) u \, \mathrm{d}x$ and $\int_\Omega z \cdot \mathrm{d}Du = z \cdot Du(\Omega)$, so basically we have to show the derivative switch property $\int_\Omega -\mathrm{div}(z) u \, \mathrm{d}x = \int_\Omega z \cdot \mathrm{d}Du$. This is true for every $z \in [C_c^1(\Omega)]^n$, then we want to prove it by a density argument. We can see that as long as $\mathrm{div}\, \phi_n \longrightarrow \mathrm{div}\, z$ in L^2, then

$$\lim_{n \longrightarrow \infty} \int_\Omega -\mathrm{div}(\phi_n) u \, \mathrm{d}x = \int_\Omega -\mathrm{div}(z) u \, \mathrm{d}x,$$

because $u \in L^2(\Omega)$. As for the other integral, let's at first notice that $\int_\Omega z \cdot \mathrm{d}Du$ makes sense since z is bounded and $u \in BV(\Omega)$. For now let's take $z = c\chi_A$, for c a constant vector and A a $Du-$measurable set. Then, by construction with mollifiers, each component of ϕ_n satisfies

$$c_i \chi_{A_n^-} \leq (\phi_n)_i \leq c_i \chi_{A_n^+},$$

with $\bigcup_n A_n^- = A = \bigcap_n A_n^+$. Therefore, we have:

$$c_i \partial_{x_i} u(A_n^-) \leq \int_\Omega \phi_n \, \mathrm{d}\partial_{x_i} u \leq c_i \partial_{x_i} u(A_n^+).$$

Moving to the limit, and exploiting the inner and outer regularity of radon measure, we get the desired result $\lim_n \int_\Omega \phi_n \cdot \mathrm{d}Du = \int_\Omega z \cdot \mathrm{d}Du$. As a generalization, we know $\phi_\varepsilon = z \star \rho_\varepsilon$ is converging in L^2 to z and also its divergence to $\operatorname{div} z$. Call $\tilde{z}_n$ the sequence of simple functions for which we can approximate $\int_\Omega z \cdot \mathrm{d}Du$ and $\tilde{\phi_n} = \tilde{z}_n \star \rho_{\varepsilon_n}$ such that $\left|\tilde{\phi_n} \cdot Du(\Omega) - \tilde{z}_n \cdot Du(\Omega)\right| < \frac{1}{n}$. Then, calling $\phi_n = \phi_{\varepsilon_n}$, we have

$$\begin{aligned} |\phi_n \cdot Du(\Omega) - z \cdot Du(\Omega)| \leq & \left|\tilde{\phi_n} \cdot Du(\Omega) - \phi_n \cdot Du(\Omega)\right| + \\ & + \left|\tilde{\phi_n} \cdot Du(\Omega) - \tilde{z}_n \cdot Du(\Omega)\right| + |z \cdot Du(\Omega) - \tilde{z}_n \cdot Du(\Omega)|, \end{aligned}$$

which is converging to 0. □

Summing up (2.5), (2.6), (2.7) and the previous preposition, we can finally write a more explicit form for the Euler - Lagrange equation:

$$\begin{cases} -\lambda \operatorname{div} z + u = g & \text{a.e. } x \in \Omega, \\ |z| \leq 1 & \text{a.e. } x \in \Omega, \\ z \cdot \nu = 0 & \text{on } \partial\Omega, \\ z \cdot Du = |Du|. \end{cases} \tag{2.8}$$

This is consistent with (2.4) with Neumann boundary conditions, since $z = \frac{\mathrm{d}Du}{\mathrm{d}|Du|}$ satisfy the constraints on z.

Before going on with some solution technique we will present a regularity result on the ROF minimizer. A complete proof of this can be found in [8].

Theorem 2.2.14 (Caselles-Chambolle-Novaga)**.** *Assume the domain $\Omega \in \mathbb{R}^N$ convex with $N \leq 7$. Call u the minimizer of ROF, then if g is uniformly continuous, also u is. In particular, given a function $\omega(t)$ continuous, nondecreasing and with $\omega(0) = 0$, then*

$$|g(x) - g(y)| \leq \omega(|x - y|) \quad \forall x, y \in \Omega \implies |u(x) - u(y)| \leq \omega(|x - y|) \quad \forall x, y \in \Omega.$$

2.2.2 Solution by level sets

In the following section we will present the relation between the (ROF) minimizer and the solution of the following problem:

$$\min_E \lambda P(E, \Omega) + \int_E s - g(x)\,\mathrm{d}x, \tag{ROFs}$$

where E is a Caccioppoli set and $s \in \mathbb{R}$ a parameter. Similarly to 1.1.30 this problem admits a solution, which we will refer to as E_s. We stress that this E_s is not uniquely defined, but we can achieve a unicity result through the following lemma.

Lemma 2.2.15. *Take $s, s' \in \mathbb{R}$ and $E_s, E_{s'}$ any pair of solution of ROF and ROFs, then*

$$s < s' \implies E_{s'} \subseteq E_s \quad \textit{(up to a negligible set of points).}$$

Proof. Let's call $J_s(E) \coloneqq \lambda P(E, \Omega) + \int_E s - g(x)\,\mathrm{d}x$. By minimality of E_s and $E_{s'}$ we have the inequality

$$J_s(E_s) \leq J_s(E_s \cup E_{s'}), \quad \bigwedge \quad J_{s'}(E_{s'}) \leq J_{s'}(E_s \cap E_{s'}),$$
$$\implies J_s(E_s) + J_{s'}(E_{s'}) \leq J_{s'}(E_s \cap E_{s'}) + J_s(E_s \cup E_{s'}).$$

However, from 1.1.8 (1) we have $P(E_s \cap E_{s'}, \Omega) + P(E_s \cup E_{s'}, \Omega) \leq P(E_s, \Omega) + P(E_{s'}, \Omega)$, therefore the previous inequality reduces to:

$$\int_{E_s} s - g(x) + \int_{E_{s'}} s' - g(x) \leq \int_{E_s \cup E_{s'}} s - g(x) + \int_{E_s \cap E_{s'}} s' - g(x),$$

$$\int_{E_{s'} \setminus E_s} s' - g(s)\,\mathrm{d}x \leq \int_{E_{s'} \setminus E_s} s - g(s)\,\mathrm{d}x \implies s'|E_{s'} \setminus E_s| \leq s|E_{s'} \setminus E_s|.$$

But, since $s < s'$, this is possible only if $|E_{s'} \setminus E_s| = 0$, which proves the lemma. ☐

Because from this monotonicity result we have that any solution E_s is contained in all the $E_{s'}$ with $s' < s$ and contains all those solutions for $s' > s$, then there must exist E_s^+ and E_s^- such that:

$$\bigcup_{s' > s} E_{s'}^+ = E_s^- \subseteq E_s \subseteq E_s^+ = \bigcap_{s' < s} E_{s'}^-.$$

We recall that the Caccioppoli sets are classes of equivalence, but we are considering the representative satisfying the (1.6) property, this makes those intersection and union of uncountable elements well defined.

Proposition 2.2.16. *The ROFs problem admits a unique solution for almost every s.*

Proof. The uniqueness comes when E_s^+ and E_s^- are equivalent, that is $|E_s^+ \setminus E_s^-| = 0$. To show this, consider the intersection of two of those set differences (for $t > s$):

$$(E_s^+ \setminus E_s^-) \cap (E_t^+ \setminus E_t^-) = (E_t^+ \cap E_s^+) \cap (\overline{E_t^-} \cap \overline{E_s^-}).$$

But $E_t^+ \subseteq E_s^- \subseteq E_s^+$, while $E_t^- \subseteq E_s^-$, consequently:

$$(E_s^+ \setminus E_s^-) \cap (E_t^+ \setminus E_t^-) = E_t^+ \cap \overline{E_s^-} \subseteq E_s^- \cap \overline{E_s^-} = \emptyset.$$

In this way we found an uncountable amount of disjoint sets, which implies that at most countably many of them have measure bigger than zero. □

Finally, the relation between ROF and ROFs is expressed as follows.

Theorem 2.2.17. *Let E_s the solution of ROFs for each s (for which we have unicity), then*

$$u(x) = \sup\{s \in \mathbb{R} : x \in E_s\}$$

solves ROF. Conversely, if u is a minimizer of ROF, then $E_s^- = \{u > s\}$ and $E_s^+ = \{u \geq s\}$.

Observation 2.2.18. Notice that indeed with the sup definition of u we have:

$$\{u > s\} = \bigcup_{s' > s} E_{s'} = E^-,$$

and conversely

$$\{u < s\} = \{x : \exists\, s' < s \text{ s.t. } x \notin E_{s'}\} = \bigcup_{s' < s} \overline{E_{s'}} = \overline{E_s^+}.$$

Observation 2.2.19. It is useful to notice that if E_s minimizes ROFs, then $\Omega \setminus E_s$ minimizes the following problem:

$$\min_E \lambda P(E, \Omega) + \int_E g(x) - s \,\mathrm{d}x. \qquad \text{(ROFs conj)}$$

Indeed, take $\Omega \setminus E$, then

$$\lambda P(\Omega \setminus E, \Omega) + \int_{\Omega \setminus E} g(x) - s \,\mathrm{d}x = P(E, \Omega) + \int_E s - g(x) \,\mathrm{d}x + \int_\Omega g(x) - s \,\mathrm{d}x,$$

which is minimized by E_s. This relation of minimizers is reflected in the ROF problem by the fact that if we change g with $-g$, then the solution move from u to $-u$. Furthermore, $\Omega \setminus E_s^+ = \{u < s\} = \{-u > -s\}$.

Proof. At first we have to show that $u \in L^2$. From minimality we have:

$$\lambda P(E_s, \Omega) + \int_{E_s} s - g(x), \mathrm{d}x \leq \lambda P(\emptyset, \Omega) + \int_{\emptyset} s - g(x), \mathrm{d}x = 0,$$

$$\implies s|E_s| \leq \int_{E_s} g(x) \, \mathrm{d}x.$$

We need to integrate this inequality between t and M, where t is such that $E_t \neq \emptyset$ and $u(x) \not\equiv t$ on E_t. Notice that.

$$\lambda P(E, \Omega) + \int_E s - g(x) \, \mathrm{d}x \leq \lambda P(E, \Omega) + (s + \|g\|_\infty)|E|,$$

hence for s negative enough $\emptyset$ is not optimal. On the other hand, as soon as exists an s for which $E_s \neq \emptyset$, then for any $t < s$ we have $u(x) \not\equiv t$ on E_t, otherwise E_s would be empty. For simplicity we will assume $t = 0$ and $s > 0$, that is $E_s \subseteq E_0 \neq \emptyset$, then we can integrate

$$\int_0^M s|E_s| \, \mathrm{d}s \leq \int_0^M \int_{E_s} g(x) \, \mathrm{d}x \, \mathrm{d}s.$$

Furthermore, applying Fubini's or Tonelli's theorem where needed, we have:

$$\int_0^M s|E_s| \, \mathrm{d}s = \int_0^M \int_{E_0} s\chi_{E_s} \, \mathrm{d}x \, \mathrm{d}s = \int_{E_0} \int_0^{M \wedge u(x)} s \, \mathrm{d}s \, \mathrm{d}x = \frac{1}{2} \int_{E_0} (M \wedge u(x))^2 \, \mathrm{d}x,$$

$$\int_0^M \int_{E_s} g(x) \, \mathrm{d}x \, \mathrm{d}s = \int_{E_0} (M \wedge u(x)) \, g(x) \, \mathrm{d}x \leq \left(\int_{E_0} (M \wedge u(x))^2, \mathrm{d}x \int_{E_0} g(x)^2 \, \mathrm{d}x \right)^{\frac{1}{2}},$$

$$\implies \frac{1}{2} \int_{E_0} (M \wedge u(x))^2 \, \mathrm{d}x \leq \left(\int_{E_0} (M \wedge u(x))^2, \mathrm{d}x \int_{E_0} g(x)^2 \, \mathrm{d}x \right)^{\frac{1}{2}}.$$

In conclusion, taking $M \longrightarrow +\infty$, we have shown that

$$\int_{\{u>0\}} u(x)^2 \, \mathrm{d}x \leq 4 \int_{\{u>0\}} g(x)^2 \, \mathrm{d}x.$$

For the other half of the integral, we can apply a similar argument to the problem presented in the observation 2.2.19, coming to the result:

$$\int_{\{u<0\}} u(x)^2 \, \mathrm{d}x \leq 4 \int_{\{u<0\}} g(x)^2 \, \mathrm{d}x,$$

proving that u is in $L^2(\Omega) \subseteq L^1(\Omega)$. We can then prove that $U \in BV(\Omega)$ and u is minimal for ROF jointly. Let v be any function in $BV(\Omega) \cap L^2(\Omega)$. Because E_s

is minimal for ROFs, it has less energy than the level set $\{v > s\}$, then we can integrate on $s \in [-M, M]$, resulting in the following inequality:

$$\int_{-M}^{M} \lambda P(E_s, \Omega) + \int_{E_s} s - g(x)\, \mathrm{d}x\, \mathrm{d}s \leq \int_{-M}^{M} \lambda P(\{v > s\}, \Omega) + \int_{\{v>s\}} s - g(x)\, \mathrm{d}x\, \mathrm{d}s.$$

However, proceeding similarly as before, we have:

$$\int_{-M}^{M} \int_{\{v>s\}} s - g(x)\, \mathrm{d}x\, \mathrm{d}s = \int_{\Omega} \int_{-M}^{M} \chi_{\{v>s\}}(x)(s - g(x))\, \mathrm{d}s\, \mathrm{d}x =$$

$$= \frac{1}{2} \int_{\Omega} (u(x) \wedge M - g(x))^2 - (u(x) \wedge (-M) - g(x))^2\, \mathrm{d}x.$$

Now moving to the limit $M \longrightarrow +\infty$, we have that $u \wedge M \longrightarrow u$ and $u \wedge -M \longrightarrow 0$ in L^2, therefore, using the coarea formula:

$$\int_{-M}^{M} \lambda P(\{v > s\}, \Omega) + \int_{\{v>s\}} s - g(x)\, \mathrm{d}x\, \mathrm{d}s \xrightarrow[M \to +\infty]{} \lambda V(v, \Omega) + \frac{1}{2} \int_{\Omega} (v - g)^2\, \mathrm{d}x.$$

Proceeding in the same way for $E_s^- = \{u > s\}$, we get

$$\lambda V(u, \Omega) + \frac{1}{2} \|u - g\|_{L^2}^2 \leq \lambda V(v, \Omega) + \frac{1}{2} \|v - g\|_{L^2}^2,$$

showing that $u \in BV(\Omega)$ and minimal for ROF.

□

Before applying this formulation by level sets to some particular cases, we present a regularity result of the minimizer's boundary, derived from a more general result in [1].

Theorem 2.2.20. *Let $g \in L^\infty(\Omega)$ and E_s the minimizer of* (ROFs)*, then $\partial E_s \backslash \partial^* E_s$ is a closed set of Hausdorff dimension* 0*. Furthermore, $\partial^* E_s$ is locally the graph of $C^{1,1}$ functions.*

2.2.3 An analytical solution

In order to test the implementations of ROF solvers, it is useful to find some analytical solutions. We will take as a reference the simple case of $g = \chi_Q$, where $Q \subset \Omega = \mathbb{R}^2$ is a square. Setting Ω to be an unbounded set is actually relevant, because, as we will see, it forces the solution to be 0 outside of Q. The case of the reference as the characteristic of a convex set has been solved in general, for instance the [9] and [10] by Alter, Caselles and Chambolle provide a very detailed explanation. They solved it through the Euler-Lagrange equations, however their results are complex to prove and technical, for this reason we will prefer the level sets approach (with proposition 2.2.17), since we are only interested in the particular simple case of a square. For a general convex set C the solution of level sets appears in the form

$$E_s = \begin{cases} \emptyset & \text{if } R_s > R^* \vee s > 1 \\ C^{R_s} & \text{if } R_s \leq R^* \wedge s \geq 0 \\ \mathbb{R}^2 & \text{if } s < 0 \end{cases}, \tag{2.9}$$

where $R_s = \frac{\lambda}{1-s}$ represent the minimal radius of curvature of ∂E, R^* is a threshold depending on the geometry, instead C^R is defined as the union of all the balls $B(x, R)$ of radius R included in C, that is $C^R = \bigcup_{B(x,R)\subset C} B(x, R)$.

Even if we are only interested in the square, we will make use of some results taken for a general convex set, to begin with the simplified formulation. At first we have to notice that we only need to study what happens for $s \in (0,1)$, because from 2.2.1 we know that the image of the minimizer must be in $[0,1]$, hence $E_s^+ = \{u \geq s\} = \emptyset$ for $s > 1$, while $E_s^- = \{u > s\} = \Omega$ for $s < 0$. Otherwise we have the following result.

Lemma 2.2.21. *Let $g = \chi_C$ for $C \subset \mathbb{R}^2$ a convex set, then ROFs, for $s \in (0,1]$, is equivalent to*

$$\min_{E \subseteq C} \lambda P(E) - (1 - s)|E|. \tag{2.10}$$

Proof. The objective function in ROFs is $\lambda P(E) + \int_E s - \chi_C \, \mathrm{d}x$, for which we can compute the integral:

$$\lambda P(E) + s|E \setminus C| - (1 - s)|E \cap C|.$$

For $s > 0$ we can suppose E bounded, otherwise the term $s|E \setminus C|$ would be $+\infty$. Now let H be an half plane containing C, we can immediately understand that $P(E \cap H) \leq P(E)$ for each E. For instance, in dimension 2, the reduced boundary of $E \cap H$ is like the one of E where a portion has been replaced by a straight line, which minimizes the perimeter. We notice that this step would not work for $\Omega \neq \mathbb{R}^2$, because $P(\partial E \cap \partial\Omega, \Omega) = 0$, so that the perimeter does mot necessarily reduce on $E \cap H$.

