Reconstruction of piecewise constant images via total variation regularization: exact support recovery and grid-free numerical methods

Romain Petit

PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Université Paris Dauphine

Reconstruction of piecewise constant images via total variation regularization: exact support recovery and grid-free numerical methods

Soutenue par
Romain PETIT
Le 12 Décembre 2022

École doctorale n°543
École doctorale SDOSE

Spécialité
Mathématiques

Composition du jury :

Gabriel Peyré Directeur de recherche CNRS & École normale supérieure, PSL	*Président*
Marc Dambrine Professeur Université de Pau et des Pays de l'Adour	*Rapporteur*
Otmar Scherzer Professeur University of Vienna	*Rapporteur*
Blanche Buet Maîtresse de conférences Université Paris-Saclay	*Examinatrice*
Antonin Chambolle Directeur de recherche CNRS & Université Paris Dauphine, PSL	*Examinateur*
Simon Masnou Professeur Université Claude Bernard Lyon 1	*Examinateur*
Vincent Duval Chargé de recherche INRIA & Université Paris Dauphine, PSL	*Directeur de thèse*
Yohann De Castro Professeur École Centrale Lyon	*Co-directeur de thèse*

Dauphine | PSL UNIVERSITÉ PARIS

Remerciements

À Vincent qui a eu bien du courage de s'embarquer dans cette galère.

À Yohann pour m'avoir suivi depuis Lyon (et pour m'avoir poussé dans le bureau de Vincent).

À Marc Dambrine et Otmar Scherzer pour leurs précieuses relectures.

À Blanche Buet, Antonin Chambolle, Simon Masnou et Gabriel Peyré pour avoir accepté d'examiner cette thèse.

À Jimmy Lamboley pour sa relecture attentive du manuscrit et sa participation à la soutenance.

Aux élèves du département MIDO pour avoir supporté mes explications bancales.

À tous les collègues et amis de MOKAPLAN, du CEREMADE et d'ailleurs.

Aux copains pour leur créativité débordante.

À dame LQ.

À Claude, Christine et Juliette (et aux futurs docteurs Petit).

Et à Loucas Pillaud-Vivien pour son template maison.

Extended abstract

This thesis is devoted to a family of inverse problems, which consist in recovering an unknown image from a set of partial measurements, possibly corrupted by some noise. Although the resolution of ill-posed inverse problems is a classical topic in signal processing, a series of works conducted over the past thirty years has generated considerable interest. They focus on the reconstruction of so-called sparse signals using variational methods. The latter rely on a functional, called the regularizer, whose minimization promotes solutions presenting the same structure as the sought-after signal.

In this thesis, our aim is to analyze a specific choice of regularizer, known as the total (gradient) variation. This functional has been extensively used in imaging since the pioneering works of Rudin Osher and Fatemi, conducted in 1992. It is mostly known for its ability to penalize oscillations while preserving discontinuities. Its minimization is known to produce piecewise constant images, which are made of a few simple shapes and hence have some form of sparsity. However, the performance of this regularizer has to our knowledge not extensively been studied from a sparse recovery viewpoint.

This thesis aims at bridging this gap. First, we consider the reconstructions obtained by minimizing the total variation in a low noise regime, and study their proximity with the sought-after image. As the latter is assumed to be sparse (i.e. made of a small number of simple shapes), we focus on the structure of the reconstructed image, and provide answers to the following questions: is it itself sparse, made of the same number of shapes, and are these shapes close to those appearing in the unknown image. Then, we address the numerical resolution of variational problems associated to this regularizer. Existing methods rely on the introduction of a fixed spatial discretization, which often yield reconstruction artifacts such as anisotropy or blur. Adapting recent ideas introduced in the context of sparse spikes recovery, we propose an algorithm which does not suffer from this grid bias, and produces a sparse representation of the reconstructed image.

Our contributions can also be considered from a calculus of variations viewpoint. We are indeed concerned with the theoretical analysis and the numerical resolution of a given variational problem. In part II, we study the convergence of its solutions as a parameter involved in its definition converges to zero. We obtain results that significantly differ from those usually derived in the field of calculus of variations. Namely, we prove the convergence of the level sets of solutions in terms of smooth normal deformations, and also show the number of non-trivial level sets (and their number of connected components) is preserved. In part III, we propose a numerical method for solving the considered problem. It does not rely on the resolution of a finite dimensional problem associated to a fixed spatial discretization. Our algorithm produces approximate solutions which have the same structure as some solutions of interest.

Keywords: inverse problems, total variation, sparsity.

Résumé détaillé

On s'intéresse dans cette thèse à une famille de problèmes inverses, qui consistent à reconstruire une image à partir de mesures partielles et possiblement bruitées. Si la résolution de problèmes inverses mal posés est une thématique classique en traitement du signal, une série de travaux menés depuis une trentaine d'années a particulièrement marqué le domaine. Ces derniers concernent la reconstruction de signaux dits parcimonieux à l'aide de méthodes variationnelles, qui reposent sur l'utilisation d'une fonctionnelle, appelée régulariseur, dont la minimisation produit des solutions de même structure que le signal inconnu.

