ICTP - The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy

ICTP Lecture Notes Series
Volume II (ISBN 92-95003-04-7) - *December 2000*

Mathematical Problems in Image Processing

All (La)TeX & PS sources Volume II: lns002 .tar.gz (2257360 bytes)
(NB: This tar-compressed file contains, LaTeX, postscript (.ps) and/or encapsulated postscript (.eps) figures and also the postscript version of each lecture note).

Mathematical Problems in Image Processing

Editor: C.E. Chidume (ICTP)

Introduction: [*Sources*: PS , PDF]

Contents

For queries or suggestions please contact via E-mail: pub_off@ictp.trieste.it

ICTP Lecture Notes Series

Introduction

This is the second volume of a new series of lecture notes of the Abdus Salam International Centre for Theoretical Physics. These new lecture notes are put onto the web pages of the ICTP to allow people from all over the world to access them freely. In addition a limited number of hard copies is printed to be distributed to scientists and institutions which otherwise do not have access to the web pages.

This volume contains the lecture notes given by A. Chambolle during the School on Mathematical Problems in Image Processing that took place at the Abdus Salam International Centre for Theoretical Physics from 4 to 22 September 2000 under the direction of L. Ambrosio (Scuola Normale Superiore di Pisa), G. Dal Maso (Scuola Internazionale Superiore di Studi Avanzati) and J.-M. Morel (Ecole Normale Supérieure, Cachan).

The topic of Chambolle's course was "Inverse problems in image processing and image segmentation: some mathematical and numerical aspects".

The School consisted of two weeks of lecture courses and one week of conference. It was financially supported by the Abdus Salam International Centre for Theoretical Physics, SISSA (Scuola Internazionale Superiore di Studi Avanzati, Trieste) and Scuola Normale Superiore di Pisa. I take this opportunity to express our gratitude to all the lecturers and speakers at the conference for their contribution towards the success of the School.

Charles E. Chidume
November, 2000

Inverse problems in Image processing and Image segmentation: some mathematical and numerical aspects

A. Chambolle*

CEREMADE (CNRS, UMR 7534),
Université de Paris-Dauphine, 75775 Paris cedex 16, France

Lecture given at the
School on Mathematical Problems in Image Processing
Trieste, 4 – 22 September 2000

LNS002001

* antonin.chambolle@ceremade.dauphine.fr

Abstract

These notes contain an introduction to some approaches to the regularization of inverse problems in image processing and to the mathematical tools that are necessary to handle correctly these approaches. The methods we consider here are variational methods. We consider mainly the minimization of two kinds of functionals: functionals based on the total variation of the image, and the so–called Mumford and Shah functional that penalizes the edge set and the gradient of the image. In both cases we study mathematically the existence of a solution in the space of functions with bounded variation (BV), and discuss then some approximations and numerical methods for computing solutions.

Keywords: Image processing, inverse problems, image segmentation, functions with bounded variation, Γ–convergence, iterative algorithms.

AMS Classification numbers: 26A45, 49J45, 49Q20, 68U10

Contents

Main notations

- $a \vee b$, $a \wedge b$: respectively, the max and the min of the two real numbers $a, b \in \mathbb{R}$.
- $\mathcal{H}^k$: the k–th dimensional Hausdorff measure. In particular, for every set $E \subseteq \mathbb{R}^N$, $\mathcal{H}^0(E)$ is the cardinality of E, also denoted by $\sharp E$.
- $\chi_E(x)$: the characteristic function of a set E, i.e., $\chi_E(x) = 1$ if $x \in E$ and $\chi_E(x) = 0$ otherwise.
- $|E| = \mathcal{L}^N(E) = \int_{\mathbb{R}^N} \chi_E(x)\, dx$: the Lebesgue measure of $E \subseteq \mathbb{R}^N$.
- $C_c(\Omega)$, $C_c^1(\Omega)$, $C_c^\infty(\Omega)$: the space of compactly supported continuous (respectively, continuously differentiable, infinitely differentiable) real–valued functions on the domain $\Omega \subseteq \mathbb{R}^N$. $C_c^\infty(\Omega)$ is also denoted by $\mathcal{D}(\Omega)$ when it is equipped with the appropriate topology in order to define the distributions by duality (see [55]). $C_c(\Omega, \mathbb{R}^N) = [C_c(\Omega)]^N$, etc.
- $C_0(\Omega)$: the space of the real–valued functions that are continuous on Ω and vanish at the boundary and/or at infinity, in the sense that if $\varphi \in C_0(\Omega)$, $\forall \varepsilon > 0$, $\exists K \subset \Omega$ compact set such that $\sup_{\Omega \setminus K} |\varphi| \leq \varepsilon$. The norm on $C_0(\Omega)$ is $\|\varphi\| = \sup_\Omega |\varphi|$. With this norm, $C_0(\Omega) = \overline{C_c(\Omega)}$. Similarly, $C_0(\Omega, \mathbb{R}^N) = [C_0(\Omega)]^N$.
- $\mathcal{M}(\Omega)$: the space of bounded Radon measures on Ω. It is isomorphic (and isometric) to the topological dual $C_0(\Omega)'$ of $C_0(\Omega)$. $\mathcal{M}(\Omega, \mathbb{R}^N) = [\mathcal{M}(\Omega)]^N = C_0(\Omega, \mathbb{R}^N)'$.
- $\langle x', x \rangle$: the Euclidean scalar product of $x, x' \in \mathbb{R}^N$, or the duality product between an element $x \in X$ of a space X and an element $x' \in X'$ of its dual (also sometimes denoted by $\langle x', x \rangle_{X',X}$). The Euclidean norm in $\mathbb{R}^N$ is usually denoted by $|\xi| = \sqrt{\langle \xi, \xi \rangle}$.
- (a, b): the set $\{t \in \mathbb{R} : a < t < b\}$. $[a, b] = \{t \in \mathbb{R} : a \leq t \leq b\}$, $(a, b] = \{t \in \mathbb{R} : a < t \leq b\}$, etc.
- $\mathbb{S}^{N-1} = \{\xi \in \mathbb{R}^N : |\xi| = 1\}$ is the $(N-1)$–dimensional sphere in $\mathbb{R}^N$.

1 Introduction: denoising and deblurring images

One fundamental branch of the image processing concerns the problem of *reconstructing* images, i.e., given some data (that may be a corrupted image but also any kind of signal, like the output of a tomography device or of a satellite aerial), how to reconstruct a clear and clean image that can be correctly understood by a human operator or post-processed by other image analysis methods.

The most basic examples of image reconstruction problems are the problems of denoising and of deblurring an image. Although they are the simplest, they share many common features with more complicated problems that are usually too specific for the purpose of short lectures. All these problems belong to the class usually known as *inverse problems*. It means that the process through which the data is obtained from the physical characteristics of the observed scene corresponds to transformations that are roughly well understood and can be more or less correctly modelized mathematically, but whose inverse either is not known or is not computable by direct methods, or whose computation is highly instable and sensitive to small changes in the data (or noise), so that the scene itself is difficult to reconstruct.

First we will describe the main classical approach to denoising and deblurring (more or less the "standard" method for solving inverse problems) and will try to explain why it is not well suited to the nature and structure of images. Then, we will introduce solutions that have been proposed in the past years to improve this approach.

1.1 The "classical approach"

Assume you observe a signal (an image) which is a matrix $G = (g_{i,j})_{1 \le i,j \le n}$ of grey level values in $[0, 1]$, and suppose you know that this signal is the sum of a "perfect world" unknown signal $U = (u_{i,j})_{1 \le i,j \le n}$ and an additive Gaussian noise $N = (n_{i,j})_{1 \le i,j \le n}$, where, for instance, all $n_{i,j}$ are independent and have mean 0 and known variance σ^2.

In a different point of view, in the continuous setting, you can assume that the signal you observe is a bounded grey-level function

$$g : \Omega \to [0, 1]$$

where Ω is "the screen", usually an open domain of $\mathbb{R}^2$ (although lower or higher dimensions may be considered), and most of the time, in the applications, a rectangle, e.g., $(0, 1) \times (0, 1)$. This function $g(x)$ will be assumed to be the sum $u(x) + n(x)$ of a "good image" $u(x)$ and an oscillation $n(x)$ that we would like to remove. We will assume that $\int_\Omega n(x)\,dx = 0$ and that $\int_\Omega n(x)^2\,dx = \sigma^2$ is known or can be correctly estimated.

The first point of view (the discrete setting) describes well the structure of digital images, and is usually adopted in the statistical approaches to image reconstruction. We will return to this setting in section 1.3 devoted to the image segmentation problem, since the origins of the approach that we will discuss in these notes are to be found in the statistical approach to image denoising. However, in the PDE or variational approach that we will usually adopt here, it is more common and more convenient to work in the continuous setting, and except otherwise mentioned we will consider this point of view in the sequel.

Up to now we have just considered an image corrupted by some noise, but usually an image also goes through all kinds of degradations, that are usually modelized by a blur of more or less known kernel. It means that instead of $g(x) = u(x) + n(x)$, the correct model should be $g(x) = Au(x) + n(x)$ where A is a linear operator, say, from $L^2(\Omega)$ into $L^2(\Omega)$ (or any kind of reasonable function space). Usually

$$Au(x) \;=\; \rho * u(x) \;=\; \int \rho(x-y)u(x)\,dy$$

is simply a blur (a convolution), with ρ some (usually non negative) kernel that is known or estimated, but one may imagine more complex operators (like tomography kernels, or all sorts of transformation).

Then, the problem we need to solve is the following: given g and an estimation of A and σ^2, is it possible to get a good approximation of u?

The first idea would be to compute $A^{-1}g = u + A^{-1}n$, however this is not feasible in practice: the operator A is often not invertible, or its inverse is impossible to compute. Consider for instance the case where $Au = \rho * u$. In the Fourier domain, we find that $\widehat{Au} = \hat{\rho}\hat{u}$ where $\hat{\cdot}$ denotes the Fourier transform. So that $u = A^{-1}v$ if and only if $\hat{u} = \hat{v}/\hat{\rho}$. But, even if $\hat{\rho}$ does not vanish, this ratio is usually not in L^2 for an arbitrary $v \in L^2$. Moreover, if ρ is a smooth low pass filter, then $\hat{\rho}(\xi)$ is very small for large frequencies $|\xi|$, so that in the case where v is the oscillatory signal n, for which $|\hat{n}(\xi)|$ remains strictly greater than zero for large $|\xi|$, the ratio $\hat{n}(\xi)/\hat{\rho}(\xi)$ will become very large and go to $+\infty$ as $|\xi|$ increases. This enhancement of the high frequencies gives birth to wild oscillations and artifacts that make the image $u + A^{-1}n$ impossible to read.

A better approach to this kind of problem, therefore, is the following: we will try to find the "best" function u among all u satisfying

$$\begin{cases} \displaystyle\int_\Omega Au(x) - g(x)\,dx \;=\; 0 \\ \displaystyle\int_\Omega |Au(x) - g(x)|^2\,dx \;=\; \sigma^2. \end{cases} \tag{1}$$

So that the main issue, now, is to find a good criterion for characterizing what the "best" function u is.

The classical approach of Tichonov consists in minimizing some quadratic norm of u, like $\int_\Omega |u|^2$ or $\int_\Omega |\nabla u|^2$ under the constraints (1). Both problems can easily be solved (using the Fourier transform) and the linear transformation of g that gives the solution u is called a Wiener filter.

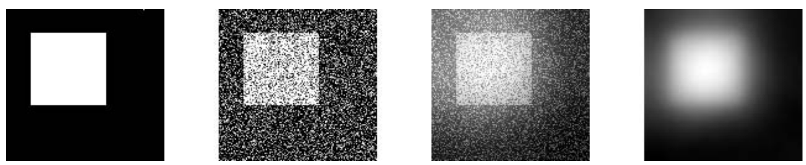

Figure 1: From left to right: a white square on a black background; the same image with noise added; the Tichonov reconstruction by minimizing $\int |u|^2$; the minimization of $\int |\nabla u|^2$.

However, while the first criterion is not regularizing enough and produces images that still look very noisy, the criterion $\int_\Omega |\nabla u|^2$ is not well suited either for the analysis of images (see Fig. 1). Indeed, if it is finite, it means that the image belongs to the Sobolev space $H^1(\Omega)$, and it is well known that a function in that space may not have discontinuities along a hypersurface, whereas the grey level of an image should be allowed to have such discontinuities that correspond to edges and boundaries of objects in the image. For instance, in dimension 1, it is well known that if $u \in H^1(I)$ (I being some interval of $\mathbb{R}$), then for every $x, y \in I$ with $x \le y$,

$$u(y) - u(x) \;=\; \int_x^y u'(s)\,ds \;\le\; \sqrt{y-x}\sqrt{\int_x^y |u'(s)|^2\,ds}$$

so that $u \in C^{0,\frac{1}{2}}(I)$ (the space of continuous $\frac{1}{2}$–Hölder functions in I) and may not have discontinuities.

This motivates the introduction of the criterion that we discuss in the next section.

1.2 The total variation criterion

In their paper [54], Rudin, Osher and Fatemi describe a different approach (see also [46, 53, 60, 61, 24, 25, 44, 33, 47]). Their idea is to try to find a criterion of minimization that corresponds better to the structure of the images. They propose to consider the "total variation" of the function u as a measure of the optimality of an image.

The total variation (that will be introduced correctly in section 2.1) is roughly the integral $\int_\Omega |\nabla u(x)|\,dx$. The main advantage is that it can be defined for functions that have discontinuities along hypersurfaces (in 2–dimensional images, along 1–dimensional curves), and this is essential to get a correct representation of the edges in an image.

The problem to solve is thus the following:

$$\min\left\{\int_\Omega |\nabla u(x)|\,dx : u \text{ satisfies (1)}\right\}. \tag{2}$$

We will show in section 2.1 that under some simple and natural assumptions, this problem has a solution. Then, we will propose a numerical approach for computing a solution.

1.3 The segmentation of images

The last approach that we will discuss in these notes can be seen as an independent problem, although historically it has the same origin. It is called the problem of image segmentation, and can be described as the problem of finding a simple representation of a given image in terms of edges and smooth areas. The proposition of D. Mumford and J. Shah [49, 50]) to solve this problem by minimizing a functional is indeed derived from statistical approaches to image denoising, introduced in particular by S. and D. Geman, that we will describe in the next section. Again, the problem of Geman and Geman was to regularize correctly an inverse problem (the problem that we have described in the previous paragraphs, written in the discrete setting), and to restore correctly the edges of the image. Thus we will briefly describe the point of view of Geman and Geman, and explain then how Mumford and Shah derived their continuous formulation.

1.3.1 A statistical approach to image denoising

The origins of the variational approaches to image segmentation are to be found in Geman and Geman's famous paper [41] in which they introduce a statistical approach for image analysis that has proved to be very efficient. First we will briefly explain how it appeared in the probabilistic setting.

We return to the discrete setting of the image denoising problem: the observed signal (or image) is a matrix $G = (g_{i,j})_{1\leq i,j\leq n}$ of grey level values in $[0,1]$, and is the combination of a "perfect world" unknown signal $U = (u_{i,j})_{1\leq i,j\leq n}$ and an additive Gaussian noise $N = (n_{i,j})_{1\leq i,j\leq n}$. The $n_{i,j}$ are independent and have mean 0 and variance σ^2. If you know the *a priori* probability $P(U)$ of the perfect world signal U, since for a given $G = U + N$, the probability of G knowing U is $P(G|U) = P(N = G - U) \sim \exp(-\|G - U\|^2/2\sigma^2)$, the Bayes' rule tells you that

$$P(U|G)P(G) = P(G|U)P(U),$$

so that $P(U|G)$, up to a constant, is $P(G|U)P(U)$, that is:

$$\frac{1}{(\sqrt{2\pi}\sigma)^{n\times n}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i,j}(g_{i,j} - u_{i,j})^2\right) P(U).$$

Geman and Geman proposed the following *a priori* probability for U: they considered that most scenes are piecewise smooth with possible discontinuities (the edges), and introduce an edge set (or *line process*) $L = (l_{i+\frac{1}{2},j})_{1\leq i<n, 1\leq j\leq n}, (l_{i,j+\frac{1}{2}})_{1\leq i\leq n, 1\leq j<n}$, where each variable $l_{\alpha,\beta}$ is either 0 or 1, and (see Fig. 2):

$$l_{i+\frac{1}{2},j} = \begin{cases} 1 & \text{if there is a break (a vertical piece of edge) between } i,j \\ & \hfill \text{and } i+1,j \\ 0 & \text{if } U \text{ has to be smooth between } i,j \text{ and } i+1,j, \end{cases}$$

$$l_{i,j+\frac{1}{2}} = \begin{cases} 1 & \text{if there is a break (a horizontal piece of edge) between } i,j \\ & \hfill \text{and } i,j+1 \\ 0 & \text{if } U \text{ has to be smooth between } i,j \text{ and } i,j+1. \end{cases}$$

They then proposed the following probability law for U, L:

$$P(U,L) = \frac{1}{Z} \exp \left\{ -\sum_{i,j} \Big(\lambda(1 - l_{i+\frac{1}{2},j})(u_{i+1,j} - u_{i,j})^2 + \mu l_{i+\frac{1}{2},j} + \lambda(1 - l_{i,j+\frac{1}{2}})(u_{i,j+1} - u_{i,j})^2 + \mu l_{i,j+\frac{1}{2}}\Big)\right\}.$$

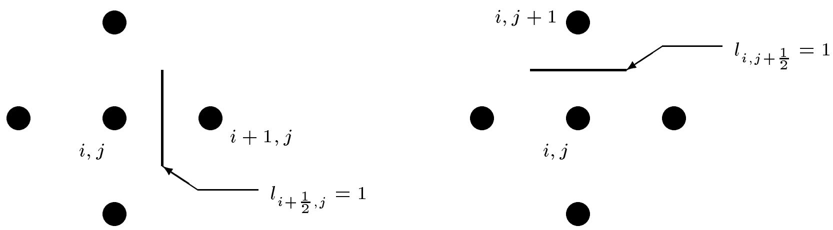

Figure 2: The *line process*: $l_{i+\frac{1}{2},j}$ and $l_{i,j+\frac{1}{2}}$.

where λ, μ are two positive weights, and Z is computed in order to have $\sum_{U,L} P(U,L) = 1$, the sum being computed over all the possible states U, L.

The problem that needs to be solved is therefore the following:
"Among all possible images U and line processes L, find the one that has the greatest probability

$$P(U,L|G) \sim e^{-E(U,L,G)}, \tag{3}$$

where the *free energy* $E(U,L,G)$ is given by

$$\begin{aligned} E(U,L,G) &= \sum_{i,j} \lambda\left((1-l_{i+\frac{1}{2},j})(u_{i+1,j}-u_{i,j})^2 + (1-l_{i,j+\frac{1}{2}})(u_{i,j+1}-u_{i,j})^2\right) \\ &\quad + \mu\left(l_{i+\frac{1}{2},j} + l_{i,j+\frac{1}{2}}\right) \\ &\quad + \frac{1}{2\sigma^2}(g_{i,j}-u_{i,j})^2 \end{aligned} \tag{4}$$

and $G = (g_{i,j})_{1\le i,j\le n}$ is the given data"

In what follows, since the observed data G will be fixed, we will drop the dependency in G in the notations and merely write $E(U,L)$.

Then, Geman and Geman proposed to maximize the probability (3) using a *simulated annealing* algorithm (see for instance [11, 12], [15], [32], the book [51] for more general segmentation models, and the book on Markov Random Field Modeling in Computer Vision by Li [45] for a general introduction to the field). This kind of method is still widely used in the computer vision community and gives good result. It has to be adapted to each particular segmentation problem (in which the problem we exposed is among the simplest, but might not be the most interesting!).

We will present other approaches, since in some simple cases it might be too costly to implement a simulated annealing algorithm. Notice that the problem of maximizing (3) is equivalent to the problem of finding a minimum to the free energy $E(U, L)$ that appears in the exponential in (3). The problem is that this energy is not convex, so that there is no known deterministic (i.e., non-probabilistic) algorithm that can be proved to surely converge to the minimum. The history of the minimization of $E(U, L)$ is therefore mostly a competition for finding a "better" algorithm, in any possible sense.

In the 80's, already, many have suggested deterministic methods to minimize directly energy (4). See for instance [38, 39], or [40], and more recently [10], but of course this list is far from being exhaustive.

In most of these papers, the problem is iteratively approximated by a sequence of simpler problems, each one becoming "less convex" as the process evolves. This is the central idea of the book Visual Reconstruction by Blake and Zisserman [14], who introduce the so–called "Graduated Non-Convexity" (GNC) algorithm.

They first noticed that, minimizing with respect to L, the energy E in (4) can be rewritten as

$$E(U) \;=\; \sum_{i,j} W_{\lambda,\mu}\left(u_{i+1,j} - u_{i,j}\right) + W_{\lambda,\mu}\left(u_{i,j+1} - u_{i,j}\right) \;+\; \frac{1}{2\sigma^2}(g_{i,j} - u_{i,j})^2 \tag{5}$$

where the non-convex potential $W_{\lambda,\mu}$ is (see Figure 3)

$$W_{\lambda,\mu}(x) \;=\; \min(\lambda x^2, \mu).$$

(We will also denote $\min(\lambda x^2, \mu)$ by $(\lambda x^2) \wedge \mu$.) Blake and Zisserman call

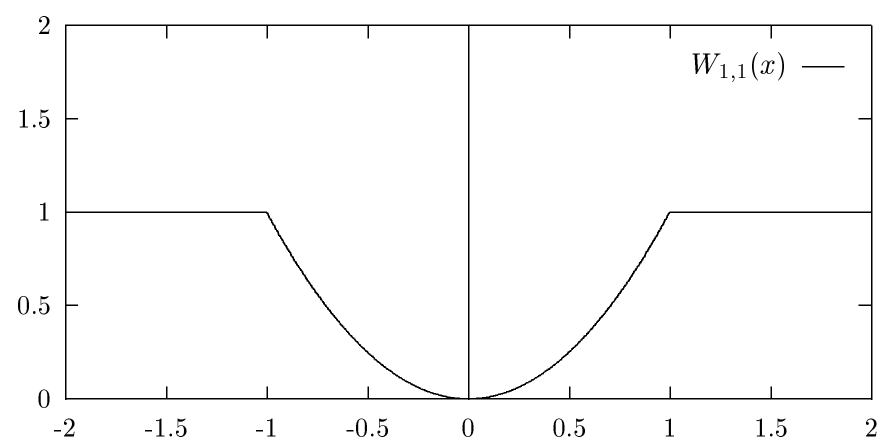

Figure 3: The function $W_{\lambda,\mu}(x)$ for $\lambda = \mu = 1$.

$E(U)$ the "weak membrane" energy, since it looks like the potential of an elastic membrane that can break when the elastic energy becomes locally too high. Notice now that their problem is very similar to the inverse problems that we have presented in the previous sections. Now, instead of minimizing a regularizing factor (that is quite more complex than Tichonov's) under a constraint like (1) (with $A = Id$), the energy has a term $\frac{1}{2\sigma^2}(g_{i,j} - u_{i,j})^2$ that could be seen as a Lagrange multiplier for the constraint $\int_\Omega |u(x)-g(x)|^2\, dx = \sigma^2$.

Their idea to minimize $E(U)$ is to replace $W_{\lambda,\mu}$ with a family of potentials $W^\sigma_{\lambda,\mu}$, $\sigma \in [0,1]$, with $W^\sigma_{\lambda,\mu}$ convex for $\sigma = 0$ and gradually going to $W_{\lambda,\mu}$ as σ increases to 1. They then propose to solve the problem for small σ, and then to increase slowly σ to improve the solution.

1.3.2 The Mumford–Shah functional

In order to study energies (4) or (5), Mumford and Shah (see [49, 50]) proposed to rewrite those in a continuous setting. They considered an observed image $g(x,y)$, with $(x,y) \in \Omega$, Ω bounded open set of $\mathbb{R}^2$, and $g(x,y) \in [0,1]$ for (almost) every (x,y), and then they noticed that the variable L, or rather the set $\{L = 1\}$, describes the "discontinuity" or "jump set" $K \subset \Omega$ of a piecewise regular function $u(x,y)$, $(x,y) \in \Omega$, whereas the finite differences $u_{i+1,j} - u_{i,j}$ (resp., $u_{i,j+1} - u_{i,j}$), are approximations of the partial derivatives $\frac{\partial u}{\partial x}(x,y)$ (resp., $\frac{\partial u}{\partial y}(x,y)$). The energy they wrote was thus (with the standard notation $\nabla u = (\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y})$ for the gradient)

$$\begin{aligned} \mathcal{E}(u,K) \;\; &= \;\; \lambda \int_{\Omega\setminus K} |\nabla u(x,y)|^2\, dxdy \;+\; \mu \cdot \text{length}(K \cap \Omega) \\ &\quad + \nu \int_\Omega (u(x,y) - g(x,y))^2\, dxdy \end{aligned} \tag{6}$$

where λ, μ, ν are positive parameters. They then proposed to study the problem of minimizing energy (6).

In these lecture notes we will try to explain briefly

(a) how this problem can mathematically be handled, in what setting, what functions space, in what sense it has a solution,

(b) a first approximation result that has been proposed in order to minimize more easily energy $\mathcal{E}(u,K)$, in a continuous setting,

(c) in what sense can one say that $\mathcal{E}(u,K)$ and $E(U,L)$ are the "same energies", in a continuous and in a discrete setting,

(d) how it is possible to approximate $\mathcal{E}(u,K)$ by discrete energies, "better", in some sense, than with energy $E(U,L)$.

What we will not describe on the other hand are the possible finite-element approaches that have also been proposed for solving the Mumford–Shah problem. It is still not clear whether they are of some interest for image processing applications or not. They are usually useful in other fields where similar problems are relevant, and in particular in fracture mechanics. The interested reader may consult [13, 37, 16], or [23, 17].

2 Some mathematical preliminaries

2.1 The functions with bounded variation (BV)

2.1.1 Why we need bounded variation functions

For the study of Rudin and Osher's problem (2), the correct mathematical setting is clearly the functions with bounded variation (the criterion they propose to minimize being simply the semi-norm defining such functions), that we will define in the next paragraph. Although it may be not as clear, this is also true for the analysis of the Mumford–Shah functional. Indeed, in order to study energy $\mathcal{E}$, Ambrosio and De Giorgi have suggested to introduce a "weak formulation" depending only on the variable u. This formulation assumes that we are able to define, given a function u, a set of discontinuities S_u and a gradient ∇u everywhere outside of S_u. The weak Mumford–Shah energy is then

$$\mathcal{E}(u) \;=\; \int_\Omega |\nabla u(x)|^2\,dx \;+\; \mathcal{H}^{N-1}(S_u) + \int_\Omega |u(x)-g(x)|^2\,dx. \qquad (7)$$

Here we consider that u, g are defined in a domain Ω of a space of arbitrary dimension N, and the set S_u is $(N-1)$–dimensional, for images you can just replace N by 2 everywhere in the notes. $\mathcal{H}^{N-1}$ denotes the $(N-1)$–*dimensional Hausdorff measure* (see for instance [35]). It is a Borel measure in $\mathbb{R}^N$ that agrees with the traditional definition of the surface for every regular hypersurface in $\mathbb{R}^N$ (any bounded part of an hyperplane, a sphere, ...).

The discontinuity set S_u can be defined for very general functions, but it usually has no kind of regularity. A correct definition of the gradient ∇u requires more regularity of u. Usually, we can define a gradient ∇u as an integrable function if u belongs to the Sobolev space $W^{1,1}(\Omega)$ (or at least $W^{1,1}_{loc}(\Omega)$). But in this case it is possible to show that S_u is almost empty (in fact, $\mathcal{H}^{N-1}(S_u) = 0$: we say that S_u is $\mathcal{H}^{N-1}$–*essentially* empty).

The space of bounded variation functions, that we are going to introduce, doesn't suffer this drawback. It contains functions for which it is possible to define correctly S_u and the gradient ∇u, in such a way that $0 < \mathcal{H}^{N-1}(S_u) < +\infty$ and $\int |\nabla u(x)|^2\, dx < +\infty$. Such a function combines some regularity, and discontinuities across the essentially $(N-1)$–dimensional set S_u

2.1.2 BV functions: definition and main properties

The space of *bounded variation functions* in Ω, denoted by $BV(\Omega)$, is defined in the following way:

$$BV(\Omega) \;=\; \Big\{u \in L^1(\Omega)\;:\; Du \text{ is a bounded Radon vector measure on } \Omega\Big\}\,, \tag{8}$$

where Du is the distributional (or weak) derivative of u, defined by

$$\langle Du, \phi\rangle_{\mathcal{D}'(\Omega,\mathbb{R}^N),\mathcal{D}(\Omega,\mathbb{R}^N)} = -\int_\Omega u(x)\mathrm{div}\, \phi(x)\, dx$$

for any vector field $\phi \in \mathcal{D}(\Omega, \mathbb{R}^N)$, *i.e.*, C^∞ with compact support in Ω.

Let us denote by $\mathcal{M}(\Omega, \mathbb{R}^N)$ the space of N–dimensional bounded (vector–valued) Radon measures on Ω. It is well known (as a consequence of Riesz' representation theorem) that $\mathcal{M}(\Omega, \mathbb{R}^N)$ is identified to the dual of $C_0(\Omega, \mathbb{R}^N)$, the space of all continuous vector fields vanishing at the boundary (this means that if $\phi \in C_0(\Omega, \mathbb{R}^N)$, for every $\varepsilon > 0$, there exists a compact set $K \subset \Omega$ such that $\sup_{x\notin K} |\phi| < \varepsilon$), on which the norm is given by $\|\phi\|_{C_0(\Omega,\mathbb{R}^N)} = \sup_{x\in\Omega} |\phi(x)|$. If $\mu \in \mathcal{M}(\Omega, \mathbb{R}^N)$ is a measure, we can define its *variation* as the Borel positive measure given by

$$|\mu|(E) \;=\; \sup\left\{\sum_{i=1}^{n} |\mu(E_i)|\;:\; \bigcup_{i=1}^{n} E_i \subseteq E\,,\; E_i \cap E_j = \emptyset\, \forall i \neq j\right\}, \tag{9}$$

for every Borel set $E \subseteq \Omega$ (here the $E_i, i = 1, \cdots, n$, are disjoint Borel sets). Saying that the measure μ is bounded is nothing else than saying that $|\mu|(\Omega) < +\infty$, the quantity $|\mu|(\Omega)$ is called the *total variation* of μ (on Ω) and defines the usual norm in the Banach space $\mathcal{M}(\Omega, \mathbb{R}^N)$.