Instead, $(E \cap H) \cap C = E \cap C$ and $(E \cap H) \setminus C \subset E \setminus C$. Summing up $E \cap H$ has a lower energy then E, for each E, that is $E_s \subset H$. But this happens for each halfspace $H \supset C$, therefore $E_s \subseteq C$. This allows us to consider only the subsets of C for the minimum problem, that with the right simplifications leads to the thesis. □

For $s = 0$, instead, $E = \mathbb{R}^2$ has finite energy equal to $-|C|$, however we notice that the objective function is $\lambda P(E) - |E \cap C| > -|E \cap C| \leq |C|$, for $E \neq \mathbb{R}^2$. Therefore, we must have $E_0 = \mathbb{R}^2$ too.

Before solving the square case, it is useful to make few observations. We can see and keep in mind that in (2.10) we are trying to reduce the perimeter while increasing the area.

A first consequence of this is that the minimizer has to be convex. Indeed consider a set E for which there is a segment external to E connecting two points on the boundary ∂E. Then calling D the portion included by this segment and ∂E, we have $\tilde{E} = E \cup D$ has a lower energy than E: we are adding more area while the length of the segment is for sure shorter than the arc on the boundary of E, reducing the perimeter. For a better understanding take as reference figure 2.2.

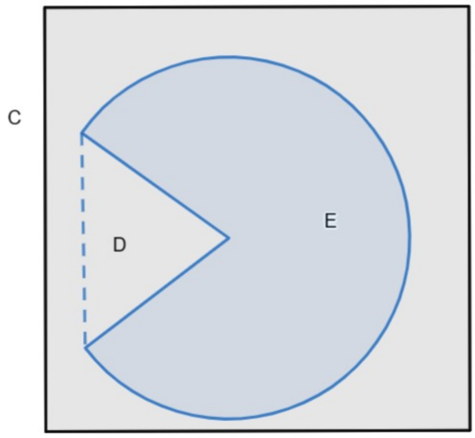

Figure 2.2: Example on how we can improve the energy convexifying the sets.

Secondly, suppose the minimizer $E_s \neq \emptyset$, then whatever shape it has, E_s must be the biggest version that fits in C. Indeed, take tE_s for t a scaling parameter greater than 1. Its energy is the parabola in t

$$p(t) = \lambda P(E_s)t - (1-s)|E_s|t^2,$$

with maximum for $t_M = \frac{\lambda P(E_s)}{2(1-s)|E_s|}$. However, because the empty set is not optimal, we must have $\lambda P(E_s) - (1-s)|E_s| \leq 0$, that is $t_M \leq 1$. Consequently, for $t > 1$ we should have

$$p(t) < p(1) = \lambda P(E_s) - (1-s)|E_s|.$$

Hence, if $tE_s \subseteq C$, then E_s would not be optimal, which is a contradiction. Then, combining this argument with proper shifts and rotation, we conclude.

Summing up we got:

Proposition 2.2.22. *The minimizer E_s of* (2.10) *satisfies the following properties:*

- *E_s is convex,*
- *E_s is either empty or the greatest scale of its shape that fits in C.*

Lastly, it is relevant to point out the connection between (2.10) and the Dido's or isoperimetric problem. The latter can be stated as: "*find the figure of maximal area, given the perimeter length B*". This is a classical problem, which induced the mathematicians to formulate some beginning ideas that have developed into calculus of variation. Its solution is notoriously the circle and in particular if we constrain a portion of the boundary to be a straight segment, the maximal area is achieved by the circular segment with perimeter B. The problem can be stated with a more mathematical notation as follows:

$$\max_{P(E)\leq B} |E|.$$

But changing the max for a min and introducing a Lagrange multiplier for the constraint, we have the equivalent saddle-point formulation:

$$\max_{\mu\geq 0} \min_{E} \mu P(E) - \mu B - |E|.$$

The minimization part is responsible for the choice of E's shape depending on μ, while the maximization selects the particular one that matches the perimeter length constraint. For example, in the case of the side fixed to be a straight line, the minimization find the general shape of the solution, that is the circular segment, while the maximization finds the optimal one among all circular segments. Therefore, the relevant problem is given by:

$$\min_{E} \mu P(E) - |E|, \tag{2.11}$$

which is visibly closed related to (2.10). Some more theoretical background on the isoperimetric problem can be found in [11] at Chapter 14.

Moving to the square example, we take $Q = \left[-\frac{L}{2}, \frac{L}{2}\right]^2$ and set $R_s = \frac{\lambda}{1-s} \in [\lambda, +\infty)$. This value is shown from the Euler-Lagrange equations to represent the maximum radius of curvature of $\partial^* E_s$, we will not prove this assertion, however the meaning of R_s will be clear at the end of the computations. Our purpose is to solve

$$\min_{E\subseteq Q} R_s P(E) - |E|. \tag{2.12}$$

Let's suppose for now that the minimizer is not empty, if the energy of the resulting minimal set is less than 0, then our guess was correct, otherwise we have $E_s = \emptyset$. The first consideration we can make is that the minimizer must have the same symmetries as the square whenever the solution is unique, otherwise at least E_s^+ and E_s^- have them, as they are respectively the union and the intersection of all possible solutions.

Taking into account the properties in 2.2.22, we must have that the optimum touches symmetrically each of the square sides (for scale maximality). Furthermore, the intersection between ∂E_s and one of the sides must be either a point or a continuous line (for convexity) and symmetrical with respect to the side central point.

Hence, we can deduce that (again by convexity) E_s must contain an octagon (or square when the contact consist only in the side central point) like the one showed in figure 2.3, even though the length of the overlapping edge has still to be determined. Before discussing the optimal length, which depends on R_s, let's figure out the optimal shape to fill the leftovers corners between the square and the octagon. That is, if we call the octagon G, we want to find the $\Delta E \subseteq \left(0, \frac{L}{2}\right)^2 \setminus G$ with lowest energy and constrained to share the side $S = \left(0, \frac{L}{2}\right)^2 \cap G$ with the octagon. Notice that in this case we assume that ΔE does not share part of the boundary with Q, except for the vertices of S, because we stated that $\partial E \cap \partial Q$ is fully captured by the octagon. In formulas, we can write the problem as:

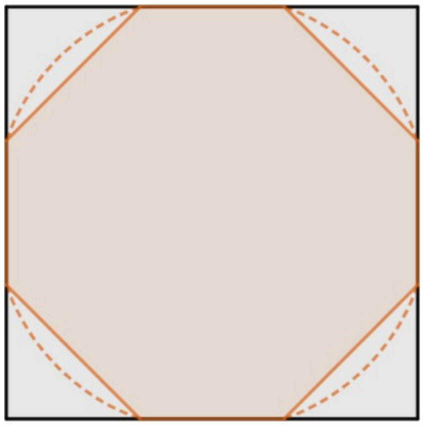

Figure 2.3: The non empty solutions of (2.12) must include a symmetrical octagon like the orange one in figure

$$\min \left\{ R_s P(\Delta E) - |\Delta E| : \; \Delta E \subseteq \left[0, \frac{L}{2}\right]^2 \setminus G, \, \partial\Delta E \supset S \right\}.$$

For the relation with the Dido's problem, this is solved by circular segments.

In this way we can describe the class of optimal shapes by only two parameters, one related to the length of the octagon side, while the other to the circular segments in the corners.

In particular we call $x \in \left[0, \frac{L}{2}\right]$ the cathetus length of the corner triangle, while $\theta \in \left[0, \frac{\pi}{2}\right]$ and r the angle and the radius representing the circular arc (see figure 2.4). the relation between these three variables is given by the length of the octagon skewed side, that is:

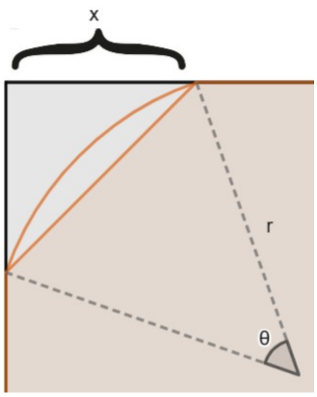

Figure 2.4

$$2r \sin \frac{\theta}{2} = \sqrt{2}x. \tag{2.13}$$

Let's call $E(x, \theta)$ the shape given with these parameters, then we can compute its perimeter and area:

$$P(E(x,\theta)) = 4(L - 2x + \theta r) = 4L - 4\left(2 - \sqrt{2}\frac{\theta/2}{\sin(\theta/2)}\right)x;$$

$$|E(x,\theta)| = L^2 - 2x^2 + 2\theta r^2 - 4r^2 \sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right) =$$

$$= L^2 - 2\left(1 - \frac{\theta/2}{\sin^2(\theta/2)} + \frac{\cos(\theta/2)}{\sin(\theta/2)}\right)x^2.$$

Consequently, E_s will be $E(x_{min}, \theta_{min})$ for the value x_{min} and θ_{min} minimizing, for $(x,\theta) \in \left[0, \frac{L}{2}\right] \times \left[0, \frac{\pi}{2}\right]$, the following function:

$$f(x,\theta) := 4R_sL - L^2 - 2b\left(\frac{\theta}{2}\right)R_sx + a\left(\frac{\theta}{2}\right)x^2,$$

with $a(\gamma) := 2\left(1 - \frac{\gamma}{\sin^2(\gamma)} + \frac{\cos(\gamma)}{\sin(\gamma)}\right)$ and $b(\gamma) := 2\left(2 - \sqrt{2}\frac{\gamma}{\sin(\gamma)}\right)$. Minimizing f on x is straightforward, therefore we have:

$$x_{min} = \frac{b(\theta_{min}/2)}{a(\theta_{min}/2)}R_s; \qquad \theta_{min} = 2\operatorname*{argmin}_{\gamma\in\left[0,\frac{\pi}{4}\right]}\left\{4R_sL - L^2 - \frac{b^2(\gamma)}{a(\gamma)}R_s^2\right\}.$$

However, with standard techniques, but with long and obnoxious computations, it is possible to show that $\frac{b^2(\gamma)}{a(\gamma)}$ is always increasing in $\left[0, \frac{\pi}{4}\right]$, which means that $\theta_{min} = \frac{\pi}{2}$. Furthermore, because $b\left(\frac{\pi}{4}\right) = a\left(\frac{\pi}{4}\right) = 4 - \pi$, we can see that

$$x_{min} = R_s.$$

It is noteworthy the fact that, from (2.13), $r = x$ for $\theta = \frac{\pi}{2}$, this means that R_s represent the radius of curvature for the rounded corners of the solution.

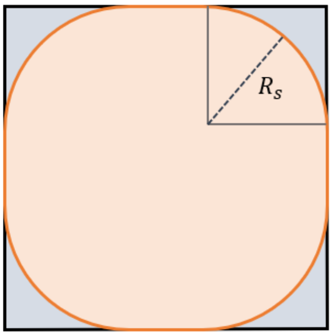

Figure 2.5: Example of a solution from (2.12)

We can see the representation of the solution set in figure 2.5. We can notice that this minimizing set can be expressed as the union of the balls of radius R_s included in the square, that is:

$$E_s = C^{R_s} = \bigcup_{B(x,R_s)\subset Q} B(x,R_s).$$

$x_{min} = R_s$, though, deny the solution validity for $R_s > \frac{L}{2}$. In this case $x_{min} = \frac{L}{2}$, for which the energy is:

$$f\left(\frac{L}{2},\theta\right) = 4R_sL - L^2 - 2b\left(\frac{\theta}{2}\right)R_s\frac{L}{2} + a\left(\frac{\theta}{2}\right)\frac{L^2}{4} =$$

$$= LR_s(4-b) + L^2\left(\frac{a}{4} - 1\right) = L\left(R_s\frac{2\sqrt{2}\gamma}{\sin\gamma} - \frac{L}{2}\left(1 + \frac{\gamma}{\sin^2(\gamma)} - \frac{\cos(\gamma)}{\sin(\gamma)}\right)\right),$$

but for $R_s > \frac{L}{2}$ and $\gamma \leq \frac{\pi}{4}$ this is for sure positive, making the empty set energetically convenient. However, we still have to determine thoroughly when this solution is actually better than the empty set for $R_s \leq \frac{L}{2}$. This happens if the associated energy is less than zero, that is:

$$f\left(R_s, \frac{\pi}{2}\right) = 4R_sL - L^2 - (4-\pi)R_s^2 < 0, \tag{2.14}$$

$$\implies R_s < \frac{L}{2+\sqrt{\pi}} =: R^*,$$

because the other possibility $R_s > \frac{L}{2-\sqrt{\pi}} > \frac{L}{2}$ is excluded.

In conclusion, consistently to what stated at the beginning of this section in (2.9), we have the solution for the level sets:

$$E_s = \begin{cases} \emptyset & \text{if } \dfrac{\lambda}{1-s} > \dfrac{L}{2+\sqrt{\pi}} \ \wedge \ s \geq 0, \\ C^{\frac{\lambda}{1-s}} & \text{if } \dfrac{\lambda}{1-s} \leq \dfrac{L}{2+\sqrt{\pi}} \ \wedge \ s \geq 0, \\ \mathbb{R}^2 & \qquad\qquad\qquad \text{if } s < 0. \end{cases} \tag{2.15}$$

Operationally, for dealing with more complicated convex sets, it is relevant to notice that $f\left(R, \frac{\pi}{2}\right) = 0 \iff \frac{P(C^R)}{|C^R|} = \frac{1}{R}$. Indeed, in general, the threshold R^* is the smallest R for which $\frac{P(C^R)}{|C^R|} = \frac{1}{R}$.

From the level sets we can reconstruct the ROF solution as

$$u(x) = \max\left\{0,\, 1 - \frac{\lambda}{r(x)}\right\}\chi_Q(x), \tag{2.16}$$

where $r(x) = R^*$ for $x \in C^{R^*}$ and $r(x) = R$ (if $R \in [0, R^*]$) for $x \in \partial C^R$. We can notice that in $Q \setminus C^\lambda$ we have $u(x) = 0$ and its maximum $u = 1 - \frac{\lambda}{R^*}$ is reached in C^{R^*}. On $C^\lambda \setminus C^{R^*}$, instead, u is decreasing smoothly to 0. On the other hand, when $\lambda > R^* = \frac{L}{2+\sqrt{\pi}}$, $u \equiv 0$, showing the harmful effect of a too high λ.

Let's now compute the energy of the solution, let's call it $F(u)$. If $\lambda \geq R^*$, the minimal energy is $F(u) = \frac{1}{2}\|\chi_Q\|_2^2 = \frac{L^2}{2}$. Otherwise, if $\lambda > R^*$, we can exploit our knowledge on the level sets. At first, let's call s^* the highest s for which $E_s \neq \emptyset$, that is

$$\frac{\lambda}{1-s^*} = R^* \implies s^* = 1 - \frac{\lambda}{R^*}.$$

Then, from the coarea formula we have

$$V(u) = \int_0^{s^*} P(E_s)\,\mathrm{d}s = \int_0^{s^*} 2\pi R_s + 4(L - 2R_s)\,\mathrm{d}s.$$

We can now do a change of variables $r = R_s = \frac{\lambda}{1-s}$, so that

$$V(U) = 4Ls^* + \int_\lambda^{R^*} (2\pi r - 8r)\frac{\lambda}{r^2}\,\mathrm{d}r = 4Ls^* - 2\lambda(4-\pi)\log\left(\frac{R^*}{\lambda}\right). \tag{2.17}$$

For the fidelity term $\frac{1}{2}\|u-g\|_2^2$, we notice that outside of Q u and g are both zero, while in $Q \setminus C^\lambda$ $u = 0$ and $g = 1$. Then, we have

$$\frac{1}{2}\|u-g\|_2^2 = \frac{1}{2}\int_Q (1-u)^2\,\mathrm{d}x = \frac{1}{2}(L^2 - \left|C^\lambda\right|) + \frac{1}{2}\int_{C^\lambda}(1-u)^2\,\mathrm{d}x.$$

The remaining integral can be solved integrating by layers, that is

$$\int_{C^\lambda}(1-u)^2\,\mathrm{d}x = \int_{(1-s^*)^2}^1 \left|C^\lambda\right| - \left|C^{R_{1-\sqrt{t}}}\right|\mathrm{d}t = \int_{(1-s^*)^2}^1 \left|C^\lambda\right| - \left|C^{\lambda/\sqrt{t}}\right|\mathrm{d}t =$$

$$\left|C^\lambda\right| - (1-s^*)^2\left|C^\lambda\right| - \int_{(1-s^*)^2}^1 L^2 - (4-\pi)\frac{\lambda^2}{t}\,\mathrm{d}t =$$

$$= \left|C^\lambda\right| - L^2 + (1-s^*)^2(L^2 - \left|C^\lambda\right|) + 2(4-\pi)\lambda^2\log\left(\frac{1}{1-s^*}\right).$$

Putting everything together we have

$$F(u) = \lambda V(u) + \frac{1}{2}\|u-g\|_2^2 =$$

$$= 4\lambda Ls^* - 2\lambda^2(4-\pi)\log\left(\frac{R^*}{\lambda}\right) + \frac{(1-s^*)^2}{2}(L^2 - \left|C^\lambda\right|) + (4-\pi)\lambda^2\log\left(\frac{1}{1-s^*}\right).$$

Recalling that $\frac{\lambda}{1-s^*} = R^* = \frac{L}{2+\sqrt{\pi}}$ and $L^2 - \left|C^\lambda\right| = (4-\pi)\lambda^2$, we can simplify the energy as a function of $\frac{R^*}{\lambda}$, that is:

$$\begin{cases} F(u) = \lambda^2\, e\left(\frac{R^*}{\lambda}\right), \\ e(r) = 4(2+\sqrt{\pi})(r-1) - (4-\pi)log(r) + \frac{4-\pi}{2r^2}. \end{cases} \tag{2.18}$$

2.3 Related models

Before concluding this chapter, we present some models derived from the ROF. A first instance consist in introducing a filter in the fidelity term, that is:

$$\min_{u\in\mathcal{U}} \lambda V(u,\Omega) + \frac{1}{2}\|Au - g\|_{L^2(\Omega)}^2,$$

where A is a linear continuous operator. This generalization can be applied in image deblurring or zooming. Indeed, if we took A as the convolution for a blurring kernel, the solver would look for an image whose blurred version is similar to the reference, but with sharper edges to minimize the total variation. Instead, if A is a blurring followed by a downsampling, this can be used to zoom in or simply to increment the image quality.

Many of the results previously mentioned can be easily generalized, for instance we can see that still the problem is convex and admits a unique solution (as in 2.2.4), and its Euler - Lagrange equation can be expressed as:

$$0 \in \lambda \partial V + Au - g.$$

(a) Original image (b) Degraded image (c) Wiener filter (d) TV-deblurring

Figure 2.6: A comparison between the classical Wiener filter for deblurring and the total variation method presented above. The edge sharpening effect of total variation is noticeable, providing better results then the the Wiener filter.

A second model related to ROF is the so called $TV - L^1$ in which the fidelity term is replaced with the L^1 norm.

$$\min_{u \in \mathcal{U}} \lambda V(u, \Omega) + \|u - g\|_{L^1(\Omega)}.$$

This choice seems particularly natural since $BV \subset L^1$, but it is also relevant for another reason. Consider a change in the contrast of g, that is using αg as reference for α a positive scaling factor. In ROF, this change will produce a completely different solution, because the total variation is $1-$homogeneous while the fidelity term is $2-$homogeneous, therefore the relation between these two changes with the scaling. On the other hand, the $TV - L^1$ model will produce the same solution, just scaled of α. This consideration is relevant not only when changing contrast, but more in general for capturing details in lower scales. Indeed, in ROF the total variation term become much more relevant when there are small oscillation, so that the areas with close shades of colors are flattened in a monochromatic chunk.

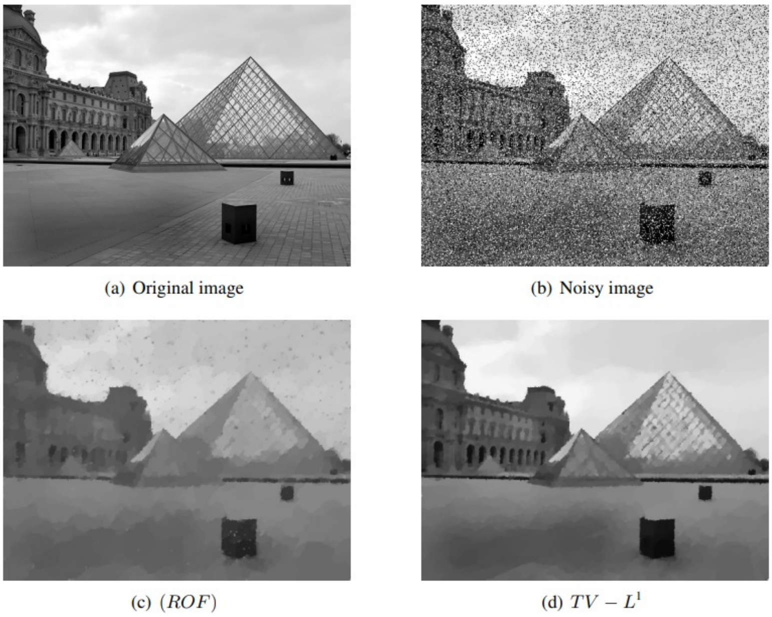

Figure 2.7: A comparison of the restoration with ROF and with $TV - L^1$ in a highly damaged image. Notice the level of details recovered in the latter.

Conversely, $TV - L^1$ is able to restore low scale details much better. In figure 2.7 we can see a comparison in a case of quite high noise (25% corrupted).

The main problem with this method is the lack of strict convexity and consequently the uniqueness is not guaranteed. The minimization has been faced in many ways (see for instance [12]), but we can use the primal-dual modified extragradient method, that will be presented in the next chapter. For this purpose, we need to rewrite the problem as a saddle-point. Fortunately this is an easy task since by definition the total variation and the absolute value can be written as a constrained maximization problem, then we get

$$\min_u \max_{\phi,\psi} \lambda \int_\Omega \phi \, dDu - \mathbb{I}_{\{\|\phi\|_\infty \leq 1\}} + \int_\Omega \psi u \, dx - \mathbb{I}_{\{\|\psi\|_\infty \leq 1\}},$$

where $\mathbb{I}_A$ is the indicator function ($\mathbb{I}_A = 0$ in A and ∞ otherwise) and ϕ is vector valued.

Chapter 3

ROF solver implementation

The purpose of this chapter is to discuss the implementation of a solver for the ROF problem. At first a discrete adaptation of the problem is needed, so that it works in the setting of pixelated digital images. Then we will introduce an algorithm able to approximate the minimizer. Despite in [13], chapter 3, several methods are actually presented and compared, we will use only the one it is reported as the best performing one in terms of execution time, that is a modified primal-dual approach. A part from its performance, this algorithm is interesting to acknowledge also because it allow to find approximately a saddle point. Finally, the solver will be validated through the analytical solution of the square found in (2.16) and some experimental results will be shown.

3.1 Discrete problem

Let's take for instance a squared image, and as domain take $\Omega = (0,1)^2$. Then we have to divide the picture in pixels, that we take again as squared. Those correspond to the tiles $Q^h_{i,j} \subset \Omega$, with $i, j = 1, \dots, N$, of side $h = \frac{1}{N}$ and covering (almost everywhere) disjointly Ω. Then we call $x_{i,j}$ the center of those squares, so that:

$$Q^h_{i,j} = x_{i,j} + \left(-\frac{h}{2}, \frac{h}{2}\right)^2; \quad x_{i,j} = \left(ih - \frac{h}{2}, jh - \frac{h}{2}\right).$$

In thus discrete setting an image u is constant on each pixel, this means that we have a double representation: on one hand u can be seen as a function in a subspace of $L^\infty(\Omega)$

$$\mathcal{Q}^h := \{u \in L^\infty(\Omega) \mid \forall\, i, j = 1, \dots, N \ u \text{ constant on } Q^h_{i,j}\},$$

on the other hand u can be seen as a matrix in $\mathbb{R}^{N\times N}$ defined as:

$$u_{i,j} := u(x_{i,j}) = u(Q^h_{i,j}).$$

Then we can write $u(x) = \sum_{i,j=1}^{N} u_{i,j}\chi_{Q^h_{i,j}}(x)$.

Now we have to discretize the objective function. For the total variation it is natural to define its discrete version as

$$V_h(u) \coloneqq \sum_{i,j=1}^{N} |\nabla_h u(x_{i,j})|_2 h^2, \tag{3.1}$$

where $\nabla_h : \mathcal{Q}^h \longrightarrow \left[\mathcal{Q}^h\right]^2$ is a linear operator that in some sense approximate the gradient. Since $\mathcal{Q}^h$ is a finite dimensional space, we can immediately appreciate that the operator ∇_h is continuous, and consequently also V_h. This is an operator that acts locally, consequently we can define it more precisely through a core linear operator A_h acting on a set of neighbour indices $\mathcal{N}$, which can be defined as a finite subset of $\mathbb{Z}^2$ containing (0,0). Then the core operator is defined as $A_h : \mathbb{R}^{\mathcal{N}} \longrightarrow \mathbb{R}^2$ and is meant to compute the discrete gradient in the center (0,0) using the value of the function on the neighbourhood $\mathcal{N}$, that is

$$\nabla_h u(x_{i,j}) = A_h u_{(i,j)+\mathcal{N}},$$

where $u_{(i,j)+\mathcal{N}}$ is a function on $\mathcal{N}$ such that $p \in \mathcal{N} \mapsto u_{(i,j)+p} = (S_p u)_{i,j}$, with S_p the shift operator. On this point of view, we can call $S_\mathcal{N}$ the collection of shift operators, that is $S_\mathcal{N} : \mathcal{Q}^h \mapsto \left[\mathcal{Q}^h\right]^{\mathcal{N}}$ so that $(S_\mathcal{N} u)_p = S_p u, \quad$ for $p \in \mathcal{N}$. In this way we have

$$\nabla_h = A_h \circ S_\mathcal{N}.$$

Dealing with shifts, it may happen that the indices $(i + p_1, j + p_2)$ exceed the boundaries 1 and N, in these cases we pose:

$$\begin{cases} \text{if } i < 1,\ u_{i,j} = u_{1,j} \\ \text{if } i > N,\ u_{i,j} = u_{N,j} \end{cases} \bigwedge \begin{cases} \text{if } j < 1,\ u_{i,j} = u_{i,1} \\ \text{if } j > N,\ u_{i,j} = u_{i,N}. \end{cases} \tag{3.2}$$

These definitions come from the fact that we would like to recover the Neumann boundary conditions of the Euler-Lagrange equation (2.8), achieved by prolonging the functions on each side of the square Ω by constants. In relation to this consideration, a property that A_h must satisfy is that for constant inputs the output must be 0, that is

$$A_h \mathbf{1}_\mathcal{N} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

where $(\mathbf{1}_\mathcal{N})_p = 1\ \forall p \in \mathcal{N}$. This means that (seeing A as a matrix) the column corresponding to (0,0) is nothing less than minus the sum of all the other columns of A. In other words, denoting a_p the column of A_h corresponding to $p \in \mathcal{N}$ we have

$$a_{(0,0)} = -\sum_{p \in \mathcal{N} \setminus (0,0)} a_p.$$

This implies that Au is actually a function of the differences $u - p - u_{0,0}$, indeed

$$A_h u = \sum_{p \in \mathcal{N} \setminus (0,0)} a_p u_p + a_{0,0} u_{0,0} = \sum_{p \in \mathcal{N} \setminus (0,0)} a_p (u_p - u_{0,0}).$$

Therefore, the discrete gradient can also be written in the form:

$$\nabla_h = \tilde{A} \circ \left(\frac{S_{\mathcal{N} \setminus (0,0)} - Id}{h} \right).$$

For the future developments it is also relevant to understand how the adjoint operator, on the L^2 scalar product, of ∇_h behaves. For reference, the L^2 scalar product on functions in $\left[\mathcal{Q}^h\right]^p$ reduces to a sum like the following

$$\langle u, v \rangle = h^2 \sum_{i,j=1}^{N} u_{i,j} \cdot v_{i,j}.$$

Let's call such adjoint operator $-\mathrm{div}_h : \left[\mathcal{Q}^h\right]^2 \longrightarrow \mathcal{Q}^h$, then it is defined such that $\langle \nabla_h u, \xi \rangle = \langle u, -\mathrm{div}_h \xi \rangle$ for every $u \in \mathcal{Q}^h$ and $\xi \in \left[\mathcal{Q}^h\right]^2$. Therefore, we have

$$\mathrm{div}_h = \left(\frac{Id - S^*_{\mathcal{N} \setminus (0,0)}}{h} \right) \circ \tilde{A}^T,$$

where $\tilde{A}^T$ is simply the transposed matrix, while $S^*_{\mathcal{N} \setminus (0,0)}$ can be seen as the vector of the adjoint operators S^*_p with $p \in \mathcal{N} \setminus (0,0)$. Hence, we only really need to understand the adjoint to the shift. For simplicity, let's take $p = (c,0)$ for $0 < c < N - 1$, then, given $u, v \in \mathcal{Q}^h$, from the notation (3.2) we have

$$\sum_{i,j=1}^{N} u_{i,j} \left(S^*_p v \right)_{i,j} = \sum_{j=1}^{N} \sum_{i=c+1}^{N+c} u_{i,j} v_{i-c,j} = \sum_{j=1}^{N} \left[\sum_{i=c+1}^{N-1} u_{i,j} v_{i-c,j} + \sum_{k=N-c}^{N} u_{N,j} v_{k,j} \right],$$

$$\Longrightarrow \left(S^*_p v \right)_{i,j} = \begin{cases} 0 & \text{if } i \le c \\ v_{i-c,j} & \text{if } c < i \le N - 1 \\ \displaystyle\sum_{k=N-c}^{N} v_{k,j} & \text{if } i = N. \end{cases}$$

A common choice for ∇_h is given as follows:

$$(\nabla_h u)_{i,j} = \left(\frac{u_{i+1,j} - u_{i,j}}{h}, \frac{u_{i,j+1} - u_{i,j}}{h} \right), \tag{3.3}$$

which in the previous notation correspond to $\mathcal{N} = \{(0,0), (1,0), (0,1)\}$ and $\tilde{A}$ the 2 dimensional identity matrix. Conversely, the discrete divergence is expressed by

$$(\mathrm{div}_h \xi)_{i,j} = \frac{\xi^1_{i,j} - \xi^1_{i-1,j}}{h} + \frac{\xi^2_{i,j} - \xi^2_{i,j-1}}{h}, \tag{3.4}$$

for $1 < i, j < N$. On the boundary layer instead we have that, if $i = 1$, then $\xi^1_{i-1,j}$ is replaced by 0 and, if $i = N$, it is replaced by $\xi^1_{i-1,j} + \xi^1_{i,j}$, which is equivalent of saying that $\xi^1_{i,j}$ is replaced by 0. Analogously if $j = 1$ or N.

This discretization leads to the discrete total variation

$$V_h(u) = h \sum_{i,j=1}^{N} \sqrt{(u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2}. \tag{3.5}$$

This formulation has some interesting properties, the first is that V_h decrease by truncation similarly to the actual total variation (see 1.1.26). Indeed, given m and M real numbers and calling $\bar{u} = m \vee (u \wedge M)$, then for any pair of indices

$$|\bar{u}_{l,m} - \bar{u}_{i,j}| \leq |u_{l,m} - u_{i,j}|,$$

because $\max\{\bar{u}_{l,m}, \bar{u}_{i,j}\} \leq \max\{u_{l,m}, u_{i,j}\}$ and $\min\{\bar{u}_{l,m}, \bar{u}_{i,j}\} \geq \min\{u_{l,m}, u_{i,j}\}$. Consequently,

$$V_h(m \vee (u \wedge M)) \leq V_h(u). \tag{3.6}$$

This is a relevant property because, as in the continuous ROF problem, it guarantee that the solution of the discrete ROF problem has the same bounds as the reference image, that is the solution will assume in each pixel values of intensity between 0 and 255 as required for digital images.

Secondly, we can notice that the actual total variation of a function in $\mathcal{Q}^h$ is made just by the jumps in values between adjacent pixels, multiplied by the length of the interface. This can be written as:

$$V(u, \Omega) = h \sum_{i,j=1}^{N} |u_{i+1,j} - u_{i,j}| + |u_{i,j+1} - u_{i,j}|.$$

Therefore, from the equivalence of the 1 and 2 norm, we can bound the total variation with (3.5) and vice verse:

$$V_h(u) \leq V(u, \Omega) \leq \sqrt{2} V_h(u), \quad \forall\, u \in \mathcal{Q}^h. \tag{3.7}$$

We have to point out that the total variation on $\mathcal{Q}^\langle$ is not an accurate choice for the discretization of the total variation, because, as explained in [14], it does not Γ-converges to the total variation in L^1, but to a functional which behaves as $\int_\Omega |\nabla u|_1 \, \mathrm{d}x$ on $W^{1,1}$ (there is a different king of modulus for the gradient).

We will now show that instead V_h as formulated in (3.5) Γ-converges to the total variation.

Proposition 3.1.1. *Let's define V_h as in* (3.5) *in $\mathcal{Q}^h$ and $+\infty$ in $L^1(\Omega) \setminus \mathcal{Q}^h$. Then*

$$V_h \xrightarrow[h\longrightarrow 0]{\Gamma} V, \quad \text{in } L^1(\Omega).$$

Proof. $\boxed{\liminf}$: Take $\phi \in \left[C_c^1(\Omega)\right]^2$ such that $|\phi(x)| \leq 1$ for all $x \in \Omega$. For continuity we can relate to ϕ a function $\phi^h \in \mathcal{Q}^h$ defined so that $\phi^h_{i,j} = \phi(x_{i,j})$. Then, given any $u \in \mathcal{Q}^h$, for the Cauchy-Schwartz inequality we have

$$V_h(u) \geq \sum_{i,j=1}^{N} h^2(\nabla_h u)_{i,j} \cdot \phi^h_{i,j} = \sum_{i,j=1}^{N} h^2 u_{i,j}(-\mathrm{div}_h \phi^h)_{i,j} = -\int_\Omega u\,\mathrm{div}_h \phi^h\,\mathrm{d}x.$$

Now, given a sequence $u^h \longrightarrow u$ in L^1, with $u^h \in \mathcal{Q}^h$ for each $h > 0$, we want to show

$$\liminf_h V_h(u^h) \geq V(u,\Omega) = \sup_\phi \int_\Omega u\,\mathrm{div}\phi\,\mathrm{d}x.$$

Therefore, we only need to show that, for each ϕ with the properties above, $\liminf_h \int_\Omega u^h\,\mathrm{div}_h\phi^h\,\mathrm{d}x \geq \int_\Omega u\,\mathrm{div}\phi\,\mathrm{d}x$, but in particular we can show that

$$\lim_h \int_\Omega u^h\,\mathrm{div}_h\phi^h\,\mathrm{d}x = \int_\Omega u\,\mathrm{div}\phi\,\mathrm{d}x.$$

Because u^h converges to u in L^1, this limit is verified if $\mathrm{div}_h\phi^h \longrightarrow \mathrm{div}\phi$ in L^∞. Similarly as before, because $\mathrm{div}\phi$ is continuous, we can define $f \in \mathcal{Q}^h$ such that $f_{i,j} = \mathrm{div}\phi(x_{i,j})$, then we have

$$\left\|\mathrm{div}\phi - \mathrm{div}_h\phi^h\right\|_\infty \leq \|\mathrm{div}\phi - f\|_\infty + \left\|f - \mathrm{div}_h\phi^h\right\|_\infty.$$

1. $\|\mathrm{div}\phi - f\|_\infty = \sup_{i,j=1,\dots,N} \sup_{x\in Q^h(x_{i,j})} |\mathrm{div}\phi(x) - \mathrm{div}\phi(x_{i,j})|$. However, $\mathrm{div}\phi$ is a continuous function on a compact domain, which means it is uniformly continuous, in other words independently from i,j

   $$\forall\varepsilon > 0,\ \exists\delta > 0 :\ |x - x_{i,j}| < \delta \implies |\mathrm{div}\phi(x) - \mathrm{div}\phi(x_{i,j})| < \varepsilon.$$

   But, because if $x \in Q^h_{i,j}$, then $|x - x_{i,j}| \leq \frac{\sqrt{2}}{2}h$, then for each $\varepsilon > 0$ we can find h small enough so that $\|\mathrm{div}\phi - f\|_\infty < \varepsilon$, proving $\|\mathrm{div}\phi - f\|_\infty \longrightarrow 0$.