As an element of the dual $C_0(\Omega,\mathbb{R}^N)'$ of $C_0(\Omega,\mathbb{R}^N)$, μ also has a norm given by

$$\|\mu\|_{C_0(\Omega,\mathbb{R}^N)'} = \sup_{\|\phi\|_{C_0(\Omega,\mathbb{R}^N)}\le 1} \langle \mu,\phi\rangle = \sup_{\|\phi\|_{C_0(\Omega,\mathbb{R}^N)}\le 1} \int_\Omega \phi(x)\mu(dx).$$

In fact, both norms coincide, which means that for every $\mu \in \mathcal{M}(\Omega,\mathbb{R}^N)$,

$$|\mu|(\Omega) = \sup\left\{\int_\Omega \phi(x)\mu(dx) \,:\, \phi\in C_0(\Omega;\mathbb{R}^N), |\phi(x)|\le 1 \ \forall x\in\Omega\right\}$$

The weak-$*$ convergence of a sequence of measures (μ_n) is understood as the weak-$*$ convergence in the dual of $C_0(\Omega,\mathbb{R}^N)$, which means that $\mu_n \rightharpoonup \mu$ weakly-$*$ if and only if

$$\int_\Omega \phi(x)\mu_n(dx) \to \int_\Omega \phi(x)\mu(dx)$$

for every $\phi \in C_0(\Omega,\mathbb{R}^N)$.

If Du is a bounded Radon measure, then since $\int_\Omega \phi Du = -\int_\Omega u \operatorname{div}\phi$ for every ϕ in $C_c^1(\Omega;\mathbb{R}^N)$, we deduce (the compactly supported C^1–regular functions being dense in the space $C_0(\Omega;\mathbb{R}^N)$) that if $u\in L^1(\Omega)$,

$$u\in BV(\Omega) \quad\Leftrightarrow\quad V(u,\Omega) := \sup\left\{\int_\Omega u(x)\operatorname{div}\phi(x)\,dx : \phi\in C_c^1(\Omega;\mathbb{R}^N),\right. \\ \left. |\phi(x)|\le 1\ \forall x\in\Omega\right\} < +\infty. \qquad (10)$$

The quantity $V(u,\Omega)$ coincides with the total variation of the measure Du, i.e., $V(u,\Omega) = |Du|(\Omega)$. In fact, saying that $V(u,\Omega)$ must be finite is an equivalent way to define the space $BV(\Omega)$. We call $V(u,\Omega) = |Du|(\Omega)$ the *total variation* of u in Ω. If $u\in C^1(\Omega)$, or u is in the Sobolev space $W^{1,1}(\Omega)$, then the notation in (2) is valid since it is simple to show that $|Du|(\Omega) = \int_\Omega |\nabla u(x)|\,dx$. The space $BV(\Omega)$, endowed with the norm $\|u\|_{BV(\Omega)} = \|u\|_{L^1(\Omega)} + |Du|(\Omega)$, is a Banach space.

Exercise. *Prove that, given $u\in L^1(\Omega)$, $Du\in\mathcal{M}(\Omega,\mathbb{R}^N)$ if and only if $V(u,\Omega)$ (given by (10)) is finite.*

The first result we can state about the total variation is the following semi-continuity property:

Theorem 1 (Semicontinuity of the total variation) *The convex functional $u\mapsto V(u,\Omega) = |Du|(\Omega)\in[0,+\infty]$ is lower semicontinuous in the $L^1_{loc}(\Omega)$ topology.*

This means that if u_n goes to u in $L^1(\Omega')$ for every $\Omega' \subset\subset \Omega$, then $|Du|(\Omega) \leq \liminf_{n\to\infty} |Du_n|(\Omega)$. The proof of theorem 1 is straightforward if we consider the definition (10) of the variation of u. Indeed, in (10), $V(u,\Omega)$ is built as the sup of the linear functionals $u \mapsto \int_\Omega u(x)\operatorname{div}\phi(x)\,dx$ for $\phi \in C_c^1(\Omega;\mathbb{R}^N)$. Since each of these functionals is continuous in the L^1_{loc} topology, we deduce that the sup is lower semicontinuous.

Next, we have the following Poincaré inequalities

Theorem 2 (Poincaré inequalities) *There exists a constant $c = c(N)$ such that if $u \in L^1_{loc}(R^N)$, then*

$$\|u\|_{L^{\frac{N}{N-1}}(\mathbb{R}^N)} \leq c|Du|(\mathbb{R}^N),$$

and if B is a ball and $u \in L^1(B)$,

$$\left\|u - \fint_B u\right\|_{L^{\frac{N}{N-1}}(B)} \leq c|Du|(B).$$

Here and everywhere in the notes $\fint_X u = \fint_X u(x)\,dx$ denotes the average $\frac{1}{|X|}\int_X u(x)\,dx$.

If Ω is a bounded Lipschitz–regular open set (this will be assumed always in what follows) we can build a continuous linear extension operator $T_{\Omega,\Omega'}$ from $BV(\Omega)$ to $BV(\Omega')$ for every Ω' with $\Omega \subset\subset \Omega'$, which means that for every $u \in BV(\Omega)$ we can find $u' \in BV(\Omega')$ with $u' \equiv u$ on Ω and $\|u'\|_{BV(\Omega')} \leq c\|u\|_{BV(\Omega)}$, the constant c depending only on Ω, Ω'. This extension allows to generalize the second inequality in the last theorem to any such Ω: we deduce that there exists a constant $c = c(\Omega)$ such that

$$\left\|u - \fint_\Omega u\right\|_{L^{\frac{N}{N-1}}(\Omega)} \leq c|Du|(\Omega) \tag{11}$$

for every $u \in BV(\Omega)$ (see [34] for details).

We state, still without any proof, the two next theorems that are fundamental for the study of the space $BV(\Omega)$.

Theorem 3 (Sobolev embeddings) *Let Ω be bounded and Lipschitz–regular. Then the space $BV(\Omega)$ is continuously embedded in $L^{N/(N-1)}(\Omega)$, and compactly embedded in $L^p(\Omega)$ for every $1 \leq p < N/(N-1)$.*

The first assertion is a consequence of the previous theorem. The second means that if a sequence of functions $(u_j)_{j\geq 1}$ is bounded in $BV(\Omega)$, *i.e.*,

$\sup_j \|u_j\|_{L^1(\Omega)} + |Du_j|(\Omega) < +\infty$, then we can extract a subsequence u_{j_k} and there exists a function $u \in BV(\Omega)$ such that, as $k \to \infty$, $Du_{j_k} \rightharpoonup Du$ weakly-$*$ as a measure and $u_{j_k} \to u$ *strongly* in $L^p(\Omega)$, for every $p < N/(N-1)$.

Theorem 4 (Approximation by smooth functions) *Let $u \in BV(\Omega)$. Then there exists a sequence $(u_n)_{n \geq 1} \subset C^\infty(\Omega)$ such that, as $n \to \infty$, $u_n \to u$ in $L^1(\Omega)$, $Du_n \rightharpoonup Du$ weakly-$*$ as measures, and*

$$|Du_n|(\Omega) = \int_\Omega |\nabla u_n(x)|\, dx \to |Du|(\Omega).$$

These properties (in fact, mainly Theorem 2) are sufficient to derive the existence for problem (2), as we are going to show in the next section.

2.1.3 Existence for the Rudin-Osher approach

The existence for problem (2) in dimension $N = 1$ or $N = 2$ is ensured provided we assume that

- the operator A satisfies $A1 = 1$ (i.e., the image of a constant function is the same function),
- the initial data satisfies $\int_\Omega |g(x) - \fint_\Omega g|^2\, dx \geq \sigma^2$,
- there exists a $\tilde{u}$ satisfying (1) such that $|Du|(\Omega) < +\infty$.

The first assumption is not absolutely necessary (we need that $A1 \not\equiv 0$) but simplifies a lot the proof, it is obviously satisfied if A corresponds to a convolution with a kernel of integral 1 ($Au = \rho * u$, $\int \rho = 1$) (provided the boundary effects are treated correctly). The second assumption is needed, observe that if the model $g = Au + n$ is correct then it should be satisfied (with n rapidly oscillating so that $\int_\Omega Au \cdot n \simeq 0$). The last assumption means that $I = \inf \{|Du|(\Omega) : u \text{ satisfies (1)}\} < +\infty$, otherwise any u satisfying (1) is a solution but the problem is of little interest. In the general continuous setting the existence of such a $\tilde{u}$ is not absolutely obvious.

The following proof is taken from [24]. We consider a minimizing sequence $(u_n)_{n \geq 1}$ for (2), of functions u_n that all satisfy the constraints and such that $|Du_n|(\Omega) \to I$ as $n \to \infty$. Such a sequence exists because of our third assumption. We assume in order to simplify the notations that $|\Omega| = 1$ (so that in particular $\fint_\Omega u = \int_\Omega u$ for every u). We show, first, that the average

$m_n = \int_\Omega u_n$ remains bounded. This is obvious if A is the identity, or has a continuous inverse. Otherwise, we can write (since $A1 = 1$)

$$\begin{aligned}\sigma^2 &= \int_\Omega |Au_n - g|^2 = \int_\Omega |Au_n - m_n + m_n - g|^2 \\ &= \int_\Omega |A(u_n - m_n) + m_n - g|^2\end{aligned}$$

so that

$$\begin{aligned}\sigma &\geq \|m_n - g\|_{L^2(\Omega)} - \|A(u_n - m_n)\|_{L^2(\Omega)} \\ &\geq \|m_n - g\|_{L^2(\Omega)} - \|A\|\,\|u_n - m_n\|_{L^2(\Omega)}\end{aligned}$$

where $\|A\|$ denotes the norm of A as a continuous operator of $L^2(\Omega)$. Since $N = 1$ or 2, $2 \leq N/(N-1)$ and by (11),

$$\|u_n - m_n\|_{L^2(\Omega)} = \left\|u_n - \int_\Omega u_n(x)\,dx\right\|_{L^2(\Omega)} \leq c|Du_n|(\Omega). \tag{12}$$

The total variation $|Du_n|(\Omega)$ remains bounded, therefore also $m_n = \int_\Omega u_n$ is bounded. This implies (using again (12)) that u_n is bounded in $L^2(\Omega)$.

Upon extracting a subsequence we may thus assume that there exists $u \in L^2(\Omega) \cap BV(\Omega)$ such that $u_n \rightharpoonup u$ weakly in L^2 and $Du_n \rightharpoonup Du$ weakly-$*$ as a measure. We also have (since A is continuous and linear) $Au_n \rightharpoonup Au$, therefore by semicontinuity we get

$$|Du|(\Omega) \leq \liminf_{n\to\infty} |Du_n|(\Omega) = I, \text{ and,}$$

$$\int_\Omega |Au(x) - g(x)|^2\,dx \leq \sigma^2, \quad \int_\Omega Au(x)\,dx = \int_\Omega g(x)\,dx.$$

(Alternatively, we could invoke Theorem 3 to deduce that some subsequence of (u_n) converges to some u *strongly* in $L^1(\Omega)$, and Theorem 1 to conclude that $|Du|(\Omega) \leq \liminf_{n\to\infty} |Du_n|(\Omega) = I$.)

We now introduce for $t \in [0,1]$ the function $u^t = tu + (1-t)\int_\Omega g$. We have for every t, $|Du^t|(\Omega) = t|Du|(\Omega) \leq tI \leq I$, $\int_\Omega Au^t = \int_\Omega g$, and we have $\int_\Omega |Au^0 - g|^2 = \int_\Omega |g - \int_\Omega g|^2 \geq \sigma^2$ (by assumption), and $\int_\Omega |Au^1 - g|^2 = \int_\Omega |Au - g|^2 \leq \sigma^2$. By continuity of the map $t \mapsto \int_\Omega |Au^t - g|^2$, there exists therefore a $t_0 \in [0,1]$ such that u^{t_0} satisfies (1), and $|Du^{t_0}|(\Omega) \leq I$. Necessarily we must have $|Du^{t_0}|(\Omega) = I$, so that $t_0 = 1$ and u is the solution of problem (2). $\square$

2.2 More properties of BV functions

In the previous section we have just introduced the very basic properties of BV functions that allowed us to state correctly problem (2) and show that it is well posed. Now, if we want to study the weak Mumford–Shah energy (7), we see that we need to know more properties of these functions. In particular, we must define correctly the discontinuity set S_u and study its regularity. We also need to describe precisely the measure Du. This will be done in the next sections. We will not prove all the results since it is too difficult for the purpose of these lectures, but we will try to give a correct idea of these results by describing with more precision the simpler one-dimensional case.

2.2.1 The jumps set S_u

Let us first introduce the *approximate limits* of a function u at some point $x \in \Omega$. Given $u : \Omega \to [-\infty, +\infty]$ a measurable function, we can define the *approximate upper limit* of u at $x \in \Omega$ as

$$u_+(x) \;=\; \inf\left\{ t \in [-\infty, +\infty] \;:\; \lim_{\rho \downarrow 0} \frac{|\{y \,:\, u(y) > t\} \cap B_\rho(x)|}{\rho^N} = 0 \right\},$$

where $B_\rho(x)$ is the ball of radius ρ centered at x and $|E|$ denotes the Lebesgue measure of the set E. $u_+(x)$ is thus the greatest lower bound of the set of values t for which the set $\{u > t\}$ has (Lebesgue) density 0 at x: on the other hand if $t < u_+(x)$, then this set must have strictly positive density at x. The *approximate lower limit* $u_-(x)$ is defined in the same way i.e.,

$$u_-(x) \;=\; -(-u)_+(x) = \sup\left\{ t \in [-\infty, +\infty] : \lim_{\rho \downarrow 0} \frac{|\{y : u(y) < t\} \cap B_\rho(x)|}{\rho^N} = 0 \right\}.$$

The set

$$S_u \;=\; \{x \in \Omega \;:\; u_-(x) < u_+(x)\},$$

is the set of essential discontinuities of u, it is a (Lebesgue–)negligible Borel set. If $x \notin S_u$, we write $\tilde{u}(x) = u_-(x) = u_+(x) = \mathrm{ap}\,\lim_{y \to x} u(y)$, and when $\tilde{u}(x) \neq \pm\infty$ we say that u is *approximately continuous* at x.

Let us first analyse the one–dimensional case, which is simpler.

2.2.2 BV functions in one dimension

In this section, we consider a (bounded) interval $I = (a, b) \subset \mathbb{R}$ (here, $a < b$ are two real numbers). In this case the total variation of a function $u \in L^1(I)$

is simply

$$V(u,I) \;=\; \sup\left\{\int_I u(x)v'(x)\,dx \;:\; v\in C_c^1(I;[-1,1])\right\}.$$

Notice that the usual "classical" definition is different:

$$\mathrm{Var}\,(u,I) \;=\; \sup\left\{\sum_{i=1}^{n-1}|u(t_{i+1})-u(t_i)| \;:\; a<t_1<\cdots<t_n<b\right\}$$

and in general we do not have $V(u,I) = \mathrm{Var}\,(u,I)$. Indeed, the second definition depends on the pointwise values of the function u and the value of $\mathrm{Var}\,(u,I)$ can be made infinite by changing u on a set of measure zero (for instance on a sequence $(x_n)_{n\geq 1}$ of points in I), whereas the first definition gives the same value for two functions that are almost everywhere equal. In fact, if $u\in C^1(I)$, then clearly for every $v\in C_c^1(I;[-1,1])$, $\int_I uv' = -\int_I u'v$, and we deduce that $V(u,I) = \int_I |u'|$, in this case it is easy to show that $V(u,I) = \mathrm{Var}\,(u,I)$.

Exercise. *Show that for every u, $V(u,I) \leq Var\,(u,I)$. [Hint: consider $v\in C_c^1(I;[-1,1])$ and remark that $\lim_{h\to 0}\frac{1}{h}\int_{I_h}(v(x+h)-v(x))u(x)\,dx = \int v'(x)u(x)\,dx$. (Here $I_h = \{x\in I \,:\, x+h\in I\}$.) Prove then that for $h>0$ sufficiently small, $\frac{1}{h}\int_{I_h}(v(x+h)-v(x))u(x)\,dx \leq Var\,(u,I)$.]*

In the general case we have that $V(u,I)\leq \mathrm{Var}\,(v,I)$ and $V(u,I) = \min\{\mathrm{Var}\,(v,I)\,:\, v=u \text{ a.e.}\}$ (see the next exercise).

The distributional derivative of $u\in L^1(I)$ is the distribution Du defined by

$$\langle Du,\varphi\rangle_{\mathcal{D}'(I),\mathcal{D}(I)} \;=\; -\int_I u(x)\varphi(x)\,dx$$

for every $\varphi\in\mathcal{D}(I)$ (i.e., the set $C_c^\infty(I)$ with the appropriate topology). The function u is in $BV(I)$ if and only if Du is a bounded Radon measure on I, which means that $Du\in\mathcal{M}(\Omega)\simeq C_0(I)'$, the dual of $C_0(I)$, which is the set of continuous functions on $\overline{I}=[a,b]$ such that $u(a)=u(b)=0$. It can be proved (quite easily) that Du is a bounded Radon measure on I if and only if $V(u,I)<+\infty$, and that in this case we have

$$V(u,I) \;=\; |Du|(I) \;=\; \sup\left\{\sum_{i=1}^n |Du(I_i)| \;:\; \bigcup_{i=1}^n I_i\subseteq I\,,\; I_i\cap I_j=\emptyset\,\forall i\neq j\right\},$$

(where the sets I_i are Borel sets) the right-hand side of the last equation being the standard definition of the total variation of the measure Du (which is also the norm of Du when it is seen as an element of the dual $C_0(I)'$). (More generally, the *variation* of a vector–valued (or real–valued) Borel measure μ is the Borel positive measure $|\mu|$ defined by equation (9).)

We now introduce two functions u_l and u_r, defined for every $x \in I$ by

$$u_l(x) \;=\; Du((a,x)) \;\;\text{and}\;\; u_r(x) \;=\; Du((a,x]).$$

Here, as usual, $(a,x) = \{y \,:\, a < y < x\}$ denotes the *open* interval of extremities a and $x > a$, which is sometimes also denoted by $]a,x[$, whereas $(a,x]$ is the interval $\{y \,:\, a < y \le x\}$.

Lemma 1 *The function u_l is left–continuous, while u_r is right–continuous. Moreover, $u_l = u_r$ except on a set at most countable.*

Proof. First of all, $u_r(x) - u_l(x) = Du(\{x\})$ so that $u_r(x) = u_l(x)$ except when x is an atom of the measure Du (a point such that $Du(\{x\}) \neq 0$), but a bounded vector (or real–valued) measure can have at most a countable number of atoms.

To show that u_l is left–continuous, for any $x \in I$ and each sequence of non–negative numbers $\rho_n \downarrow 0$ we must show that $u_l(x-\rho_n)$ goes to $u_l(x)$ as $n \to \infty$. But $|u_l(x-\rho_n) - u_l(x)| = |Du([x-\rho_n, x))| \le |Du|([x-\rho_n, x))$, and by standard properties of positive measures we know that (assuming, without loss of generality, that ρ_n is a decreasing sequence) $\lim |Du|([x-\rho_n, x)) = |Du|(\cap_{n\ge 1}[x-\rho_n, x)) = |Du|(\emptyset) = 0$. Therefore u_l is left–continuous. For the same reason, u_r is right–continuous (indeed, $u_r(x) = Du(I) - Du((x,b))$). $\square$

Remark. More precisely, we can show in the same way that

$$\begin{aligned}\lim_{\substack{\rho\to 0\\ \rho>0}} u_l(x-\rho) \;&=\; \lim_{\substack{\rho\to 0\\ \rho>0}} u_r(x-\rho) \;=\; u_l(x), \quad \text{and} \quad \lim_{\substack{\rho\to 0\\ \rho>0}} u_l(x+\rho)\\ &=\; \lim_{\substack{\rho\to 0\\ \rho>0}} u_r(x+\rho) \;=\; u_r(x).\end{aligned}$$

Lemma 2 *The distributional derivatives of u_l, u_r and u are equal ($Du_l = Du_r = Du$).*

Proof. Let us show, for instance, that $Du_l = Du$. Consider $\varphi \in \mathcal{D}(I)$. We have (using Fubini's theorem)

$$\int_I \varphi Du_l \;=\; -\int_a^b \varphi'(x) Du((a,x))\,dx \;=\; -\int_a^b \varphi'(x)\left\{\int_{y\in(a,x)} Du(dy)\right\} dx \;=$$

$$= \;-\int_{y\in(a,b)}\left\{\int_{x\in(y,b)} \varphi'(x)\,dx\right\} Du(dy) \;=\; -\int_I -\varphi(y)Du(dy) \;=\; \int_I \varphi Du,$$

showing the desired equality. □

In particular, we deduce from the last lemma that $D(u-u_l) = D(u-u_r) = D(u_l-u_r) = 0$ so that the functions u, u_l, u_r can differ at most by a constant. We can redefine the functions u_l and u_r by adding the appropriate constant so that $u_l = u$ and $u_r = u$ almost everywhere in I (*i.e.*, now, $u_l(x) = c + Du((a,x))$ and $u_r(x) = c + Du((a,x])$ with $c \in \mathbb{R}$ appropriately chosen to have $u_l = u_r = u$ a.e.) We have shown so far the following proposition.

Proposition 1 *Every $u \in BV(I)$ has a left–continuous and a right-continuous representant*[1]

Exercise. *Show that $|Du|(I) = Var(u_l, I) = Var(u_r, I)$.*

We now introduce the function

$$\dot{u} \;=\; \frac{Du}{\mathcal{L}^1}$$

which is the Radon–Nykodym derivative of the measure Du with respect to the Lebesgue measure $\mathcal{L}^1$ on I (in particular, $\dot{u} \in L^1(I)$). The Radon–Nykodym derivation theorem states that for $\mathcal{L}^1$–a.e. $x \in I$,

$$\dot{u}(x) \;=\; \lim_{\rho\to 0} \frac{Du((x-\rho, x+\rho))}{2\rho} \;=\; \lim_{\rho\to 0} \frac{Du([x-\rho, x+\rho])}{2\rho}$$

and we can write the measure Du as

$$Du \;=\; \dot{u}(x)\,dx \;+\; D^s u$$

with $D^s u \perp \mathcal{L}^1$, which means that there exists a Borel set $E \subset I$ such that $|E| = \mathcal{L}^1(E) = 0$ and $|D^s u|(I \setminus E) = 0$. In particular the Radon–Nykodym

[1] We recall that a *representant* of a function $u \in L^1$ is a function $\tilde{u}$ a.e. equal to u, or more precisely belonging to the equivalence class of a.e. equal functions defining u.

derivative $\frac{|D^s u|}{\mathcal{L}^1}$ is zero, so that for $\mathcal{L}^1$–a.e. $x \in I$, $\lim_{\rho\to 0} |D^s u|((x-\rho, x+\rho))/2\rho = 0$.

Consider now x a Lebesgue point of $\dot u$, *i.e.*, such that $\lim_{\rho\to 0} \frac{1}{\rho}\int_{x-\rho}^{x+\rho} |\dot u(y) - \dot u(x)|\, dy = 0$ (a.e. $x \in I$ satisfies this property), and also assume that $\lim_{\rho\to 0} |D^s u|([x-\rho, x+\rho])/2\rho = 0$. Then:

$$\limsup_{\rho\downarrow 0} \left| \frac{u_l(x+\rho) - u_l(x)}{\rho} - \dot u(x) \right| = \limsup_{\rho\downarrow 0} \left| \frac{1}{\rho} \left(\int_x^{x+\rho} \dot u(y)\, dy \right.\right.$$
$$\left.\left. + D^s u([x, x+\rho)) \right) - \dot u(x) \right| \le \limsup_{\rho\downarrow 0} \frac{1}{\rho} |D^s u|([x, x+\rho))$$
$$+ \limsup_{\rho\downarrow 0} \frac{1}{\rho} \int_x^{x+\rho} |\dot u(y) - \dot u(x)|\, dy = 0.$$

In the same way, we can prove that $\limsup_{\rho\downarrow 0} |u_l(x-\rho) - u_l(x)/\rho - \dot u(x)| = 0$, showing that u_l has a (classical) derivative at x which is $\dot u(x)$. We have shown the following proposition.

Proposition 2 *The functions u_l and u_r have a derivative a.e. in I, and $u_l'(x) = u_r'(x) = \dot u(x)$ for a.e. $x \in I$.*

Remark. In a similar way we can show that at a.e. x,

$$\limsup_{\rho\to 0} \frac{1}{2\rho} \int_{|y-x|<\rho} \frac{|u(y) - u(x) - \dot u(x)(y-x)|}{|y-x|}\, dy = 0, \tag{13}$$

which expresses the fact that $\dot u(x)$ is the *approximate derivative* of u at x. This property will have a generalization in higher dimension.

2.2.3 The jumps set and the singular part of Du

Now we will try do describe better the singular part $D^s u$ of the measure Du, and the set S_u. The first property is the following.

Proposition 3 *At every $x \in I$, $u_+(x) = u_l(x) \vee u_r(x)$ and $u_-(x) = u_l(x) \wedge u_r(x)$. In particular, $S_u = \{x \in I \,:\, u_l(x) \neq u_r(x)\}$.*

Remark. By Lemma 1 we deduce that the set S_u is at most countable. In fact, since $u_r(x) - u_l(x) = Du(\{x\})$, it is the set of the atoms of the measure Du.

Proof. Let us first show that $u_+(x) \geq u_l(x)$ for every $x \in I$. Let $t < u_l(x)$: by the left–continuity of u_l there exists $\rho > 0$ such that $x - \rho < y \leq x \Rightarrow u_l(y) > t$. Therefore $\{y : u_l(y) > t\} \supseteq (x - \rho, x)$ so that if $\rho' \leq \rho$, $\rho' = |(x - \rho', x)| \leq |\{y : u_l(y) > t\} \cap B_{\rho'}(x)| = |\{y : u(y) > t\} \cap B_{\rho'}(x)|$, where the last equality comes from the fact that $u = u_l$ a.e. in I. We deduce that $\liminf_{\rho' \downarrow 0} |\{y : u(y) > t\} \cap B_{\rho'}(x)| \geq 1$ so that (by the definition of u_+) $t \leq u_+(x)$. Thus $u_l(x) \leq u_+(x)$. In the same way we get that $u_r(x) \leq u_+(x)$.

Conversely let $t > u_l(x) \vee u_r(x)$. By left– and right–continuity we know that there exists $\rho > 0$ such that $x - \rho < y < x \Rightarrow u_l(y) < t$ and $x < y < x + \rho \Rightarrow u_r(y) < t$. As before we deduce this time that $\limsup_{\rho' \downarrow 0} |\{y : u(y) > t\} \cap B_{\rho'}(x)| = 0$. Thus, $u_+(x) \leq u_l(x) \vee u_r(x)$. This proves that $u_+ = u_l \vee u_r$ on I. The proof of the equality $u_- = u_l \wedge u_r$ is identical. □

Now, we split the measure $D^s u$ into two parts, called respectively Ju ("J" for "jumps") and Cu ("C" for "Cantor"):

$$Ju \;=\; D^s u \llcorner S_u \quad\text{and}\quad C_u \;=\; D^s u \llcorner (I \setminus S_u).$$

(Notice that, since $|S_u| = 0$ (S_u is finite or countable), J_u is also $Du \llcorner S_u$).

Since S_u is the set of the atoms of the measure Du, we have

$$\begin{aligned} J_u &= \sum_{x \in S_u} Du(\{x\}) \delta_x \\ &= \sum_{x \in S_u} (u_r(x) - u_l(x)) \delta_x . \end{aligned}$$

(δ_x stands for the Dirac mass at x.) This measure represents the jumps of u across its discontinuities. It can also be written as

$$J_u \;=\; \sum_{x \in S_u} (u_+(x) - u_-(x)) \nu_u(x) \delta_x \;=\; (u_+ - u_-)\, \nu_u\, \mathcal{H}^0 \llcorner S_u \tag{14}$$

where $\nu_u(x) \in \{-1, +1\}$ represents the direction of the jump of u at x: $\nu_u(x) = +1$ if $u_l(x) = u_-(x)$, $u_r(x) = u_+(x)$, so that u is "increasing" at x ($u_l(x) < u_r(x)$), whereas $\nu_u(x) = -1$ when $u_l(x) = u_+(x)$, $u_r(x) = u_-(x)$, meaning u is "decreasing" at x ($u_l(x) > u_r(x)$). This last expression (14) will be generalized in higher dimension.

Consider now the measure Cu. It has no atoms (*i.e.*, $Cu(\{x\}) = 0$ for every $x \in I$) since $Du(\{x\}) = 0$ and $D^s u(\{x\}) = 0$ for every $x \in I \setminus S_u$. On the other hand, it is singular with respect to the Lebesgue measure $\mathcal{L}^1$ (*i.e.*,

$Cu \perp \mathcal{L}^1$). It is called the *Cantor part* of u. We will soon show an example of a function u with Du having a Cantor part.

Let us now return for a while to the weak Mumford–Shah functional (7). In one–dimension, we can write it

$$\mathcal{E}(u) \;=\; \int_I |\dot{u}(x)|^2\,dx \;+\; \mathcal{H}^0(S_u) \;+\; \int_I |u(x)-g(x)|^2\,dx.$$

(Here the zero–dimensional Hausdorff measure of S_u is simply the cardinality $\sharp S_u$ of the set S_u.)

In our definitions of $\dot{u}(x)$ and S_u, we see that the weak energy $\mathcal{E}(u)$ is correctly defined. However, if we try to find a minimum of $\mathcal{E}(u)$ in the class of all functions with bounded variation, we realize that $\inf\{\mathcal{E}(u) : u \in BV(I)\} = 0$ and that it is in general not reached! This happens because it is possible to approximate every function in $L^2(I)$ (here, g) by BV functions such that $S_u = \emptyset$, $\dot{u}(x) = 0$ a.e., and all the derivatives are Cu. A typical example of such a function is the "Cantor-Vitali" function, defined as the (uniform) limit of the continuous functions in $[0,1]$

$$u_k(x) \;=\; \frac{|C_k \cap [0,x]|}{|C_k|} \text{ where } C_0 = [0,1],\; C_k = C_{k-1} \setminus \bigcup_{n=1}^{3^{k-1}} \left(\frac{n}{3^k}, \frac{n+1}{3^k}\right) \text{ for } k \geq 1,$$

see Figure 4. The set $C = \cap_{k=0}^{\infty} C_k = \lim_k C_k$ is the Cantor set, it has zero length. The function u is continuous, and $u' = 0$ in $[0,1] \setminus C$, i.e., almost everywhere in $(0,1)$. The derivative Du is entirely supported by the negligible set C, and is therefore singular with respect to the Lebesgue measure. Thus $Du = Cu$.

Exercise. *Show that any function $f \in L^2(0,1)$ can be approximated in L^2 norm by a sequence f_n of functions in $BV(0,1)$ with $Df_n = Cf_n$ ($\dot{f}_n = 0$, $S_{f_n} = \emptyset$ for every n).*

If we want to minimize $\mathcal{E}(u)$, we have to restrict ourselves to the set of function we want to consider. We will therefore introduce a new subspace of $BV(I)$, made of the functions for which Cu is zero.