2. Let's call $\tilde{\Omega}_h$ the set generated by subtracting from Ω the layer of pixels of side h on the boundary. Then, because ϕ is compactly supported in Ω, for $\bar{h}$ small enough we have $\left\|f - \mathrm{div}_h\phi^h\right\|_{L^\infty(\Omega)} = \left\|f - \mathrm{div}_h\phi^h\right\|_{L^\infty(\tilde{\Omega}_{\bar{h}})}$. But for $h < \bar{h}$ we can generalize div_h as a functional $\left[C^0(\Omega)\right]^2 \longrightarrow C^0(\tilde{\Omega}_{\bar{h}})$ using the expression in (3.4):

   $$\mathrm{div}_h\phi(x_1,x_2) = \frac{\phi^1(x_1,x_2) - \phi^1(x_1-h,x_2)}{h} + \frac{\phi^2(x_1,x_2) - \phi^2(x_1,x_2-h)}{h}.$$

In this case $(\mathrm{div}_h \phi^h)_{i,j} = \mathrm{div}_h \phi(x_{i,j})$. Hence, we have

$$\begin{aligned}\left\| f - \mathrm{div}_h \phi^h \right\|_{L^\infty(\tilde{\Omega}_{\bar{h}})} &= \sup_{x_{i,j} \in \tilde{\Omega}_{\bar{h}}} \{\mathrm{div}\phi(x_{i,j}) - \mathrm{div}_h \phi(x_{i,j})\} \\ &\leq \sup_{x \in \tilde{\Omega}_{\bar{h}}} \{\mathrm{div}\phi(x) - \mathrm{div}_h \phi(x)\}.\end{aligned}$$

But for C^1 functions the incremental ratio converges uniformly on compacts to the derivative, therefore we must also have that $\left\| f - \mathrm{div}_h \phi^h \right\|_\infty \longrightarrow 0$.

$\boxed{\limsup}$: At first consider $u \in C^1(\Omega)$ and the linear operator

$$\begin{aligned}\bar{\circ}_h : \ L^1(\Omega) &\longrightarrow \mathcal{Q}^h \\ u \mapsto (\bar{u}_h)_{i,j} &= \frac{1}{h^2} \int_{Q^h_{i,j}} u \,\mathrm{d}x\end{aligned}$$

that corresponds to taking the integral mean on each pixel. For h going to 0, $\bar{u}_h$ is known to converge to u in L^1, so we will prove that such $\bar{u}_h$ is the recovery sequence, in other words that $V_h(\bar{u}_h) \xrightarrow[h \longrightarrow 0]{} V(u, \Omega)$. Similarly to div_h, we can define ∇_h for continuous functions as in (3.3), substituting $x_{i,j}$ for a general $x \in \Omega$. Then, using the linearity of $\bar{\circ}_h$ and ∇_h, we have

$$V_h(\bar{u}_h) = \sum_{i,j=1}^{N} h^2 |(\nabla_h \bar{u}_h)_{i,j}| = \sum_{i,j=1}^{N} \left| \int_{Q^h_{i.j}} \overline{(\nabla_h u)}_h \,\mathrm{d}x \right| \leq \int_\Omega |\nabla_h u| \,\mathrm{d}x.$$

$$\Longrightarrow \ V_h(\bar{u}_h) - V(u, \Omega) \leq \int_\Omega |\nabla_h u| - |\nabla u| \,\mathrm{d}x \leq \int_\Omega |\nabla_h u - \nabla u| \,\mathrm{d}x.$$

But it is well known that $\nabla_h u \longrightarrow \nabla u$ in L^1, consequently

$$\limsup_h V_h(\bar{u}_h) \leq V(u, \Omega).$$

Now we can generalize this result from $u \in C^1(\Omega)$ to $BV(\Omega)$ by density. Indeed, from 1.1.13, for each $u \in BV(\Omega)$ we can find $\{u^k\} \subset C^1(\Omega)$ such that

$$\begin{cases} u^k \longrightarrow u \quad \text{in } L^1(\Omega), \\ \lim\limits_{k \to +\infty} V(u^k, \Omega) = V(u, \Omega). \end{cases}$$

But for each k $\lim_{h \longrightarrow 0} V_h(\overline{u^k}_h) = V(u^k, \Omega)$. Therefore, for any sequence of indices h_n we can find a subsequence (that I call h_k) for which $\left| V_{h_k}(\overline{u^k}_{h_k}) - V(u^k, \Omega) \right| \leq \frac{1}{k}$, consequently

$$\left| V_{h_k}(\overline{u^k}_{h_k}) - V(u, \Omega) \right| \leq \left| V_{h_k}(\overline{u^k}_{h_k}) - V(u^k, \Omega) \right| + \left| V(u^k, \Omega) - V(u, \Omega) \right| \xrightarrow[k \longrightarrow \infty]{} 0.$$

$\square$

After this discussion on the discretization of the total variation, we need to deal with the reference term. The L^2 distance can be kept, however we have to approximate the reference image $g \in L^\infty(\Omega)$ in the ROF problem with a $g^h \in \mathcal{Q}^h$. As for the recovery sequence in the previous Γ-limit, a good choice of g^h is give by averaging on the pixels:

$$g^h_{i,j} \coloneqq \frac{1}{h^2} \int_{Q^h_{i,j}} g \,\mathrm{d}x. \tag{3.8}$$

This definition of g^h satisfies the following two properties:

- $g^h \longrightarrow g$ in L^2 for $h \longrightarrow 0$,
- $\left\| g^h \right\|_\infty \leq \|g\|_\infty$.

Those are actually the only required properties, therefore different choices of g^h can be made, but we will prove that as long as they satisfies those properties we have a correct approximation of the ROF problem.

We have now all the instruments to define the discrete ROF problem:

$$\min_{u \in \mathcal{Q}^h} V_h(u) + \frac{1}{2\lambda} \left\| u - g^h \right\|^2_{L^2}. \tag{ROF_h}$$

For the next results we will refer to the objective functional of (ROF) as F and the one of (ROF_h) as F_h.

As told previously, the maximum and minimum of the (ROF_h) are controlled by those of g^h, this is expressed by the following proposition.

Proposition 3.1.2. *Take $g \in L^\infty(\Omega)$, g^h as* (3.8)*, V_h decreasing by truncation and $\mathcal{U}$ as in* (2.3)*, then*

$$\inf_{u \in \mathcal{Q}^h} F_h(u) = \inf_{u \in \mathcal{Q}^h \cap \mathcal{U}} F_h(u).$$

Proof. The proof follows as in preposition 2.2.1. I just want to stress that here the property of V_h to decrease by truncation is needed. More precisely we obtain that if u_h is the minimizer, then $\min g^h \leq u_h \leq \max g^h$. But because $\left\| g^h \right\|_\infty \leq \|g\|_\infty$ we must have $\|u_h\|_\infty \leq \|g\|_\infty$. To completely gain that the minimizers belong to $\mathcal{U} \subset BV(\Omega)$, we further need $V(u_h)$ to be bounded for each h. But we can even show that they are equibounded, indeed

$$F_h(u_h) \leq F_h(0) = \frac{1}{2\lambda} \left\| g^h \right\|^2_2 \leq \frac{1}{2\lambda} \|g\|^2_2,$$

where $\|g_h\|_2 \leq \|g\|_2$ comes from the Jensen inequality. This implies:

$$\frac{1}{2\lambda} \|g\|^2_2 \geq V_h(u_h) \geq V(u_h).$$

$\square$

This preposition further shows that the sequence of functionals $\{F_h\}$ is equi-mildly coercive, which is needed to apply the fundamental theorem of Γ-convergence. We also have the existence and uniqueness of the (ROF_h) minimizer.

Proposition 3.1.3. *The ROF_h problem admits a unique solution.*

Proof. Proceed exactly like in the proof of Prop. 2.2.4. □

Now we can finally show that (ROF_h) is a good approximation of the ROF problem.

Proposition 3.1.4. *F_h Γ-converges to F and in particular, calling u_h^* the minimizers of F_h and u^* the one of F, then*

$$u_h^* \longrightarrow u^* \quad \text{in } L^1 \text{ and } L^2.$$

Proof. Because we can restrict the minimization on the compact $\mathcal{U}$, if we have the Γ-convergence, we also have the convergence of the minimizers. Thus, we only need to prove this Γ-convergence. For the liminf, take any $u_h \in \mathcal{Q}^h$ converging in L^1 to $u \in BV(\Omega)$. By the observation 2.2.3 we must also have the L^2 convergence, therefore $u_h - g^2 \longrightarrow u - g$ in L^2 for h going to 0. Then, exploiting that V_h Γ-converges to V, we conclude

$$\liminf_h F_h(u_h) \geq \liminf_h V_h(u_h) + \liminf_h \frac{1}{2\lambda}\left\|u_h - g^h\right\|_2^2 \geq F(u).$$

For the limsup inequality, we take as recovery sequence the one of V_h, then we proceed like in the liminf part. □

There are also some results on the convergence rate of the minimizers and bounds on the error of the discrete solution. For example in [15] is shown that if $g \in W^{1,2}$, then $\|u_h^* - u^*\|_2^2 \sim O(h^{1/2})$; while [16] proves that if $g \in Lip(\alpha)$, then $\|u_h^* - u^*\|_2^2 \sim O\left(h^{\frac{\alpha}{1+\alpha}}\right)$.

3.2 Solver algorithm

In this section it will be presented an iterative algorithm to solve (ROF_h). In [13] some possible algorithms are presented and compared, but in terms of run-time the best one seems to be an Arrow-Hurwicz type method, or also called primal-dual approach.

At first, let's discuss in general the method, but for more insights we invite the reader to consult [17]. This method has been developed for dealing with saddle point problems and basically consists in performing jointly a gradient ascent on the variable to maximize and a gradient descent for the other one. For our purposes we are interested in problem of the form:

$$\min_{x\in X} F(Dx) + G(x), \tag{3.9}$$

where X is an Hilber space, F and G belong to Γ^0 on their respective domains (see the the definition above 2.2.7) and D is a linear continuous operator from X to the domain of F, which we call Y. We can move to a saddle point problem using the definition 2.2.7 of Legendre conjugate and recalling that $F = F^{**}$, because $F \in \Gamma^0$. Hence, the problem (3.9) is equivalent to

$$\min_{x\in X} \max_{y\in Y} \langle y, Dx\rangle - F^*(y) + G(x). \tag{3.10}$$

By swapping the minimization and maximization, we can get the dual problem

$$\max_{y\in Y} \min_{x\in X} \langle y, Dx\rangle - F^*(y) + G(x) = \max_{y\in Y} -\left(\max_{x\in X}\langle -D^T y, x\rangle - G(x)\right) - F^*(y) =$$

$$= \max_{y\in Y} -G^*(-D^T y) - F^*(y).$$

In general a min max problem has a greater or equal value than the corresponding max min, however under very weak assumption, that for an unconstrained problem as (ROF_h) are satisfied, we actually have the equality, then

$$\min_{x\in X} F(Dx) + G(x) = \max_{y\in Y} -G^*(-D^T y) - F^*(y).$$

This suggest us to introduce the always positive quantity, called *dual gap*

$$\mathcal{G}(x,y) = F(Dx) + G(x) - G^*(-D^T y) + F^*(y),$$

and it satisfies that $\mathcal{G}(\hat{x},\hat{y}) = 0$ if and only if $(\hat{x},\hat{y})$ is a solution for (3.10). Furthermore, if the min and the max are exchangeable, we can have the equivalent definition of a saddle point $(\hat{x},\hat{y})$: a point such that for every $(x,y) \in X \times Y$ the following inequality is satisfied:

$$\langle y, D\hat{x}\rangle - F^*(y) + G(\hat{x}) \leq \langle \hat{y}, D\hat{x}\rangle - F^*(\hat{y}) + G(\hat{x}) \leq \langle \hat{y}, Dx\rangle - F^*(\hat{y}) + G(x). \tag{3.11}$$

The Arrow-Hurwicz methods on (3.10) develops as follows: we define the starting point (x^0, y^0) and fix two step parameters (small) $\tau, \sigma > 0$, then for each iteration we upgrade our guess applying firstly an implicit gradient ascent step on the dual variable y, secondly a descent step on x. We can write the iteration explicitly as:

$$\begin{cases} y^{n+1} = y^n + \sigma(Dx^n - \partial F^*(y^{n+1})), \\ x^{n+1} = x^n - \tau(D^T y^{n+1} + \partial G(x^{n+1})). \end{cases}$$

The first idea to improve this algorithm is to introduce an acceleration, that is an adaptative way of selecting the step sizes σ and τ for each iteration, see [18]. This idea has been applied, however we are not going to implement it because in may knowledge there is not yet a proof of its convergence. Here, instead are going to apply another variation proposed in [19] for the Mumford-Shah problem. We basically introduce on the primal variable a mid-step $\tilde{x}^n$ between x^n and x^{n+1} such that $\tilde{x}^0 = x^0$ and the iterative step assume the form

$$\begin{cases} y^{n+1} = (Id + \sigma \partial F^*)^{-1}(y^n + \sigma D\tilde{x}^n), \\ x^{n+1} = (Id + \tau \partial G)^{-1}(x^n - \tau D^T y^{n+1}), \\ \tilde{x}^{n+1} = 2x^{n+1} - x^n. \end{cases} \tag{3.12}$$

For this algorithm we have a convergence result, that we are going to show soon, but first we need to introduce some notation. Let's define the partial primal dual gap in $B_1 \subseteq X$ and $B_2 \subseteq Y$ as

$$\mathcal{G}_{B_1 \times B_2}(x, y) = \max_{y' \in B_2} \{\langle y', Dx\rangle - F^*(y') + G(x)\} + \\ - \min_{x' \in B_1} \{\langle y, Dx'\rangle - F^*(y) + G(x')\}.$$

We notice that $\mathcal{G}_{X\times Y} = \mathcal{G}$ and that, similarly to $\mathcal{G}$, $\mathcal{G}_{B_1\times B_2} \geq 0$ and it vanishes only in the saddle points (if there are any in $B_1 \times B_2$).

Theorem 3.2.1. *Let $L = \|D\|$ and assume* (3.10) *has a saddle point $(\hat{x}, \hat{y})$. Consider the iterative algorithm defined in* (3.12)*, with $\tilde{x}^0 = x^0$. If $\tau\sigma L^2 < 1$, then:*

1. *for any N,*

$$\frac{\left\|y^N - \hat{y}\right\|^2}{2\sigma} + \frac{\left\|x^N - \hat{x}\right\|^2}{2\tau} \leq \frac{1}{1 - \tau\sigma L^2}\left(\frac{\|y^0 - \hat{y}\|^2}{2\sigma} + \frac{\|x^0 - \hat{x}\|^2}{2\tau}\right);$$

2. *Let $\bar{x}_N = \left(\sum_{n=1}^N x^n\right)/N$ and $\bar{y}_N = \left(\sum_{n=1}^N y^n\right)/N$, then for each B_1, B_2 bounded there exists a constant $C(B_1, B_2)$, depending on B_1 and B_2, such that*

$$\mathcal{G}_{B_1\times B_2}(\bar{x}_N, \bar{y}_N) \leq \frac{C(B_1, B_2)}{N}.$$

This means that any accumulation point of $(\bar{x}_N, \bar{y}_N)$ is a saddle point, and in particular also the accumulation point under weak convergence.

3. *If X and Y are finite dimensional, then there exists a saddle point (x^*, y^*) which (x^n, y^n) converges to.*

Proof. The first two equations in (3.12) can be rewritten as

$$\frac{y^n - y^{n+1}}{\sigma} + D\tilde{x}^n \in \partial F^*(y^{n+1}),$$

$$\frac{x^n - x^{n+1}}{\tau} - D^T y^{n+1} \in \partial G(x^{n+1}).$$

This means that for any $(x, y) \in X \times Y$ we have

$$F^*(y) \geq F^*(y^{n+1}) + \left\langle \frac{y^n - y^{n+1}}{\sigma} + D\tilde{x}^n, y - y^{n+1} \right\rangle,$$

$$G(x) \geq G(x^{n+1}) + \left\langle \frac{x^n - x^{n+1}}{\tau} - D^T y^{n+1}, x - x^{n+1} \right\rangle.$$

From the properties of scalar product we have

$$\langle y^n - y^{n+1}, y - y^{n+1} \rangle = \frac{1}{2}\left\|y^n - y^{n+1}\right\|^2 + \frac{1}{2}\left\|y - y^{n+1}\right\|^2 - \frac{1}{2}\|y - y^n\|^2,$$

$$\langle x^n - x^{n+1}, x - x^{n+1} \rangle = \frac{1}{2}\left\|x^n - x^{n+1}\right\|^2 + \frac{1}{2}\left\|x - x^{n+1}\right\|^2 - \frac{1}{2}\|x - x^n\|^2.$$

We are interested in summing the two inequalities above, then we need to deal the sum of the components with the D operator.

$$\langle D\tilde{x}^n, y - y^{n+1} \rangle - \langle x - x^{n+1}, D^T y^{n+1} \rangle = \langle D\tilde{x}^n, y - y^{n+1} \rangle + \langle D(x^{n+1} - x), y^{n+1} \rangle =$$

$$= \langle D(x^{n+1} - x - \tilde{x}^n), y^{n+1} - y \rangle + \langle D(x^{n+1} - x), y \rangle =$$
$$= \langle D(x^{n+1} - \tilde{x}^n), y^{n+1} - y \rangle + \langle Dx^{n+1}, y \rangle - \langle Dx, y^{n+1} \rangle.$$

Replacing $\tilde{x}^n = 2x^n - x^{n-1}$ and adding and subtracting y^n we have

$$\langle D(x^{n+1} - \tilde{x}^n), y^{n+1} - y \rangle = \langle D(x^{n+1} - x^n), y^{n+1} - y \rangle - \langle D(x^n - x^{n-1}), y^{n+1} - y \rangle =$$

$$= \langle D(x^{n+1} - x^n), y^{n+1} - y \rangle - \langle D(x^n - x^{n-1}), y^{n+1} - y^n \rangle - \langle D(x^n - x^{n-1}), y^n - y \rangle,$$

But we also have the following inequalities:

$$\langle D(x^n - x^{n-1}), y^{n+1} - y^n \rangle \leq L\left\|x^n - x^{n-1}\right\|\left\|y^{n+1} - y^n\right\|$$

$$\leq \frac{L}{2}\sqrt{\frac{\sigma}{\tau}}\left\|x^n - x^{n-1}\right\|^2 + \frac{L}{2}\sqrt{\frac{\tau}{\sigma}}\left\|y^{n+1} - y^n\right\|^2 =$$
$$= L\sqrt{\sigma\tau}\left(\frac{\|x^n - x^{n-1}\|^2}{2\tau} + \frac{\|y^{n+1} - y^n\|^2}{2\sigma}\right);$$

Then putting all togather we have

$$\begin{aligned}F^*(y)+G(x) \geq & F^*(y^{n+1})+G(x^{n+1})+\\ & +\frac{\|y^n-y^{n+1}\|^2}{2\sigma}+\frac{\|y-y^{n+1}\|^2}{2\sigma}-\frac{\|y-y^n\|^2}{2\sigma}+\\ & +\frac{\|x^n-x^{n+1}\|^2}{2\tau}+\frac{\|x-x^{n+1}\|^2}{2\tau}-\frac{\|x-x^n\|^2}{2\tau}+\\ & +\langle D(x^{n+1}-x^n),y^{n+1}-y\rangle-\langle D(x^n-x^{n-1}),y^n-y\rangle\\ & -L\sqrt{\sigma\tau}\left(\frac{\|x^n-x^{n-1}\|^2}{2\tau}+\frac{\|y^{n+1}-y^n\|^2}{2\sigma}\right)+\\ & +\langle Dx^{n+1},y\rangle-\langle Dx,y^{n+1}\rangle,\end{aligned}$$

that can be rearranged as

$$\begin{aligned}\frac{\|y-y^n\|^2}{2\sigma}+\frac{\|x-x^n\|^2}{2\tau} \geq & \frac{\|y-y^{n+1}\|^2}{2\sigma}+\frac{\|x-x^{n+1}\|^2}{2\tau}+\\ & +(1-\sqrt{\sigma\tau}L)\frac{\|y^n-y^{n+1}\|^2}{2\sigma}+\\ & +\frac{\|x^n-x^{n+1}\|^2}{2\tau}-\sqrt{\sigma\tau}L\frac{\|x^n-x^{n-1}\|^2}{2\tau}+\\ & +\langle D(x^{n+1}-x^n),y^{n+1}-y\rangle-\langle D(x^n-x^{n-1}),y^n-y\rangle+\\ & +[G(x^{n+1})-F^*(y)+\langle Dx^{n+1},y\rangle]+\\ & -[G(x)+\langle Dx,y^{n+1}\rangle-F^*(y^{n+1})].\end{aligned} \tag{3.13}$$

We can now sum from $n=0$ up to $n=N-1$, with $x^{-1}=2x^0-\tilde{x}^0=x^0$. The first row of the equation above can be simplified

$$\sum_{n=0}^{N-1}\frac{\|y-y^n\|^2}{2\sigma}+\frac{\|x-x^n\|^2}{2\tau} \geq \sum_{n=0}^{N-1}\frac{\|y-y^{n+1}\|^2}{2\sigma}+\frac{\|x-x^{n+1}\|^2}{2\tau}+\dots$$

$$\implies \frac{\|y-y^0\|^2}{2\sigma}+\frac{\|x-x^0\|^2}{2\tau} \geq \frac{\left\|y-y^N\right\|^2}{2\sigma}+\frac{\left\|x-x^N\right\|^2}{2\tau}+\dots$$

The third row $\frac{1}{2\tau}\sum_{n=0}^{N-1}\|x^n-x^{n+1}\|^2-\sqrt{\sigma\tau}L\|x^n-x^{n-1}\|^2$ can also be rewritten as

$$L\frac{\left\|x^N-x^{N-1}\right\|^2}{2\tau}+(1-\sqrt{\sigma\tau}L)\sum_{n=1}^{N-1}\frac{\|x^n-x^{n-1}\|^2}{2\tau}.$$

The fourth row is a telescopic sum, therfore it reduces to

$$\langle D(x^N-x^{N-1}),y^N-y\rangle-\langle D(x^0-x^{-1}),y^0-y\rangle=\langle D(x^N-x^{N-1}),y^N-y\rangle,$$

but similarly as before we further have

$$\langle D(x^N - x^{N-1}), y^N - y\rangle \geq -L\left\|x^N - x^{N-1}\right\|\left\|y^N - y\right\| \geq$$

$$\geq -\frac{L}{2}\frac{1}{L\tau}\left\|x^N - x^{N-1}\right\|^2 - \frac{L}{2}L\tau\left\|y^N - y\right\|^2 = -L\frac{\left\|x^N - x^{N-1}\right\|^2}{2\tau} - L^2\sigma\tau\frac{\left\|y^N - y\right\|^2}{2\sigma}.$$

Hence, after the sum the inequality become

$$\begin{aligned}\frac{\|y - y^0\|^2}{2\sigma} + \frac{\|x - x^0\|^2}{2\tau} \geq (1 - L^2\sigma\tau)&\frac{\left\|y - y^N\right\|^2}{2\sigma} + \frac{\left\|x - x^N\right\|^2}{2\tau} + \\ &+(1 - \sqrt{\sigma\tau}L)\sum_{n=1}^{N-1}\frac{\|x^n - x^{n-1}\|^2}{2\tau} + \frac{\|y^n - y^{n-1}\|^2}{2\sigma} + \\ &+\sum_{n=1}^{N}[G(x^n) + \langle Dx^n, y\rangle - F^*(y) - G(x) - \langle Dx, y^n\rangle + F^*(y^n)].\end{aligned} \tag{3.14}$$

Notice that $\mathcal{G}(x^n, y^n) = G(x^n) + \langle Dx^n, \hat{y}\rangle - F^*(\hat{y}) - G(\hat{x}) - \langle D\hat{x}, y^n\rangle + F^*(y^n) \geq 0$, then, taking $(x, y) = (\hat{x}, \hat{y})$, we have

$$\frac{\|y - y^0\|^2}{2\sigma} + \frac{\|x - x^0\|^2}{2\tau} \geq (1 - L^2\sigma\tau)\frac{\left\|y - y^N\right\|^2}{2\sigma} + \frac{\left\|x - x^N\right\|^2}{2\tau},$$

which proves the first statement of the theorem.

The second statement follows as well from (3.14), indeed for convexity of F^* and G we have

$$\frac{1}{N}\sum_{n=1}^{N}F^*(y^n) \geq F^*(\bar{y}_N); \qquad \frac{1}{N}\sum_{n=1}^{N}G(x^n) \geq G(\bar{x}_N).$$

Therefore, we must have that for every $(x, y) \in X \times Y$

$$G(\bar{x}_N) + \langle D\bar{x}_N, y\rangle - F^*(y) - G(x) - \langle Dx, \bar{y}_N\rangle + F^*(\bar{y}_N)$$

$$\leq \frac{1}{N}\left(\frac{\|y - y^0\|^2}{2\sigma} + \frac{\|x - x^0\|^2}{2\tau}\right),$$

but then maximizing for $y \in B_2$ and minimizing for $x \in B_1$ we conclude that there exists a C such that

$$\mathcal{G}_{B_1\times B_2}(\bar{x}_N, \bar{y}_N) \leq \frac{C}{N}. \tag{3.15}$$

Notice that the boundedness of B_1 and B_2 is needed to preserve the meaning of the inequality through the maximization and minimization. Take now (x^*, y^*) a weak accumulation point for $(\bar{x}_N, \bar{y}_N)$, notice that such a point always exists because from the first statement we deduce that $(\bar{x}_N, \bar{y}_N)$ is bounded, hence weakly compact. Then, because G and F^* are lower semicontinuous and convex, they are also weakly lower semicontinuous, which means that also $\mathcal{G}_{B_1\times B_2}$ is weakly lower semicontinuous. Thus, moving (3.15) to the limit, we conclude that $\mathcal{G}_{B_1\times B_2}(x^*, y^*) = 0$, proving that (x^*, y^*) is a saddle point.

Finally, we need to prove the convergence of the algorithm for finite dimensional spaces. From the first statement we have that the sequence (x^N, y^N) is bounded, therefore it admits a subsequence (x^{n_k}, y^{n_k}) strongly (we are in finite dimension) converging to a point (x^*, y^*). But plugging $(x, y) = (\hat{x}, \hat{y})$ in (3.14) we can deduce that $\sum_{n=1}^{N-1} \frac{\left\|x^n - x^{n-1}\right\|^2}{2\tau} + \frac{\left\|y^n - y^{n-1}\right\|^2}{2\sigma}$ has to remain bounded for each N, consequently we must have

$$\lim_N x^n - x^{n-1} = \lim_n y^n - y^{n-1} = 0.$$

This means that (x^{n_k-1}, y^{n_k-1}) converges too to (x^*, y^*), in other words (x^*, y^*) is a fixed point of (3.12), that is a saddle point. To prove the actual convergence of (x^N, y^N), take $n_k < N$ and sum (3.13) from $n = n_k$ up to N for $(x, y) = (x^*, y^*)$. In this way we get an inequality similar to (3.14), but the last row is the sum of duality gaps, which are positive, thus they can be omitted, as well for the second row. The lower term in the telescopic sum from the fourth row of (3.13), instead, can not be erased, this leaves the inequality

$$\langle D(x^{n_k} - x^{n_k-1}), y^{n_k} - y^*\rangle + \frac{\|y^* - y^{n_k}\|^2}{2\sigma} + \frac{\|x^* - x^{n_k}\|^2}{2\tau} \geq$$

$$\geq (1 - L^2\sigma\tau)\frac{\left\|y^* - y^N\right\|^2}{2\sigma} + \frac{\left\|x^* - x^N\right\|^2}{2\tau}.$$

But the left hand side is going to 0, then also $(x^N, y^N) \longrightarrow (x^*, y^*)$.

□

This result shows also that the order of convergence is $O\left(\frac{1}{N}\right)$, which is quite slow. However, [13] shows how the introduction of the acceleration can impressively boost the algorithm speed in the case of (ROF_h), even though there are not clear theoretical explanations, it could even be happening only for the specific problem we are facing, and not in general for all those in the (3.10) form.

We can now see how to apply this algorithm specifically for (ROF_h). The space X is represented by $\mathcal{Q}^h$, while $Y = [\mathcal{Q}^2]^2$. The role of $F(Dx)$ is taken by the discrete total variation V_h, which indeed can be written as $V_h(u) = \|\nabla_h u\|_{L^1}$, that is $F = \|\cdot\|_{L^1}$ and $D = \nabla_h$. Instead the operator G is represented by the similarity term $\frac{1}{2\lambda}\|u - g\|_{L^2}^2$, where to lighten the notation we assumed $g = g^h$. In order to apply (3.12) we need $\nabla_h^T = -\mathrm{div}_h$, F^* and the inverse of $Id + \sigma\partial F^*$ and $Id + \tau\partial G$. Let's begin dealing with the latter.

$$\partial G(u) = \frac{1}{2\lambda}\partial\left(\|u-g\|_{L^2}^2\right) = \frac{1}{\lambda}(u-g).$$

Therefore, to invert $Id + \tau\partial G$ we have to sole for u in the following equation

$$u + \frac{\tau}{\lambda}(u-g) = v; \implies u = \frac{\lambda}{\lambda+\tau}\left(v + \frac{\tau}{\lambda}g\right). \tag{3.16}$$

Instead, we can compute F^* noticing how it can be expressed as supremum

$$F(\xi) = h^2\sum_{i,j=1}^{N}|\xi_{i,j}|_2 = h^2\sum_{i,j=1}^{N}\sup\{p_{i,j}\cdot\xi_{i,j} \mid |p_{i,j}| \le 1\} = \sup_{p\in K}\langle p,\xi\rangle_{L^2},$$

where $K = \{p \in \left[\mathcal{Q}^h\right]^2 : \forall i,j = 1,\dots,N \quad |p_{i,j}|_2 = 1\}$. This means that $F = \mathbb{I}_K^*$, where $H(p) = 0$ if $p \in K$ and $\mathbb{I}_K(p) = +\infty$ otherwise. We can notice that K is a non empty, closed and convex set, because it can be seen as the cartesian product of the closed balls $\{|p_{i,j}| \le 1\}$. As a consequence, H is proper, lower semicontinuous and convex, then $\mathbb{I}_K = \mathbb{I}_K^{**} = F^*$. Therefore, we have to compute $\partial\mathbb{I}_K$. On $\mathrm{int}(K)$ $\mathbb{I}_K$ is constant, which means $\partial\mathbb{I}_K = 0$. For $p \in \partial K$, instead, from the definition 2.2.5 we have

$$\partial\mathbb{I}_K(p) = \{v \mid \langle v, q-p\rangle \le 0 \ \forall q \in K\}.$$

But the v satisfying this condition are the outward normal vectors to K in p, that is

$$\partial\mathbb{I}_K(p) = \{\alpha n(p) \mid \alpha \ge 0,\ n(p) \text{ outward normal vector in } p\}.$$

While for $p \in \mathcal{Q}^h \setminus K$, $\partial\mathbb{I}_K(p) = \emptyset$. We need now to invert $Id + \sigma\partial\mathbb{I}_K$, that is we have to solve for ξ the equation

$$p \in \xi + \sigma\partial\mathbb{I}_K(\xi) \implies \frac{1}{\sigma}(p-\xi) \in \partial\mathbb{I}_K(\xi).$$

Let's divide the analysis in two cases: $p \in \mathrm{int}(K)$ and $p \notin \mathrm{int}(K)$.

- $p \in \mathrm{int}(K)$: for any $\xi \in \mathrm{int}(K)$, $\partial\mathbb{I}_K = \{0\}$, therefore in this case we have only one possible solution, $p = \xi$. If $\xi \in \partial K$ we have no possible solutions, because $p - \xi$ would always be an inward vector.

- $p \notin \mathrm{int}(K)$: clearly ξ can not be in the interior of K, but now $p - \xi$ is an outward vector, which is also normal if ξ is an orthogonal projection of p on ∂K. Though, Such projection is unique on convex sets.

In conclusion we found that $(Id + \sigma\partial F^*)^{-1}$ is the orthogonal projection on K: Π_K. This is easy to compute explicitly, because the set K is built as a cartesian product, therefore $(\Pi_K(p))_{i,j}$ will be the projection of $p_{i,j}$ on the (i,j)-th constraint $\{|p_{i,j}| \leq 1\}$, that is

$$(\Pi_K(p))_{i,j} = \begin{cases} p_{i,j} & \text{if } |p_{i,j}| \leq 1 \\ \dfrac{p_{i,j}}{|p_{i,j}|} & \text{otherwise.} \end{cases}$$

Summing everything up we have to use the following iterative step:

$$\begin{cases} \xi^{n+1} = \Pi_K(\xi^n + \sigma\nabla_h\tilde{u}^n) \\ u^{n+1} = \dfrac{\lambda}{\lambda+\tau}\left(u^n + \tau\mathrm{div}_h\xi^{n+1} + \dfrac{\tau}{\lambda}g\right) \\ \tilde{u}^{n+1} = 2u^{n+1} - u^n, \end{cases}$$

where, as before, we called u the primal variable and ξ the dual. The condition of convergence is $\sigma\tau\|\nabla_h\|^2 < 1$, and we can have a rough estimation of $\|\nabla_h\|$

$$\sum_{i,j=1}^{N}\left|(\nabla_h u)_{i,j}\right|^2 h^2 = \frac{1}{h^2}\sum_{i,j=1}^{N} h^2[(u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2] \leq \frac{8}{h^2}\sum_{i,j=1}^{N} h^2 u_{i,j}^2,$$

$$\implies \|\nabla_h\| \leq \frac{2\sqrt{2}}{h}.$$

In order to avoid the inconvenience of scaling σ and τ with h, it could be useful to use as linear operator $D = h\nabla_h$. This is like substituting λ with $\frac{\lambda}{h}$, a part from the change of operator D, but practically (doing few simplifications) it turns to be analogous of taking $h\tau$ and $h\sigma$ in place of τ and σ. Then we have the routine

$$\begin{cases} \xi^{n+1} = \Pi_K(\xi^n + \sigma h\nabla_h\tilde{u}^n) \\ u^{n+1} = \dfrac{\lambda}{\lambda+h\tau}\left(u^n + \tau h\mathrm{div}_h\xi^{n+1} + \dfrac{h\tau}{\lambda}g\right) \\ \tilde{u}^{n+1} = 2u^{n+1} - u^n. \end{cases} \tag{3.17}$$

3.3 Experimental results

In this last section we are going to show the effect of the ROF minimization on an image, furthermore the algorithm and its convergence will be validated using the analytical solution found in (2.16). The routine (3.17) has been implemented in Matlab to deal with gray scale images, which are square of side length of 1 covered with $N \times N$ pixels, so that we define $h = N^{-1}$. For the algorithm parameters, we generally set $\tau = \sigma = 0.1$, while as starting configuration we took the dual variable

ξ^0 to be (0,0) everywhere and the primal variable $u^0 = g$, because we expect the solution to be close to the reference image, at least for a rather accurate choice of λ. We usually run 1500 iterations, which are often more than enough for reaching a satisfying state of convergence. We also set a break if the relative difference in energy between two consecutive steps was below a given tolerance. However we used very fine values of it like 10^{-10} or even 10^{-12}, the reason why is that, because this is not properly a gradient descent method, the energy oscillate through the first iterations, and in proximity of an apex u^n may face very little changes for few iterations, that may trigger a break if the tolerance is too low, preventing to reach the convergence.