2.2.4 Special BV functions, in dimension one

Definition. We say that a function $u \in BV(I)$ is a *special function with bounded variation* if $Cu = 0$, which means that the singular part $D^s u$ of the

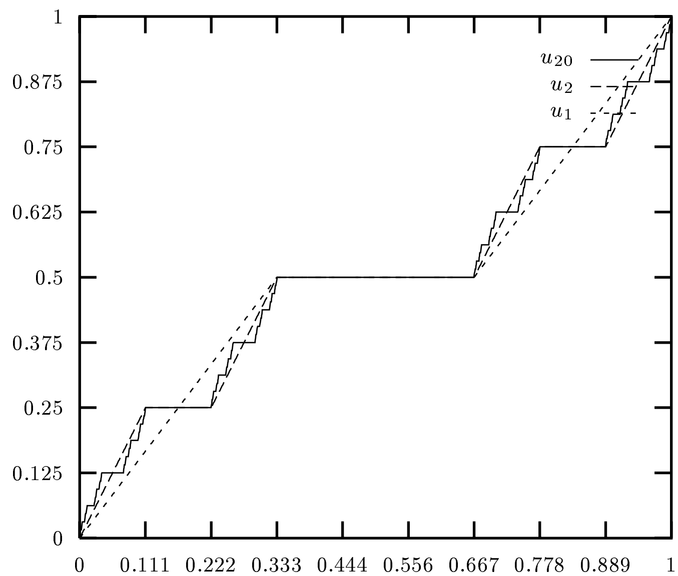

Figure 4: The Cantor-Vitali function.

distributional derivative Du is concentrated on the jump set S_u. We denote by $SBV(I)$ the space of such functions.

The main tool in order to prove the existence of a minimizer for the weak Mumford–Shah energy $\mathcal{E}$ is the following compactness and semicontinuity theorem, due to Ambrosio.

Theorem 5 (Ambrosio, one dimensional version) *Let $I \subset \mathbb{R}$ be an open and bounded interval and (u_j) be a sequence in $SBV(I)$. Suppose that*

$$\sup_j \int_I \dot{u}_j(x)^2\, dx \;+\; \mathcal{H}^0(S_{u_j}) \;+\; \|u_j\|_{L^\infty(I)} \;<+\infty.$$

Then there exist a subsequence (not relabeled) and a function $u \in SBV(I)$ such that

$$\begin{aligned} &u_j(x) \to u(x) \text{ a.e. in } I, \\ &\dot{u}_j \rightharpoonup \dot{u} \text{ weakly in } L^2(I), \\ &\mathcal{H}^0(S_u) \;\le\; \liminf_{j\to\infty} \mathcal{H}^0(S_{u_j}). \end{aligned} \tag{15}$$

Proof. Consider such a sequence u_j. In what follows we will extract several subsequences from u_j that will all still be denoted by u_j. Remark that since $\sup_j \mathcal{H}^0(S_{u_j}) = \sup_j \sharp S_{u_j} < +\infty$, there exists an integer k such that $k = \liminf_j \sharp S_{u_j}$ and we can extract a first subsequence such that $\sharp S_{u_j} = k$ for every j. We let $S_{u_j} = \{x_j^1, \cdots, x_j^k\}$, with $a < x_j^1 < x_j^2 < \cdots < x_j^k < b$. Extracting a further subsequence we may assume that each x_j^n converges to some $x^n \in \bar{I} = [a,b]$. For $t \geq 0$ we will set $I_t = I \setminus \cup_{n=1}^k [x^n - t, x^n + t]$. For a fixed $\ell \geq 1$, if j is large enough we have that $x_j^n \notin I_{1/\ell}$ for every $n = 1, \cdots, k$. In this case, $u_j \in H^1(I_{1/\ell})$ and is uniformly bounded:

$$\sup_j \int_{I_{1/\ell}} |u_j'|^2 \, dx \; = \; \sup_j \int_{I_{1/\ell}} |\dot{u}_j|^2 \, dx \; < \; +\infty \, , \;\; \text{and}$$

$$\sup_j \|u_j\|_{L^\infty(I_{1/\ell})} \; < \; +\infty.$$

We can therefore extract a subsequence such that u_j converges to some function $u \in H^1(I_{1/\ell})$, uniformly on $I_{1/\ell}$, and $\dot{u}_j = u_j' \rightharpoonup u'$ weakly in $L^2(I_{1/\ell})$.

Using a diagonal procedure, since $\cup_{\ell \geq 1} I_{1/\ell} = I_0$, we can in this way build a function $u \in H^1_{loc}(I_0)$ such that $u_j \to u$ locally uniformly on I_0 and $\dot{u}_j \rightharpoonup u'$ weakly in $L^2_{loc}(I_0)$.

But since $\dot{u}_j$ is bounded in $L^2(I) = L^2(I_0)$, we deduce that $\dot{u}_j \rightharpoonup u'$ weakly in $L^2(I)$.

In particular, $u' \in L^2(I)$ and $u \in H^1(I \setminus \{x^1, \cdots, x^k\}) \cap L^\infty(I)$, so that $u \in SBV(I)$, $\dot{u} = u'$, and $S_u \subseteq \{x^1, \cdots, x^k\}$, showing also that $\sharp S_u \leq k$ and achieving the proof of Theorem 5. $\square$

Exercise. *Show that* $u \in H^1(I \setminus \{x^1, \cdots, x^k\}) \cap L^\infty(I) \Rightarrow u \in SBV(I)$, $\dot{u} = u'$, *and* $S_u \subseteq \{x^1, \cdots, x^k\}$.

Exercise. *Use Theorem 5 to show that the weak Mumford–Shah energy* $\mathcal{E}$ *has a minimizer in* $SBV(I)$.

2.2.5 The general, N–dimensional case

We return to the general case of functions defined on an open set $\Omega \subseteq \mathbb{R}^N$, $N \geq 1$.

If $u \in BV(\Omega)$, it can been shown that the set S_u is *countably* $(\mathcal{H}^{N-1}, N-1)$*–rectifiable*, i.e.,

$$S_u = \bigcup_{i=1}^{\infty} K_i \cup \mathcal{N}$$

where $\mathcal{H}^{N-1}(\mathcal{N}) = 0$ and each K_i is a compact subset of a C^1–hypersurface Γ_i. Note that this is a very weak notion of regularity: the set S_u could still be, for instance, dense in Ω.

There exists a Borel function $\nu_u : S_u \to \mathbb{S}^{N-1}$ such that $\mathcal{H}^{N-1}$-a.e. in S_u the vector $\nu_u(x)$ is normal to S_u at x in the sense that it is normal to Γ_i if $x \in K_i$. For every $u, v \in BV(\Omega)$, we must therefore have $\nu_u = \pm\nu_v$ $\mathcal{H}^{N-1}$-a.e. in $S_u \cap S_v$.

As in the one–dimensional case, the derivative Du of every $u \in BV(\Omega)$ can be decomposed as follows:

$$\begin{array}{rcl} Du &=& \nabla u(x)\,dx \;+\; J_u \;+\; Cu \\ &=& \nabla u(x)\,dx \;+\; (u_+ - u_-)\nu_u \mathcal{H}^{N-1} \llcorner S_u \;+\; Cu \end{array}$$

where $\nabla u = \frac{Du}{\mathcal{L}^N}$, the Radon–Nykodym derivative of Du with respect to the Lebesgue measure $\mathcal{L}^N$, is also the *approximate gradient* of u, defined a.e. in Ω by

$$\operatorname{ap}\lim_{y \to x} \frac{u(y) - u(x) - \langle \nabla u(x), y - x\rangle}{|y - x|} = 0,$$

(remember equation (13)). $\mathcal{H}^{N-1} \llcorner S_u$ is the restriction of the $(N-1)$–dimensional Hausdorff measure to the set S_u so that $J_u = (u_+ - u_-)\nu_u \mathcal{H}^{N-1} \llcorner S_u$ is the "jump" part of the measure Du, that is carried by the discontinuity set of u (compare with equation (14)). Eventually, Cu is the *Cantor part* of the measure Du, which is singular with respect to the Lebesgue measure and such that $|Cu|(E) = 0$ for any $(N-1)$–dimensional set E with $\mathcal{H}^{N-1}(E) < +\infty$.

With these definitions of $\nabla u(x)$ and S_u, we see here again that the weak energy (7), $\mathcal{E}(u)$, is correctly defined. Here again as in the one–dimensional case we have $\inf\{\mathcal{E}(u) : u \in BV(\Omega)\} = 0$ and the infimum is usually not reached. We must consider as previously the functions $u \in BV(\Omega)$, such that Cu is zero.

2.2.6 Special BV functions

Definition. We say that a function $u \in BV(\Omega)$ is a *special function with bounded variation* if $Cu = 0$, which means that the singular part of the distributional derivative Du is concentrated on the jump set S_u. We denote by $SBV(\Omega)$ the space of such functions. We also define the space $GSBV(\Omega)$ of *generalized* SBV functions as the set of all measurable functions $u : \Omega \to [-\infty, +\infty]$ such that for any $k > 0$, $u^k = (-k \vee u) \wedge k \in SBV(\Omega)$ (where $X \wedge Y = \min(X, Y)$ and $X \vee Y = \max(X, Y)$) (This follows Ambrosio's definition in [2], notice that sometimes $GSBV(\Omega)$ is defined as the space we call hereafter $GSBV_{loc}(\Omega)$, which is the space of functions that belongs to $GSBV(A)$ for any open set $A \subset\subset \Omega$, i.e., such that $\overline{A}$ is compact and included in Ω.)

If $u \in GSBV_{loc}(\Omega) \cap L^1_{loc}(\Omega)$, u has an approximate gradient a.e. in Ω, moreover, as $k \uparrow \infty$, the function $u^k = (-k \vee u) \wedge k$ satisfies

$$\nabla u^k \to \nabla u \ \text{ a.e. in } \Omega, \qquad \text{and} \qquad |\nabla u^k| \uparrow |\nabla u| \ \text{ a.e. in } \Omega; \tag{16}$$

$$S_{u^k} \subseteq S_u, \ \ \mathcal{H}^{N-1}(S_{u^k}) \to \mathcal{H}^{N-1}(S_u) \quad \text{and} \quad \nu_{u^k} = \nu_u \ \ \mathcal{H}^{N-1}\text{-a.e. in } S_{u^k}. \tag{17}$$

2.2.7 Ambrosio's compactness theorem

We mention the following compactness and lower semi-continuity result that was proved in [2]:

Theorem 6 (Ambrosio) *Let Ω be an open subset of $\mathbb{R}^N$ and let (u_j) be a sequence in $GSBV(\Omega)$. Suppose that there exist $p \in [1, \infty]$ and a constant C such that*

$$\int_\Omega |\nabla u_j|^2 \, dx + \mathcal{H}^{N-1}(S_{u_j}) + \|u_j\|_{L^p(\Omega)} \le C < +\infty$$

for every j. Then there exist a subsequence (not relabeled) and a function $u \in GSBV(\Omega) \cap L^p(\Omega)$ such that

$$\begin{aligned} & u_j(x) \to u(x) \ \textit{a.e. in } \Omega, \\ & \nabla u_j \rightharpoonup \nabla u \ \textit{weakly in } L^2(\Omega, \mathbb{R}^N), \\ & \mathcal{H}^{N-1}(S_u) \le \liminf_{j \to \infty} \mathcal{H}^{N-1}(S_{u_j}). \end{aligned} \tag{18}$$

Moreover

$$\int_{S_u} |\langle \nu_u, \xi \rangle| \, d\mathcal{H}^{N-1} \le \liminf_{j \to \infty} \int_{S_{u_j}} \left|\left\langle \nu_{u_j}, \xi \right\rangle\right| \, d\mathcal{H}^{N-1} \tag{19}$$

for every $\xi \in \mathbb{S}^{N-1}$.

There exist variants of this theorem, with different proofs (see [3, 4, 5]). We need however in these lectures to consider this version, since the conclusion (19) will be useful in order to study the anisotropic variants of the Mumford–Shah functional that appear in the finite differences discretizations that are common in image processing.

Remark. By a standard diagonalization technique Theorem 6 also holds if u_j and u are only in $GSBV_{loc}(\Omega)$.

In this setting we are now able to show the existence of the weak Mumford–Shah functional of Ambrosio and De Giorgi (*cf* section 2.3.1). However, first we end this section on functions with bounded variation with a paragraph about some useful additional properties.

2.2.8 Slicing

We now explain how a (special) bounded variation function can be described and its properties recovered from its 1-dimensional "slices", i.e., its restrictions to 1-dimensional lines. Many results of the sections 2.2.2–2.2.4 can be extended to the N–dimensional case using the following properties. In fact, most of Theorem 6 (in the case $p = \infty$) can be recovered from Theorem 3 and Theorem 5 in this way, the very difficult part being to show that $\nabla u_j \rightharpoonup \nabla u$. Many of the following results will be needed in order to study the variational approximations of the Mumford–Shah functional.

We consider for $\xi \in \mathbb{S}^{N-1}$ the sets $\xi^\perp = \{x \in \mathbb{R}^N \,:\, \langle \xi, x\rangle = 0\}$ and for any $z \in \xi^\perp$, $\Omega_{z,\xi} = \{t \in \mathbb{R} \,:\, z + t\xi \in \Omega\}$. On $\Omega_{z,\xi}$ we define a function $u_{z,\xi} : \Omega_{z,\xi} \to [-\infty, +\infty]$ by $u_{z,\xi}(s) = u(z + s\xi)$. If $u \in BV(\Omega)$, we have the following classical representation (see for instance [2, 7]): for $\mathcal{H}^{N-1}$-a.e. $z \in \xi^\perp$, $u_{z,\xi} \in BV(\Omega_{z,\xi})$ and for any Borel set $B \subseteq \Omega$

$$\langle Du, \xi\rangle(B) \;=\; \langle Du(B), \xi\rangle \;=\; \int_{\xi^\perp} d\mathcal{H}^{N-1}(z) Du_{z,\xi}(B_{z,\xi})$$

where $B_{z,\xi}$ is defined in the same way as $\Omega_{z,\xi}$; conversely if $u_{z,\xi} \in BV(\Omega_{z,\xi})$ for at least N independent vectors $\xi \in \mathbb{S}^{N-1}$ and $\mathcal{H}^{N-1}$-a.e. $z \in \xi^\perp$, and if

$$\int_{\xi^\perp} d\mathcal{H}^{N-1}(z)|Du_{z,\xi}|(\Omega_{z,\xi}) \;<\; +\infty$$

then $u \in BV(\Omega)$. Now (see [3, 2]), if $u \in SBV_{loc}(\Omega)$, then for almost every $z \in \xi^\perp$, $u_{z,\xi} \in SBV_{loc}(\Omega_{z,\xi})$ (the converse is true provided this property is satisfied for at least N independent vectors ξ and u has locally bounded variation), and the approximate derivative satisfies

$$\dot{u}_{z,\xi}(s) \;=\; \langle \nabla u(z + s\xi), \xi\rangle$$

for a.e. $s \in \Omega_{z,\xi}$, moreover

$$S_{u_{z,\xi}} \;=\; \{s \in \Omega_{z,\xi} \;:\; z + s\xi \in S_u\}\,,$$

$$(u_{z,\xi})_\pm(s) \;=\; u_\pm(z+s\xi) \;\; \forall s \in S_{u_{z,\xi}},$$

and for any Borel set $B \subseteq \Omega$

$$\int_{\xi^\perp} d\mathcal{H}^{N-1}(z)\mathcal{H}^0(B_{z,\xi}\cap S_{u_{z,\xi}}) \;=\; \int_{B\cap S_u} |\langle \nu_u(x),\xi\rangle|\, d\mathcal{H}^{N-1}(x).$$

The reader interested in knowing more about the space BV and how the results in this section are proved should consult, for instance, the books [6, 34, 36, 42, 62].

2.3 Back to the Mumford–Shah functional

2.3.1 Existence for the weak formulation

Now, in this setting, it is clear that the weak Mumford–Shah functional (7) has a minimum in $GSBV(\Omega)$ (which, in fact, is in $SBV(\Omega)$). Indeed, consider a minimizing sequence $(u_j)_{j\geq 1}$ for the problem $\inf_u E(u)$, in $GSBV(\Omega)$. Then, this sequence satisfies the conditions of Theorem 6 (with $p = 2$). Therefore, some subsequence (still denoted by u_j) converges almost everywhere to a function $u \in GSBV(\Omega)$ with

$$\int_\Omega |\nabla u(x)|^2\, dx \;\leq\; \liminf_{j\to\infty} \int_\Omega |\nabla u_j(x)|^2\, dx$$

(since ∇u_j goes to ∇u weakly in $L^2(\Omega)$, by (18)),

$$\mathcal{H}^{N-1}(S_u) \;\leq\; \liminf_{j\to\infty} \mathcal{H}^{N-1}(S_{u_j}), \text{ and}$$

$$\int_\Omega |u(x)-g(x)|^2\, dx \;\leq\; \liminf_{j\to\infty} \int_\Omega |u_j(x)-g(x)|^2\, dx$$

(by Fatou's lemma). Therefore $E(u) \leq \liminf_{j\to\infty} E(u_j)$ and u is a minimizer for the weak Mumford–Shah functional. Notice that since g is bounded, we can always replace u with its truncation at level $\|g\|_\infty$, $(-\|g\|_\infty \vee u(x)) \wedge \|g\|_\infty$, and decrease the energy, so that the minimum u has to satisfy $\|u\|_\infty \leq \|g\|_\infty$ and is in $SBV(\Omega)\cap L^\infty(\Omega)$.

In the next section we will explain how the weak problem is then related to the strong original one (that is, the minimization of $\mathcal{E}(u,K)$).

2.3.2 From the weak to the strong formulation

Once we have proved the existence of a minimizer for the weak Mumford–Shah energy $E(u)$ using Theorem 6, we need to show that it can also be considered as a minimizer for the original energy $\mathcal{E}(u, K)$ defined by (6). In arbitrary dimension N the general definition for $\mathcal{E}$ is

$$\mathcal{E}(u,K) \;=\; \int_{\Omega\setminus K} |\nabla u(x)|^2\,dx \;+\; \mathcal{H}^{N-1}(K\cap\Omega) + \int_\Omega |u(x)-g(x)|^2\,dx,$$

where the length has been replaced with the $(N-1)$–dimensional Hausdorff measure. (We have also dropped the constant parameters $\lambda, \mu\nu$.)

The natural way to associate a set K to $u \in SBV(\Omega)$ is to set $K = \overline{S_u}$. However, if u is arbitrary, we could have $\mathcal{H}^{N-1}(K\cap\Omega) > \mathcal{H}^{N-1}(S_u)$. For instance, the function

$$v(x) \;=\; \sum_{k=1}^{\infty} \frac{1}{2^k}\chi_{B_{2^{-k}}(x_k)},$$

where the sequence $(x_k)_{k\geq 1}$ is the set of all points in Ω with rational coordinates, is such that $S_v = \Omega \cap \cup_{k=1}^{\infty} \partial B_{2^{-k}}(x_k)$. This has finite length, but is dense in Ω. Thus $\mathcal{H}^{N-1}(\overline{S_v}\cap\Omega) = +\infty$.

A minimizer u of $E(u)$ will be a minimizer for $\mathcal{E}$ if and only if we can prove that $\mathcal{H}^{N-1}(\overline{S_u}\cap\Omega) = \mathcal{H}^{N-1}(S_u)$. (Conversely, it is "simple" to show that if $u \in H^1(\Omega\setminus K)$ and K is a closed set with $\mathcal{H}^{N-1}(K) < +\infty$, then $u \in SBV(\Omega)$ and $\mathcal{H}^{N-1}(S_u \setminus K) = 0$, that is, S_u is included in K up to a $\mathcal{H}^{N-1}$–negligible set.)

This difficult result was proved by De Giorgi, Carriero and Leaci [29], and independently in dimension $N = 2$ by Dal Maso, Morel and Solimini [28] (see also the book [48] for a general overview of the problem). They proved that if u minimizes $E(u)$, then $\mathcal{H}^{N-1}(\Omega\cap\overline{S_u}\setminus S_u) = 0$ and $u \in C^1(\Omega\setminus\overline{S_u})$, so that $(u, \overline{S_u})$ minimizes $\mathcal{E}$.

We make here the observation that if we slightly change the problem, introducing an anisotropy in the energy, then these results still hold. Indeed, if we consider a weak functional

$$E'(u) = \int_\Omega Q(\nabla u(x))\,dx \;+\; \int_{S_u} N(\nu_u(x))\,d\mathcal{H}^{N-1}(x), + \int_\Omega |u(x)-g(x)|\,dx, \tag{20}$$

where Q is a positive definite quadratic form in $\mathbb{R}^N$ and N is a norm in $\mathbb{R}^N$ (a 1-homogeneous convex function with $0 < \min_{\xi\in\mathbb{S}^{N-1}} N(\xi) \leq \max_{\xi\in\mathbb{S}^{N-1}} N(\xi)$

$< +\infty$), then E' has a minimizer u in $GSBV(\Omega)$ (exercise, you need to use inequality (19) in Theorem 6), moreover, it is possible to adapt the proofs in [29] and show that $\mathcal{H}^{N-1}(\Omega \cap \overline{S_u} \setminus S_u) = 0$ and $u \in C^1(\Omega \setminus \overline{S_u})$.

2.4 Variational approximations and Γ–convergence

In these lectures we will describe a few ways to approximate the Mumford–Shah problem, or variants of this problem. This has to be done because numerically, it is difficult to deal with a jump set K. We introduce in this part a special notion of convergence that is adapted to variational problems. As a matter of fact, if you are looking for the minimizer of a function $F(x)$, $x \in X$ (where X is some space), and want to approximate it with minimizers (x_n) of approximate problems $\min_{x\in X} F_n(x)$, then when can you say that (x_n) converge to a minimizer of F? If you consider the classical notions of limits of functions, then only the uniform convergence seems suitable to handle this problem. However, this notion of convergence is far too strong for most applications. This motivates the introduction of the following definition of Γ–*convergence*, specially invented for studying the limit of variational problems.

We will limit ourselves to the case where X is a metric space. For more details we refer to [27].

Given a metric space (X, d) and $F_k : X \rightarrow [-\infty, +\infty]$ a sequence of functions, we define for every $u \in X$ the Γ-*lim inf* of F

$$F'(u) = \Gamma - \liminf_{k\to\infty} F_k(u) = \inf_{u_k \to u} \liminf_{k\to\infty} F_k(u_k)$$

and the Γ-*lim sup* of F

$$F''(u) = \Gamma - \limsup_{k\to\infty} F_k(u) = \inf_{u_k \to u} \limsup_{k\to\infty} F_k(u_k),$$

and we say that F_k Γ–converges to $F : X \rightarrow [-\infty, +\infty]$ if $F' = F'' = F$. F', F'', and F (if they exist) are lower semi-continuous on X. We have the following two properties:

1. *F_k Γ–converges to F if and only if for every $u \in X$,*

(i) for every sequence u_k converging to u, $F(u) \leq \liminf_{k\to\infty} F_k(u_k)$;

(ii) there exists a sequence u_k that converges to u and such that $\limsup_{k\to\infty} F_k(u_k) \leq F(u)$;

2. *If $G : X \to \mathbb{R}$ is continuous and F_k Γ–converges to F, then $F_k + G$ Γ–converges to $F + G$.*

The following result makes clear the interest of the notion of Γ–convergence:

Theorem 7 *Assume F_k Γ–converges to F and for every k let u_k be a minimizer of F_k over X. Then, if the sequence (or a subsequence) u_k converges to some $u \in X$, u is a minimizer for F and $F_k(u_k)$ converges to $F(u)$.*

Eventually, we give the following definition of Γ–convergence in the case where $(F_h)_{h>0}$ is a family of functionals on X indexed by a continuous parameter h: we say that F_h Γ–converges to F in X as $h \downarrow 0$ if and only if for every sequence (h_j) that converges to zero as $j \to \infty$, F_{h_j} Γ–converges to F.

The reader who would like to know more about the Γ–convergence may consult the books [9, 27]. Also, the excellent notes [1] by G. Alberti contain a good introduction to this theory as well as to the applications to phase transition problems, that are very close (at least technically) to the methods and techniques of section 4.

3 The numerical analysis of the total variation minimization

3.1 The discrete energy

Let us consider problem (2), in dimension 2, and let us try to find a way to compute a solution. We will discuss the approach studied by Vogel and Oman [60, 61] (see also [24, 31]).

Although it is not absolutely obvious we will first assume that there exists a Lagrange multiplier $\lambda > 0$ such that problem (2) is equivalent to the problem

$$\min_{u \in BV(\Omega)} |Du|(\Omega) \,+\, \lambda \int_\Omega |Au(x) - g(x)|^2 \, dx \tag{21}$$

(see [24] for details, we must assume here $A1 = 1$, so that a minimizer of (21) automatically satisfies $\int_\Omega Au = \int g$, as well as the other assumptions of section 2.1.3). The problem of determining the correct λ is also difficult, we will not consider it in this short section.

First we must discretize (21). For simplicity we assume that u and g are discretized on the same square lattice, $i, j = 1, \cdots, L$. (This is the case in some situations, but there exist other common situations like the reconstruction of tomographic data, or the zooming, where it is not true.) The functions u and g are approximated by discrete matrices $U = (U_{i,j})_{1\le i,j\le L}$ and $G = (G_{i,j})_{1\le i,j\le L}$. The term $\lambda\int_\Omega |Au(x) - g(x)|^2\,dx$ is replaced, in the discrete setting, by a term $\lambda\sum_{i,j}|(AU)_{i,j} - g_{i,j}|^2$. (We omit the scale factor, but it is important in the practical applications.) In the discrete formula A denotes a linear operator of $\mathbb{R}^N = \mathbb{R}^{L\times L}$ (we set $N = L^2$) and $(AU)_{i,j}$ is the component i,j of AU.

There are several ways to approximate $|Du|(\Omega)$. The simplest (which, however, has several drawbacks), is to consider the variation along the horizontal and vertical directions

$$\sum_{1\le i<L}\sum_{1\le j\le L}|U_{i+1,j} - U_{i,j}| \;+\; \sum_{1\le i\le L}\sum_{1\le j<L}|U_{i,j+1} - U_{i,j}|$$

(here again we omitted the scale factor). Due to the strong anisotropy of this approximation the results it gives are not very good (in fact it is an approximation of $|D_1u|(\Omega) + |D_2u|(\Omega)$ —where in $\mathbb{R}^2$ D_iu is the derivative of u along the ith direction, $i = 1, 2$— which is a semi–norm in $BV(\Omega)$ that is equivalent to $|Du|(\Omega)$ but not invariant under rotations in the plane), and many other authors try to consider a "more isotropic" approximation of the total variation (see for instance [60, 33, 47].

Therefore the discrete energy we need to minimize is the following

$$E(U) \;=\; \sum_{i,j}\left(|U_{i+1,j} - U_{i,j}| \;+\; |U_{i,j+1} - U_{i,j}|\right) + \lambda\sum_{i,j}|(AU)_{i,j} - g_{i,j}|^2. \quad (22)$$

3.2 The method

Due to the strong nonlinearity of (22) (or rather of its derivative D_UE), it is difficult (although feasible) to minimize it by a straightforward gradient descent method. The nonexistence of the derivative of the absolute value $|x|$ at $x = 0$ (a problem that is often overcome by replacing $|x|$ with $\sqrt{\beta + x^2}$, with β a small parameter) is not the only difficulty. Another approach would be to define a dual problem, using convex duality. In the one–dimensional case and when A is the identity, it leads to a very simple and efficient algorithm, but in other situations it is not very practical. (If you know a bit of convex analysis you may think about it!) For a similar approach see [25].

The solution we will study here is common in the image processing literature (see [10, 15, 38, 39, 40, 60, 61]). It is closely related to the method we will use in section 6, or to the approach in section 4.

It consists in noticing that for every $x \in \mathbb{R}$, $x \neq 0$,

$$|x| \;=\; \min_{v>0} \left(\frac{v}{2} x^2 \;+\; \frac{1}{2v} \right),$$

the minimum being reached for $v = 1/|x|$. We thus introduce the function $f(x,v) = vx^2/2 + 1/(2v)$, a new field $V = (V_{i+1/2,j})_{1\le i<L, 1\le j\le L} \cup (V_{i,j+1/2})_{1\le i\le L, 1\le j<L} \in \mathbb{R}_+^{(L-1)\times L + L\times(L-1)}$ (of positive real numbers) and a new energy,

$$F(U,V) \;=$$
$$\sum_{i,j} \Big(f(|U_{i+1,j} - U_{i,j}|, V_{i+\frac12,j}) \;+\; f(|U_{i,j+1} - U_{i,j}|, V_{i,j+\frac12}) \Big)$$
$$+ \lambda \sum_{i,j} |(AU)_{i,j} - g_{i,j}|^2 = \; \sum_{i,j} \left(\frac12 V_{i+\frac12,j} |U_{i+1,j} - U_{i,j}|^2 \;+\; \frac12 V_{i,j+\frac12} |U_{i,j+1} - U_{i,j}|^2 \right.$$
$$\left. + \frac{1}{2V_{i+\frac12,j}} \;+\; \frac{1}{2V_{i,j+\frac12}} \right) \;+\; \lambda \sum_{i,j} |(AU)_{i,j} - g_{i,j}|^2$$

and we notice that

$$\min_V F(U,V) \;=\; E(U),$$

the minimum being reached for $V_{i+1/2,j} = 1/|U_{i+1,j} - U_{i,j}|$ (or at $+\infty$ if $U_{i+1,j} = U_{i,j}$) and $V_{i,j+1/2} = 1/|U_{i,j+1} - U_{i,j}|$.

We choose some starting values U^0, V^0 and compute for every $n \ge 1$

$$U^n \;=\; \arg\min_U F(U, V^{n-1})\,, \text{ and}$$

$$V^n \;=\; \arg\min_V F(U^n, V).$$

The idea is that as n becomes large, U^n will converge to the minimizer of (22). This is actually true if we slightly modify the algorithm (and the function E we minimize).

We choose an $\varepsilon > 0$ and introduce the convex closed set $K_\varepsilon = \{V \,:\, \varepsilon \le V_{i+1/2,j} \le 1/\varepsilon \text{ and } \varepsilon \le V_{i,j+1/2} \le 1/\varepsilon \;\forall i,j\}$ in $\mathbb{R}^M$ ($M = (L-1)\times L + L \times (L-1)$). We define a new energy $E_\varepsilon(U) = \min_{V\in K_\varepsilon} F(U,V)$. It is possible to show that as ε becomes small, the minimizer of E_ε approaches the minimizer of E. Moreover, it is easy to compute explicitly E_ε:

$$E_\varepsilon \;=\; \sum_{i,j} \left(j_\varepsilon(U_{i+1,j} - U_{i,j}) \;+\; j_\varepsilon(U_{i,j+1} - U_{i,j}) \right) \;+\; \lambda \sum_{i,j} |(AU)_{i,j} - g_{i,j}|^2$$

where

$$j_\varepsilon(x) = \min_{\varepsilon \le v \le 1/\varepsilon} f(x,v) = \begin{cases} \frac{1}{2\varepsilon}x^2 + \frac{\varepsilon}{2} & \text{if } |x| \le \varepsilon, \\ |x| & \text{if } \varepsilon \le |x| \le \frac{1}{\varepsilon}, \\ \frac{\varepsilon}{2}x^2 + \frac{1}{2\varepsilon} & \text{if } |x| \ge \frac{1}{\varepsilon}. \end{cases}$$

Define $\phi_\varepsilon(x) = (\varepsilon \vee 1/|x|) \wedge 1/\varepsilon$ ($= 1/|x|$ if $\varepsilon \le |x| \le 1/\varepsilon$, $1/\varepsilon$ if $|x| \le \varepsilon$, and ε if $|x| \ge 1/\varepsilon$). Then $\phi_\varepsilon(x)$ is the unique value in $[\varepsilon, 1/\varepsilon]$ such that $j_\varepsilon(x) = f(x, \phi_\varepsilon(x))$. We deduce that the unique $V \in K_\varepsilon$ for which $E_\varepsilon(U) = \min_{K_\varepsilon} F(U,\cdot) = F(U,V)$ is given by $V_{i+\frac{1}{2},j} = \phi_\varepsilon(x_{i+1,j} - x_{i,j})$ and $V_{i,j+\frac{1}{2}} = \phi_\varepsilon(x_{i,j+1} - x_{i,j})$ for every i,j. In this case we set $\Phi_\varepsilon(U) = V$ and this defines a continuous function $\Phi_\varepsilon : \mathbb{R}^N \to K_\varepsilon \subset \mathbb{R}^M$.

The algorithm, now, consists in computing for every $n \ge 1$, the starting values U^0, V^0 being chosen,

$$U^n = \arg\min_U F(U, V^{n-1})., \text{ and}$$

$$V^n = \arg\min_{V \in K_\varepsilon} F(U^n, V) = \Phi_\varepsilon(U^n).$$

3.3 Proof of the convergence of the algorithm

We assume (as in the continuous formulation) that $A1_N = 1_N$, where 1_N is the vector in $\mathbb{R}^N$ defined by $(1_N)_{i,j} = 1$ for every $1 \le i,j \le L$ (remember $N = L \times L$ is the dimension of the space where U lives). Then we have the following proposition.

Proposition 4 *There exist $\bar{U}$, $\bar{V} = \Phi_\varepsilon(\bar{U})$ such that as $n \to \infty$, $U^n \to \bar{U}$ and $V^n \to \bar{V}$, and $\bar{U}$ is a (the) minimizer of E_ε.*

Proof. First we claim that the following holds

Lemma 3 *There exist $0 < \alpha < \beta < +\infty$ such that the second derivatives $D^2_{UU}F$ and $D^2_{VV}F$ satisfy*

$$\alpha I_N \le D^2_{UU}F(U,V) \le \beta I_N \quad \textit{and} \quad \alpha I_M \le D^2_{VV}F(U,V) \le \beta I_M$$

for every $U \in \mathbb{R}^N$ and $V \in K_\varepsilon$.

This is equivalent to saying that for every $U \in \mathbb{R}^N$, $V \in K_\varepsilon$, $\xi \in \mathbb{R}^N$ and $\eta \in \mathbb{R}^M$, $\alpha|\xi|^2 \le \langle D^2_{UU}F(U,V)\xi, \xi\rangle \le \beta|\xi|^2$ and $\alpha|\eta|^2 \le \langle D^2_{VV}F(U,V)\eta, \eta\rangle \le \beta I_M|\eta|^2$. Here α and β both depend on ε.

Proof. We will leave to the reader the proof of three of the inequalities of the lemma and will prove the first one, which is the more difficult. We first recall the following "Poincaré inequality" (in finite dimension): there exists a constant $c > 0$ such that for every $\xi \in \mathbb{R}^N = \mathbb{R}^{L\times L}$ such that $\sum_{i,j} \xi_{i,j} = 0$,

$$\sum_{1\le i,j\le L} |\xi_{i,j}|^2 \;\le\; c \left(\sum_{1\le i<L,j} |\xi_{i+1,j} - \xi_{i,j}|^2 \;+\; \sum_{i,1\le j<L} |\xi_{i,j+1} - \xi_{i,j}|^2 \right). \quad (23)$$

Exercise. *Prove this inequality. Hint: suppose it is not true and consider a sequence ξ^n such that for every n, $\sum_{i,j} \xi^n_{i,j} = 0$ and $1 = \sum_{i,j} |\xi^n_{i,j}|^2 \ge n \sum_{i,j}(|\xi^n_{i+1,j} - \xi^n_{i,j}|^2 + |\xi^n_{i,j+1} - \xi^n_{i,j}|^2)$. Then, if ξ is the limit of a subsequence ξ^{n_k} of ξ^n, find a contradiction on ξ.*

Notice that for every U, $V \in K_\varepsilon$ and $\xi \in \mathbb{R}^N$,

$$\begin{aligned}\left\langle D^2_{UU}F(U,V)\xi,\xi\right\rangle \;&=\; \sum_{i,j} \left(V_{i+\frac12,j}|\xi_{i+1,j} - \xi_{i,j}|^2 \;+\; V_{i,j+\frac12}|\xi_{i,j+1} - \xi_{i,j}|^2\right) \;+\; |A\xi|^2 \\ &\ge\; \varepsilon \sum_{i,j} \left(|\xi_{i+1,j} - \xi_{i,j}|^2 \;+\; |\xi_{i,j+1} - \xi_{i,j}|^2\right) \;+\; |A\xi|^2\end{aligned}$$

In particular, letting $m(\xi) = (1/N)\sum_{i,j}\xi_{i,j}$ be the average of ξ, we have (since $A1_N = 1_N$) $\sqrt{\langle D^2_{UU}F(U,V)\xi,\xi\rangle} \ge |A\xi| = |A(\xi - m(\xi)1_N) + m(\xi)1_N| \ge |m(\xi)1_N| - |A||\xi - m(\xi)1_N|$. But by (23), $|\xi - m(\xi)1_N|^2 \le c\sum_{i,j}(|\xi_{i+1,j} - \xi_{i,j}|^2 + |\xi_{i,j+1} - \xi_{i,j}|^2) \le (1/\varepsilon)\langle D^2_{UU}F(U,V)\xi,\xi\rangle$, therefore $|m(\xi)1_N| \le c\sqrt{\langle D^2_{UU}F(U,V)\xi,\xi\rangle}$ (here c denotes any positive constant that does not depend on U,V,ξ). Moreover, using again (23), $c\langle D^2_{UU}F(U,V)\xi,\xi\rangle \ge |\xi - m(\xi)1_N|^2$. Since 1_N and $\xi - m(\xi)1_N$ are orthogonal we deduce that $|\xi|^2 \le c\langle D^2_{UU}F(U,V)\xi,\xi\rangle$. □

Remark. Observe the identity between this proof and the proof of the coerciveness of the energy in section 2.1.3.

We next prove the following lemma.

Lemma 4 *For every $n \ge 1$,*

$$E_\varepsilon(U^{n-1}) \;-\; E_\varepsilon(U^n) \;\ge\; \frac{\alpha}{2}\left(|U^{n-1} - U^n|^2 \;+\; |V^{n-1} - V^n|^2\right).$$

Proof. For every $n \ge 1$, we have $D_U F(U^n, V^{n-1}) = 0$ while

$$\langle D_V F(U^n, V^n), V - V^n\rangle \ge 0$$

for every $V \in K_\varepsilon$. We deduce that (using Lemma 3)

$$\begin{aligned} F(U^n, V^{n-1}) &= F(U^n, V^n) + \left\langle D_V F(U^n, V^n), V^{n-1} - V^n \right\rangle + \int_0^1 (1-t) \\ &\quad \times \left\langle D^2_{VV} F(U^n, V^n + t(V^{n-1} - V^n))(V^{n-1} - V^n), V^{n-1} - V^n \right\rangle dt \\ &\geq F(U^n, V^n) + \frac{\alpha}{2} |V^{n-1} - V^n|^2. \end{aligned}$$

In a similar way we prove that $F(U^{n-1}, V^{n-1}) \geq F(U^n, V^{n-1}) + \frac{\alpha}{2} |U^{n-1} - U^n|^2$. Since $E_\varepsilon(U^n) = F(U^n, V^n)$, the lemma is proved. □

Since by construction the sequence $E_\varepsilon(U^n) = F(U^n, V^n)$ must decrease and is bounded from below, it must have a limit in $\mathbb{R}_+$ and $E_\varepsilon(U^{n-1}) - E_\varepsilon(U^n) \to 0$, therefore $U^{n-1} - U^n$ and $V^{n-1} - V^n$ go to zero as $n \to \infty$.

Now, we notice that E_ε is coercive, which means that for every $c > 0$ the set $\{E_\varepsilon \leq c\}$ is bounded in $\mathbb{R}^N$ (and closed), hence compact (this can be deduced from Lemma 3). Thus we may extract a subsequence U^{n_k} and find a $\bar{U} \in \mathbb{R}^N$ such that as $k \to \infty$, $U^{n_k} \to \bar{U}$. By continuity $V^{n_k} = \Phi_\varepsilon(U^{n_k}) \to \Phi_\varepsilon(\bar{U})$, and we let $\bar{V} = \Phi_\varepsilon(\bar{U})$. We also have $D_U F(U^{n_k}, V^{n_k - 1}) = 0$ and since $V^{n_k - 1} - V^{n_k} \to 0$ by Lemma 4, $V^{n_k - 1} \to \bar{V}$ so that by continuity, $D_U F(\bar{U}, \bar{V}) = 0$.

We conclude using the following lemma

Lemma 5 *Let $\bar{U}, \bar{V}$ satisfy $D_U F(\bar{U}, \bar{V}) = 0$ and $\bar{V} = \arg\min_{V \in K_\varepsilon} F(\bar{U}, V) = \Phi_\varepsilon(\bar{U})$. Then $D_U E_\varepsilon(\bar{U}) = 0$.*

Proof. Let $h \in \mathbb{R}^N$ and $t > 0$. Letting $V_t = \Phi_\varepsilon(\bar{U} + th)$ (that goes to $\bar{V}$ as $t \downarrow 0$), we have

$$\begin{aligned} E_\varepsilon(\bar{U} + th) - E_\varepsilon(\bar{U}) &= F(\bar{U} + th, \Phi_\varepsilon(\bar{U} + th)) - F(\bar{U}, \bar{V}) \\ &= (F(\bar{U} + th, V_t) - F(\bar{U}, V_t)) \\ &\quad + (F(\bar{U}, V_t) - F(\bar{U}, \bar{V})) \end{aligned}$$

Since $V_t \in K_\varepsilon$, $F(\bar{U}, V_t) \geq F(\bar{U}, \bar{V}))$ so that $E_\varepsilon(\bar{U} + th) - E_\varepsilon(\bar{U}) \geq F(\bar{U} + th, V_t) - F(\bar{U}, V_t)$. Since $F(\bar{U} + th, V_t) - F(\bar{U}, V_t) = t \langle D_U F(\bar{U}, V_t), h \rangle + \int_0^t (t - s) \langle D^2_{UU} F(\bar{U}, V_t) h, h \rangle \, dt$ and $\int_0^t (t - s) \langle D^2_{UU} F(\bar{U}, V_t) h, h \rangle \, dt \leq \beta t^2 |h|^2 / 2$, we deduce that

$$\langle D_U E_\varepsilon(\bar{U}), h \rangle = \lim_{t \downarrow 0} \frac{E_\varepsilon(\bar{U} + th) - E_\varepsilon(\bar{U})}{t} \geq \langle D_U F(\bar{U}, \bar{V}), h \rangle = 0.$$

Since h is arbitrary, $D_U E_\varepsilon(\bar{U}) = 0$. □

Since E_ε is (C^1 and) strictly convex[2] (meaning that for every U, U' and $0 < \theta < 1$, $E_\varepsilon(\theta U + (1-\theta)U') < \theta E_\varepsilon(U) + (1-\theta)E_\varepsilon(U')$ unless $U = U'$), it has a unique minimizer characterized by the equation $D_U E = 0$. We deduce that $\bar{U}$ is the unique minimizer of E_ε. This achieves the proof of Proposition 4, since by uniqueness of this minimizer any subsequence of (U^n) must converge to the same value $\bar{U}$, so that the whole sequence (U^n) converges to $\bar{U}$. □

Exercise. *Prove that as $\varepsilon \to 0$, E_ε Γ–converges to E, so that the minimizer of E_ε tends to the minimizer of E.*

3.4 Two examples

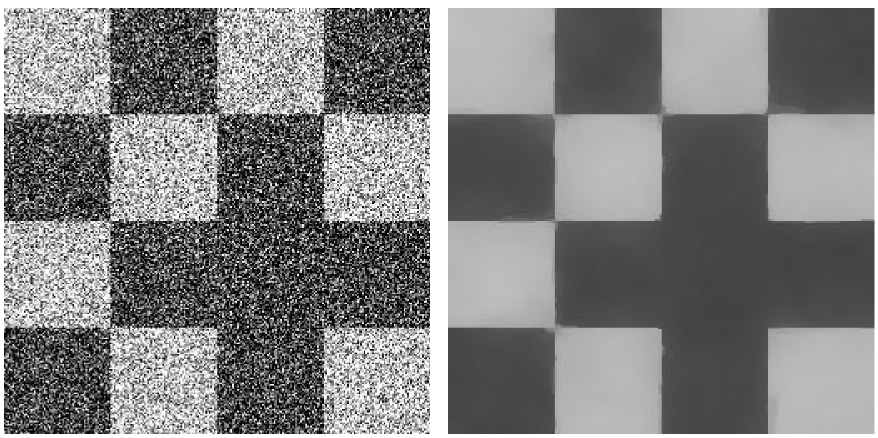

Figure 5: A noisy image and the reconstruction.

We just show two examples of image denoising (i.e., A is the identity matrix) obtained with this method (these are taken from [24]). The first one (Fig. 5) represents the reconstruction of a piecewise constant image on which a Gaussian noise (of standard deviation 60 for values between 0 and 255) has been added. On the left, the noisy image is presented, while on the right the reconstruction with an appropriate value of λ is shown. The second example (Fig. 6) shows a "true picture" that has been corrupted by some Gaussian noise (here the standard deviation is approximately 30, always for values between 0 and 255). The reconstruction (on the right) is less good,

[2]We have to assume that A is injective. Otherwise, $\bar{U}$ is still a minimizer of E_ε but might be not unique.

Figure 6: Another example of total variation minimization.

since the total variation tends to favor piecewise constant images, so that the result is a bit "blocky".

4 The numerical analysis of the Mumford–Shah problem (I)

4.1 Ambrosio and Tortorelli's approximate energy

Now we will describe the first attempt that has been made to provide an approximation of the Mumford–Shah functional by simpler (elliptic) variational problems. This result is due to Ambrosio and Tortorelli (see [7] and [8]). They have proposed to replace the set K by a function $v(x)$, and design energies that depend on a scale parameter ε, so that as ε goes to zero the function $1 - v(x)$, in some sense, becomes an approximation of the characteristic function of the discontinuity set K.

Their approximation $F_\varepsilon(u, v)$, defined over the space $L^2(\Omega) \times L^2(\Omega)$, is the following

$$\begin{aligned} F_\varepsilon(u,v) &= \int_\Omega (v(x)^2 + k_\varepsilon)|\nabla u(x)|^2\,dx + \int_\Omega \varepsilon|\nabla v(x)|^2 \\ &\quad + \frac{(1-v(x))^2}{4\varepsilon}\,dx + \int_\Omega |u(x)-g(x)|^2\,dx \end{aligned} \tag{24}$$

for $u, v \in H^1(\Omega)$, and they set $F(u, v) = +\infty$ if u or v is in $L^2(\Omega) \setminus H^1(\Omega)$. The parameter $k_\varepsilon > 0$ is needed in order to have that for $\varepsilon > 0$, F is coercive in $H^1(\Omega) \times H^1(\Omega)$ (i.e., greater than a constant times the norm of (u,v) in this space). It has to go to zero faster than ε as ε goes to zero (i.e., $\lim_{\varepsilon\downarrow 0} k_\varepsilon/\varepsilon = 0$), the reason for this will be made clear in the proof.

In order to show that F_ε Γ–converges to an energy such as the weak Mumford–Shah energy E, they need to redefine it on the same space as F_ε, that is, $L^2(\Omega) \times L^2(\Omega)$. To do this, they consider the fact that as ε goes to zero, they want v to become almost everywhere equal to 1, and thus they set for every $u, v \in L^2(\Omega)$:

$$F(u,v) = \begin{cases} \displaystyle\int_\Omega |\nabla u(x)|^2\,dx + \mathcal{H}^{N-1}(S_u) + \int_\Omega (u(x)-g(x))^2\,dx & \text{if } \begin{cases} u \in GSBV(\Omega), \\ v(x) = 1 \text{ a.e. in } \Omega, \end{cases} \\ +\infty & \text{otherwise.} \end{cases}$$

Notice that $F(u,v)$ is just $E(u)$ when $v = 1$ a.e., and $+\infty$ otherwise. Then, they are able to show the following theorem.

Theorem 8 (Ambrosio–Tortorelli) *As ε goes to zero, F_ε Γ–converge to F.*

In particular, this means that for small ε, the minimizers of F_ε will be close to minimizers of F.

This approximation has been used in efforts to compute image segmentations and for other application (see [16, 37, 52], that are based on a finite-element version of Theorem 8 established by Bellettini and Coscia [13], and [57, 56, 58, 59] where in particular Shah considers the gradient flow of $F_\varepsilon(u,v)$ and of similar energies). It works quite well, however, it has the feature that the approximation of the Mumford–Shah functional will be correct only if the discretization step (the pixels' size) is much smaller than the scale parameter ε. This is not very convenient for most applications. For this reason, we will study in the following sections the problem from a different point of view, considering the original discrete energies ($E(U,L)$ or $E(U)$) and explain how one can find the energy they approximate in the continuous setting (which will be a variant of the Mumford–Shah energy).

We quickly explain in one dimension how the proof of Theorem 8 goes. In dimension greater than one, the proof is obtained mostly through a localization and a slicing argument, that we will briefly mention in the subsequent section.

4.2 Sketch of the proof of Ambrosio and Tortorelli's theorem, in dimension one

Basically, in order to show Theorem 8, you have to choose any sequence $(\varepsilon_j)_{j\geq 1}$ of positive numbers with $\lim_{j\to\infty}\varepsilon_j = 0$, and to prove that for any (u,v),

i) if (u_j, v_j) goes to (u,v) (in L^2-norm) as $j\to\infty$, then $F(u,v) \leq \liminf_{j\to\infty} F_{\varepsilon_j}(u_j,v_j)$, and

ii) there exists (u_j,v_j) that converges to (u,v) such that $\limsup_{j\to\infty} F_{\varepsilon_j}(u_j,v_j) \leq F(u,v)$

4.2.1 Proof of (i)

To prove (i), we consider a sequence (u_j,v_j). We of course can assume that $\liminf_{j\to\infty} F_{\varepsilon_j}(u_j,v_j) < +\infty$ (otherwise there is nothing to prove), and we can extract a subsequence (unless really needed —i.e., if we need to keep the track of the original sequence or compare two different subsequences— we will always denote the subsequences like the original sequence) such that $\liminf_{j\to\infty} F_{\varepsilon_j}(u_j,v_j) = \lim_{j\to\infty} F_{\varepsilon_j}(u_j,v_j)$. In particular we have $c = \sup_j F_{\varepsilon_j}(u_j,v_j) < +\infty$, thus $\int_\Omega (1-v_j(x))^2\,dx \leq c\varepsilon_j \to 0$ as $j\to\infty$. Then, since v_j goes to 1 in $L^2(\Omega)$, we must have $v=1$ a.e.. We then just need to show that $E(u) \leq \liminf_{j\to\infty} F_{\varepsilon_j}(u_j,v_j)$. Since it is clear that $\int_\Omega (u_j(x)-g(x))^2\,dx$ goes to $\int_\Omega (u(x)-g(x))^2\,dx$ as $j\to\infty$, we have to show that $u \in GSBV(\Omega)$ and estimate the other two terms, $\mathcal{H}^0(S_u) = \sharp S_u$ (notice that the measure $\mathcal{H}^0(S_u)$ is the cardinality of the set S_u, that we also denote by $\sharp S_u$), and $\int_\Omega \dot{u}(x)^2\,dx$.

Now let Σ be the set of points $x\in\Omega$ such that for every $\delta>0$ (small, so that $(x-\delta,x+\delta)\in\Omega$), $u \notin H^1(x-\delta,x+\delta)$. We will show that this set is finite. Indeed, choose $x_1,\dots,x_k \in \Sigma$ with $x_1 < x_2 < \cdots < x_k$. Choose also δ such that $x_i+\delta < x_{i+1}-\delta$ for every $i=1,\dots,k-1$ and $(x_i-\delta,x_i+\delta)\subset\Omega$ for every i.

Then, we must have for every i that $\limsup_{j\to\infty}\inf_{B_{\delta/2}(x_i)} v_j = 0$. Otherwise, there exists i and a subsequence u_{j_k}, v_{j_k} such that $\lim_{k\to\infty}\inf_{B_{\delta/2}(x_i)} v_{j_k} = \alpha > 0$. And $v_{j_k} \geq \alpha/2$ in $B_{\delta/2}(x_i)$ for k large enough. But then, $\int_{x_i-\delta/2}^{x_i+\delta/2}(v_{j_k} + k_{\varepsilon_{j_k}})u'_{j_k}(x)^2\,dx \leq c$ implies that $\int_{x_i-\delta/2}^{x_i+\delta/2} u'_{j_k}(x)^2\,dx \leq 2c/\alpha$, so that u_{j_k} is uniformly bounded in $H^1(x_i-\delta/2, x_i+\delta/2)$. In this case, its limit u also has to be in $H^1(x_i-\delta/2,x_i+\delta/2)$, and this is in contradiction with the choice of $x_i\in\Sigma$.

Now, for every i we know that $\limsup_{j\to\infty}\inf_{B_{\delta/2}(x_i)} v_j = 0$. We also know that $v_j\to 1$ in $L^2(\Omega)$. Therefore if we fix i and choose $\eta>0$ small, there will exist for j large enough a point $x_i(j)\in(x_i-\delta/2,x_i+\delta/2)$ with $v_j(x_i(j))<\eta$, and $x_i'(j)\in(x_i-\delta,x_i-\delta/2)$ and $x_i''(j)\in(x_i+\delta/2,x_i+\delta)$ with $v_j(x_i'(j))>1-\eta$ and $v_j(x_i''(j))>1-\eta$. We then have (using the fact that $A^2+B^2\geq 2AB$)

$$\begin{aligned}\int_{x_i-\delta}^{x_i+\delta}\varepsilon_j v_j'(x)^2+\frac{\left(1-v_j(x)\right)^2}{4\varepsilon_j}\,dx &\geq \int_{x_i-\delta}^{x_i+\delta}|v_j'(x)||1-v_j(x)|\,dx\\ &\geq \int_{x_i'(j)}^{x_i(j)}|1-v_j(x)||v_j'(x)|\,dx\\ &\quad+\int_{x_i(j)}^{x_i''(j)}|1-v_j(x)||v_j'(x)|\,dx\\ &\geq \frac{(1-\eta)^2-\eta^2}{2}+\frac{(1-\eta)^2-\eta^2}{2} = 1-2\eta.\end{aligned}$$

In particular, we get that for j large enough $\int_\Omega \varepsilon_j v_j'(x)^2+\frac{(1-v_j(x))^2}{4\varepsilon_j}\,dx\geq(1-2\eta)k$.

Since this is valid for an arbitrary finite subset $\{x_1,\dots,x_k\}$ of Σ, it shows that $\sharp\Sigma<+\infty$, and more precisely,

$$\sharp\Sigma(1-2\eta)\leq\liminf_{j\to\infty}\int_\Omega\varepsilon_j v_j'(x)^2+\frac{(1-v_j(x))^2}{4\varepsilon_j}\,dx\leq c=\sup_j F_{\varepsilon_j}(u_j,v_j)<+\infty.$$

This is true for every $\eta>0$, so that eventually

$$\sharp\Sigma\leq\liminf_{j\to\infty}\int_\Omega\varepsilon_j v_j'(x)^2+\frac{(1-v_j(x))^2}{4\varepsilon_j}\,dx$$

holds.

Now, by the definition of Σ, $u\in H^1_{loc}(\Omega\setminus\Sigma)$, and we need to find an estimate for $\int_{\Omega\setminus\Sigma}u'(x)^2\,dx$. Indeed, if we knew that $\int_{\Omega\setminus\Sigma}u'(x)^2\,dx<\infty$ (i.e., $u\in H^1(\Omega\setminus\Sigma)$), then it would yield $u\in SBV(\Omega)$ and $S_u=\Sigma$, $\dot{u}=u'$.

Notice that the proof we have just written could easily be transformed to lead to the following lemma:

Lemma 6 *Let $\eta>0$. Then, for every $\delta>0$, there exists J such that for every $j\geq J$, if $x_1<x_2<\cdots<x_k$ are such that $v_j(x_i)<\eta$ and $x_{i+1}-x_i>\delta$, then $k\leq c/(1-2\eta)$.*

We leave the proof of this lemma to the reader (use the same arguments as in the proof above, after having chosen J such that if $j \geq J$, $|\{x : v_j(x) < 1-\eta\}| < \delta$).

In this case, once $\eta > 0$ is chosen, we can choose $\delta > 0$ and select for all $j \geq J$ a maximal set $x_1(j) < x_2(j) < \cdot < x_{k(j)}(j)$, with $x_{i+1}(j) - x_i(j) > \delta$ and $v_j(x_i(j)) < \eta$, and we have $k(j) \leq c/(1-2\eta)$. Therefore, there exist $k \leq c/(1-2\eta)$, a subsequence (u_{j_l}, v_{j_l}), and k points $x_1 < x_2 < \cdots < x_k$, such that $k(j_l) = k$ for all l and $x_i(j_l) \to x_i$ as $l \to \infty$ for every $i = 1, \ldots, k$.

If l is large enough we thus have (by the maximality of the set $\{x_1(j), \ldots, x_{k(j)}(j)\}$) that $v_{j_l} \geq \eta$ in the open set $\Omega^\delta = \Omega \setminus \cup_{i=1}^k [x_i - 2\delta, x_i + 2\delta]$, so that $\int_{\Omega^\delta} u'_{j_l}(x)^2\,dx \leq c/\eta$ and since u_{j_l} goes to u in $L^2(\Omega)$ we know that this implies the convergence of u'_{j_l} to u' weakly in $L^2(\Omega^\delta)$.

If $w \in L^2(\Omega^\delta)$, the functions $w\sqrt{k_{\varepsilon_{j_l}} + v_{j_l}^2}$ go to w strongly in $L^2(\Omega^\delta)$, since $v_{j_l} \to 1$ in $L^2(\Omega)$.

Thus, $\displaystyle\lim_{l\to\infty} \int_{\Omega^\delta} \sqrt{k_{\varepsilon_{j_l}} + v_{j_l}(x)^2}\, u'_{j_l}(x) w(x)\,dx =$

$$\lim_{l\to\infty} \int_{\Omega^\delta} u'_{j_l}(x) \left(w(x) \sqrt{k_{\varepsilon_{j_l}} + v_{j_l}(x)^2} \right) dx = \int_{\Omega^\delta} u'(x) w(x)\,dx$$

and $\sqrt{k_{\varepsilon_{j_l}} + v_{j_l}(x)^2}\, u'_{j_l}(x)$ goes weakly to u' in $L^2(\Omega^\delta)$. This yields

$$\int_{\Omega^\delta} u'(x)^2\,dx \leq \liminf_{l\to\infty} \int_{\Omega^\delta} (k_{\varepsilon_{j_l}} + v_{j_l}(x)^2) u'_{j_l}(x)^2\,dx \leq c < +\infty.$$

Since δ is arbitrary we can deduce that

$$\int_\Omega u'(x)^2\,dx \leq \liminf_{j\to\infty} \int_\Omega (k_{\varepsilon_j} + v_j(x)^2) u'_j(x)^2\,dx$$

and in particular $u \in H^1(\Omega \setminus \Sigma)$. Therefore $u \in SBV(\Omega)$, $S_u = \Sigma$, $\dot{u} = u'$, and

$$\int_\Omega \dot{u}(x)^2\,dx + \sharp\Sigma + \int_\Omega (u(x) - g(x))^2\,dx \leq$$

$$\leq \liminf_{j\to\infty} \int_\Omega (k_{\varepsilon_j} + v_j(x)^2) u'_j(x)^2\,dx + \int_\Omega \varepsilon_j v'_j(x)^2 + \frac{(1-v_j(x))^2}{4\varepsilon_j}\,dx + \int_\Omega (u_j(x) - g(x))^2\,dx$$

which was the thesis we wanted to prove, and (i) is true.

4.2.2 Proof of (ii)

To prove (ii), we consider $u \in SBV(\Omega)$ with $E(u) = F(u,1) < +\infty$ (otherwise there is nothing to prove). In this case, S_u is a finite set, and $u \in H^1(\Omega \setminus S_u)$ (in particular it is continuous everywhere but in S_u).

In order to simplify the proof we will consider that $\Omega = (-1,1)$ and that u has just one jump at point 0, but the study of the general case can be localized in small intervals around each discontinuity of u so that it is not very different.

Consider the function $\gamma(t) = 1 - \exp(-t/2)$ for $t \geq 0$. We leave the following result to the reader:

Exercise. *Prove that* $v_\varepsilon(x) = \gamma(x/\varepsilon)$ *minimizes*

$$\int_0^\infty \varepsilon v'(x)^2 + \frac{(1-v(x))^2}{4\varepsilon}\,dx$$

on the set $\{v \in L^2_{loc}(0,+\infty), v' \in L^2(0,+\infty), v(0) = 0\}$, *and that the value of the minimum is* $1/2$.

Now, we set for every $\varepsilon > 0$ $v_\varepsilon(x) = 0$ if $|x| < a_\varepsilon$, and $v_\varepsilon(x) = \gamma\left(\frac{|x|-a_\varepsilon}{\varepsilon}\right)$ otherwise, where a_ε goes to zero with ε and will be fixed later on, and we set $u_\varepsilon(x) = u(-a_\varepsilon) + \frac{u(a_\varepsilon)-u(-a_\varepsilon)}{2a_\varepsilon}(x+a_\varepsilon)$ if $|x| < a_\varepsilon$, and $u_\varepsilon(x) = u(x)$ otherwise.