To get used to the behaviour of ROF, let's right away see it in action. We are going to apply the method on the Matlab repertoire image `cameraman.tif`, of size 256×256, on which we applied a white noise with standard deviation 25 (as a reference we remind that the gray gradient goes from 0, black, to 255, white). Four different value of λ: $0.4 \cdot 255$, $0.2 \cdot 255$, $0.1 \cdot 255$, $0.05 \cdot 255$. We have to point out that we are representing λ as the standardized value, that would be used if $\mathrm{Im}(g)$ was in [0,1], multiplied by the scaling factor. Indeed, as pointed out in the section 2.3, the fidelity term scale quadratically, while the total variation linearly, this means that in order to recover the same results we need to rescale λ the same as for g. In figure 3.1 you can see the original image and the noisy version, while in figure 3.2 we have the results of ROF-denoising.

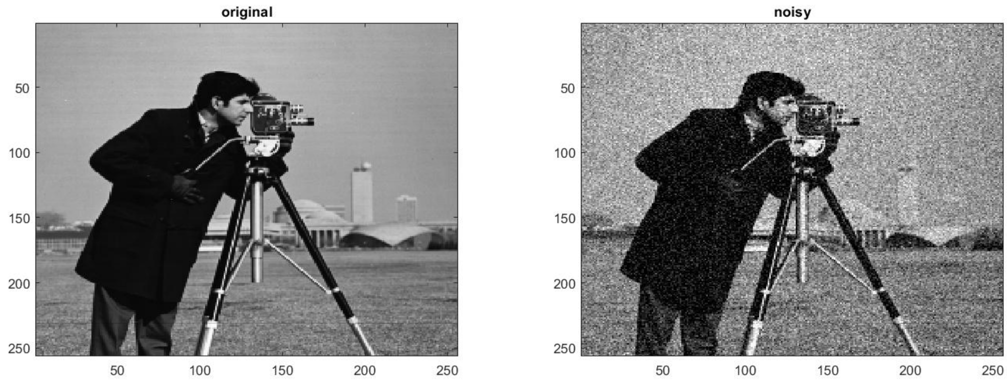

Figure 3.1: original cameraman image (left), noisy version (right).

Reasonably the effect of the total variation is much more visible in the solution for the highest $\lambda = 0.4 * 255$. Visibly this term is trying to flatten the image as much as possible, creating huge spots of uniform colors, while still preserving the contours of the objects in high contrast with the background. Indeed, we can clearly distinguish the camera and the man profile, even though a lot of facial details are lost and the man's gloves completely blends with its jackets. The

Figure 3.2: Resulting images after applying the ROF minimization with different $\lambda/255$: 0.4 (up left), 0.2 (up right), 0.1 (down left), 0.05 (down right).

distinction between the sky and the ground is marked, despite the latter became almost monochromatic, with no signs of grass left. The buildings in the background are not really recognizable and the two skyscrapers completely disappear.

As expected, reducing λ, the level of details grows, although we keep this impressionistic kind of look, typical of the ROF model. For $\lambda = 0.1 * 255$ all the objects are fully recognizable, with the exception of the shorter skyscraper. The grass is finally visible, however not with a comparable level of detail as the original image. Recovering it is a hard task, because, as we can see from the noisy image, the grass pattern is not very discernible from the noise. We can finally notice that for λ too low, in the example $0.05 * 255$, the effect of the total variation is so weak that the noise is not completely erased.

We will now have a look to the example of the square, explored analytically at the end of section 2.2. However, the solution shown is not totally repeatable, because it assumed an infinite domain. This actually change the solution, even if the square is totally included in the bounded domain Ω. Indeed, let's take $\lambda > R^*$, case for which the analytical solution (2.16) results identically zero, and consider as a competitor a constant function c, then its energy is:

$$\frac{1}{2}\int_\Omega (c-\chi_Q)^2 \,\mathrm{d}x = \frac{c^2|\Omega \setminus Q| + (1-c)^2|Q|}{2},$$

which is a quadratic function in c with minimum for $c = \frac{|Q|}{|\Omega|}$. If $\Omega = (0,1)^2$ and Q is of side L, then the constant $c = L^2$ is better than the 0 function. Interestingly, this is also close to the solution found by the solver. We put $L = 0.5$, this means that $R^* = \frac{1}{2(2+\sqrt{\pi})} \in (0.1, 0.2)$, so we can chose $\lambda = 0.2$ to gain the desired effect. The solution found was slightly increasing (not enough to be visible) in both coordinate at a comparable rate, this may be due to numerical reasons of for discretization factors as the anisotropy introduced by the discrete derivative, which links the increment in the cell in position (i,j) only with those in $(i+1,j)$ and $(i,j+1)$. However its average was $L^2 = 0.25$. Even more interesting is the fact that this kind of solution remains if we choose $\lambda = 0.1$. My guess is that the solution in this case is

$$u(x) = \max\left\{s_{min}, 1 - \frac{\lambda}{r(x)}\right\}, \tag{3.18}$$

where $r(x)$ is defined as in (2.16), while s_{min} is the lowest s for which C^{R_s} (as defined in (2.9)) is convenient with respect to Ω in the problem (ROFs). From all the observation on the numerical solver, such a solution seems plausible, although I was not able to prove or disprove its correctness formally. In order to compute s_{min} or $R_{min} = R_{s_{min}} = \frac{\lambda}{1-s_{min}}$, we need to find the smallest s or R_s for which

$$s|\Omega| - L^2 \geq (1-s)(R_s P(C^{R_s}) - \left|C^{R_s}\right|) = (1-s)(4R_s L - L^2 - (4-\pi)R_s^2),$$

where the last equality comes from (2.14), while on the left hand side there is the energy of Ω for (ROFs). But, because $s|\Omega| - L^2 = |\Omega| - L^2 - (1-s)|\Omega|$, we can multiply the equation by R_s to get

$$(|\Omega| - L^2)R_s - \lambda|\Omega| \geq 4L\lambda R_s - \lambda L^2 - (4-\pi)\lambda R_s^2,$$

$$\begin{aligned}
\implies R_{min} &= \frac{\sqrt{(|\Omega| - L^2 - 4L\lambda)^2 + 4\lambda^2(4-\pi)(|\Omega| - L^2)} - (|\Omega| - L^2 - 4L\lambda)}{2\lambda(4-\pi)} = \\
&= \frac{2\lambda(|\Omega| - L^2)}{\sqrt{(|\Omega| - L^2 - 4L\lambda)^2 + 4\lambda^2(4-\pi)(|\Omega| - L^2)} + (|\Omega| - L^2 - 4L\lambda)} = \\
&= \frac{2\lambda}{\sqrt{(1-\zeta)^2 + (4-\pi)\lambda\zeta/L} + 1 - \zeta},
\end{aligned}$$

where $\zeta = \frac{4L\lambda}{|\Omega|-L^2}$. Then, $s_{min} = 1 - \frac{\lambda}{R_{min}}$ can be computed as follows:

$$s_{min} = \frac{1}{2}\left(1 + \zeta - \sqrt{(1-\zeta)^2 + (4-\pi)\frac{\lambda\zeta}{L}}\right).$$

In the case of $\lambda = 0.1$ and $L = 0.5$ we can indeed compute that $s_{min} \approx 0.2514$, while $1 - \frac{\lambda}{R^*} \approx 0.2455$, making the constant solution ($u \equiv L^2$) better. The actual bound for λ can be found imposing $s_{min} < 1 - \frac{\lambda}{R^*}$ which implies

$$\zeta - 1 + 2\frac{\lambda}{R^*} < \sqrt{(1-\zeta)^2 + (4-\pi)\frac{\lambda\zeta}{L}},$$

$$\implies (4-\pi)\frac{\lambda\zeta}{L} > 4\frac{\lambda}{R^*}(\zeta - 1) + 4\frac{\lambda^2}{R^{*2}}.$$

We notice that in taking the second power of the inequality we assumed $\zeta - 1 + 2\frac{\lambda}{R^*} \geq 0$ for the validity of the following steps. If $\zeta - 1 + 2\frac{\lambda}{R^*} < 0$, then for sure $s_{min} < 1 - \frac{\lambda}{R^*}$, but can be verified afterward the the upper bound we are going to find is more strict than this one anyway. We can now multiply the inequality above by $\frac{R^{*2}}{4\lambda} > 0$ and substitute ζ to get

$$\lambda\left(1 - \frac{R^{*2}}{|\Omega| - L^2}(4 - \pi - 4(2 + \sqrt{\pi}))\right) < R^*,$$

$$\lambda\left(1 + \frac{L^2}{|\Omega| - L^2}\right) < R^*, \quad \implies \lambda < R^*\left(1 - \frac{L^2}{|\Omega|}\right).$$

For $L = 0.5$ and $|\Omega| = 1$, this condition become, approximately, $\lambda < 0.0994$, and indeed using $\lambda = 0.099$ we can start, even so barely, seeing a non constant solution, as in figure 3.3. Furthermore, with the choice of λ precisely $\lambda = R^*\left(1 - \frac{L^2}{|\Omega|}\right)$, we have $s_{min} = 1 - \frac{\lambda}{R^*} = \frac{L^2}{|\Omega|}$, coming back to the optimal constant.

Even though we do not know the for sure if it is the solution, from an experimental point of view it seems quite close to the actual one, this justifies that we are going to assume that function to choose λ. Because we want to see if it behaves as expected from the continuous solution, we have to reduce s_{min} under a threshold α for which the difference in brightness is barely visible. We choose $\alpha = 5\%$. Now, we develop the computations:

$$s_{min} \leq \alpha \implies (1 + \zeta - 2\alpha)^2 \leq (1-\zeta)^2 + (4-\pi)\frac{\lambda\zeta}{L},$$

$$2\zeta - 4\alpha - 4\alpha\zeta + 4\alpha^2 \leq -2\zeta + (4-\pi)\frac{\lambda\zeta}{L},$$

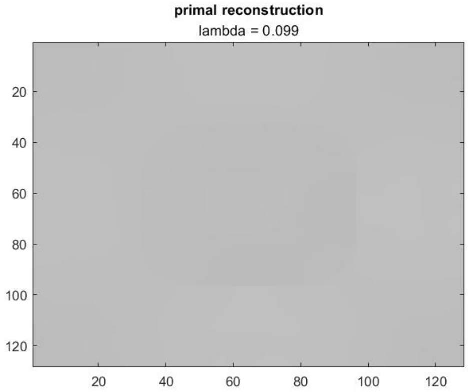

Figure 3.3

$$(1-\alpha)\zeta - \frac{4-\pi}{4L}\lambda\zeta \leq \alpha(1-\alpha), \implies 4L(1-\alpha)\lambda - (4-\pi)\lambda^2 \leq \alpha(1-\alpha)(|\Omega| - L^2);$$

$$\lambda \leq b - \sqrt{b^2 - c}, \tag{3.19}$$

with $b = \frac{2L(1-\alpha)}{4-\pi}$ and $c = \alpha(1-\alpha)\frac{|\Omega|-L^2}{4-\pi}$. When $L = 0.5$ and $\alpha = 0.05$, we find $\lambda \leq 0.018912$, this upper bound is what we are going to use for λ is the experiments.

We are going to compare the results using 5 resolutions, identified by the number N of pixels per dimension, $N = 32, 64, 128, 256, 384$. Take as reference the image 3.4. At first we notice that the background is not totally white, as in the reference square, even if it assume values in average below 0.05 (I am considering 0 white and 1 black). This is due to value of the threshold α which is not too low, however reducing it would have brought to a λ too little, making the decay on the square corners too slow and not very visible. Discussing further the behaviour of the background, we observed higher values near the corners of the frame, especially the upper left one as a consequence of the anisotropy from the derivative discretization. On the other hand there were a bit of wavering near the inner square corners probably due to some numerical error. This phenomenon is mildly visible in image 3.5, taken from the case of $N = 384$. Talking about the square corners, we observe the expected rounding, progressively more evident increasing N. We can also appreciate, especially for smaller N, the

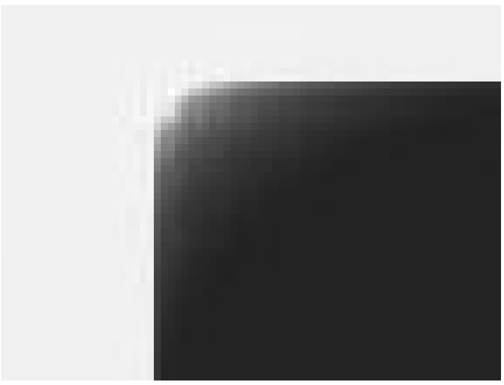

Figure 3.5

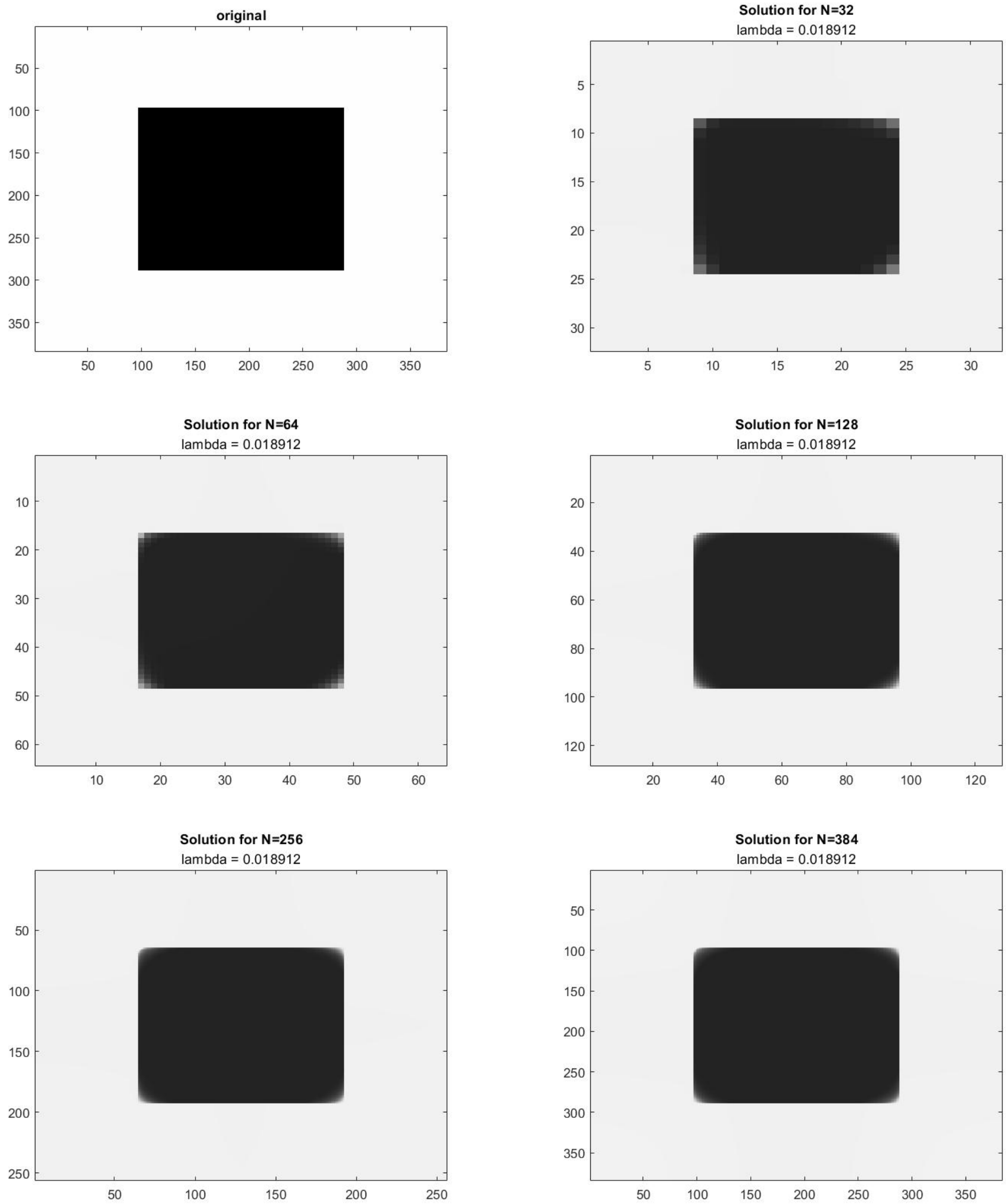

Figure 3.4: Example of the square: reference square (notice the axes' different scaling) image compared with the computational solution with different resolution.

asymmetry of the solution, which results in lower values for higher indices of row and column. We can in conclusion compare the maximum values, in table 3.1. These promisingly seem to approach the analytical maximum $1 - \lambda/R^* \approx 0.8573$.

	N=32	N=64	N=128	N=256	N=384	analyt.
max val	0.8612	0.8593	0.8582	0.8577	0.8576	0.8573

Table 3.1

86

Appendix A

Radon-Nikodym derivatives on Radon measures

In this appendix some basic results on Radon measures and their Radon-Nikodym derivatives. We begin defining a Radon measure.

Definition A.0.1. Let X be a locally compact and separable metric space, we define:

- $\mathcal{B}(X)$, the *Borel σ-algebra* as the σ-algebra generated by the open sets (or equivalently the balls).
- We call a measure defined on the measurable space $(X, \mathcal{B}(X))$ as a *Borel measure*.
- A Borel finite on the compacts is called *Radon measure*.

The Radon measures on $\mathbb{R}^n$, paired with the euclidean distance, are of particular interest because satisfy the following properties.

Proposition A.0.2 (inner regularity)**.** *Let μ be a σ-finite Borel measure on X, then for E measurable:*

$$\mu(E) = \sup\{\mu(K) : \ K \subset E, \, K \textit{ compact}\}.$$

Proposition A.0.3 (outer regularity)**.** *Let μ a Borel measure on X s.t. we can partitionate $X = \bigcup_h X_h$, where for all h $\mu(X_h) < \infty$, then for E measurable:*

$$\mu(E) = \inf\{\mu(A) : \ E \subset A, \, A \textit{ open}\}.$$

We call $C_0(\Omega)$ the closure under infinity norm of the space $C_c(\Omega)$.

Proposition A.0.4 (Riesz representation)**.** *Let Ω open in $\mathbb{R}^n$ and L be a linear and continuous functional on $[C_0(\Omega)]^m$ with respect to the infinity norm. Then there exists a unique $\mathbb{R}^m$-valued Radon measure such that:*

$$L(u) = \sum_{h=1}^{n} \int_{\Omega} u_h \, \mathrm{d}\mu_h.$$

The aim of this appendix is showing that Radon measures on $\mathbb{R}^n$ have a characterization of a derivative as a limit of a difference quotient. We begin defining absolute continuity and the Radon-Nikodym derivative.

Definition A.0.5. We say that a measure ν is *absolutely continuous* with respect to a positive measure μ if, given a measurable set B:

$$\mu(B) = 0 \implies \nu(B) = 0.$$

We denote this relation as: $\nu \ll \mu$.

Note that ν can also be $\mathbb{R}^N$-valued.

Definition A.0.6. Given a scalar measure μ on $(X, \mathcal{E})$ and a measurable function $f : X \to \mathbb{R}^m$, we denote the measure $f\mu$ such that for every $B \in \mathcal{E}$:

$$f\mu(B) = \int_B f \, \mathrm{d}\mu = \left(\int_B f_1 \, \mathrm{d}\mu, \ldots, \int_B f_m \, \mathrm{d}\mu \right).$$

We observe that always $f\mu << \mu$, as the integral of μ-negligible sets is 0. But actually every absolutely continuous measure can be written uniquely as one of the kind $f\mu$.

Proposition A.0.7 (Radon-Nikodym derivative)**.** *Let ν a $\mathbb{R}^N$-valued measure on a measurable space $(X, \mathcal{E})$ and μ a positive σ-finite measure on $(X, \mathcal{E})$. If $\nu \ll \mu$, then there is a unique f measurable such that $\nu = f\mu$.*

We call such f the Radon-Nikodym derivative, sometimes it is also denoted as $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ or $D_\mu \nu$. Before dealing with the case of Radon measures, we need to recall some results. From now on we will take the space X to be metric.

Theorem A.0.8 (Besicovitch covering)**.** *Let A be a bounded set in $\mathbb{R}^n$ and $\mathcal{F}$ a fine cover of A, that is $\forall x \in A$, $\exists \bar{\rho} > 0$ s.t. $\forall \rho \leq \bar{\rho}$, $\bar{B}(x, \rho) \in \mathcal{F}$. Then we can find an at most countable subcover $\mathcal{F}'$, for which every $x \in A$ belongs to at most b_n balls.*

Now for ν vector valued Radon measure and μ positive Radon measure, we call:

$$D_\mu^+ \nu(x) = \limsup_{\rho \to 0^+} \frac{\nu(B(x, \rho))}{\mu(B(x, \rho))}; \qquad D_\mu^- \nu(x) = \liminf_{\rho \to 0^+} \frac{\nu(B(x, \rho))}{\mu(B(x, \rho))};$$

where $B(x,\rho)$ is the ball centered in x of radius ρ. We notice that for a Radon measure λ, the function $\lambda(B(x,\rho))$ is continuous in x, indeed

$$|\lambda(B(x+h,\rho)) - \lambda(B(x,\rho))| = \left| \int_X \chi_{B(x+h,\rho)} - \chi_{B(x,\rho)} \, d\lambda \right|.$$

$\chi_{B(x+h,\rho)}$ is converging pointwise, for $h \to 0$, to $\chi_{B(x,\rho)}$, while $\left|\chi_{B(x+h,\rho)} - \chi_{B(x,\rho)}\right| \leq 2\chi_{B(x,\rho+1)} \in L^1(X,\lambda)$ as for radon measures $\lambda(B(x,\rho+1))$ is always finite. Therefore, for dominated convergence we have the continuity. Then, $\nu(B(x,\rho))$ and $\mu(B(x,\rho))$ are continuous and consequently measurable, this implies that also $D^+_\mu\nu$ and $D^-_\mu\nu$ are measurable. We want to show that $D^+_\mu\nu = D^-_\mu\nu = D_\mu\nu$ μ-a.e. and $\nu = D_\mu\nu\mu$. In order to do it we need the next proposition.

Proposition A.0.9. *Let μ and ν positive Radon measures on $\mathbb{R}^n$ and $E \subset \operatorname{supp}\mu$ Borel set, then*

$$D^-_\mu\nu(x) \leq t, \quad \forall x \in E \implies \nu(E) \leq t\mu(E),$$
$$D^+_\mu\nu(x) \geq t, \quad \forall x \in E \implies \nu(E) \geq t\mu(E).$$

Observation A.0.10. $\nu(E)$ finite $\implies \mu\left(\left\{x \in E : \; D^+_\mu\nu(x) = +\infty\right\}\right) = 0.$

Proof. Since we are working with Radon measures, we can take E bounded, then we can generalize using the inner regularity. Take $A \supset E$ open and bounded, then consider $\varepsilon > 0$. We define:

$$\mathcal{F} = \left\{\bar{B}(x,\rho) : \; x \in E, \, \bar{B}(x,\rho) \subset A, \, \nu(\bar{B}(x,\rho)) < (t+\varepsilon)\mu(\bar{B}(x,\rho))\right\}.$$

Since $D^+_\mu\nu \leq t$, then $\forall x \in E \; \exists \bar{\rho}$ s.t. $\forall \rho \leq \bar{\rho}$, $\bar{B}(x,\rho) \in \mathcal{F}$. This means $\mathcal{F}$ is a fine cover of E, then by Besicovitch theorem we can find a countable subcovering $\mathcal{F}'$ such that each point is in at most b_n balls. Then we have:

$$\nu(E) \leq \sum_{B \in \mathcal{F}'} \nu(B) \leq \sum_{B \in \mathcal{F}'} (t+\varepsilon)\mu(B) \leq (t+\varepsilon)\mu(A),$$

but from the arbitrariness of ε and the outer regularity of μ we can conclude $\nu(E) \leq t\mu(E)$.

Similarly, if $D^+_\mu\nu \geq t$, for any $A \subset E$ compact we can take:

$$\mathcal{F} = \left\{\bar{B}(x,\rho) : \; x \in A, \, \bar{B}(x,\rho) \subset E, \, \nu(\bar{B}(x,\rho)) > (t-\varepsilon)\mu(\bar{B}(x,\rho))\right\}.$$

Again, this is a fine covering and consequently we can extract a $\mathcal{F}'$ as above and conclude:

$$\nu(E) \geq \sum_{B \in \mathcal{F}'} \nu(B) \geq \sum_{B \in \mathcal{F}'} (t-\varepsilon)\mu(B) \geq (t-\varepsilon)\mu(A).$$

Then, using the inner regularity, we can conclude.

□

Theorem A.0.11. *Let ν be an $\mathbb{R}^n$-valued Radon measure absolutely continuous with respect to μ a positive Radon measure, then the limit*

$$D_\mu \nu(x) = \lim_{\rho \to 0^+} \frac{\nu(B(x,\rho))}{\mu(B(x,\rho))}$$

exists finite for μ-almost every $x \in$supp μ. Furthermore, $D_\mu \nu$ is the Radon-Nikodym derivative of ν with respect to μ, in other words $\nu = D_\mu \nu \, \mu$.

Proof. We can assume ν to be a one dimensional positive measure. Indeed we can immediately generalize to non positive measures with the decomposition in positive and negative part $\nu = \nu^+ - \nu^-$, while with multidimensional measures we can deal componentwise.

ν, as a Radon measure, is finite on the compacts, then from the observation A.0.10 the set $\{D^+_\mu \nu = +\infty\}$ is locally μ-negligible, that is:

$$\forall K \,\text{compact}, \quad \mu\left(\{D^+_\mu \nu = +\infty\} \cap K\right) = 0.$$

Then from inner regularity, $\{D^+_\mu \nu = +\infty\}$ is μ-negligible. We now define, for B measurable set:

$$\lambda^+(B) = \int_B D^+_\mu \nu \, \mathrm{d}\mu; \qquad \lambda^-(B) = \int_B D^-_\mu \nu \, \mathrm{d}\mu;$$

In order to show that $D_\mu \nu$ exists and it is the Radon-Nikodym derivative, We want to show that $\lambda^+ \leq \nu \leq \lambda^-$. For the first inequality, given a measurable set B in supp μ and an arbitrary $t > 1$, we define the disjoint sets:

$$N^+_B = \left\{x \in B : \, D^+_\mu \nu(x) = 0\right\}; \qquad B^+_n = \left\{x \in B : \, D^+_\mu \nu(x) \in (t^n, t^{n+1}]\right\}, \; n \in \mathbb{Z}.$$

Since on B^+_n $D^+_\mu \nu \geq t^n$, from the previous proposition we can conclude:

$$\lambda^+(B^+_n) = \int_{B^+_n} D^+_\mu \nu \, \mathrm{d}\mu \leq t^{n+1} \mu(B^+_n) \leq t \nu(B^+_n),$$

while $\lambda^+(N^+_B) = 0 \leq t\nu(B^+_n)$. Hence, as N^+_B and B^+_n are a disjoint cover of B, we get:

$$\lambda^+(B) = \lambda^+(N^+_B) + \sum_{n \in \mathbb{Z}} \lambda^+(B^+_n) \leq t\nu(N^+_B) + \sum_{n \in \mathbb{Z}} t\nu(B^+_n) = t\nu(B).$$

But for arbitrariness of $t > 1$ we have the first inequality $\lambda^+ \leq \nu$.

Similarly, we can define:

$$N^-_B = \left\{x \in B : \, D^-_\mu \nu(x) = 0\right\}; \qquad B^-_n = \left\{x \in B : \, D^-_\mu \nu(x) \in (t^n, t^{n+1}]\right\}, \; n \in \mathbb{Z}.$$

Thus, $\lambda^-(N_B^-) = 0$ and

$$\lambda^-(B_n^-) = \int_{B_n^-} D_\mu^- \nu \, \mathrm{d}\mu \geq t^n \mu(B_n^-) \geq \frac{1}{t} \nu(B_n^-),$$

$$\implies \lambda^-(B) \geq \frac{1}{t} \nu(B).$$

Finally for arbitrariness of $t > 1$ we can conclude. $\square$

Appendix B

Equivalence of BV definitions

In this section we will prove that this two definitions 1.1.3 and 1.1.4 are equivalent and provide some more insight on the total variation of measure. For more details see [5] the first two chapters. Before the proof, we need to introduce some measure tools.

Definition B.0.1. Let μ a vector valued measure on the measurable space $(X, \mathcal{E})$, we define its *total variation* as:

$$|\mu|(E) := \sup\left\{\sum_{h\in H}|\mu(E_h)| \ : \ H \subseteq \mathbb{N} \ , \{E_h\}_{h\in H} \subset \mathcal{E}, \ E = \bigsqcup_{h\in H} E_h\right\}$$

where $\sqcup$ denotes a disjoint union, therefore the supremum is made over the disjoint numerable partitions of E. We now show that this total variation is still a measure.

Proposition B.0.2. *$|\mu|$ is a positive measure and μ is finite if and only if $|\mu|$ is finite.*

Proof. $|\mu|$ clearly is positive, we have Let $\{E_h\}_{h\in\mathbb{N}}, \{F_j\}_{j\in\mathbb{N}} \subset \mathcal{E}$ two disjoint coverings of E. Therefore, $F'_{hj} = E_h \cap F_j$ is a countable disjoint covering of E_h, F_j and E. So by σ-additivity we have:

$$\sum_j |\mu(F_j)| = \sum_j \left|\sum_h \mu(F'_{hj})\right| \leq \sum_{h,j} \left|\mu(F'_{hj})\right| \leq \sum_h |\mu|(E_h),$$

Thus, if we take the supremum over the partitions $\{F_j\}$ we have $|\mu|(E) \leq$

$\sum_h |\mu|(E_h)$. On the other hand, since F'_{hj} is a countable partition of E, we have:

$$|\mu|(E) \geq \sum_{h,j} \left|\mu(F'_{hj})\right| \geq \sum_h \left|\sum_j \mu(F'_{hj})\right|.$$

Then, considering $\{F'_{hj}\}$ as a covering of E_h, we can extract the supremum to get $|\mu|(E) \geq \sum_h |\mu|(E_h)$. Therefore, we have the σ-additivity:

$$|\mu|(E) = \sum_h |\mu|(E_h).$$

Instead, for the second statement, we see immediately that $|\mu|(X) \geq |\mu(X)|$, so $|\mu|$ finite implies μ finite.

We will prove the reverse implication at first for $N = 1$ Now suppose $|\mu|(X) = +\infty$ but $|\mu(X)| < +\infty$, then there must exist a partition $\{X_h\}$ s.t.

$$\sum_{h=0}^{\infty} |\mu(X_h)| > 2(|\mu(X)| + 1).$$

Let's call I_+ the set of h such that $\mu(X_h) > 0$ and I_- the set of those for which $\mu(X_h) < 0$. Then for at least one of these sets, lets say I_+ it must be verified:

$$\sum_{h \in I_+} |\mu(X_h)| > |\mu(X)| + 1 > 1.$$

If we call $E = \bigcup_{h \in I_+} X_h$, then:

$$|\mu(E)| = \left|\sum_{h \in I_+} \mu(X_h)\right| = \sum_{h \in I_+} |\mu(X_h)| > |\mu(X)| + 1,$$

because all the $\mu(X_h)$ share the same sign.

$$|\mu(X \setminus E)| = |\mu(X) - \mu(E)| \geq |\mu(E)| - |\mu(X)| > 1.$$

However, since $|\mu|(X) = +\infty$, either $|\mu|(E) = +\infty$ or $|\mu|(X \setminus E) = +\infty$. We call F_1 a set between E and $X \setminus E$ has infinity total variation and E_1 the other one. Then we apply iteratively the same argument, but with F_i instead of X. In this way we find a succession of disjointed sets E_i such that $|\mu(E_i)| > 1$ for all i. If I had at this succession $E_0 = \bigcap_{i=1}^{+\infty} F_i$, we have a partition of X, therefore $\mu(X) = \sum_{i=0}^{+\infty} \mu(E_i)$. However, since $|\mu(E_i)| > 1$, this sum cannot converge, that is a contradiction.

If $N \geq 2$, then there is a constant C such that $|\mu(X)| \leq C \sum_{j=1}^{N} |\mu_j(X)|$. Then:

$$|\mu|(X) = \sup\left\{\sum_h |\mu(E_h)|\right\} \leq \sup\left\{C \sum_h \sum_{j=1}^{N} |\mu_j(E_h)|\right\} \leq$$

$$\leq C\sum_{j=1}^{N}\sup\left\{\sum_{h}|\mu_j(E_h)|\right\} = C\sum_{j=1}^{N}|\mu_j|(X),$$

where all the sup have to be intended as over the partitions of X. Thus, μ finite implies that each component is finite, then also all the $|\mu_j|$ are finite. This concludes that $|\mu|$ is finite.

□

We will use this definition to show that for $f \in BV(\Omega)$, $V(f,\Omega) = |Df|(\Omega)$. Let's now focus on the relation between μ and $|\mu|$. We have a helpful expression for $|f\mu|$.

Proposition B.0.3. *If μ is positive and $f \in [L^1(X;\mu)]^m$, then*

$$|f\mu|(B) = |f|\mu(B); \quad \forall B \in \mathcal{E}.$$

Proof.

- $|f\mu|(B) \leq |f|\mu(B)$: given a partition of B in disjoint measurable sets $\{B_h\}$, from the Jensen inequality we have:

$$\sum_h|f\mu(B_h)| = \sum_h\left|\int_{B_h} f\,\mathrm{d}\mu\right| \leq \sum_h\int_{B_h}|f|\,\mathrm{d}\mu = \int_B|f|\,\mathrm{d}\mu.$$

- $|f\mu|(B) \geq |f|\mu(B)$: consider a succession $\{z_h\}_{h\in\mathbb{N}} \subset \mathbb{R}^m$ dense in $\mathcal{S}^{m-1}$ (the spherical surface embedded in $\mathbb{R}^m$). For $\epsilon > 0$ we define:

$$\sigma_\epsilon(x) := \min\left\{h \in \mathbb{N} \mid \langle f(x), z_h\rangle \geq (1-\epsilon)|f(x)|\right\}.$$

We can notice that the functions $\psi_h(x) = \langle f(x), z_h\rangle - (1-\epsilon)|f(x)|$ are measurable and

$$\sigma_\epsilon^{-1}(h) = \psi_h^{-1}([0,+\infty)) \setminus \left(\bigcup_{j=1}^{h-1}\psi_j^{-1}([0,+\infty))\right).$$

Therefore, $B_h = \sigma_\epsilon^{-1}(h) \cap B$ are measurable and form a disjoint partition of B.

$$(1-\epsilon)|f|\mu(B) = \sum_h\int_{B_h}(1-\epsilon)|f|\,\mathrm{d}\mu \leq \sum_h\int_{B_h}\langle f(x), z_h\rangle\,\mathrm{d}\mu =$$

$$= \sum_h\langle\int_{B_h} f(x)\,\mathrm{d}\mu, z_h\rangle = \sum_h\langle f\mu(B_h), z_h\rangle \leq \sum_h|f\mu(B-h)| \leq |f\mu|(B);$$

$$\Longrightarrow (1-\epsilon)|f|\mu(B) \leq |f\mu|(B),$$

but for the arbitrariness of ϵ we get the result.

□

Now we can finally get the *polar decomposition* for measures.

Proposition B.0.4 (Polar decomposition)**.** *let μ a $\mathbb{R}^N$-valued measure on $(X, \mathcal{E})$.*

$$\exists! f : X \to \mathcal{S}^{N-1}, \text{ integrable, s.t. } \mu = f|\mu|.$$

Proof. Clearly μ is absolutely continuous with respect to $|\mu|$. Therefore, $\exists! f \in [L^1(X,\mu)]^N$ s.t. $\mu = f|\mu|$, and from B.0.3 we have:

$$|\mu| = |f|\mu|| = |f||\mu|,$$

$$\implies |f| = 1 \quad |\mu| - a.e.$$

□

Observation B.0.5. If we look at radon measures as dual of $[C_0(X)]^m$, then we can notice that its dual norm is bounded by the total variation:

$$\left|\int_X g \,\mathrm{d}\mu\right| = \left|\int_X gf \,\mathrm{d}|\mu|\right| \leq \int_X |g| \,\mathrm{d}|\mu| \leq |\mu|(X)\|g\|_\infty.$$

However, it can be proven that actually $\|\mu\| = |\mu|(X)$.