Then (using the result of the previous exercise),

$$\begin{aligned}
E_\varepsilon(u_\varepsilon, v_\varepsilon) &= \int_{|x|\geq a_\varepsilon} (k_\varepsilon + v_\varepsilon(x))u'(x)^2\,dx \;+\; 2a_\varepsilon k_\varepsilon \left(\frac{u(a_\varepsilon) - u(-a_\varepsilon)}{2a_\varepsilon}\right)^2 \\
&\quad + \int_{|x|\geq a_\varepsilon} \frac{1}{\varepsilon}\gamma'\left(\frac{|x|-a_\varepsilon}{\varepsilon}\right)^2 + \frac{1}{4\varepsilon}\left(1 - \gamma\left(\frac{|x|-a_\varepsilon}{\varepsilon}\right)\right)^2 dx \\
&\quad + \frac{2a_\varepsilon}{4\varepsilon} + \int_{|x|\geq a_\varepsilon} (u(x) - g(x))^2\,dx \;+\; \int_{-a_\varepsilon}^{a_\varepsilon} (u_\varepsilon(x) - g(x))^2\,dx \\
&\leq (1 + k_\varepsilon)\int_{-1}^{1} u'(x)^2\,dx \;+\; 2\|u\|_\infty^2 \frac{k_\varepsilon}{a_\varepsilon} \\
&\quad + 2 \times \frac{1}{2} + \frac{2a_\varepsilon}{4\varepsilon} \\
&\quad + \int_{-1}^{1} (u(x) - g(x))^2\,dx \;+\; 2a_\varepsilon(\|u\|_\infty + \|g\|_\infty)^2.
\end{aligned}$$

We see that we will get the result if both $k_\varepsilon/a_\varepsilon$ and $a_\varepsilon/\varepsilon$ go to zero as ε goes to zero. This is possible if and only if $k_\varepsilon/\varepsilon \to 0$, since in this case we can let $a_\varepsilon = \sqrt{k_\varepsilon \varepsilon}$. Then, we have $\limsup_{\varepsilon\downarrow 0} E_\varepsilon(u_\varepsilon, v_\varepsilon) \leq F(u)$ and point (ii) is proved.

4.3 Higher dimensions

In dimension greater than one, the proof is obtained through a localization and a slicing argument. The reader, if interested, should report himself to Ambrosio and Tortorelli's paper, to [19] where a similar argument is used, or to Braides' book [18]. We will briefly explain, without giving too many details, how the proof goes.

4.3.1 The first inequality

Consider a sequence $\varepsilon_j \downarrow 0$ and u^j, v^j such that $u^j \to u$ and $v^j \to v$ strongly in $L^2(\Omega)$ as $j \to \infty$. Again, it is clear that $v = 1$ a.e., and we need to show that $E(u) \le \liminf_{j\to\infty} F_{\varepsilon_j}(u^j, v^j)$. To prove this inequality we first *localize* the energy $F_{\varepsilon_j}(u^j, v^j)$ by letting, for every $A \subseteq \Omega$ open,

$$F_{\varepsilon_j}(u^j, v^j, A) = \int_A \left((v^j(x)^2 + k_{\varepsilon_j})|\nabla u^j(x)|^2 + \varepsilon_j |\nabla v^j(x)|^2 + \frac{(1 - v^j(x))^2}{4\varepsilon_j} + |u^j(x) - g(x)|^2 \right) dx. \tag{25}$$

Then, we fix an open set A and a unit vector $\xi \in \mathbb{S}^{N-1}$, and write

$$\begin{aligned} F_{\varepsilon_j}(u^j, v^j, A) = & \int_{\xi^\perp} d\mathcal{H}^{N-1}(z) \int_{A_{z,\xi}} dt \Big((v^j(z + t\xi)^2 + k_{\varepsilon_j})|\nabla u^j(z + t\xi)|^2 \\ & + \varepsilon_j |\nabla v^j(z + t\xi)|^2 + \frac{(1 - v^j(z + t\xi))^2}{4\varepsilon_j} \\ & + |u^j(z + t\xi) - g(z + t\xi)|^2 \Big) \end{aligned}$$

so that

$$F_{\varepsilon_j}(u^j, v^j, A) \ge \int_{\xi^\perp} F^{1D}_{\varepsilon_j}(u^j_{z,\xi}, v^j_{z,\xi}, A_{z,\xi}, g_{z,\xi})\, d\mathcal{H}^{N-1}(z).$$

(We follow the notations in section 2.2.8: $A_{z,\xi} = \{t \in \mathbb{R} : z + t\xi \in A\}$, and for every $t \in A_{z,\xi}$, $u^j_{z,\xi}(t) = u^j(z + t\xi)$, $v^j_{z,\xi}(t) = v^j(z + t\xi)$, $g_{z,\xi}(t) = g(z + t\xi)$.) Here $F^{1D}_{\varepsilon_j}$ denotes the localized Ambrosio–Tortorelli energy (25) in dimension 1 (given $I \subset \mathbb{R}$ an open set and $w, r \in H^1(I)$, $h \in L^\infty(I)$):

$$\begin{aligned} F^{1D}_{\varepsilon_j}(w, r, I, h) = & \int_I (r(t)^2 + k_{\varepsilon_j}) w'(t)^2\, dt + \int_I \varepsilon_j r'(t)^2 + \frac{(1 - r(t))^2}{4\varepsilon_j}\, dt \\ & + \int_A |w(t) - h(t)|^2\, dt. \end{aligned}$$

Since as $j\to\infty$,

$$\int_\Omega |u^j(x)-u(x)|^2\,dx \;=\; \int_{\xi^\perp}\int_{\Omega_{z,\xi}} |u^j_{z,\xi}(t)-u_{z,\xi}(t)|\,dt d\mathcal{H}^{N-1}(z)\to 0$$

we may assume we have extracted a subsequence (not relabeled) such that $u^j_{z,\xi}\to u_{z,\xi}$ for almost every $z\in\xi^\perp$ (such that $\Omega_{z,\xi}\neq\emptyset$).

Then, the one–dimensional result states that for such a ξ,

$$E^{1D}(u_{z,\xi},A_{z,\xi},g_{z,\xi}) \le \liminf_{j\to\infty} F^{1D}_{\varepsilon_j}(u^j_{z,\xi},v^j_{z,\xi},A_{z,\xi},g_{z,\xi}),$$

where again if $I\subset\mathbb{R}$ is open,

$$E^{1D}(w,I,h) \;=\; \int_I \dot w(t)^2\,dt \;+\; \mathcal{H}^0(S_w\cap I) \;+\; \int_I (w(t)-h(t))^2\,dt$$

if $w\in SBV(I)$, and $E^{1D}(w,I,h)=+\infty$ otherwise. Using Fatou's lemma, we deduce that

$$\int_{\xi^\perp}\left(\int_{A_{z,\xi}} \dot u_{z,\xi}(t)^2\,dt \;+\; \mathcal{H}^0(S_{u_{z,\xi}}\cap A_{z,\xi}) \;+\; \int_{A_{z,\xi}} (u_{z,\xi}(t)-g_{z,\xi}(t))^2\,dt\right) d\mathcal{H}^0(z)$$
$$\le\; \liminf_{j\to\infty} F_{\varepsilon_j}(u^j,v^j,A).$$

In particular we get that if $\liminf_{j\to\infty} F_{\varepsilon_j}(u^j,v^j,A)<+\infty$, $u_{z,\xi}\in SBV(A_{z,\xi})$ for a.e. every $z\in\xi^\perp$, and since this is true for every ξ we deduce that $u\in SBV(A)$. Thanks to the results of section 2.2.8, the last inequality can then be rewritten as

$$E_\xi(u,A) \;=\;$$
$$\int_A \langle\nabla u(x),\xi\rangle^2\,dx \;+\; \int_{S_u\cap A} |\langle\nu_u(x),\xi\rangle|\,d\mathcal{H}^{N-1}(x) \;+\; \int_A |u(x)-g(x)|^2\,dx$$
$$\le\; \liminf_{j\to\infty} F_{\varepsilon_j}(u^j,v^j,A).$$

To conclude, we admit that (see [19, Prop. 6.5]), if $(\xi_n)_{n\ge1}$ is a dense sequence of points in $\mathbb{S}^{N-1}$,

$$E(u) \;=\; \sup\left\{\sum_{n=1}^k E_{\xi_n}(u,A_n)\;:\;k\in\mathbb{N}, (A_n)_{n=1,\cdots,k} \text{ disjoint open subsets of } \Omega\right\}$$

and we observe that if $(A_n)_{n=1,\cdots,k}$ is such a family, then

$$\sum_{n=1}^k E_{\xi_n}(u,A_n) \;\le\; \sum_{n=1}^k \liminf_{j\to\infty} F_{\varepsilon_j}(u^j,v^j,A_n) \;\le\; \liminf_{j\to\infty} F_{\varepsilon_j}(u^j,v^j),$$

so that $E(u)\le\liminf_{j\to\infty} F_{\varepsilon_j}(u^j,v^j)$. □

4.3.2 The second inequality

Now, given $u \in SBV(\Omega)$ such that $E(u) < +\infty$, let us build functions u_ε and v_ε such that $u_\varepsilon \to u$, $v_\varepsilon \to 1$ as $\varepsilon \downarrow 0$, and such that $\limsup_{\varepsilon \downarrow 0} F_\varepsilon(u_\varepsilon, v_\varepsilon) \le E(u)$. We will also assume that u is bounded and that S_u is essentially closed in Ω, which means that $\mathcal{H}^{N-1}(\Omega \cap (\overline{S_u} \setminus S_u)) = 0$. This, in fact, is not restrictive, since it is possible to approximate every $u \in SBV(\Omega)$ with a sequence of bounded functions (u_j) such that for every j, $\mathcal{H}^{N-1}(\Omega \cap (\overline{S_{u_j}} \setminus S_{u_j})) = 0$, in such a way that $\lim_{j \to \infty} E(u_j) = E(u)$. This is a consequence of the essential–closedness of the jumps set of the minimizers of the Mumford–Shah functional, mentioned in section (2.3.2).

Exercise. *Show this approximation property.*

If S_u (which is rectifiable) is essentially closed in Ω, then the limit of the quantities

$$L_\delta(S_u) \;=\; \frac{|\{x \in \Omega \,:\, \mathrm{dist}(x, S_u) \le \delta\}|}{2\delta}$$

as $\delta \downarrow 0$, called the *Minkowsky content* of S_u, is exactly $\mathcal{H}^{N-1}(S_u)$ (see [36]). Notice moreover that since $\delta \mapsto L_\delta(S_u)$ is continuous on $(0, +\infty)$ and bounded by $|\Omega|/(2\delta)$, it is bounded, so that there exists a constant c_L such that

$$|\{x \in \Omega \,:\, \mathrm{dist}(x, S_u) \le \delta\}| \le 2c_L\delta \tag{26}$$

for every $\delta \ge 0$.

We consider the function $\gamma(\cdot)$ defined in section 4.2.2, and, again, $a_\varepsilon = \sqrt{k_\varepsilon \varepsilon}$, and let $S^\varepsilon = \{x \in \Omega \,:\, \mathrm{dist}(x, S_u) \le a_\varepsilon\}$. We set

$$v_\varepsilon(x) \;=\; \begin{cases} \gamma\left(\dfrac{\mathrm{dist}(x, S_u) - a_\varepsilon}{\varepsilon}\right) & \text{if } x \notin S^\varepsilon\,, \\ 0 & \text{otherwise.} \end{cases}$$

and

$$u_\varepsilon(x) \;=\; u(x)\left(\frac{\mathrm{dist}(x, S_u)}{a_\varepsilon} \wedge 1\right)$$

so that $u_\varepsilon = u$ on $\Omega \setminus S^\varepsilon$. Then $u_\varepsilon \to u$ and $v_\varepsilon \to 1$ as $\varepsilon \downarrow 0$. We will denote in what follows $\mathrm{dist}(x, S_u) = d(x)$. Out of S^ε, $\nabla u_\varepsilon = \nabla u$, whereas if $x \in S^\varepsilon$, $\nabla u_\varepsilon(x) = \nabla u(x) d(x)/a_\varepsilon + u(x) \nabla d(x)/a_\varepsilon$ so that $|\nabla u_\varepsilon(x)| \le |\nabla u(x)| + \|u\|_\infty / a_\varepsilon$ (we admit that ∇d exists a.e. and that $|\nabla d| \equiv 1$). Therefore,

$$\int_\Omega (v_\varepsilon(x)^2+k_\varepsilon)|\nabla u_\varepsilon(x)|^2\,dx \;\le\; (1+k_\varepsilon)\int_{\Omega\setminus S^\varepsilon}|\nabla u(x)|^2\,dx + 2k_\varepsilon\left(\int_{S^\varepsilon}|\nabla u(x)|^2\,dx + \frac{|S^\varepsilon|\|u\|_\infty^2}{a_\varepsilon^2}\right).$$

Since $|S^\varepsilon| = 2a_\varepsilon L_{a_\varepsilon}(S_u) \sim 2a_\varepsilon \mathcal{H}^{N-1}(S_u)$ as $\varepsilon\to 0$ and $k_\varepsilon/a_\varepsilon\to 0$, we deduce that $\limsup_{\varepsilon\downarrow 0}\int_\Omega (v_\varepsilon^2+k_\varepsilon)|\nabla u_\varepsilon|^2 \le \int_\Omega |\nabla u|^2$.

Exercise. *Show that $u_\varepsilon\in H^1(\Omega)$ and that $\nabla u_\varepsilon = (d\nabla u + u\nabla d)/a_\varepsilon$ in S^ε.*

Let us now show that $\limsup_{\varepsilon\downarrow 0}\int_\Omega \varepsilon|\nabla v_\varepsilon|^2 + (1-v_\varepsilon)^2/(4\varepsilon) \le \mathcal{H}^{N-1}(S_u)$. Out of S^ε, $\nabla v_\varepsilon(x) = \gamma'\left((d(x)-a_\varepsilon)/\varepsilon\right)\nabla d(x)/\varepsilon$, so that

$$\int_\Omega \varepsilon|\nabla v_\varepsilon(x)|^2 + \frac{(1-v_\varepsilon(x))^2}{4\varepsilon}\,dx \;=\; \frac{|S^\varepsilon|}{4\varepsilon} + \frac{1}{4\varepsilon}\int_{\Omega\setminus S^\varepsilon} 4\gamma'\left(\frac{d(x)-a_\varepsilon}{\varepsilon}\right)^2 + \left(1-\gamma\left(\frac{d(x)-a_\varepsilon}{\varepsilon}\right)\right)^2\,dx.$$

The ratio $|S^\varepsilon|/(4\varepsilon)$ is of order $a_\varepsilon/\varepsilon$ and goes to zero as $\varepsilon\downarrow 0$. Since $\gamma'(t) = (1-\gamma(t))/2 = \exp(-t/2)/2$, the second integral is

$$\frac{1}{2\varepsilon}\int_{\Omega\setminus S^\varepsilon}\exp\left(-\frac{d(x)-a_\varepsilon}{\varepsilon}\right)\,dx.$$

We notice that

$$\begin{aligned}\int_{\Omega\setminus S^\varepsilon}\exp\left(-\frac{d(x)-a_\varepsilon}{\varepsilon}\right)\,dx &= \int_{\Omega\setminus S^\varepsilon}\int_0^1 \chi_{\{t\,:\,t\le \exp(-(d(x)-a_\varepsilon)/\varepsilon)\}}\,dt dx\\ &= \int_0^1 |\{x\in\Omega\,:\,a_\varepsilon<d(x)\le a_\varepsilon-\varepsilon\log t\}|\,dt.\end{aligned}$$

Let $h_\varepsilon(t) = |\{x\in\Omega\,:\,a_\varepsilon<d(x)\le a_\varepsilon-\varepsilon\log t\}|/(2\varepsilon) = (a_\varepsilon/\varepsilon - \log t)L_{(a_\varepsilon-\varepsilon\log t)}(S_u) - a_\varepsilon/\varepsilon L_{a_\varepsilon}(S_u)$. By (26), $|h_\varepsilon(t)|\le c_L(a_\varepsilon/\varepsilon-\log t)\le c_L(1-\log t)$ (if ε is small enough) for every $t\in(0,1)$ and the latter function is integrable on $(0,1)$. Moreover, we know that as $\varepsilon\downarrow 0$ $\lim_{\varepsilon\downarrow 0} h_\varepsilon(t) = (-\log t)\mathcal{H}^{N-1}(S_u)$. By Lebesgue's dominated convergence theorem we deduce that $\lim_{\varepsilon\downarrow 0}\int_0^1 h_\varepsilon(t)\,dt = \mathcal{H}^{N-1}(S_u)$, so that

$$\lim_{\varepsilon\downarrow 0}\frac{1}{2\varepsilon}\int_{\Omega\setminus S^\varepsilon}\exp\left(-\frac{d(x)-a_\varepsilon}{\varepsilon}\right)\,dx \;=\; \mathcal{H}^{N-1}(S_u)$$

This shows that $\lim_{\varepsilon\downarrow 0}\int_\Omega \varepsilon|\nabla v_\varepsilon|^2 + (1-v_\varepsilon)^2/(4\varepsilon) = \mathcal{H}^{N-1}(S_u)$, and achieves the proof of Theorem 8 in arbitrary dimension. □

5 The numerical analysis of the Mumford–Shah problem (II)

Now we consider a different problem, that can be stated in this way: "in what sense is Blake and Zisserman's energy $E(U)$ (or Geman and Geman's energy $E(U, L)$) a discrete approximation of the Mumford–Shah functional?" We will see that in fact, it is *not* an approximation of this functional, but of a slightly different functional in which the length of the discontinuity set is measured in a different (anisotropic) way.

5.1 Rescaling Blake and Zisserman's functional

But first of all, if we want to see $E(U)$ as a discrete approximation of something, we have to introduce a discretization step (or scale parameter) $h > 0$ and explain how the parameters λ and μ in (5) must vary with h in order to get some result. How can this be done?

Consider the simpler, one–dimensional Blake and Zisserman's energy

$$E^1(U) = \sum_{i=1}^{n-1} W_{\lambda,\mu}(u_{i+1} - u_i) \, + \, \sum_{i=1}^{n} (u_i - g_i)^2$$

where $U = (u_i)_{i=1}^n$ and $G = (g_i)_{i=1}^n$ are 1–dimensional signals.

First of all, if we want the last term of the energy to be an approximation of an integral $\int_0^1 (u(x) - g(x))^2 \, dx$, then we need to assume that the signal G is in fact some discretization $G^h = (g_i^h)_{i=1}^n$ at step $h = 1/n$ of the function $g \in L^2(0,1)$. For instance, we can let $g_i^h = (1/h) \int_{(i-1)h}^{ih} g(x) \, dx$. Then, we know that $\sum_{i=1}^n g_i^h \chi_{[(i-1)h, ih)}$ will converge to g in L^2. In this case, if we consider for all $h = 1/n$ a signal $U = U^h = (u_i^h)_{i=1}^n$, we will have that

$$\sum_{i=1}^{n} h \, (u_i^h - g_i^h)^2 \longrightarrow \int_0^1 (u(x) - g(x))^2 \, dx$$

as h goes to 0 (n to $+\infty$) if the functions $\sum_{i=1}^n u_i^h \chi_{[(i-1)h, ih)}$ converge to u in $L^2(0,1)$.

But if this is true, then it is easy to show that, at least in the distributional sense,

$$\sum_{i=1}^{n-1} \frac{u_{i+1}^h - u_i^h}{h} \chi_{[(i-1)h, ih)} \rightharpoonup Du$$

where Du denotes the distributional derivative of u. In the regions where u is differentiable, it is therefore reasonable to ask that $(u^h_{i+1} - u^h_i)/h \sim u'(x)$ if $x \sim ih$, and the sum $\sum_{i=1}^{n-1} W_{\lambda,\mu}(u^h_{i+1} - u^h_i)$ will be an approximation of $\int u'(x)^2\, dx$ in these regions if when $u^h_{i+1} - u^h_i$ is small,

$$W_{\lambda,\mu}(u^h_{i+1} - u^h_i) \;\simeq\; h \cdot \left(\frac{u^h_{i+1} - u^h_i}{h}\right)^2 \;=\; \frac{(u^h_{i+1} - u^h_i)^2}{h}\,.$$

This implies that we must choose $\lambda \sim 1/h$.

On the other hand, when u has a jump, if $(i+1)h$ is on one side of the jump and ih on the other side, the difference $u^h_{i+1} - u^h_i$ should go to $u_r - u_l = \pm(u_+ - u_-)$ and thus have the order of magnitude of a constant. In this case, we want to count "1" in the energy. Since the value of $W_{\lambda,\mu}(t)$ for large t is μ, it means that we must choose $\mu \sim 1$. The rescaled energy then becomes (recording that $W_{1/h,1}(t) = \min(t^2/h, 1)$)

$$E^1_h(U^h) \;=\; \sum_{i=1}^{n-1} \min\left(\frac{(u^h_{i+1} - u^h_i)^2}{h}, 1\right) \;+\; \sum_{i=1}^{n} h\,(u^h_i - g^h_i)^2.$$

Letting $f(t) = \min(t,1) = t \wedge 1$, this can be written as

$$E^1_h(U^h) \;=\; h\left(\sum_{i=1}^{n-1} \frac{1}{h} f\left(\frac{(u^h_{i+1} - u^h_i)^2}{h}\right) \;+\; \sum_{i=1}^{n} (u^h_i - g^h_i)^2\right). \qquad (27)$$

In a similar way, the rescaled 2–dimensional Blake and Zisserman energy will be

$$\begin{aligned} E^2_h(U^h) \;&=\; h^2\left(\sum_{i,j} \frac{1}{h} f\left(\frac{(u^h_{i+1,j} - u^h_{i,j})^2}{h}\right) \;+\; \frac{1}{h} f\left(\frac{(u^h_{i,j+1} - u^h_{i,j})^2}{h}\right)\right. \\ &\qquad \left. +\; \sum_{i,j} (u^h_{i,j} - g^h_{i,j})^2\right) \qquad (28) \end{aligned}$$

where now $(g^h_{i,j})_{1 \le i,j \le n}$ is the correct discretization of an image $g \in L^2(\Omega)$ with $\Omega = (0,1) \times (0,1)$: for instance, we can let

$$g^h_{i,j} \;=\; \frac{1}{h^2} \int_{(i-1)h}^{ih} \int_{(j-1)h}^{jh} g(x,y)\, dx dy.$$

5.2 The Γ–limit of the rescaled 1–dimensional functional

In order to state a Γ–convergence result for the energy $E_h^1(U^h)$ defined by (27), we must first consider E_h^1 as a functional over $L^2(0,1)$. This is done by defining it for every $u^h \in L^2(0,1)$ as

$$E_h^1(u^h) = \begin{cases} E_h^1(U^h) & \text{if } u^h = \sum_{i=1}^n u_i^h \chi_{[(i-1)h, ih)}, U^h = (u_i^h)_{i=1}^n, \\ +\infty & \text{otherwise,} \end{cases}$$

which means that $E_h^1(u_h)$ has a finite value only when u^h is a piecewise constant function at scale h. Then, we have the following result ([20]).

Theorem 9 *E_h^1 Γ–converges to E_0^1 as h goes to zero (n goes to infinity), where*

$$E_0^1(u) = \begin{cases} \int_0^1 \dot{u}(x)^2\,dx + \mathcal{H}^0(Su) + \int_0^1 (u(x) - g(x))^2\,dx \;\; \textit{if } u \in SBV(0,1), \\ +\infty \;\; \textit{if } u \in L^2(0,1) \setminus SBV(0,1). \end{cases} \tag{29}$$

We will sketch a proof of this result. We need to prove that

i. if (u^h) goes to u (in L^2-norm) as $h \to 0$, then $E_0^1(u) \le \liminf_{h\to 0} E_h^1(u^h)$, and

ii. there exists (u^h) that converges to u such that $\limsup_{h\to 0} E_h^1(u^h) \le E_0^1(u)$

5.2.1 Proof of (i)

To prove (i), we consider a sequence (u^h) converging to u as h goes to zero. We can assume that $\liminf_{h\to 0} E_h^1(u^h) < +\infty$ (otherwise there is nothing to prove), and even, by extracting a subsequence, that $\sup_h E_h^1(u^h) < +\infty$. In particular, for every h, there is a discrete signal $U^h = (u_i^h)_{i=1}^n$ ($n = 1/h$) such that $u^h = \sum_{i=1}^n u_i^h \chi_{[(i-1)h, ih)}$ and $E_h^1(u^h) = E_h^1(U^h)$. Then, we build a new function v^h in the following way:

- if $x \in [0, h)$ we let $v^h(x) = u_1^h$;
- then, for $x \in [ih, (i+1)h)$ $(1 \le i \le n-1)$:
  - if $|u_{i+1}^h - u_i^h| \le \sqrt{h}$, we let $v^h(x) = u_i^h + (x - ih)(u_{i+1}^h - u_i^h)/h$,

– otherwise, if $|u^h_{i+1} - u^h_i| > \sqrt{h}$, we let $v^h(x) = u^h_i$ if $x \in [ih, (i + 1/2)h)$ and $v^h(x) = u^h_{i+1}$ if $x \in [(i+1/2)h, (i+1)h)$.

We see that $v^h(ih) = u^h_i$ for all i, and that v^h is affine in the intervals $[ih, (i+1)h)$ such that $f((u^h_{i+1} - u^h_i)^2/h) = (u^h_{i+1} - u^h_i)^2/h$, and piecewise constant with exactly one jump in the intervals such that $f((u^h_{i+1} - u^h_i)^2/h) = 1$.

With this construction, we have that $v^h \in SBV(0,1)$, and that

$$\int_0^1 \dot{v}^h(x)^2\,dx \;+\; \mathcal{H}^0(S_{v^h}) \;=\; h\sum_{i=1}^{n-1} \frac{1}{h} f\left(\frac{(u^h_{i+1} - u^h_i)^2}{h}\right).$$

We can show that $\int_0^1 (v^h(x) - g(x))^2\,dx$ is less than some constant times $h\sum_{i=1}^n (u^h_i - g^h_i)^2$, so that we may apply Theorem 6 to deduce that some subsequence of v^h converges a.e. to a function $v \in SBV(0,1)$, with

$$\int_0^1 \dot{v}(x)^2\,dx \;+\; \mathcal{H}^0(S_v) \;\le \liminf_{h\to 0} \int_0^1 \dot{v}^h(x)^2\,dx \;+\; \mathcal{H}^0(S_{v^h}).$$

Since it is easy to show that v^h must converge (at least) weakly to u in $L^2(0,1)$, and since $h\sum_{i=1}^n (u^h_i - g^h_i)^2 \to \int_0^1 (u(x) - g(x))^2\,dx$, we get that $v = u$, $u \in SBV(0,1)$ and

$$\int_0^1 \dot{u}(x)^2\,dx \;+\; \mathcal{H}^0(S_u) \;+\; \int_0^1 (u(x) - g(x))^2\,dx \;\le \liminf_{h\to 0} E^1_h(u^h),$$

which is the thesis we wanted to show.

Remark. We have shown slightly more than just the point (i). Notice indeed that we can easily deduce that if u^h is bounded uniformly in $L^\infty(0,1)$ and $\sup_h E^1_h(u^h) < +\infty$, then some subsequence of u^h converges (weakly in L^2: easy, strongly in L^2: first show that (u^h) is bounded in $BV(0,1)$, hence compact in L^2) to a function u with $E^1_0(u) \le \liminf_{h\to 0} E^1_h(u^h)$. In particular, we can deduce from Theorem 9 and this remark that if u_h is for every h a minimizer of E^1_h, then it has subsequences that converge to a minimizer of E^1_0. (See Theorem 12 in section 5.4 below for a more general statement.)

5.2.2 Proof of (ii)

Now we consider proving (ii). This is very simple. Choose $u \in SBV(0,1)$ with $E^1_0(u) < +\infty$. The function u is piecewise continuous and has a finite number of jumps $x_1 < x_2 < \cdots < x_k$. We define for every $n \geq 1$ and $h = 1/n$ the discrete signal $U = (u_i^h)_{i=1}^n$ by $u_i^h = u(ih - 0)$, which is the left limit $\lim_{\varepsilon \downarrow 0} u(ih - \varepsilon)$ of u at ih (thus $u_i^h = u(ih)$ if $ih \notin S_u$). It is standard that $u^h = \sum_{i=1}^n u_i^h \chi_{[(i-1)h, ih)}$ goes to u in $L^2(0,1)$, in fact it is easy to prove that u^h goes to u uniformly on $(x_l + \delta, x_{l+1} - \delta)$ for every $l = 1, \ldots, k$ and small $\delta > 0$.

If $[ih, (i+1)h) \cap S_u = \emptyset$, we have (using the Cauchy-Schwarz inequality)

$$u_{i+1}^h - u_i^h = \int_{ih}^{(i+1)h} \dot{u}(x)\, dx \leq \sqrt{h} \left(\int_{ih}^{(i+1)h} \dot{u}(x)^2\, dx \right)^{\frac{1}{2}},$$

so that $(u_{i+1}^h - u_i^h)^2/h \leq \int_{ih}^{(i+1)h} \dot{u}(x)^2\, dx$. Thus, denoting I_h the set $\{i \in [1,n] : [ih, (i+1)h) \cap S_u \neq \emptyset\}$, we have

$$\begin{aligned} h \sum_{i=1}^{n-1} \frac{1}{h} f\left(\frac{(u_{i+1}^h - u_i^h)^2}{h} \right) &\leq \sum_{i \notin I_h} \int_{ih}^{(i+1)h} \dot{u}(x)^2\, dx + \sharp I_h \\ &\leq \int_0^1 \dot{u}(x)^2\, dx + \sharp S_u. \end{aligned}$$

Therefore $\limsup E_h^1(u_h) \leq E_0^1(u)$ and point (ii) is proved.

5.3 The Γ–limit of the rescaled 2–dimensional functional

For the 2–dimensional functional, we have the same kind of result. We also define the functional $E_h^2(U^h)$ as a functional over $L^2(\Omega)$, with $\Omega = (0,1) \times (0,1)$, in the following way: for every $1 \leq i,j \leq n$ we let $C_{i,j}^h$ be the square $[(i-1)h, ih) \times [(j-1)h, jh)$, and

$$E_h^2(u^h) = \begin{cases} E_h^2(U^h) & \text{if } u^h = \sum_{1 \leq i,j \leq n} u_i^h \chi_{C_{i,j}^h}, U^h = (u_{i,j}^h)_{1 \leq i,j \leq n}, \\ +\infty & \text{otherwise.} \end{cases}$$

Then, we define E_0^2 as

$$E_0^2(u) = \begin{cases} \int_\Omega |\nabla u(x)|^2\, dx + \int_{S_u} |\nu_u^1(x)| \quad + |\nu_u^2(x)|\, d\mathcal{H}^1 + \int_\Omega (u(x) - g(x))^2\, dx \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } u \in SBV(\Omega), \\ +\infty \qquad\qquad\qquad\qquad\qquad\qquad \text{if } u \in L^2(\Omega) \setminus SBV(\Omega). \end{cases} \tag{30}$$

Here the vector $\nu_u(x) = (\nu_u^1(x), \nu_u^2(x))$ is the normal vector to the jump set S_u at x. Notice that in this case E_0^2 is not the Mumford and Shah functional: it is slightly different and measures the length of the jump set in an anisotropic way. We point out the fact that it is of the form of E' in definition (20), so that it still admits minimizers and the jump set S_u of a minimizer u is still essentially closed (i.e., $\mathcal{H}^1(\Omega \cap \overline{S_u} \setminus S_u) = 0$).