Finally, we are ready to prove the equivalence between the two definitions of BV.

Proposition B.0.6. *The definitions 1.1.3 and 1.1.4 are equivalent and $\forall u \in BV(\Omega) \quad V(u,\Omega) = |Du|(\Omega)$.*

Proof. ($\Longleftarrow$): Take u as in 1.1.4 and let $g \in C_c^1(\Omega)$ with $\|g\|_\infty \leq 1$, then:

$$\int_\Omega u \cdot \mathrm{div}(g) \,\mathrm{d}x = \sum_{j=1}^N -\int_\Omega g_j \,\mathrm{d}D_j u.$$

Call f the function like in B.0.4 such that $Du = f|Du|$.

$$\implies \int_\Omega u \cdot \mathrm{div}(g) \,\mathrm{d}x = \int_\Omega -\sum_j g_j f_j \,\mathrm{d}|Du| \leq \int_\Omega |g| \cdot |f| \,\mathrm{d}|Du|$$

$$\leq \int_\Omega \mathrm{d}|Du| = |Du|(\Omega).$$

Extracting the supremum on g we get: $V(u,\Omega) \leq |Du|(\Omega) < +\infty$.

($\implies$): Take u as in 1.1.3, then for every $\phi \in [C_c^\infty(\Omega)]^N$:

$$\left|\int_\Omega u \,\mathrm{div}\phi \,\mathrm{d}x\right| \leq \|\phi\|_\infty V(u,\Omega).$$

Therefore, we can define on $[C_c^\infty(\Omega)]^N$ a linear and continuous functional as:

$$L(\phi) = \int_\Omega u \,\mathrm{div}\phi \,\mathrm{d}x.$$

As C_c^∞ is dense in C_0, we can uniquely extend L to $[C_0(\Omega)]^N$, preserving $\|L\| \leq V(u,\Omega)$. Hence, by Riesz representation theorem (see [5]), exists a unique radon measure μ s.t.:

$$L(\phi) = \sum_{j=1}^{N} \int_\Omega \phi_j \,\mathrm{d}\mu_j, \quad \forall \phi \in [C_0(\Omega)]^N,$$

with $|\mu|(\Omega) = \|L\| \leq V(u,\Omega) < +\infty$. But, since for every $\phi \in [C_c^\infty(\Omega)]^N$ this measure must verify:

$$\sum_{j=1}^{N} \int_\Omega \phi_j \,\mathrm{d}\mu_j = \int_\Omega u \,\mathrm{div}\phi \,\mathrm{d}x = -\sum_{j=1}^{N} \langle D_j u, \phi_j \rangle,$$

we have $Du = -\mu$, which implies $|Du|(\Omega) = |\mu|(\Omega) \leq V(u,\Omega)$.

Therefore, the two definitions imply each other and $|Du|(\Omega) = V(u,\Omega)$.

□

Now we will present some further insight on the relation between a measure μ and its total variation $|\mu|$, in particular in relation with the weak $\star$ convergence.

Definition B.0.7. We say that a sequence of Radon measures $\{\mu_h\}$ on X converge weakly $\star$ to a Radon measure μ, denoted $\mu_h \xrightharpoonup{*} \mu$, if $\forall g \in C_0(X)$:

$$\lim_{h\to\infty} \int_X g \,\mathrm{d}\mu_h = \int_X g \,\mathrm{d}\mu.$$

The following is an important results of the total variation compactness with respect to the weak $\star$ convergence.

Theorem B.0.8 (De La Valée Poussin)**.** *Take $\{\mu_h\}$ a sequence of finite Radon measure on a metric space X, then if $\sup_h |\mu_h|(X) < +\infty$ there exists a subsequence μ_{h_k} converging weakly $\star$.*

Proof. $[C_0(\mathbb{R}^n)]^m$ is separable, that is there exists a countable set $\mathcal{B}$ which is dense (in uniform norm) in $[C_0(X)]^m$. Then, given any $g_1 \in \mathcal{B}$, we have, from B.0.5 that the sequence $\{\langle \mu_h, g_1 \rangle\}$ is bounded in h and consequently it admits a subsequence $\{\langle \mu_h^1, g_1 \rangle\}$ converging to some value a_1 so that

$$\left| \langle \mu_h^1, g_1 \rangle - a_1 \right| < \frac{1}{h}.$$

Now given another element $g_2 \in \mathcal{B} \setminus \{g_1\}$, since $\mu_h^1(g_2)$ is bounded, we can extract a subsequence $\{\langle \mu_h^2, g_2 \rangle\}$ converging to a_2 and

$$\left| \langle \mu_h^2, g_2 \rangle - a_2 \right| < \frac{1}{h}.$$

Iterating we find nested sequences $\{\mu_h^k\}$ and values a_k such that

$$\left| \langle \mu_h^k, g_k \rangle - a_k \right| < \frac{1}{h},$$

and $\{g_k\}_{k \in \mathbb{N}} = \mathcal{B}$. Proceeding with a diagonal argument, we take μ_h^h, which satisfy:

$$\left| \langle \mu_h^h, g_k \rangle - a_k \right| < \frac{1}{h}, \qquad \forall h > k.$$

$$\implies \lim_{h \to 0} \langle \mu_h^h, g_k \rangle = a_k.$$

For Banach-Steinhaus, we can define on $\mathcal{B}$ a measure μ such that $\langle \mu, g_k \rangle = a_k$ for all k, then we can extend it by density. Now, for every $g \in [C_0(X)]^m$ consider $g^\varepsilon \in \mathcal{B}$ such that $\|g - g^\varepsilon\|_\infty < \varepsilon$, then, from the boundedness of $|\mu_h|(X)$ there exists a C independent from h such that

$$\left| \langle \mu - \mu_h^h, g \rangle \right| \leq |\langle \mu, g - g^\varepsilon \rangle| + \left| \langle \mu - \mu_h^h, g^\varepsilon \rangle \right| + \left| \langle \mu_h^h, g - g^\varepsilon \rangle \right| \leq$$

$$\leq \left| \langle \mu - \mu_h^h, g^\varepsilon \rangle \right| + C\|g - g^\varepsilon\|_\infty \longrightarrow C\|g - g^\varepsilon\|_\infty < C\varepsilon.$$

Thus, by arbitrariness of $\varepsilon > 0$ we conclude. □

We can generalize to non finite measures

Corollary B.0.9. *Take $\{\mu_h\}$ a sequence Radon measure on a metric space X, then if $\sup_h |\mu_h|(K) < +\infty$ for every compact K, there exists a subsequence μ_{h_k} locally converging weakly $\star$.*

We can also say that the total variation is lower semicontinuous with respect to weak $\star$ convergence.

Proposition B.0.10 (Semicontinuity)**.** *Let $\{\mu_h\}$ be a sequence converging locally weakly $\star$ to μ, then*

$$|\mu_h| \overset{*}{\rightharpoonup} \lambda \implies \lambda \geq |\mu|,$$

Proof. Because $|\mu|$ is the norm of μ as functional on $[C_0]^m$, the semicontinuity property is given. □

In conclusion we present a theorem, useful to prove some results on the reduced boundary. At first we need the definition of a Lebesgue point.

Definition B.0.11. We say that x is a Lebesgue point of a function f integrable on a positive measure μ if

$$\lim_{\rho\to 0}\frac{1}{\mu(B(x,\rho))}\int_{B(x,\rho)}|f(x)-f(y)|\,\mathrm{d}\mu = 0.$$

Observation B.0.12. We notice that if μ is a vector valued Radon measure with polar decomposition $\mu = f|\mu|$, then every x on which f is defined is a Lebesgue point. Indeed, from the Jensen inequality

$$\left(\frac{1}{|\mu|(B(x,\rho))}\int_{B(x,\rho)}|f(x)-f(y)|\,\mathrm{d}|\mu|(y)\right)^2 \leq$$
$$\leq \frac{1}{|\mu|(B(x,\rho))}\int_{B(x,\rho)}|f(x)-f(y)|^2\,\mathrm{d}|\mu|(y) =$$
$$= 2\left(1-\left\langle f(x), \frac{\mu(B(x,\rho))}{|\mu|(B(x,\rho))}\right\rangle\right).$$

However, from A.0.11 and the fact that $|f(x)| = 1$, this must converge to 0.

Theorem B.0.13. *Take μ a Radon measure with polar decomposition $\mu = f|\mu|$ and let x be a Lebesgue point. We define the measure $\mu_{x,\rho}$ as*

$$\mu_{x,\rho}(A) = \mu(x+\rho A).$$

Then, given a sequence $\{\rho_i\}$ converging to 0,

$$\exists\,\nu = \lim_i \frac{\mu_{x,\rho_i}}{|\mu|(B(x,\rho_i))} \iff \exists\,\sigma = \lim_i \frac{|\mu|_{x,\rho_i}}{|\mu|(B(x,\rho_i))}$$

where the limits are intended as weakly $\star$. Furthermore, we have $\sigma = |\nu|$ and the polar decomposition $\nu = f|\nu|$.

Proof. Let $B = B(0{,}1)$ and take $\phi \in C_0(B)$.

$$\int_B \phi(z)\,\mathrm{d}|\mu|_{x,\rho} - \left\langle f(x), \int_B \phi(z)\,\mathrm{d}\mu_{x,\rho}\right\rangle = \int_B \phi(z)(1-\langle f(x), f(x+\rho z)\rangle)\,\mathrm{d}|\mu|_{x,\rho} =$$
$$= \int_B \phi\left(\frac{y-x}{\rho}\right)(1-\langle f(x), f(y)\rangle)\,\mathrm{d}|\mu| = o(|\mu|(B(x,\rho))),$$

where the last step depends on the fact that ϕ is bounded and $|1-\langle f(x), f(y)\rangle| \leq |f(x)-f(y)|$, then as x is a Lebesgue point we have the convergence behaviour as above. To summarize we concluded:

$$\frac{|\mu|_{x,\rho_i}}{|\mu|(B(x,\rho_i))} \overset{\star}{\rightharpoonup} \langle f, \nu\rangle = \sigma.$$

Then, if we define g the function for the polar decomposition of ν, we have

$$\sigma = \langle f, g\rangle |\nu|$$

and then $|\sigma| \leq |\nu|$. But from B.0.10 we must also have $|\nu| \leq |\sigma|$. This means $|\nu| = |\sigma| = |\langle f, g\rangle||\nu|$, that is $\langle f, g\rangle = \pm 1$ almost everywhere. But since σ is a positive measure, we have:

$$\langle f, g\rangle = 1 \implies f = g,$$

and therefore also $\sigma = |\nu|$.

The reverse implication instead is a direct consequence of the De La Vallée Poussin Theorem:

$$\frac{|\mu|_{x,\rho_i}}{|\mu|(B(x,\rho_i))} \overset{\star}{\rightharpoonup} \sigma \implies \frac{\mu_{x,\rho_i}}{|\mu|(B(x,\rho_i))} \overset{\star}{\rightharpoonup} \nu.$$

Then the relation between ν and σ is satisfied as before.

□

Observation B.0.14. We conclude observing that in the proof the denominator $|\mu|(B(x,\rho))$ is used only on to balance the limit behaviour $o(|\mu|(B(x,\rho))$. Therefore, if we know that there is a function $g(\rho)$ with the same limit behaviour as $|\mu|(B(x,\rho))$, we can substitute it in the denominator. For example we can state:

Take μ a Radon measure with polar decomposition $\mu = f|\mu|$ and let x be a Lebesgue point. Suppose we know $|\mu|(B(x,\rho)) \leq \alpha\rho^{n-1}$, for some real positive α. Then, given a sequence $\{\rho_i\}$ converging to 0,

$$\exists\, \nu = \lim_i \frac{\mu_{x,\rho_i}}{\alpha\rho_i^{n-1}} \iff \exists\, \sigma = \lim_i \frac{|\mu|_{x,\rho_i}}{\alpha\rho_i^{n-1}};$$

where the limits are intended as weakly $\star$. Furthermore, we have the polar decomposition $\nu = f\sigma$.

Acronyms

ROF

Rudin-Osher-Fatemi problem

ROFs

Rudin-Osher-Fatemi problem by level sets

ROFh

discretized Rudin-Osher-Fatemi problem

Bibliography

[1] Enrico Giusti. *Minimal Surfaces and Functions of Bounded Variation.* Birkhäuser, Boston, MA, 1984 (cit. on pp. 28, 52).

[2] O. Dovgoshey, O. Martio, V. Ryazanov, and M. Vuorinen. «The Cantor function». In: *Expositiones Mathematicae* 24.1 (2006), pp. 1–37. ISSN: 0723-0869. DOI: `https://doi.org/10.1016/j.exmath.2005.05.002`. URL: `https://www.sciencedirect.com/science/article/pii/S0723086905000502` (cit. on p. 32).

[3] Gianni Maso. *An Introduction to* Γ*-Convergence.* Birkhäuser Boston, MA, 1993 (cit. on p. 32).

[4] David Mumford and Jayant Shah. «Optimal approximations by piecewise smooth functions and associated variational problems». In: *Communications on Pure and Applied Mathematics* 42 (1989), pp. 577–685 (cit. on p. 40).

[5] Luigi Ambrosio, Diego Pallara, and Nicola Fusco. «Functions of Bounded Variation and Free Discontinuity Problems». In: 2000 (cit. on pp. 41, 93, 97).

[6] Luigi Ambrosio and Vincenzo Maria Tortorelli. «Approximation of functional depending on jumps by elliptic functional via Γ-convergence». In: *Communications on Pure and Applied Mathematics* 43.8 (1990), pp. 999–1036. DOI: `https://doi.org/10.1002/cpa.3160430805`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.3160430805`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160430805` (cit. on p. 41).

[7] Ivar Ekeland and Roger Témam. *Convex Analysis and Variational Problems.* Society for Industrial and Applied Mathematics, 1999. DOI: `10.1137/1.9781611971088`. eprint: `https://epubs.siam.org/doi/pdf/10.1137/1.9781611971088`. URL: `https://epubs.siam.org/doi/abs/10.1137/1.9781611971088` (cit. on p. 43).

[8] Matteo Novaga Vicent Caselles Antonin Chambolle. «Regularity for solutions of the total variation denoising problem». In: *Rev. Mat. Iberoam* 27.1 (2011), pp. 233–252. DOI: `https://doi.org/10.4171/RMI/634` (cit. on p. 48).

[9] François Alter, Vincent Caselles, and A. Chambolle. «Evolution of characteristic functions of convex sets in the plane by the minimizing total variation flow». In: *Interfaces and Free Boundaries* 7 (2005), pp. 29–53 (cit. on p. 52).

[10] A.Chambolle F.Alter V.Caselles. «A characterization of convex calibrable sets in $\mathbb{R}^N$». In: *Mathematische Annalen* 332.2 (2005), pp. 329–366. DOI: 10.1007/s00208-004-0628-9 (cit. on p. 52).

[11] Francesco Maggi. *Sets of Finite Perimeter and Geometric Variational Problems: An Introduction to Geometric Measure Theory*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2012. DOI: 10.1017/CBO9781139108133 (cit. on p. 55).

[12] Ernie Esser. «Applications of Lagrangian-Based Alternating Direction Methods and Connections to Split Bregman». In: *CAM Rep* 9 (Jan. 2009) (cit. on p. 61).

[13] Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, and Thomas Pock. «An Introduction to Total Variation for Image Analysis». In: *Theoretical Foundations and Numerical Methods for Sparse Recovery*. Ed. by Massimo Fornasier. Berlin, New York: De Gruyter, 2010, pp. 263–340. ISBN: 9783110226157. DOI: doi:10.1515/9783110226157.263. URL: https://doi.org/10.1515/9783110226157.263 (cit. on pp. 63, 70, 76).

[14] Antonin Chambolle, Alessandro Giacomini, and Luca Lussardi. *Continuous limits of discrete perimeters*. 2009. DOI: 10.48550/ARXIV.0902.2313. URL: https://arxiv.org/abs/0902.2313 (cit. on p. 66).

[15] Ming-Jun Lai, Bradley Lucier, and Jingyue Wang. «The Convergence of a Central-Difference Discretization of Rudin-Osher-Fatemi Model for Image Denoising». In: *Scale Space and Variational Methods in Computer Vision*. Ed. by Xue-Cheng Tai, Knut Mørken, Marius Lysaker, and Knut-Andreas Lie. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 514–526. ISBN: 978-3-642-02256-2 (cit. on p. 70).

[16] JINGYUE WANG and BRADLEY J. LUCIER. «ERROR BOUNDS FOR FINITE-DIFFERENCE METHODS FOR RUDIN—OSHER—FATEMI IMAGE SMOOTHING». In: *SIAM Journal on Numerical Analysis* 49.1/2 (2011), pp. 845–868. ISSN: 00361429. URL: http://www.jstor.org/stable/23074424 (visited on 07/29/2023) (cit. on p. 70).

[17] A.M. Duguid. «Studies in linear and non-linear programming, by K. J. Arrow, L. Hurwicz and H. Uzawa. Stanford University Press, 1958. 229 pages. 7.50.» In: *Canadian Mathematical Bulletin* 3.3 (1960), pp. 196–198. DOI: 10.1017/S0008439500025522 (cit. on p. 70).

[18] Ernie Esser, Xiaoqun Zhang, and Tony F. Chan. «A General Framework for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science». In: *SIAM Journal on Imaging Sciences* 3.4 (2010), pp. 1015–1046. DOI: `10.1137/09076934X`. eprint: `https://doi.org/10.1137/09076934X`. URL: `https://doi.org/10.1137/09076934X` (cit. on p. 72).

[19] Thomas Pock, Daniel Cremers, Horst Bischof, and Antonin Chambolle. «An algorithm for minimizing the Mumford-Shah functional». In: *2009 IEEE 12th International Conference on Computer Vision*. 2009, pp. 1133–1140. DOI: `10.1109/ICCV.2009.5459348` (cit. on p. 72).