The anisotropy in E_0^2 is unavoidable since the discrete energy E_h^2 is not isotropic either, as illustrated by the following exercise.

Exercise. *Assume $g = 0$. Let $C = [a,b] \times [c,d] \subset \Omega = (0,1) \times (0,1)$, with $\ell = b - a = c - d > 0$, be a square in Ω, and let C' be the same square rotated by $45°$. Let $u_{i,j}^h$ be 1 if $(ih, jh) \in C$ (i.e., if $a \le ih \le b$ and $c \le jh \le d$) and 0 otherwise (u^h is an approximation of the characteristic function of C). Similarly, let ${u'}_{i,j}^h$ be 1 if $(ih, jh) \in C'$ and 0 otherwise. Show that as h goes to zero, $E_h^2(u^h) \sim 4\ell$. On the other hand, show that $E_h^2({u'}^h) \sim 4\sqrt{2}\ell$.*

With these definitions of E_h^2 and E_0^2, we have the following theorem [21]:

Theorem 10 *As $h = 1/n$ goes to zero, E_h^2 Γ–converges to E_0^2 in $L^2(\Omega)$.*

The proof of Theorem 10 in [21] is based on the same ideas as the proof of Theorem 9 that was just given, and we will not repeat it here. On the other hand, this result is a particular case of Theorem 11 that will be proved in the next section.

5.4 More general finite-differences approximations

Now, we will introduce a general result for finite difference discrete approximations of the Mumford–Shah functional, of which Theorems 9 and 10 are particular cases. What follows is derived from [22].

In 1995, De Giorgi imagined the following non-local functional, defined for any measurable function u on $\mathbb{R}^N$,

$$F_\varepsilon(u) = \iint_{\mathbb{R}^N \times \mathbb{R}^N} \frac{1}{\varepsilon} \mathrm{arctg} \left(\frac{(u(x) - u(x + \varepsilon\xi))^2}{\varepsilon} \right) e^{-|\xi|^2} \, dx d\xi$$

as a possible approximation, as ε goes to zero, of the first-order part of the Mumford–Shah functional (the part that depends on Du)

$$F(u) = \lambda \int_{\mathbb{R}^N} |\nabla u(x)|^2 \, dx + \mu \mathcal{H}^{N-1}(S_u),$$

defined on functions $u \in GSBV_{loc}(\mathbb{R}^N)$. Here λ, μ are two positive parameters. Notice that the function arctg (t) looks like the function $f(t) = \min(t, 1)$

that we introduce in the previous sections: in the neighborhood of 0, it behaves as t, whereas as $t \to \infty$, it behaves like a constant ($\frac{\pi}{2}$ instead of 1). We will see in the sequel that we could consider any non-decreasing function f such that $f(0) = 0$, $f'(0) = 1$ (or some positive constant), and $\lim_{t\to+\infty} f(t) = 1$ (or any other positive constant).

This conjecture of De Giorgi was proved by Gobbino [43]. He established that, as $\varepsilon \downarrow 0$, F_ε Γ–converges to F in the strong $L^p(\mathbb{R}^N)$ topology for any $1 \le p < +\infty$, for $\lambda = \frac{\sqrt{\pi}^N}{2}$ and $\mu = \lambda\sqrt{\pi}$.

This result is very close to the Theorems stated in the previous section. We will show here how we can formulate a discrete version of Gobbino's theorem, and give its complete proof, based on Gobbino's proof.

Let us give some details. Let $\Omega \subseteq \mathbb{R}^N$ be an open domain with Lipschitz boundary, and for every $h > 0$ and every $u : \Omega \cap h\mathbb{Z}^N \to \mathbb{R}$ let

$$F_h(u,\Omega) = h^N \sum_{\begin{cases} x \in h\mathbb{Z}^N \\ x \in \Omega \end{cases}} \sum_{\begin{cases} \xi \in \mathbb{Z}^N \\ x + h\xi \in \Omega \end{cases}} \frac{1}{h} f_\xi \left(\frac{(u(x) - u(x+h\xi))^2}{h} \right) \phi(\xi) \ \in [0,+\infty], \tag{31}$$

where:

- $\phi : \mathbb{Z}^N \to [0,+\infty)$ is even, satisfies $\phi(0) = 0$, $\sum_{\xi\in\mathbb{Z}^N} |\xi|^2 \phi(\xi) < +\infty$, and $\phi(e_i) > 0$ for any $i = 1,\dots,N$ where $(e_i)_{1\le i\le N}$ is the canonical basis of $\mathbb{R}^N$ (in practical applications the support of ϕ will have to be finite and small);

- for any ξ with $\phi(\xi) > 0$, $f_\xi : [0,+\infty) \to [0,+\infty)$ is a non-decreasing bounded function with $f_\xi \equiv f_{-\xi}$, $f_\xi(0) = 0$, $f'_\xi(0) = \alpha_\xi > 0$, and $\lim_{t\to+\infty} f_\xi(t) = \beta_\xi$, and we assume that f_ξ is below (or equal to) the function $t \mapsto \alpha_\xi t \wedge \beta_\xi$. We also assume both $\sup_{\xi\in\mathbb{Z}^N} \alpha_\xi$ and $\sup_{\xi\in\mathbb{Z}^N} \beta_\xi$ are finite;

- we will adopt in the sequel the convention that any term in the sum above is zero whenever either x or $x + h\xi$ is not in Ω even if we do not explicitly write these conditions under the summation signs (this convention will be adopted everywhere in what follows unless otherwise stated), as well, we'll usually write $F_h(u)$ instead of $F_h(u,\Omega)$ when not ambiguous.

Fix $p \in [1,+\infty)$ (but you can assume $p = 2$, as in the previous sections), and let $\ell^p(\Omega \cap h\mathbb{Z}^N)$ be the vector space of functions $u : \Omega \cap h\mathbb{Z}^N \to \mathbb{R}$ such that the norm

$$\|u\|_p \;=\; \left\{h^N \sum_{x\in\Omega\cap h\mathbb{Z}^N} |u(x)|^p\right\}^{\frac{1}{p}}$$

is finite. In the sequel we will always identify a function u in $\ell^p(\Omega\cap h\mathbb{Z}^N)$ and the piecewise constant function in $L^p(\mathbb{R}^N)$ equal to $u(x)$ on $x + \left(-\frac{h}{2},\frac{h}{2}\right)^N$ for any $x\in\Omega\cap h\mathbb{Z}^N$ (and to 0 elsewhere), so that $\|u\|_p = \|u\|_{L^p(\mathbb{R}^N)}$ and that a sentence such as "$u_h\in\ell^p(\Omega\cap h\mathbb{Z}^N)$ converges to $u\in L^p(\Omega)$ as $h\downarrow 0$" will have a natural sense. We also set $F_h(u)=+\infty$ for any $u\in L^p(\Omega)$ that is not the restriction to Ω of the piecewise constant extension of a function in $\ell^p(\Omega\cap h\mathbb{Z}^N)$.

Let now, for any $u\in L^p(\Omega)\cap GSBV_{loc}(\Omega)$,

$$\begin{aligned} F(u) \;=\; & \int_\Omega \sum_{\xi\in\mathbb{Z}^N} \phi(\xi)\alpha_\xi |\langle\nabla u(x),\xi\rangle|^2\,dx \\ & + \int_{S_u} \sum_{\xi\in\mathbb{Z}^N} \phi(\xi)\beta_\xi |\langle\nu_u(x),\xi\rangle|\,d\mathcal{H}^{N-1}(x) \;\in[0,+\infty], \end{aligned} \tag{32}$$

and set $F(u)=+\infty$ if $u\in L^p(\Omega)\setminus GSBV_{loc}(\Omega)$. We will also denote sometimes

$$\begin{aligned} F(u,B) \;=\; & \int_B \sum_{\xi\in\mathbb{Z}^N} \phi(\xi)\alpha_\xi |\langle\nabla u(x),\xi\rangle|^2\,dx \\ & + \int_{B\cap S_u} \sum_{\xi\in\mathbb{Z}^N} \phi(\xi)\beta_\xi |\langle\nu_u(x),\xi\rangle|\,d\mathcal{H}^{N-1}(x) \end{aligned}$$

when $B\subseteq\Omega$ is a Borel set. We have the following theorem.

Theorem 11 *F_h Γ–converges to F as $h\downarrow 0$ in $L^p(\Omega)$ (endowed with its strong topology), for any $p\in[1,+\infty)$.*

We also have the following compactness result:

Theorem 12 *Let $p\in[1,+\infty)$, $g\in L^p(\Omega)\cap L^\infty(\Omega)$, and for any $h>0$ let u_h be a minimizer over $\ell^p(\Omega\cap h\mathbb{Z}^N)$ of*

$$F_h(u) \;+\; \int_\Omega |u(x)-g(x)|^p\,dx \tag{33}$$

(or, equivalently, of

$$F_h(u) \;+\; \left(\|u-g^h\|_p\right)^p \tag{34}$$

where $g^h \in \ell^p(\Omega \cap h\mathbb{Z}^N)$ is a suitable discretization of g at scale h, with $g^h \to g$ in $L^p(\Omega)$ as $h \downarrow 0$ and $\|g^h\|_\infty \le \|g\|_\infty$ for all h). Then (u_h) is relatively compact in $L^p(\Omega)$ and if some subsequence u_{h_j} goes to u as $j \to \infty$, $u \in SBV_{loc}(\Omega) \cap L^p(\Omega)$ is a minimizer of

$$F(u) + \int_\Omega |u(x) - g(x)|^p \, dx.$$

These theorems, for $p = 2$, provide a generalization of the previous Theorems 9 and 10. For instance, Theorem 10 is the case where $N = 2$, $p = 2$, $\Omega = (0,1) \times (0,1)$, $\phi \equiv 0$ on $\mathbb{Z}^2$ except $\phi(0,1) = \phi(1,0) = \phi(0,-1) = \phi(-1,0) = 1/2$, and $f_\xi(t) = t \wedge 1$, so that

$$F(u) = \int_\Omega |\nabla u(x)|^2 \, dx + \int_{S_u} |\nu_u^1(x)| + |\nu_u^2(x)| \, d\mathcal{H}^1(x).$$

Remark. The condition $\phi(e_i) > 0$ for $i = 1, \ldots, N$ is necessary only for the coercivity, i.e. to establish Lemma 7 and Theorem 12. This is important in practical applications for the stability of the numerical schemes. Even if we have not discussed it in the previous sections (except in the Remark on page 56), a similar coercivity and compactness result also hold for Theorems 8, 9 and 10.

If we wanted only to prove the Γ–convergence of F_h to F, it would be sufficient to assume that $\phi(\xi_i) > 0$, $i = 1, \ldots, N$ for some basis $(\xi_i)_{1 \le i \le N} \in \mathbb{Z}^{N \times N}$ of $\mathbb{R}^N$.

We will first describe the implementation of these energies. The proofs of Theorem 11 and Theorem 12 will then be given in the last section of these notes. The next sections are extracted from [22].

6 A numerical method for minimizing the Mumford–Shah functional

In this section we describe a numerical method for the implementation of the energies we have introduced in these lectures. We will describe the minimization of problem (34) for $p = 2$, since the energies E_h^1 and E_h^2 are a particular case. In particular, we will show how the choice of ϕ, α_ξ and β_ξ influences the results.

6.1 An iterative procedure for minimizing (34)

Let us quickly describe a standard procedure for minimizing energies such as (34). Of course we do not pretend to compute an exact minimizer of the energy, since the high non-convexity of the problem does not allow this. However, the iterative algorithm we describe gives satisfactory results. A variant has been successfully implemented in the case of the approximation of [23] (see [17]). Many other similar implementations have been made for solving image reconstruction problems (see for instance [60, 10], and the pioneering work [40] by D. Geman and G. Reynolds).

We assume Ω is bounded so that the discrete problem is finite-dimensional for every fixed $h > 0$ (in the applications Ω will be a rectangle). The non-convexity in the energy F_h comes from the non-convexity of the functions f_ξ, $\xi \in \mathbb{Z}^N$. In order to simplify the computations we will assume that the f_ξ are all identical, up to a rescaling:

$$f_\xi(t) \;=\; \beta_\xi f\left(\frac{\alpha_\xi}{\beta_\xi} t\right)$$

for all $\xi \in \mathbb{Z}^{\mathbb{N}}$ (with $\phi(\xi) > 0$) and $t \geq 0$. The function f is nondecreasing, and satisfies $f(0) = 0$, $f(+\infty) = 1$, and $f'(0) = 1$. It could be, of course, the function $f(t) = t \wedge 1$ of section 5, except that a differentiable function provides better numerical results. An interpretation of Blake and Zisserman's GNC algorithm would correspond to approximate gradually $t \wedge 1$ with smooth functions.

We will thus assume, as well, that f is concave, and differentiable. Thus, $-f$ is convex (we extend it with the value $+\infty$ on $\{t < 0\}$), and lower semi-continuous. Let

$$\psi(-v) \;=\; \sup_{t \in \mathbb{R}} tv - (-f)(t) \;=\; (-f)^*(v)$$

be the Legendre-Fenchel transform of $-f$, by a classical result $(-f)^{**} = -f$ so that

$$-f(t) \;=\; \sup_{v \in \mathbb{R}} tv - \psi(-v) \;=\; -\inf_{v \in \mathbb{R}} tv + \psi(v).$$

It is well known that the first sup in this equation is attained at v such that $t \in \partial(-f)^*(v)$ (the subdifferential of $(-f)^*$ at t), and that this is equivalent to $v \in \partial(-f)(t)$, and since $\partial(-f)(t) = \{-f'(t)\}$ for $t > 0$ and $(-\infty, -1]$ for $t = 0$ we deduce that the sup is reached at some $v \in [-1, 0]$ (since for $t = 0$

we check that $(-f)^*(-1) = 0$ and thus the sup is reached at $v = -1$). Hence

$$f(t) \;=\; \min_{v\in[0,1]} tv + \psi(v)$$

and the min is reached for $v = f'(t)$. (If $f(t) = t \wedge 1$, all of this is still true except that for $t = 1$, the min is reached for any $v \in [0,1]$.) We may therefore rewrite F_h in the following way:

$$F_h(u) \;=\; \min_{v(\cdot,\cdot)} F_h(u,v)$$

for $v : (\Omega \cap h\mathbb{Z}^N) \times (\Omega \cap h\mathbb{Z}^N) \to [0,1]$ and

$$\begin{aligned} F_h(u,v) \;&=\; h^N \sum_{x\in h\mathbb{Z}^N} \sum_{\xi\in\mathbb{Z}^N} \left\{ \alpha_\xi v(x, x+h\xi) \left| \frac{u(x) - u(x+h\xi)}{h} \right|^2 \right. \\ &\qquad \left. + \beta_\xi \frac{\psi(v(x,x+h\xi))}{h} \right\} \phi(\xi). \end{aligned} \tag{35}$$

The algorithm consists in minimizing alternatively $F_h(u,v) + \|u - g^h\|_2^2$ with respect to u and v. The minimization with respect to v is straightforward, since it just consists in computing for each $x, y \in \Omega \cap h\mathbb{Z}^N$

$$v(x,y) \;=\; f'\left(\alpha_\xi \frac{(u(x) - u(y))^2}{\beta_\xi h}\right),$$

with $\xi = \pm(y-x)/h$. The minimization with respect to u is also a simple (linear) problem, since the energy is convex and quadratic with respect to u. Of course there is no way of knowing whether the algorithm converges to a solution or not, what is certain is that the energy decreases and goes to some critical level, while the function u converges to either a critical point or, if it exists, a continuum of critical points. Notice that if f is strictly increasing, v is everywhere strictly positive.

In the applications shown in these notes we considered $f(t) = \frac{2}{\pi}\mathrm{arctg}\,\frac{\pi x}{2}$, so that

$$f'(t) \;=\; \frac{1}{1 + \frac{\pi^2 x^2}{4}}.$$

Notice that one never has to compute explicitly the position of the edges during the minimization. Once a minimizer of the energy has been found, it is possible to extract the edges out of the segmented image by standard algorithms (using Canny's or more sophisticated edge detectors, with a very narrow kernel since the images on which the edges have to be found are piecewise smooth). The value of the auxiliary function v is also a good indicator for the position of the edges (it is "large" on the edges and close to zero everywhere else), and should be taken into account. An elementary method may be for instance to consider the zero-crossings of the (discretized) operator $d^2u(\nabla u, \nabla u)$ in the regions where v is large.

6.2 Anisotropy of the length term

In some of the above mentioned image processing papers it had been noticed that the segmentations could be improved by trying to modify slightly the energy, making it "less anisotropic". Here we illustrate how the result of Theorem 11 allows to control this anisotropy and find explicitly the correct parameters for the "best" energies.

In this section, like in section 5, n will be an integer ($n > 1$), we will set $h = 1/n$ and the functions u and g^h (defined on $[0,1) \times [0,1) \cap h\mathbb{Z}^2$) will be denoted as $n \times n$ matrices $(u_{i,j})_{0 \le i,j < n}$ and $(g^h_{i,j})_{0 \le i,j < n}$. We will compare the following two cases (pay attention to the fact that the notations here for the energies are different from the notations in section 5, in fact, the following E^1_n is similar to E^2_h in sec. 5, the other energies are new)

$$E^1_n(u) = h^2 \sum_{i,j} \frac{\beta_1}{h} f\left(\alpha_1 \frac{|u_{i+1,j} - u_{i,j}|^2}{\beta_1 h}\right) + \frac{\beta_1}{h} f\left(\alpha_1 \frac{|u_{i,j+1} - u_{i,j}|^2}{\beta_1 h}\right) + |u_{i,j} - g^h_{i,j}|^2,$$

(which is Blake and Zisserman's "weak membrane" energy, except that f is smoother than $t \mapsto t \wedge 1$) and

$$E^2_n(u) = h^2 \sum_{i,j} \frac{\beta_2}{h} f\left(\alpha_2 \frac{|u_{i+1,j} - u_{i,j}|^2}{\beta_2 h}\right) + \frac{\beta_2}{h} f\left(\alpha_2 \frac{|u_{i,j+1} - u_{i,j}|^2}{\beta_2 h}\right) +$$

$$+ \frac{\beta'_2}{h} f\left(\alpha'_2 \frac{|u_{i+1,j+1} - u_{i,j}|^2}{\beta'_2 h}\right) + \frac{\beta'_2}{h} f\left(\alpha'_2 \frac{|u_{i-1,j+1} - u_{i,j}|^2}{\beta'_2 h}\right) + |u_{i,j} - g^h_{i,j}|^2.$$

By Theorem 12, the limit points of the minimizers of E^1_n and E^2_n, as $n \to \infty$, will be minimizers of respectively

$$E^1_\infty(u) = \lambda_1 \int_\Omega |\nabla u(x)|^2\, dx + \mu_1 \Lambda_1(S_u) + \int_\Omega |u(x) - g(x)|^2\, dx$$

and

$$E^2_\infty(u) = \lambda_2 \int_\Omega |\nabla u(x)|^2 \, dx \; + \; \mu_2 \Lambda_2(S_u) \; + \; \int_\Omega |u(x) - g(x)|^2 \, dx,$$

(for $u \in L^2(\Omega) \cap GSBV(\Omega)$, and $+\infty$ otherwise) with $\Omega = (0,1) \times (0,1)$, $\lambda_1 = \alpha_1$, $\lambda_2 = \alpha_2 + 2\alpha'_2$, and

$$\mu_1 \Lambda_1(S_u) \; = \; \int_{S_u} \beta_1(|\nu_1(x)| + |\nu_2(x)|) \, d\mathcal{H}^1(x),$$

$$\mu_2 \Lambda_2(S_u) \; = \; \int_{S_u} \beta_2(|\nu_1(x)|+|\nu_2(x)|)+\beta'_2(|\nu_1(x)-\nu_2(x)|+|\nu_1(x)+\nu_2(x)|) \, d\mathcal{H}^1(x),$$

where $(\nu_1(x), \nu_2(x))$ is the normal vector to S_u at x. Simple computations

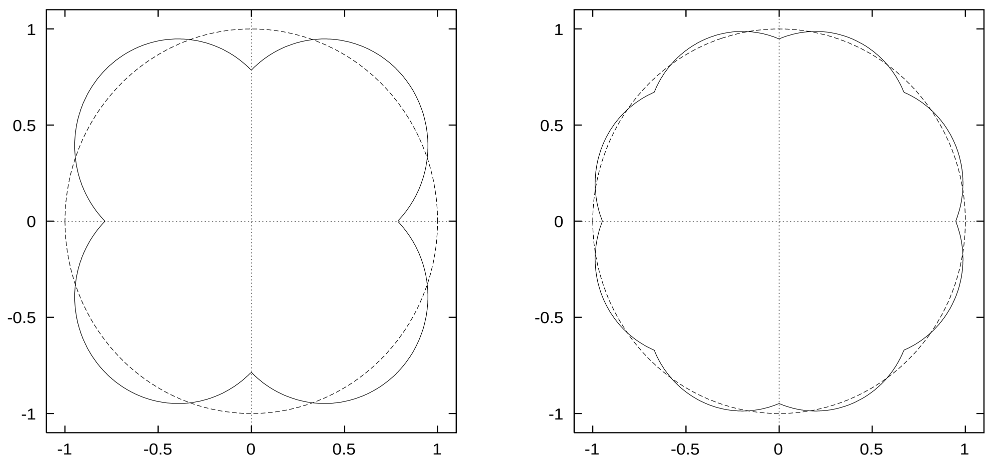

Figure 7: The solid line represents the length of a unit vector, as a function of the angle. Left: for Λ_1, right: for Λ_2. The dashed line is the unit circle.

show that $\beta'_2 = \beta_2/\sqrt{2}$ is an optimal choice, since it minimizes the ratio of the length Λ_2 of the longest (with length Λ_2) vector in $\mathbb{S}^1$ over the length of the shortest. For any rectifiable 1-set $E \subset \mathbb{R}^2$ with normal vector (ν_1, ν_2) at x we define the lengths

$$\Lambda_1(E) \; = \; \frac{\pi}{4} \int_E (|\nu_1(x)| + |\nu_2(x)|) \, d\mathcal{H}^1(x)$$

and

$$\Lambda_2(E) \; = \; \frac{\pi}{8} \int_E \left(|\nu_1(x)| + |\nu_2(x)| + \frac{|\nu_1(x) - \nu_2(x)| + |\nu_1(x) + \nu_2(x)|}{\sqrt{2}} \right) \, d\mathcal{H}^1(x).$$

(Notice that $\Lambda_2 = (\Lambda_1 + \Lambda_1 \circ R_{\frac{\pi}{4}})/2$ where $R_{\frac{\pi}{4}}$ is the rotation of angle $\pi/4$ in $\mathbb{R}^2$.) The choice of the parameters $\pi/4$ and $\pi/8$ is made in order to ensure that a "random" set of lines has in average the same length Λ_1 and Λ_2 (and Euclidean length), in other words, the unit circle has length 2π in both cases. This is of course not the only possible choice. For instance, one could prefer to parameterize these lengths in such a way that the error (with respect to the Euclidean length) $\Lambda_i(e_{max}) - 1$ on the measure of the longest (for Λ_i) vector $e_{max} \in \mathbb{S}^1$ is equal to the error $1 - \Lambda_i(e_{min})$ on the measure of the shortest vector. In this case, one should choose as parameters $\frac{2}{1+\sqrt{2}}$ instead of $\pi/4$ for Λ_1 and $\frac{2}{1+\sqrt{2}+\sqrt{4+2\sqrt{2}}}$ instead of $\pi/8$ for Λ_2.

With the choice we made, we get that $\mu_1 = 4\beta_1/\pi$ and $\mu_2 = 8\beta_2/\pi$. In both cases the limit energy is anisotropic, what is interesting is that the second length Λ_2 is far "less anisotropic" than the first length Λ_1. As a matter of fact, the longest vector in $\mathbb{S}^1$ for Λ_1 is about 41.4% longer than the shortest (the ratio is $\sqrt{2}$) while it is only 8.2% longer for the length Λ_2 (the ratio is $2\sqrt{2}\cos\frac{\pi}{8}/(1+\sqrt{2}) = \sqrt{4+2\sqrt{2}}/(1+\sqrt{2})$). The difference of anisotropy of both lengths is striking in figure 7.

Figure 8: Original images for the next examples.

6.3 Numerical experiments

We show here a few experiments, so that the reader can see for himself the difference of behaviour of the lengths Λ_1 and Λ_2. Notice that in all of our comparisons we of course always choose $\lambda_1 = \lambda_2$ and $\mu_1 = \mu_2$. In Figure 9 and Figure 10 (see original pictures in Figure 8), one notices that the edges are usually nicer when length Λ_2 is used, whereas images obtained by minimization of energy E_n^1 are more “blocky”. Notice in particular the

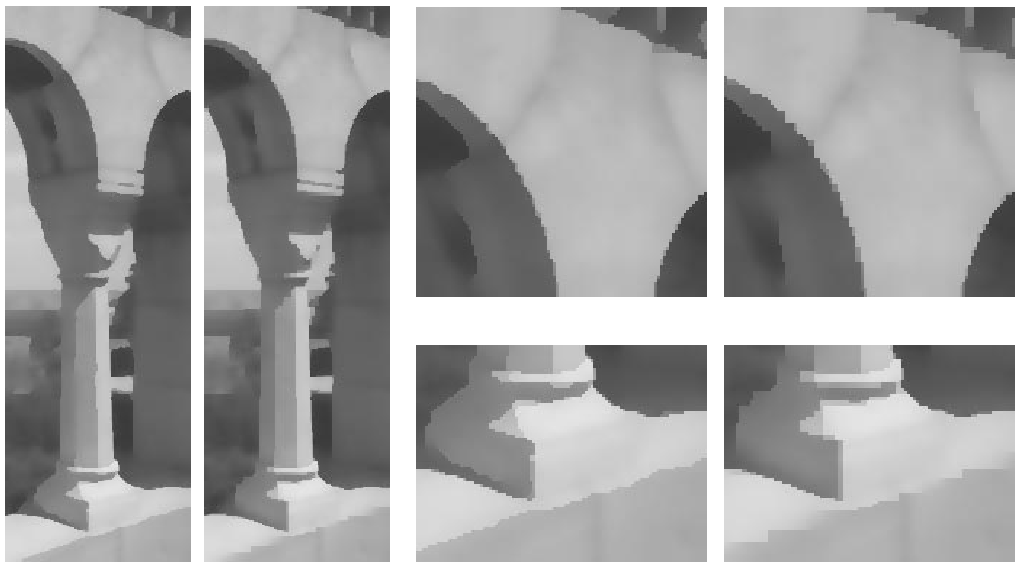

Figure 9: The segmented column. Left, by using energy E_n^2, then with energy E_n^1. The details appear in the same order.

diagonal line at the bottom of the image. However, the vertical edges on the column (Figure 9, second picture) are nicer with energy E_n^1. The reason is clear: these edges are vertical, and the vertical and horizontal lines have a much lower costs than lines with other orientations with this energy.

In Figures 10 and 11 the results are similar: the edges look much nicer when energy E_n^2 is minimized. The other two figures (Figs. 12 and 13) show segmentations in presence of noise. In the two segmentations of the disk, the total length of the edges found was $6.56 \times R$ with energy E_n^2 and $6.40 \times R$ with E_n^1. These lengths are slightly overestimated because a few spurious edges were found, and also because of some oscillation of the boundary, that is due to the noise. Again in Figure 13 the result is more blocky with energy E_n^1.

Figure 10: Segmented lady with energy E_n^2 and two details.

Figure 11: Segmented lady with energy E_n^1 and two details.

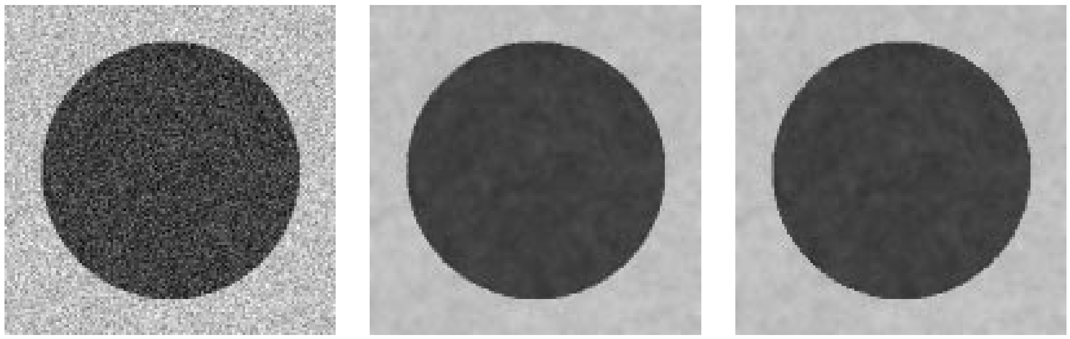

Figure 12: The noisy disk (grey level values 64 (disk) and 192 (background), std. dev. of noise 40). Middle, the segmented disk with energy E_n^2. Right, with E_n^1.

Figure 13: A noisy image (std. dev.= 25 for values between 0 and 255) and the segmented outputs, by minimizing E_n^2 (middle) and E_n^1 (right).

In the last two experiments we used another energy, namely,

$$E_n^3(u) \;=\; h^2\sum_{i,j}\frac{\beta_3}{h}f\left(\alpha_3\frac{|u_{i+1,j}-u_{i,j}|}{\beta_3 h}^2\right) \;+\; \frac{\beta_3}{h}f\left(\alpha_3\frac{|u_{i,j+1}-u_{i,j}|}{\beta_3 h}^2\right) \;+$$

$$+\;\frac{\beta_3'}{h}f\left(\alpha_3'\frac{|u_{i\pm1,j+1}-u_{i,j}|}{\beta_3' h}^2\right)\;+$$

$$+\;\frac{\beta_3''}{h}f\left(\alpha_3''\frac{|u_{i+2,j\pm1}-u_{i,j}|}{\beta_3'' h}^2\right) \;+\; \frac{\beta_3''}{h}f\left(\alpha_3''\frac{|u_{i\pm1,j+2}-u_{i,j}|}{\beta_3'' h}^2\right) \;+\; |u_{i,j}-g_{i,j}^h|^2.$$

We choose $\beta_3' = \beta_3/\sqrt{2}$ and $\beta_3'' = \beta_3/\sqrt{5}$. Now, as $n\to\infty$, the limit points of the minimizers of E_n^3 minimize

$$E_\infty^3(u) = \lambda_3\int_\Omega |\nabla u(x)|^2\,dx \;+\; \mu_3\Lambda_3(S_u) \;+\; \int_\Omega |u(x)-g(x)|^2\,dx,$$

with $\lambda_3 = \alpha_3 + 2\alpha_3' + 10\alpha_3''$,

$$\Lambda_3(E) \;=\; \frac{\pi}{16}\int_E \Bigg(|\nu_1(x)| + |\nu_2(x)| + \frac{|\nu_1(x)-\nu_2(x)| + |\nu_1(x)+\nu_2(x)|}{\sqrt{2}} +$$

$$+\frac{|2\nu_1(x)-\nu_2(x)| + |\nu_1(x)+2\nu_2(x)| + |2\nu_1(x)+\nu_2(x)| + |\nu_1(x)-2\nu_2(x)|}{\sqrt{5}}\Bigg)\,d\mathcal{H}^1(x)$$

(this time $\Lambda_3 = (\Lambda_1 + \Lambda_1\circ R_{\frac{\pi}{4}} + \Lambda_1\circ R_\theta + \Lambda_1\circ R_{-\theta})/4$ with $\theta = \operatorname{arctg} 2$), and $\mu_3 = 16\beta_3/\pi$. Figure 14 illustrates how "isotropic" the measure Λ_3 is, and Figures 15 and 16 show examples. (Now, the length of the longest vector in $\mathbb{S}^1$ is about 5.0% greater than the length of the shortest.) The results look slightly better than the ones obtained with energy E_n^2, however, the computational cost is quite higher.

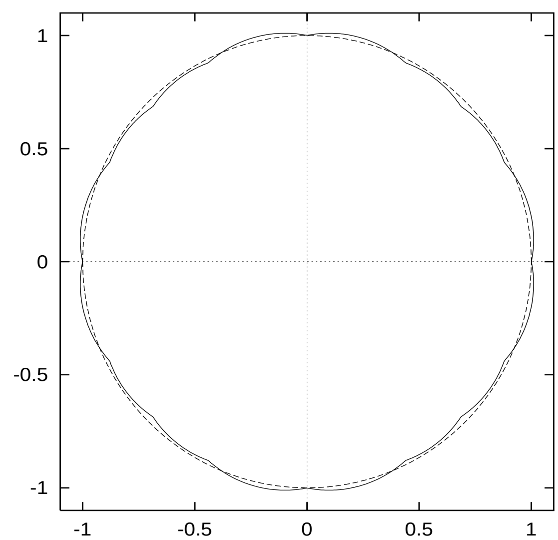

Figure 14: Same as Figure 7, this time for Λ_3.

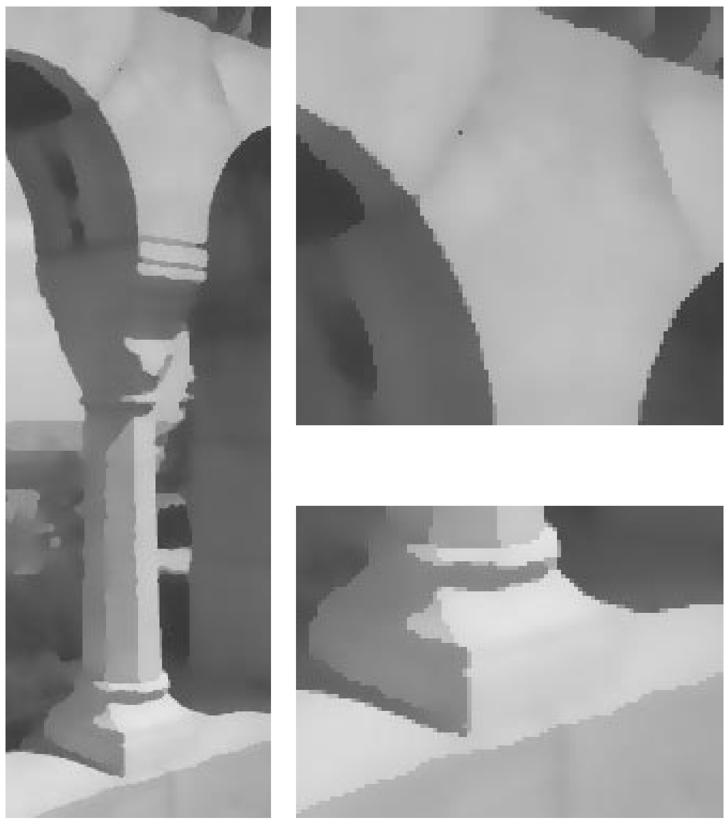

Figure 15: Segmentation with energy E_n^3 (the column).

Figure 16: Segmentation with energy E_n^3(the lady).

A Proof of Theorems 11 and 12

This last section is entirely devoted to the proof of the theorems of section 5.4. Most of it relies on Gobbino's work [43], except for some adaptations and a slight difference in the proof (which avoids the use of Gobbino's technical Lemmas 3.1 and 3.2 in [43], and makes it simpler).

We let for any $u \in L^p(\Omega)$

$$F'(u) = (\Gamma - \liminf_{h\downarrow 0} F_h)(u)$$

and

$$F''(u) = (\Gamma - \limsup_{h\downarrow 0} F_h)(u).$$

In the next section A.1 we will prove a preliminary lemma that will be helpful in the sequel. Then, the aim of the following two sections A.2 and A.3 will be to prove Theorem 11, i.e., to prove that $F'(u) \geq F(u)$ and $F''(u) \leq F(u)$ for all $u \in L^p(\Omega)$. Eventually in section A.4, we will prove Theorem 12.

A.1 A compactness lemma

The lemma we show in this section will be needed to establish Theorem 12, but it will also give some a priori information on the regularity of functions $u \in L^p(\Omega)$ such that $F'(u) < +\infty$.

Lemma 7 *Let $h_j \downarrow 0$ and $u_{h_j} \in \ell^p(\Omega \cap h_j\mathbb{Z}^N)$ such that*

$$\sup_j F_{h_j}(u_{h_j}) < +\infty \quad \textit{and} \quad \sup_j \|u_{h_j}\|_\infty < +\infty.$$

Then there exist a subsequence (not relabeled) u_{h_j} and $u \in SBV_{loc}(\Omega)$ such that

$$u_{h_j}(x) \to u(x) \quad \textit{a.e. in } \Omega$$

as j goes to infinity, and

$$\int_\Omega |\nabla u(x)|^2\, dx + \mathcal{H}^{N-1}(S_u) \; < \; +\infty$$

Proof. In order to simplify the notations we drop the subscript j. Let $f = \min_{i=1,\dots,N} f_{e_i}$, $c = \min_{i=1,\dots,N} \phi(e_i) > 0$, and choose $\alpha, \beta > 0$ such that $\alpha t \wedge \beta \leq f(t)$ for all $t \geq 0$. We have:

$$F_h(u_h) \geq 2c\, h^N \sum_{x\in h\mathbb{Z}^N} \sum_{i=1}^N \alpha \left| \frac{u_h(x) - u_h(x+he_i)}{h} \right|^2 \wedge \frac{\beta}{h}$$

(remember we only sum on x such that $x, x + he_i \in \Omega$). We first show that the sequence (u_h) is bounded in $BV_{loc}(\Omega)$ (so that it is compact in $L^1_{loc}(\Omega)$). Choose $R > 0$, $i \in \{1, \dots, N\}$, and write (with $\Omega_{\frac{1}{2}\sqrt{N}h} = \left\{x \in \Omega \ : \ \mathrm{dist}(x, \partial\Omega) > \frac{1}{2}\sqrt{N}h\right\}$)

$$
\begin{aligned}
|D_i u_h| \, (\Omega_{\frac{1}{2}\sqrt{N}h} \cap B_R(0)) \quad &\le \sum_{x \in h\mathbb{Z}^N \cap B_R(0)} h^{N-1} |u_h(x) - u_h(x + he_i)| \\
&\le \sum_{x \in X_+} 2\|u_h\|_\infty h^{N-1} \\
&\quad + h^N \sum_{x \in X_-} \left| \frac{u_h(x) - u_h(x + he_i)}{h} \right|
\end{aligned}
$$

where $X_+ = \left\{x \in h\mathbb{Z}^N \cap B_R(0) \ : \ |u_h(x) - u_h(x + he_i)| > \sqrt{\beta h / \alpha}\right\}$ and $X_- = h\mathbb{Z}^N \cap B_R(0) \setminus X_+$. Of course, we only consider points $x \in h\mathbb{Z}^N$ such that x and $x + he_i$ belong to Ω. Then, using the Cauchy-Schwarz inequality,

$$
\begin{aligned}
|D_i u_h| \, (\Omega_{\frac{1}{2}\sqrt{N}h} \cap B_R(0)) \quad &\le \quad 2\|u_h\|_\infty h^{N-1} \sharp X_+ \\
&\quad + C R^{\frac{N}{2}} \left\{ h^N \sum_{x \in X_-} \left| \frac{u_h(x) - u_h(x + he_i)}{h} \right|^2 \right\}^{\frac{1}{2}} \\
&\le \quad \frac{\|u_h\|_\infty F_h(u_h)}{\beta c} \ + \ C R^{\frac{N}{2}} \sqrt{\frac{F_h(u_h)}{2\alpha c}}
\end{aligned}
$$

with C some constant depending only on N, so that eventually, for any $\eta > 0$,

$$
\sup_h |D u_h| (\Omega_\eta \cap B_R(0)) \ < \ +\infty.
$$

This shows that upon extracting a subsequence we may assume that u_h converges almost everywhere in Ω (and in $L^p_{loc}(\Omega)$, as well) to some function u that belongs to $BV(\Omega \cap B_R(0))$ for any $R > 0$.

Now consider the extension of u_h (on $\mathbb{R}^N$, $u_h(x)$ being considered to be 0 outside of Ω)

$$
v_h(y) = \sum_{x \in h\mathbb{Z}^N} u_h(x) \Delta_N \left(\frac{y - x}{h} \right),
$$

where $\Delta(t) = (1 - |t|)^+$ for any $t \in \mathbb{R}$ and $\Delta_N(y) = \prod_{i=1}^N \Delta(y_i)$ for any $y \in \mathbb{R}^N$. We estimate $\int |\nabla v_h|^2$ on an "elementary cell", for instance $(0, h)^N$:

$$
\begin{aligned}
&\int_{(0,h)^N} |\partial_1 v_h(y)|^2 \, dy \; = \\
&= \int_{(0,h)^N} \left| \sum_{x\in\{0,h\}^N} u_h(x) \frac{1}{h} \Delta' \left(\frac{y_1 - x_1}{h} \right) \prod_{i=2}^{N} \Delta \left(\frac{y_i - x_i}{h} \right) \right|^2 dy \\
&= h \int_{(0,h)^{N-1}} dy_2 \ldots dy_N \left| \sum_{x\in\{0,h\}^{N-1}} \frac{u_h(h,x) - u_h(0,x)}{h} \prod_{i=2}^{N} \Delta \left(\frac{y_i - x_i}{h} \right) \right|^2 \\
&= h \int_{(0,h)^{N-2}} dy_3 \ldots dy_N \int_0^h \left| \left(\frac{y_2}{h} \right) \sum_{x\in\{0,h\}^{N-2}} \frac{u_h(h,h,x) - u_h(0,h,x)}{h} \prod_{i=3}^{N} \Delta \left(\frac{y_i - x_i}{h} \right) \right. \\
&\qquad \left. + \left(1 - \frac{y_2}{h} \right) \sum_{x\in\{0,h\}^{N-2}} \frac{u_h(h,0,x) - u_h(0,0,x)}{h} \prod_{i=3}^{N} \Delta \left(\frac{y_i - x_i}{h} \right) \right|^2 dy_2 \\
&\le \frac{h}{2}^2 \left\{ \int_{(0,h)^{N-2}} dy_3 \ldots dy_N \left| \sum_{x\in\{0,h\}^{N-2}} \frac{u_h(h,h,x) - u_h(0,h,x)}{h} \prod_{i=3}^{N} \Delta \left(\frac{y_i - x_i}{h} \right) \right|^2 \right. \\
&\qquad \left. + \int_{(0,h)^{N-2}} dy_3 \ldots dy_N \left| \sum_{x\in\{0,h\}^{N-2}} \frac{u_h(h,0,x) - u_h(0,0,x)}{h} \prod_{i=3}^{N} \Delta \left(\frac{y_i - x_i}{h} \right) \right|^2 \right\};
\end{aligned}
$$

by induction we deduce that

$$
\int_{(0,h)^N} |\partial_1 v_h(y)|^2 \, dy \; \le \; \frac{h^N}{2^{N-1}} \sum_{x\in\{0,h\}^{N-1}} \left| \frac{u_h(h,x) - u_h(0,x)}{h} \right|^2 .
$$

Notice that we could therefore conclude that

$$
\int_{\Omega_{\sqrt{N}h}} |\nabla v_h(y)|^2 \, dy \; \le \; h^N \sum_{x\in h\mathbb{Z}^N} \sum_{i=1}^{N} \left| \frac{u_h(x) - u_h(x+he_i)}{h} \right|^2
$$

(with $\Omega_{\sqrt{N}h} = \left\{ x \in \Omega \; : \; \mathrm{dist}(x, \partial\Omega) > \sqrt{N}h \right\}$, since we control the gradient of v_h only on the cubes $x + (0,h)^N$, $x \in h\mathbb{Z}^N$ whose 2^N vertices all belong to Ω), but since we cannot control the right-hand side of this expression if it is summed over all x we must introduce a slight modification of v_h: we thus define $\hat{v}_h = v_h$, except each time

$$
|u_h(x) - u_h(x+he_i)| > \sqrt{\frac{\beta h}{\alpha}}, \tag{A.1}
$$

in which case we set $\hat{v}_h \equiv 0$ on $(x, x+he_i) \times \prod_{i' \neq i}(x - he_{i'}, x + he_{i'}) = U^h_{x,e_i}$. The new function $\hat{v}_h$ is in $SBV_{loc}(\Omega)$, and $S_{\hat{v}_h} \subseteq \bigcup_{(x,e_i)\in X_h} \partial U^h_{x,e_i}$ where the union is taken on $X_h = \{(x, e_i) \,:\, \text{(A.1) holds}\}$. Now, we can write

$$\int_{\Omega_{\sqrt{N}h}} |\nabla \hat{v}_h(y)|^2 \, dy \;\leq\; \sum_{(x,e_i)\in h\mathbb{Z}^N \setminus X_h} \left| \frac{u_h(x) - u_h(x+he_i)}{h} \right|^2 \;\leq\; \frac{F_h(u_h)}{2c\alpha}, \tag{A.2}$$

moreover since $\mathcal{H}^{N-1}(\partial U^h_{x,e_i}) = \kappa h^{N-1}$ (with $\kappa = 2^{N-1}(N+1)$),

$$\mathcal{H}^{N-1}(\Omega_{\sqrt{N}h} \cap S_{\hat{v}_h}) \;\leq\; \sharp X_h \kappa h^{N-1} \;\leq\; \kappa \frac{F_h(u_h)}{2c\beta}. \tag{A.3}$$

From (A.2) and (A.3) and since $\sup_h \|\hat{v}_h\|_\infty < +\infty$, we deduce invoking Ambrosio's Theorem 6 (see section 2.2.6) that some subsequence of $\hat{v}_h$ converges to a function $v \in L^\infty(\Omega) \cap SBV_{loc}(\Omega)$, with

$$\begin{aligned}\int_\Omega |\nabla v(x)|^2 \, dx + \mathcal{H}^{N-1}(S_v) \;&=\; \sup_{A \subset\subset \Omega} \int_A |\nabla v(x)|^2 \, dx + \mathcal{H}^{N-1}(A \cap S_v) \\ &\leq\; \tfrac{1}{2c}\left(\tfrac{1}{\alpha} + \tfrac{\kappa}{\beta}\right) \liminf_{h \downarrow 0} F_h(u_h) \;<\; +\infty.\end{aligned}$$

The proof of the lemma is achieved once we notice that v must be equal to u (as for instance by the construction of v_h and $\hat{v}_h$ it is simple to check that for any $A \subset\subset \Omega$ with regular boundary $\int_A (u_h(y) - \hat{v}_h(y)) \, dy \to 0$ as $h \downarrow 0$). $\square$

Remark. If we drop the condition $\phi(e_i) > 0$ for $i = 1, \ldots, N$, the result may be false. For instance, if $N = 1$, $\phi \equiv 0$ except at -2 and 2 where $\phi(-2) = \phi(2) = 1$, the family $(u_h)_{h>0}$ defined by

$$u_h(kh) \;=\; \begin{cases} 0 & \text{if } k \in 2\mathbb{Z} \\ 1 & \text{if } k \in 2\mathbb{Z}+1 \end{cases} \qquad \text{for every } k \in \mathbb{Z}$$

satisfies the assumptions of Lemma 7 but is not compact.

A.2 Estimate from below the Γ–limit

In this section we wish to prove that for all $u \in L^p(\Omega)$,

$$F(u) \leq F'(u). \tag{A.4}$$

We must therefore prove that for any $u \in L^p(\Omega)$ and any sequence (u_{h_j}) that converges to u in $L^p(\Omega)$ as $j \to \infty$ (with $\lim_{j\to\infty} h_j = 0$) we have,

$$F(u) \leq \liminf_{j\to\infty} F_{h_j}(u_{h_j}). \tag{A.5}$$

Let $u \in L^p(\Omega)$, and we will suppose first that it is bounded. Choose also an arbitrary decreasing sequence $h_j \downarrow 0$ and functions u_{h_j} that converge to u in $L^p(\Omega)$. We can assume that $\|u_{h_j}\|_\infty \le \|u\|_\infty$, as truncating u_{h_j} we decrease its energy $F_{h_j}(u_{h_j})$. It is clearly not restrictive to consider, as well, that the lim inf is in fact a limit, and that $\sup_j F_{h_j}(u_{h_j}) < +\infty$ (since if $\liminf_{j\to\infty} F_{h_j}(u_{h_j}) = +\infty$ the result is obvious). In view of Lemma 7 we deduce that

$$u \in SBV_{loc}(\Omega) \quad \text{and} \quad \int_\Omega |\nabla u(x)|^2\, dx + \mathcal{H}^{N-1}(S_u) < +\infty. \tag{A.6}$$

In the sequel we will drop the subscripts j and write "$h \downarrow 0$" for "$j \to \infty$". We prove (A.5) following Gobbino's method in [43], with a few modifications and adaptations. Let

$$\hat{\Omega}_h = \bigcup_{x \in h\mathbb{Z}^N \cap \Omega} x + \left[-\frac{h}{2}, \frac{h}{2}\right]^N$$

and notice that $\Omega_{\frac{1}{2}\sqrt{N}h} = \left\{x \in \Omega \ : \ \operatorname{dist}(x, \partial\Omega) > \frac{1}{2}\sqrt{N}h\right\} \subset \hat{\Omega}_h$. We have (still using the convention that we only consider in the sums the points that fall inside Ω)

$$\begin{aligned} F_h(u_h) &= h^N \sum_{x \in h\mathbb{Z}^N} \sum_{\xi \in \mathbb{Z}^N} \frac{1}{h} f_\xi\left(\frac{(u_h(x) - u_h(x+h\xi))^2}{h}\right)\phi(\xi) \\ &= \int_{\hat{\Omega}_h} dy \sum_{\xi \in \mathbb{Z}^N \cap \frac{1}{h}(\hat{\Omega}_h - y)} \frac{1}{h} f_\xi\left(\frac{(u_h(y) - u_h(y+h\xi))^2}{h}\right)\phi(\xi) \\ &= \sum_{\xi \in \mathbb{Z}^N} \phi(\xi) \int_{\hat{\Omega}_h \cap (\hat{\Omega}_h - h\xi)} \frac{1}{h} f_\xi\left(\frac{(u_h(y) - u_h(y+h\xi))^2}{h}\right) dy. \end{aligned}$$

For every $\xi \in \mathbb{Z}^N$ we let

$$\hat{F}_h(u_h, \xi) = \int_{\hat{\Omega}_h \cap (\hat{\Omega}_h - h\xi)} \frac{1}{h} f_\xi\left(\frac{(u_h(y) - u_h(y+h\xi))^2}{h}\right) dy.$$

Inequality (A.5) will follow by Fatou's lemma if we prove that for any ξ,

$$\liminf_{h\downarrow 0} \hat{F}_h(u_h, \xi) \ge \alpha_\xi \int_\Omega |\langle \nabla u(x), \xi\rangle|^2\, dx + \beta_\xi \int_{S_u} |\langle \nu_u(x), \xi\rangle|\, d\mathcal{H}^{N-1}(x). \tag{A.7}$$

We choose $A \subset\subset \Omega$. If h is small enough (i.e., $h \le \mathrm{dist}(A, \partial\Omega)/(|\xi| + \frac{1}{2}\sqrt{N})$) then

$$\hat{F}_h(u_h, \xi) \ \ge \ \int_A \frac{1}{h} f_\xi \left(\frac{(u_h(y) - u_h(y + h\xi))^2}{h} \right) dy$$

and it will be sufficient to show that

$$\liminf_{h \downarrow 0} \ \int_A \frac{1}{h} f_\xi \left(\frac{(u_h(y) - u_h(y + h\xi))^2}{h} \right) dy \ \ge \\ \ge \ \alpha_\xi \int_A |\langle \nabla u(x), \xi \rangle|^2 \, dx \ + \ \beta_\xi \int_{S_u \cap A} |\langle \nu_u(x), \xi \rangle| \, d\mathcal{H}^{N-1}(x), \tag{A.8}$$

as the supremum of the right-hand side of (A.8) for all $A \subset\subset \Omega$ is the right-hand side of (A.7). This is part of Gobbino's result [43], but we present a slightly different approach, still based on the "slicing" (see section 2.2.6 for technical details) of the functions u_h in the direction ξ.

Let $\xi^\perp = \left\{ z \in \mathbb{R}^N \ : \ \langle z, \xi \rangle = 0 \right\}$, and for every $z \in \xi^\perp$, $A_{z,\xi} = \{ s \in \mathbb{R} \ : z + s\xi \in A \}$, $(u_h)_{z,\xi}(s) = u_h(z + s\xi)$. We rewrite the first integral over A in (A.8):

$$\int_{\xi^\perp} d\mathcal{H}^{N-1}(z) \int_{A_{z,\xi}} \frac{1}{h} f_\xi \left(\frac{((u_h)_{z,\xi}(s) - (u_h)_{z,\xi}(s+h))^2}{h} \right) |\xi| \, ds \ = \\ = \ |\xi| \int_{\xi^\perp} d\mathcal{H}^{N-1}(z) \sum_{k \in \mathbb{Z}} \int_{A_{z,\xi} \cap [kh, kh+h)} \frac{1}{h} f_\xi \left(\frac{((u_h)_{z,\xi}(s) - (u_h)_{z,\xi}(s+h))^2}{h} \right) ds \\ = \ |\xi| \int_{\xi^\perp} d\mathcal{H}^{N-1}(z) \int_{[0,h)} dt \left\{ \sum_{k \in \mathbb{Z}} \frac{1}{h} f_\xi \left(\frac{((u_h)_{z,\xi}(t + kh) - (u_h)_{z,\xi}(t + (k+1)h))^2}{h} \right) \right\}$$

(by the change of variable $t + kh = s$) where the sum is taken only on the $k \in \mathbb{Z}$ such that $t + kh \in A_{z,\xi}$. Now, with the change of variable $t = h\tau$, this becomes

$$|\xi| \int_{\xi^\perp} d\mathcal{H}^{N-1}(z) \int_0^1 d\tau \left\{ h \sum_{k \in \mathbb{Z}} \frac{1}{h} f_\xi \left(\frac{((u_h)_{z,\xi}((\tau + k)h) - (u_h)_{z,\xi}((\tau + k + 1)h))^2}{h} \right) \right\}.$$

We will prove that for a.e. $(z, \tau) \in \xi^\perp \times (0,1)$,

$$\liminf_{h \downarrow 0} \ h \sum_{\begin{cases} k \in \mathbb{Z} \\ (\tau + k)h \in A_{z,\xi} \end{cases}} \frac{1}{h} f_\xi \left(\frac{((u_h)_{z,\xi}((\tau + k)h) - (u_h)_{z,\xi}((\tau + k + 1)h))^2}{h} \right) \ \ge \\ \ge \ \alpha_\xi \int_{A_{z,\xi}} |\dot{u}_{z,\xi}(x)|^2 \, dx \ + \ \beta_\xi \mathcal{H}^0(S_{u_{z,\xi}} \cap A_{z,\xi}). \tag{A.9}$$

In order to prove (A.9), we need some information on the limit of $((u_h)_{z,\xi}((\tau+k)h))_{k\in\mathbb{Z}}$ as $h\downarrow 0$. Since, using the same changes of variables,

$$\begin{aligned}\int_\Omega |u_h(y)-u(y)|^p\,dy &= \int_{\xi^\perp} d\mathcal{H}^{N-1}(z)\int_{\Omega_{z,\xi}} |(u_h)_{z,\xi}(s)-u_{z,\xi}(s)|^p\,|\xi|\,ds \\ &= |\xi|\int_{\xi^\perp} d\mathcal{H}^{N-1}(z)\int_0^1 d\tau\left\{h\sum_{k\in\mathbb{Z}}|(u_h)_{z,\xi}((\tau+k)h)\right. \\ &\qquad \left. -u_{z,\xi}((\tau+k)h)|^p\right\}\end{aligned}$$

(where in the sum we consider only k such that $(\tau+k)h\in\Omega_{z,\xi}$) we may assume (upon extracting a subsequence) that for a.e. $(z,\tau)\in\xi^\perp\times(0,1)$,

$$\lim_{h\downarrow 0} h\sum_{k\in\mathbb{Z}}|(u_h)_{z,\xi}((\tau+k)h)-u_{z,\xi}((\tau+k)h)|^p \;=\; 0. \tag{A.10}$$

Choose a (z,τ) such that (A.10) holds. By (A.6) we may also assume when choosing z that

$$u_{z,\xi}\in SBV_{loc}(\Omega_{z,\xi}) \quad\text{and}\quad \int_{\Omega_{z,\xi}}|\dot u_{z,\xi}(s)|^2\,ds\;+\;\mathcal{H}^0(S_{u_{z,\xi}})\;<\;+\infty,$$

so that $u_{z,\xi}$ is continuous except at a finite number of points. Thus, for almost all $s\in\Omega_{z,\xi}$,

$$\lim_{h\downarrow 0} u_{z,\xi}\left(\left(\tau+\left[\frac{s}{h}\right]\right)h\right)\;=\;u_{z,\xi}(s)$$

(where $[\cdot]$ denotes the integer part). We easily deduce from this and (A.10) that the piecewise constant function $v_h:\Omega_{z,\xi}\to\mathbb{R}$ defined by

$$v_h(s)=(u_h)_{z,\xi}\left(\left(\tau+\left[\frac{s}{h}\right]\right)h\right)$$

converges to $u_{z,\xi}$ in $L^p_{loc}(\Omega_{z,\xi})$.

Remark. Following Gobbino (proof of Lemma 3.3, Step 2 in [43]) we could also prove that for a.e. $\tau\in(0,1)$, $u_{z,\xi}((\tau+[s/h])h)\to u_{z,\xi}(s)$ in $L^1_{loc}(\Omega_{z,\xi})$, so that the a priori information on the regularity of u is not really needed.

We return to the proof of inequality (A.9). Notice that if $f(t)=t\wedge 1$, it is simply a consequence of Theorem 9. The proof that follows is needed because we want to consider more general functions f, and provide a generalization to such functions of the thesis of Theorem 9. For any $I \subset\subset A_{z,\xi}$, we denote

$$\begin{aligned} G(v_h, I) &= \int_I \frac{1}{h} f_\xi \left(\frac{|v_h(s+h) - v_h(s)|^2}{h} \right) ds \\ &= \sum_{k \in \mathbb{Z}} |(kh, kh+h) \cap I| \frac{1}{h} f_\xi \left(\frac{|v_h((k+1)h) - v_h(kh)|^2}{h} \right). \end{aligned}$$

If h is small enough, $(\tau + [s/h])h \in A_{z,\xi}$ for every $s \in I$ so that the lim inf in (A.9) is greater than $\liminf_{h\downarrow 0} G(v_h, I)$. Therefore, we just need to prove that for any $I \subset\subset A_{z,\xi}$,

$$\liminf_{h\downarrow 0} G(v_h, I) \ \geq \ \alpha_\xi \int_I |\dot{u}_{z,\xi}(s)|^2 \, ds \ + \ \beta_\xi \mathcal{H}^0(S_{u_{z,\xi}} \cap I); \tag{A.11}$$

indeed, taking then the lowest greater bound of the right-hand term of (A.11) for all I, we will get (A.9). Because of the super-additivity of $\liminf_{h\downarrow 0} G(v_h, \cdot)$ we may assume without loss of generality that I is an interval. To prove (A.11), we then choose $\alpha, \beta > 0$ such that $\alpha t \wedge \beta \leq f_\xi(t)$ for all $t \geq 0$ (noticing that α—respectively, β—may be chosen as close as wanted to α_ξ—resp., β_ξ), and we write

$$G(v_h, I) \ \geq \sum_{(kh, kh+h) \subset I} h\alpha \left| \frac{v_h((k+1)h) - v_h(kh)}{h} \right|^2 \wedge \beta.$$

Redefining a function $\tilde{v}_h$ with $\tilde{v}_h(kh) = v_h(kh)$ for $kh \in I$, affine on the intervals $(kh, kh+h) \subset I$ such that $h\alpha \left| \frac{v_h((k+1)h) - v_h(kh)}{h} \right|^2 \leq \beta$ and piecewise constant, jumping once on the intervals with the reverse inequality (just like in the proof of Theorem 9), we get

$$G(v_h, I) \ \geq \ \alpha \int_{I_h} |\dot{\tilde{v}}_h(s)|^2 \, ds \ + \ \beta \mathcal{H}^0(S_{\tilde{v}_h} \cap I_h)$$

with $I_h = \{x \in I \ : \ \operatorname{dist}(x, \mathbb{R} \setminus I) > h\}$, so that invoking Theorem 6 (section 2.2.6) we get the existence of a function $\tilde{v}$ such that some subsequence of $\tilde{v}_h$ goes to $\tilde{v}$ a.e., and that satisfies

$$\alpha \int_I |\dot{\tilde{v}}(s)|^2 \, ds \ + \ \beta \mathcal{H}^0(S_{\tilde{v}} \cap I) \ \leq \ \liminf_{h\downarrow 0} G(v_h, I). \tag{A.12}$$

We check then that $\tilde{v}$ has to be equal to $u_{z,\xi}$ (noticing easily, for instance, that $(v_h - \tilde{v}_h) \rightharpoonup 0$ weakly in L^p). If $\alpha \to \alpha_\xi$ we deduce from (A.12)

$$\alpha_\xi \int_I |\dot{u}_{z,\xi}(s)|^2 \, ds \ \leq \ \liminf_{h\downarrow 0} G(v_h, I), \tag{A.13}$$

whereas sending β to β_ξ we get

$$\beta_\xi \mathcal{H}^0(S_{u_{z,\xi}} \cap I) \;\le\; \liminf_{h\downarrow 0} G(v_h, I). \tag{A.14}$$

Inequality (A.11) is deduced from the last two inequalities by subdividing the interval I into suitable subintervals (the connected components of a small neighborhood of $S_{u_{z,\xi}}$ and its complement) and using the appropriate inequality (A.13) or (A.14) in each subinterval. Hence (A.9) holds, and using Fatou's lemma we deduce (A.8), as

$$|\xi| \int_{\xi^\perp} d\mathcal{H}^{N-1}(z) \left(\alpha_\xi \int_{A_{z,\xi}} |\dot u_{z,\xi}(s)|^2 \;+\; \beta_\xi \mathcal{H}^0(S_{u_{z,\xi}} \cap A_{z,\xi}) \right) \;=$$

$$=\; \alpha_\xi \int_A |\langle \nabla u(x), \xi\rangle|^2\, dx \;+\; \beta_\xi \int_{S_u \cap A} |\langle \nu_u(x), \xi\rangle| d\mathcal{H}^{N-1}(x).$$

Inequality (A.5) therefore holds in the case $u \in L^\infty(\Omega)$.

Now, if $u \in L^p(\Omega)$ is not bounded, choose again $u_{h_j} \to u$ in $L^p(\Omega)$. Consider $u^k = (-k \vee u) \wedge k$ and $u^k_{h_j} = (-k \vee u_{h_j}) \wedge k$, clearly $u^k_{h_j} \to u^k$ in $L^p(\Omega)$, so that

$$F(u^k) \;\le\; \liminf_{j\to\infty} F_{h_j}(u^k_{h_j}).$$

But as f is increasing, $F_{h_j}(u^k_{h_j}) \le F_{h_j}(u_{h_j})$ so that

$$F(u^k) \;\le\; \liminf_{j\to\infty} F_{h_j}(u_{h_j}).$$

If this is finite, we conclude by noticing that $\lim_{k\to\infty} F(u^k) = F(u)$ (by (16), (17)); so that the proof of (A.4) is achieved. □

Remark. Notice that if $u_{h_j} \to u$ in $L^p_{loc}(\Omega)$, the result still holds. Indeed, for any $A \subset\subset \Omega$ we have $u_{h_j} \to u$ in $L^p(A)$ and since the result holds in this case we can write

$$F(u, A) \;\le\; \liminf_{j\to\infty} F_{h_j}(u_{h_j}, A) \;\le\; \liminf_{j\to\infty} F_{h_j}(u_{h_j}, \Omega).$$

Then, as $F(u, \Omega) = \sup_{A\subset\subset\Omega} F(u, A)$ we get (A.5). (Thus the F_h also Γ–converge to F in $L^p(\Omega)$ endowed with the $L^p_{loc}(\Omega)$ topology.)

A.3 Estimate from above the Γ–limit

Given $u \in GSBV_{loc}(\Omega) \cap L^p(\Omega)$ with $F(u) = F(u, \Omega) < +\infty$, we want to build $u_h \in \ell^p(\Omega \cap h\mathbb{Z}^N)$ such that $u_h \to u$ in $L^p(\Omega)$ and

$$\limsup_{h \downarrow 0} F_h(u_h) \;\leq\; F(u). \tag{A.15}$$

In order to be able to assume some regularity on the function u we first prove the following lemma. It is a (simpler) variant of the results in [30] and [26] that are usually needed to show the $\Gamma - \limsup$ inequality for most approximations the Mumford–Shah functional, like Ambrosio and Tortorelli's. For F_ε, however, a very strong regularity of the jump set is not needed, and this lemma is sufficient.

Lemma 8 *Let $u \in GSBV_{loc}(\Omega) \cap L^p(\Omega)$ with $F(u) < +\infty$. There exists a sequence $(u_k)_{k \geq 1} \subseteq SBV(\Omega)$ of bounded functions with bounded supports, that are almost everywhere continuous in Ω and such that*

- *$u_k \to u$ in $L^p(\Omega)$ as k goes to infinity,*
- *$\lim_{k \to \infty} F(u_k) = F(u)$.*

Remark. The information on the support of u_k makes sense only when Ω is unbounded.

Proof. For every integer $k \geq 1$ first let $u^k = (-k \vee u) \wedge k$ be the truncated of u at level k. We choose in $L^p(\mathbb{R}^N)$ a minimizer v_k of

$$v \;\mapsto\; F(v) + k \int_{\mathbb{R}^N} |v(x) - u^k(x)|^p \, dx.$$

Then,

$$\begin{aligned}
\|v_k - u\|_{L^p(\mathbb{R}^N)} \;&\leq\; \|v_k - u^k\|_{L^p(\mathbb{R}^N)} \;+\; \|u^k - u\|_{L^p(\mathbb{R}^N)} \;\leq \\
&\leq\; \left(\frac{1}{k} F(u^k)\right)^{\frac{1}{p}} \;+\; \left(\int_{\{|u|>k\}} (|u(x)| - k)^p \, dx\right)^{\frac{1}{p}} \\
&\leq\; \left(\frac{1}{k} F(u)\right)^{\frac{1}{p}} \;+\; \left(\int_{\{|u|>k\}} |u(x)|^p \, dx\right)^{\frac{1}{p}} \to 0
\end{aligned}$$

as $k \to \infty$, moreover (see the observation about functional E' defined by (20) in section 2.3.2), we know that $\mathcal{H}^{N-1}(\Omega \cap \overline{S_{v_k}} \setminus S_u) = 0$ and $v_k \in C^1(\Omega \setminus \overline{S_{v_k}})$.

In particular v_k is almost everywhere continuous. We also have that $F(v_k) \leq F(u^k) \leq F(u)$ and $|v_k(x)| \leq k$ for all $x \in \Omega$. Set now for every integer $n > 1$ and $x \in \Omega$

$$v_{k,n}(x) = \begin{cases} v_k(x) - \dfrac{1}{n} & \text{if } v_k(x) > 1/n; \\ 0 & \text{if } |v_k(x)| \leq 1/n; \text{ and} \\ v_k(x) + \dfrac{1}{n} & \text{if } v_k(x) < -1/n. \end{cases}$$

Clearly $v_{k,n}$ is still a.e. continuous and goes to v_k in $L^p(\Omega)$ as $n \to \infty$, so that we can choose n_k such that $\|v_{k,n_k} - v_k\|_{L^p(\Omega)} \leq 1/k$. We set $w_k = v_{k,n_k}$.

We also have $S_{w_k} \subseteq S_{v_k}$,

$$\nabla w_k = \begin{cases} \nabla v_k & \text{a.e. in } \{x \in \Omega \,:\, |v_k(x)| > 1/n_k\}; \\ 0 & \text{a.e. in the complement,} \end{cases}$$

and

$$|\{w_k \neq 0\}| = \left| \left\{ |v_k| > \frac{1}{n_k} \right\} \right| \leq n_k^p \int_\Omega |v_k(x)|^p \, dx < +\infty$$

so that in particular $w_k \in L^q(\Omega)$ for any $q \in [1, +\infty]$.

Choose at last $\zeta \in C_0^\infty(\mathbb{R}^N)$ with $0 \leq \zeta \leq 1$ and $\zeta \equiv 1$ on $B_1(0)$, and set for $R > 0$ and any $x \in \Omega$ $w_{k,R}(x) = \zeta\left(\frac{x}{R}\right) w_k(x)$. For any R,

$$S_{w_{k,R}} \subseteq S_{w_k} \subseteq S_{v_k}$$

and if $\xi \in \mathbb{Z}^N$,

$$\begin{aligned} \int_\Omega |\langle \nabla w_{k,R}(x), \xi \rangle|^2 \, dx \, = \\ = \; & \int_{B_R(0) \cap \Omega} |\langle \nabla w_k(x), \xi \rangle|^2 \, dx \\ & + \int_{\Omega \setminus B_R(0)} \left| \zeta\left(\frac{x}{R}\right) \langle \nabla w_k(x), \xi \rangle + \frac{w_k(x)}{R} \left\langle \nabla \zeta\left(\frac{x}{R}\right), \xi \right\rangle \right|^2 \, dx \\ & \leq \int_{B_R(0) \cap \Omega} |\langle \nabla v_k(x), \xi \rangle|^2 \, dx \\ & + 2 \int_{\Omega \setminus B_R(0)} |\langle \nabla v_k(x), \xi \rangle|^2 \, dx + \frac{C}{R^2} |\xi|^2 \int_{\Omega \setminus B_R(0)} |w_k(x)|^2 \, dx \end{aligned}$$

with $C = 2\|\nabla\zeta\|^2_{L^\infty(\mathbb{R}^N)}$. Hence

$$F(w_{k,R}) \le F(v_k) + \left(\sum_{\xi\in\mathbb{Z}^N} \alpha_\xi \phi(\xi)|\xi|^2\right)\left\{\int_{\Omega\setminus B_R(0)} |\nabla v_k(x)|^2\,dx + \frac{C}{R^2}\int_{\Omega\setminus B_R(0)} |w_k(x)|^2\,dx\right\}.$$

Since w_k and ∇v_k are in $L^2(\Omega)$, we can choose R large enough in order to have

$$F(w_{k,R}) \le F(v_k) + \frac{1}{k}. \tag{A.16}$$

Choose R_k large enough so that (A.16) holds and $\|w_{k,R_k} - w_k\|_{L^p(\Omega)} \le 1/k$, and set $u_k = w_{k,R_k}$. Clearly u_k is still a.e. continuous. Moreover, $F(u_k) \le F(u)+1/k$, u_k goes to u in $L^p(\Omega)$ as $k\to\infty$, and by Theorem 6 (section 2.2.6) we deduce that

$$F(u) \le \liminf_{k\to\infty} F(u_k)$$

so that $\lim_{k\to\infty} F(u_k) = F(u)$ and the lemma is true. □

We now establish (A.15). First consider the case $\Omega = \mathbb{R}^N$. Given $u \in GSBV_{loc}(\mathbb{R}^N) \cap L^p(\mathbb{R}^N)$ with $F(u) < +\infty$, we build invoking Lemma 8 a sequence of compactly supported, bounded and a.e. continuous functions u_k converging to u such that $F(u_k)\to F(u)$ as k goes to infinity. By a standard diagonalization procedure, if we know how to build for every k a sequence $((u_k)_h)_{h>0}$ converging to u_k in $L^p(\mathbb{R}^N)$ as $h\downarrow 0$, such that

$$\limsup_{h\downarrow 0} F_h((u_k)_h) \le F(u_k),$$

we will be able to find u_h with $u_h\to u$ and satisfying (A.15). In the sequel we may therefore assume that u is bounded, compactly supported, and continuous at almost every $x \in \mathbb{R}^N$.

For $y \in (0,h)^N$ define $u_h^y \in \ell^p(h\mathbb{Z}^N)$ by $u_h^y(x) = u(y+x)$ for any $x \in h\mathbb{Z}^N$. We compute the mean of $F_h(u_h^y)$ over $(0,h)^N$:

$$\begin{aligned} h^{-N}\int_{(0,h)^N} F_h(u_h^y)\,dy &= \sum_{\xi\in\mathbb{Z}^N}\phi(\xi)\sum_{x\in h\mathbb{Z}^N}\int_{(0,h)^N}\frac{1}{h}f_\xi\left(\frac{(u(y+x)-u(y+x+h\xi))^2}{h}\right)dy \\ &= \sum_{\xi\in\mathbb{Z}^N}\phi(\xi)\int_{\mathbb{R}^N}\frac{1}{h}f_\xi\left(\frac{(u(y)-u(y+h\xi))^2}{h}\right)dy. \end{aligned}$$

At this point (following exactly Gobbino's proof), we write

$$\int_{\mathbb{R}^N} \frac{1}{h} f_\xi \left(\frac{(u(y)-u(y+h\xi))}{h}^2 \right) dy \; =$$

$$\begin{aligned}
&= \int_{\xi^\perp} d\mathcal{H}^{N-1}(z) \int_{\mathbb{R}} dt\, \frac{1}{h} f_\xi \left(\frac{(u(z+t\frac{\xi}{|\xi|}) - u(z+t\frac{\xi}{|\xi|}+h\xi))}{h}^2 \right) \\
&= |\xi| \int_{\xi^\perp} d\mathcal{H}^{N-1}(z) \int_{\mathbb{R}} ds\, \frac{1}{h} f_\xi \left(\frac{(u(z+s\xi) - u(z+(s+h)\xi))}{h}^2 \right) \\
&= |\xi| \int_{\xi^\perp} d\mathcal{H}^{N-1}(z) F^1_{\xi,h}(u_{z,\xi})
\end{aligned}$$

where $u_{z,\xi}(s) = u(z+s\xi)$ and we have set

$$F^1_{\xi,h}(v) \; = \; \int_{\mathbb{R}} \frac{1}{h} f_\xi \left(\frac{(v(s)-v(s+h))}{h}^2 \right) ds$$

for any measurable function v. Since we assumed $f_\xi(t) \le \alpha_\xi t \wedge \beta_\xi$, we have

$$F^1_{\xi,h}(v) \; \le \; \int_{\mathbb{R}} \alpha_\xi \left| \frac{v(s)-v(s+h)}{h} \right|^2 \wedge \frac{\beta_\xi}{h}\, ds \tag{A.17}$$

and as shown in [43] by M. Gobbino, this is less than

$$\alpha_\xi \int_{\mathbb{R}} |\dot{v}(s)|^2\, ds \; + \; \beta_\xi \mathcal{H}^0(S_v)$$

provided $v \in SBV_{loc}(\mathbb{R})$ and this expression is finite.

Exercise. *Check this fact, by computing the integral in (A.17) separately over $S_v^h = \bigcup_{s\in S_v}[s-h, s]$ and over $\mathbb{R}\setminus S_v^h$.*

Therefore,

$$h^{-N} \int_{(0,h)^N} F_h(u_h^y)\, dy \; \le$$

$$\begin{aligned}
&\le \sum_{\xi\in\mathbb{Z}^N} \phi(\xi)|\xi| \int_{\xi^\perp} d\mathcal{H}^{N-1}(z) F^1_{\xi,h}(u_{z,\xi}) \\
&\le \sum_{\xi\in\mathbb{Z}^N} \phi(\xi)|\xi| \int_{\xi^\perp} d\mathcal{H}^{N-1}(z) \left\{ \alpha_\xi \int_{\mathbb{R}} |\langle \nabla u(z+s\xi), \xi\rangle|^2\, ds \; + \; \beta_\xi \mathcal{H}^0(S_{u_{z,\xi}}) \right\} \\
&= \sum_{\xi\in\mathbb{Z}^N} \phi(\xi) \left\{ \int_{\mathbb{R}^N} \alpha_\xi |\langle \nabla u(x), \xi\rangle|^2\, dx \; + \; \int_{S_u} \beta_\xi |\langle \nu_u(x), \xi\rangle|\, d\mathcal{H}^{N-1}(x) \right\} \; = \; F(u).
\end{aligned}$$

Thus, for y in some set of positive measure in $(0,h)^N$,

$$F_h(u_h^y) \leq F(u). \tag{A.18}$$

For all h we choose y_h such that inequality (A.18) holds and set $u_h = u_h^{y_h}$. We easily check that if u is continuous at $x \in \mathbb{R}^N$ then $u_h(x) \to u(x)$ as $h \downarrow 0$ (since $u_h(x) = u(x')$ for some x' such that $|x - x'| < \frac{3}{2}\sqrt{N}h$). Since u is almost everywhere continuous, u_h converges to u a.e. in $\mathbb{R}^N$. We also have $\|u_h\|_{L^\infty(\mathbb{R}^N)} \leq \|u\|_{L^\infty(\mathbb{R}^N)}$ and the functions u_h, u are zero outside some compact set so that by Lebesgue's theorem $u_h \to u$ in $L^p(\mathbb{R}^N)$. Since clearly, (A.15) holds for this sequence u_h, the proof of the case $\Omega = \mathbb{R}^N$ is achieved.

We now return to the general case where Ω is a Lipschitz domain. The method used in order to localize the previous result is adapted from [23]. We choose a function $u \in GSBV_{loc}(\Omega) \cap L^p(\Omega)$, and once again invoking Lemma 8 we see that it is not restrictive to assume that u is bounded with bounded support. Since we assumed that $\partial\Omega$ is Lipschitz, (and since u is zero outside some bounded set) we can extend u outside of Ω (using the same reflection procedure as for instance in [34] for the extension of $W^{1,p}$ functions) into a bounded compactly supported SBV function (still denoted by u) such that $\mathcal{H}^{N-1}(\partial\Omega \cap S_u) = 0$ and $F(u, \mathbb{R}^N) < +\infty$. Then, we build (u_h) like previously, such that u_h goes to u in $L^p(\mathbb{R}^N)$ and

$$\limsup_{h\downarrow 0} F_h(u_h, \mathbb{R}^N) \leq F(u, \mathbb{R}^N).$$

We can write

$$F_h(u_h, \mathbb{R}^N) \geq F_h(u_h, \Omega) + F_h(u_h, \overline{\Omega}^c)$$

where $\overline{\Omega}^c$ is the complement of $\overline{\Omega}$ in $\mathbb{R}^N$. Notice that we have dropped all terms involving differences of values of u_h at one point in $\overline{\Omega}$ and another in Ω^c. Sending h to zero we get

$$\limsup_{h\downarrow 0} F_h(u_h, \mathbb{R}^N) \geq \limsup_{h\downarrow 0} F_h(u_h, \Omega) + \liminf_{h\downarrow 0} F_h(u_h, \overline{\Omega}^c),$$

and we deduce from (A.4) that

$$\limsup_{h\downarrow 0} F_h(u_h, \Omega) + F(u, \overline{\Omega}^c) \leq \limsup_{h\downarrow 0} F_h(u_h, \mathbb{R}^N) \leq F(u, \mathbb{R}^N).$$

Thus, u being extended in such a way that $F(u, \overline{\Omega}^c) < +\infty$,

$$\limsup_{h\downarrow 0} F_h(u_h, \Omega) \leq F(u, \overline{\Omega}).$$

Since $\mathcal{H}^{N-1}(\partial\Omega \cap S_u) = 0$, $F(u, \overline{\Omega}) = F(u, \Omega)$ and we get the thesis. This achieves the proof of Theorem 11. □

A.4 Proof of Theorem 12

For any $h > 0$ let $(u_h)_{h>0}$ be a minimizer in $\ell^p(\Omega \cap h\mathbb{Z}^N)$ of

$$F_h(u) \;+\; \int_\Omega |u(x) - g(x)|^p \, dx \tag{A.19}$$

where $g \in L^\infty(\Omega) \cap L^p(\Omega)$.

Replacing u_h with $(-\|g\|_{L^\infty(\Omega)} \vee u_h) \wedge \|g\|_{L^\infty(\Omega)}$ we decrease the energy, thus in fact $\|u_h\|_{L^\infty(\Omega)} \le \|g\|_{L^\infty(\Omega)}$. In view of Lemma 7, since $\sup_{h>0} F_h(u_h) < +\infty$, some subsequence $(u_{h_j})_{j\ge1}$ of $(u_h)_{h>0}$ converges to a function $u \in SBV_{loc}(\Omega)$ a.e. in Ω. From the uniform bound on $\|u_h\|_\infty$ we deduce that $u_{h_j} \to u$ in $L^p_{loc}(\Omega)$.

If $|\Omega| < +\infty$, the convergence is in $L^p(\Omega)$ and we simply conclude invoking Theorem 7 (section 2.4). Otherwise, we know (by the remark at the end of section A.2 and Fatou's lemma) that

$$F(u) \;+\; \int_\Omega |u(x) - g(x)|^p \, dx \;\le\; \liminf_{j\to\infty} F_{h_j}(u_{h_j}) \;+\; \int_\Omega |u_{h_j}(x) - g(x)|^p \, dx.$$

For any $v \in L^p(\Omega)$, we consider $(v_{h_j})_{j\ge1}$ a sequence converging to v in $L^p(\Omega)$ such that

$$\limsup_{j\to\infty} F_{h_j}(v_{h_j}) \;\le\; F(v).$$

For all j we have that

$$F_{h_j}(u_{h_j}) \;+\; \int_\Omega |u_{h_j}(x) - g(x)|^p \, dx \le F_{h_j}(v_{h_j}) \;+\; \int_\Omega |v_{h_j}(x) - g(x)|^p \, dx,$$

so that at the limit we get

$$F(u) \;+\; \int_\Omega |u(x) - g(x)|^p \, dx \le F(v) \;+\; \int_\Omega |v(x) - g(x)|^p \, dx,$$

showing the minimality of u. If we choose $v = u$, we also deduce that

$$\lim_{j\to\infty} \|u_{h_j} - g\|_{L^p(\Omega)} \;=\; \|u - g\|_{L^p(\Omega)},$$

thus, by equi-integrability, $u_{h_j} \to u$ strongly in $L^p(\Omega)$, since we had the convergence in $L^p_{loc}(\Omega)$. In the case where we minimize

$$F_h(u) \;+\; \left(\|u - g^h\|_p\right)^p$$

instead of (A.19) the proof is not different. □

References

[1] G. Alberti. Variational models for phase transitions, an approach via gamma-convergence. In G. Buttazzo *et al.*, editor, *Differential Equations and Calculus of Variations*. Springer–Verlag, 2000. (Also available at `http://cvgmt.sns.it/papers/`).

[2] L. Ambrosio. A compactness theorem for a new class of functions with bounded variation. *Boll. Un. Mat. Ital. (7)*, 3-B:857–881, 1989.

[3] L. Ambrosio. Variational problems in SBV and image segmentation. *Acta Appl. Math.*, 17:1–40, 1989.

[4] L. Ambrosio. Existence theory for a new class of variational problems. *Arch. Rat. Mech. Anal.*, 111(1):291–322, 1990.

[5] L. Ambrosio. A new proof of the SBV compactness theorem. *Calc. Var. Partial Differential Equations*, 3(1):127–137, 1995.

[6] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford mathematical monographs. Oxford Clarendon Press, 2000.

[7] L. Ambrosio and V.M. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via Γ-convergence. *Comm. Pure Appl. Math.*, 43(8):999–1036, 1990.

[8] L. Ambrosio and V.M. Tortorelli. On the approximation of free discontinuity problems. *Boll. Un. Mat. Ital. (7)*, 6-B:105–123, 1992.

[9] H. Attouch. *Variational convergence for functions and operators*. Applicable Mathematics Series. Pitman (Advanced Publishing Program), Boston, Mass.–London, 1984.

[10] G. Aubert, M. Barlaud, P. Charbonnier, and L. Blanc-Féraud. Deterministic edge-preserving regularization in computed imaging. Technical Report TR#94-01, I3S, CNRS URA 1376, Sophia-Antipolis, France, 1994.

[11] R. Azencott. Image analysis and Markov fields. In *SIAM Proceedings of 1st Int. Conf. Appl. Math., Paris*, 1987.

[12] R. Azencott. Markov fields and image analysis. In *Proceedings AFCET Congress, Antibes, France*, 1987.

[13] G. Bellettini and A. Coscia. Discrete approximation of a free discontinuity problem. *Numer. Funct. Anal. Optim.*, 15(3-4):201–224, 1994.

[14] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.

[15] L. Blanc-Féraud and M. Barlaud. Restauration d'images bruitées par analyse multirésolution et champs de Markov. In *proceedings of "13^e colloque GRETSI sur le traitement du signal et des images", Juan-les-Pins, France*, pages 829–832, 1991.

[16] B. Bourdin. Image segmentation with a finite element method. *M2AN Math. Model. Numer. Anal.*, 33(2):229–244, 1999.

[17] B. Bourdin and A. Chambolle. Implementation of a finite-elements approximation of the Mumford–Shah functional. *Numer. Math.*, 85(4):609–646, 2000.

[18] A. Braides. *Approximation of free-discontinuity problems*. Number 1694 in Lecture Notes in Mathematics. Springer–Verlag, Berlin, 1998.

[19] A. Braides and G. Dal Maso. Non-local approximation of the Mumford–Shah functional. *Calc. Var. Partial Differential Equations*, 5(4):293–322, 1997.

[20] A. Chambolle. Un théorème de Γ–convergence pour la segmentation des signaux. *C. R. Acad. Sci. Paris*, t. 314 Série I:191–196, 1992.

[21] A. Chambolle. Image segmentation by variational methods: Mumford and Shah functional and the discrete approximations. *SIAM J. Appl. Math.*, 55(3):827–863, 1995.

[22] A. Chambolle. Finite-differences discretizations of the Mumford–Shah functional. *M2AN Math. Model. Numer. Anal.*, 33(2):261–288, 1999.

[23] A. Chambolle and G. Dal Maso. Discrete approximation of the Mumford–Shah functional in dimension two. *M2AN Math. Model. Numer. Anal.*, 33(4):651–672, 1999.

[24] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numer. Math.*, 76(2):167–188, 1997.

[25] T. F. Chan, G. H. Golub, and P. Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM J. Sci. Comput.*, 20(6):1964–1977, 1999.

[26] G. Cortesani. Strong approximation of $GSBV$ functions by piecewise smooth functions. *Ann. Univ. Ferrara Sez. VII (N.S.)*, 43:27–49, 1997.

[27] G. Dal Maso. *An introduction to* Γ*-convergence*. Birkhäuser, Boston, 1993.

[28] G. Dal Maso, J.-M. Morel, and S. Solimini. A variational method in image segmentation: Existence and approximation results. *Acta Math.*, 168:89–151, 1992.

[29] E. De Giorgi, M. Carriero, and A. Leaci. Existence theorem for a minimum problem with free discontinuity set. *Arch. Rational Mech. Anal.*, 108:195–218, 1989.

[30] F. Dibos and E. Séré. An approximation result for the minimizers of Mumford–Shah functional. *Boll. Un. Mat. Ital. (7)*, 11-A:149–162, 1997.

[31] D. C. Dobson and C. R. Vogel. Convergence of an iterative method for total variation denoising. *SIAM J. Numer. Anal.*, 34(5):1779–1791, 1997.

[32] R. C. Dubes, A. K. Jain, S. G. Nadabar, and C. C. Chen. MRF model-based algorithms for image segmentation. In *proc. 10th IEEE Int. Conf on Pattern Recognition, Atlantic City*, pages 808–814, 1990.

[33] S. Durand, F. Malgouyres, and B. Rougé. Image de-blurring, spectrum interpolation and application to satellite imaging. Technical Report 9916, CMLA, ENS Cachan, 1999.

[34] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[35] K. J. Falconer. *The geometry of fractal sets*. Cambridge University Press, Cambridge, 1985.

[36] H. Federer. *Geometric Measure Theory*. Springer–Verlag, New–York, 1969.

[37] S. Finzi-Vita and P. Perugia. Some numerical experiments on the energy minimization problem. In *Proc. of the Second European Workshop on Image Processing and Mean Curvature Motion*, pages 233–240, Palma de Mallorca, September 1995.

[38] D. Geiger and F. Girosi. Parallel and deterministic algorithms for MRFs: surface reconstruction. *IEEE Trans. PAMI*, PAMI-13(5):401–412, May 1991.

[39] D. Geiger and A. Yuille. A common framework for image segmentation. *Internat. J.Comp. Vision*, 6(3):227–243, August 1991.

[40] D. Geman and G. Reynolds. Constrained image restoration and the recovery of discontinuities. *IEEE Trans. PAMI*, PAMI-3(14):367–383, 1992.

[41] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. PAMI*, PAMI-6(6), November 1984.

[42] E. Giusti. *Minimal surfaces and functions of bounded variation.* Birkhäuser, Boston, 1984.

[43] M. Gobbino. Finite difference approximation of the Mumford–Shah functional. *Comm. Pure Appl. Math.*, 51(2):197–228, 1998.

[44] F. Guichard and F. Malgouyres. Total variation based interpolation. In *Proceedings of the European Signal Processing Conference*, volume 3, pages 1741–1744, 1998.

[45] S. Z. Li. *Markov Random Field Modeling in Computer Vision.* Spinger Verlag, 1995. (see also `http://markov.eee.ntu.ac.sg:8000/~szli/MRF_Book/MRF_Book.html`).

[46] P.-L. Lions, S. J. Osher, and L. Rudin. Denoising and deblurring using constrained nonlinear partial differential equations. Technical report, Cognitech Inc., Santa Monica, CA, 1992.

[47] F. Malgouyres and F. Guichard. Edge direction preserving image zooming: a mathematical and numerical analysis. Technical Report 9930, CMLA, ENS Cachan, 1999.

[48] J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation*. Birkhäuser, Boston, 1995.

[49] D. Mumford and J. Shah. Boundary detection by minimizing functionals, I. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition, San Francisco*, 1985. (also *Image Understanding*, 1988).

[50] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.

[51] M. Nitzberg, D. Mumford, and T. Shiota. *Filtering, segmentation and depth*. Number 662 in Lecture Notes in Computer Science. Springer–Verlag, Berlin, 1993.

[52] T. J. Richardson and S. K. Mitter. A variational formulation-based edge focussing algorithm. *Sādhanā*, 22(4):553–574, 1997.

[53] L. Rudin and S. J. Osher. Total variation based image restoration with free local constraints. In *Proceedings of the IEEE ICIP'94*, volume 1, pages 31–35, Austin, Texas, 1994.

[54] L. Rudin, S. J. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D.*, 60:259–268, 1992. [also in *Experimental Mathematics: Computational Issues in Nonlinear Science* (Proc. Los Alamo Conf. 1991)].

[55] L. Schwartz. *Théorie des distributions*. Hermann, Paris, 1966.

[56] J. Shah. Properties of segmentations which minimize energy functionals. *(Preprint Northeastern Univ. Math. Dept., Boston)*, December 1988.

[57] J. Shah. Parameter esimation, multiscale representation and algorithms for energy-minimizing segmentations. In *proc. 10th IEEE Int. Conf. on Pattern Recognition, Atlantic City*, pages 815–819, 1990.

[58] J. Shah. Segmentation by nonlinear diffusion. In *proc. IEEE Comp. Soc. Conf. on Pattern Recognition*, pages 202–207, 1991.

[59] J. Shah. Segmentation by nonlinear diffusion, II. In *proc. IEEE Comp. Soc. Conf. on Pattern Recognition, Champaign, IL*, pages 644–647, June 1992.

[60] C. R. Vogel and M. E. Oman. Iterative methods for total variation denoising. *SIAM J. Sci. Comput*, 17(1):227–238, 1996. Special issue on iterative methods in numerical linear algebra (Breckenridge, CO, 1994).

[61] C. R. Vogel and M. E. Oman. Fast, robust total variation-based reconstruction of noisy, blurred images. *IEEE Trans. Image Process.*, 7(6):813–824, 1998.

[62] W. P. Ziemer. *Weakly Differentiable Functions*. Springer–Verlag, Berlin, 1989.