La principale contribution de cette thèse est l'étude d'un choix particulier de régulariseur, la variation totale (du gradient). Cette fonctionnelle est utilisée en imagerie depuis les travaux de Rudin Osher et Fatemi, menés en 1992, notamment pour sa capacité à pénaliser les oscillations tout en préservant les discontinuités. Alors qu'il est bien connu que sa minimisation produit des images constantes par morceaux, présentant une forme de parcimonie (elles sont composées d'un petit nombre de formes simples), ce point de vue n'a à notre connaissance pas été privilégié pour analyser les performances de ce régulariseur.

Dans cette thèse, on se propose de mener cette étude. Dans la partie II, on considère les reconstructions obtenues par minimisation de la variation totale dans un régime de faible bruit, et on étudie leur proximité avec l'image inconnue. Cette dernière étant supposée parcimonieuse (c'est à dire composée d'un petit nombre de formes simples), on s'intéresse particulièrement à la structure de la reconstruction: est-elle elle-même parcimonieuse, est-elle composée du même nombre de formes, et ces formes sont-elles proches de celles présentes dans l'image inconnue. Dans la partie III, on propose une méthode numérique pour résoudre les problèmes variationnels associés à ce régulariseur. On adapte des travaux récents sur la reconstruction de mesures discrètes, afin d'obtenir un algorithme ne reposant pas sur l'introduction d'une discrétisation spatiale fixe. Ceci a l'avantage, contrairement aux techniques existantes, de n'introduire ni flou ni anisotropie dans les images reconstruites, et d'en produire une représentation parcimonieuse.

Les contributions de cette thèse peuvent aussi être considérées du point de vue du calcul des variations. On s'intéresse en effet à une problème variationnel, qu'on analyse théoriquement et pour lequel on propose une méthode numérique de résolution. Dans la partie II, on s'intéresse à la convergence des solutions de ce problème lorsque l'un des paramètres le définissant tend vers zéro. Le type de convergence qu'on obtient est bien plus fort que ceux considérés usuellement en calcul des variations. Plus précisément, on montre la convergence des ensembles de niveau des solutions en termes de déformations normales, ainsi que celle du nombre d'ensemble de niveaux non triviaux (et de leur nombre de composantes connexes). Dans la partie III, on résout ce problème numériquement sans se ramener à un problème de dimension finie via l'introduction d'une discrétisation spatiale fixe. On propose une méthode itérative qui construit des solutions approximées ayant la même structure que certaines solutions d'intérêt.

Mots-clés: problèmes inverses, variation totale, parcimonie.

Contents

Contributions and outline

Part I. We begin this manuscript with a general introduction to inverse problems and variational regularization techniques. We then introduce the total (gradient) variation, which is the specific regularizer this thesis is dedicated to. We review existing theoretical results on total variation regularization, and discuss numerical methods for solving the associated variational problems. We argue that these subjects have not extensively been studied from a sparse recovery viewpoint, and state that the main aim of this thesis is to conduct such an analysis. We also stress that our contributions are attempts to adapt recently introduced ideas for sparse spikes recovery in a continuous domain.

Part II. This part is devoted to the noise robustness analysis of total variation regularization, and more specifically to the proof of an *exact support recovery* result.

- **Piecewise constant functions.** The main purpose of this section is to introduce the class of sparse objects we aim at recovering, which are piecewise constant functions with a few specific properties. We call them M*-simple functions*. We first begin by collecting results on the faces of the total variation unit ball, which were first proved in [Duval, 2022]. As we focus on exposed faces (instead of general linearly closed faces in the above-mentioned work), we are able to provide slightly different (and often simpler) proofs of some results. It turns out that the elements of finite dimensional exposed faces are M-simple functions. This analysis in particular allows to decompose M-simple functions belonging to the same face using a common set of atoms, which is crucial to prove our support recovery result. As a byproduct, we are also able to state an abstract sufficient identifiability condition, which can be seen as a *strengthened source condition*.

- **The prescribed curvature problem.** We then focus on a geometric variational problem called the *prescribed curvature problem*, which naturally appears in the analysis conducted in the previous section. Our aim is to study the behaviour of its solutions under variations of the associated curvature functional. We prove that, for two sufficiently close curvature functionals, solutions of the two problems are close in terms of C^2 normal deformations. Assuming some solution of one problem is *strictly stable*, we show that it has a neighborhood (in terms of C^2 normal deformations) which contains at most one solution of the other problem.

- **Exact support recovery.** Using the results of the first two sections, we are finally able to prove our support recovery result. To achieve this, we first show that the dimension of the exposed faces of the total variation unit ball is stable in some sense. We then define a so-called *non-degenerate source condition*, under which exact support recovery is guaranteed.

We stress that the term *exact* refers to the estimation of some measure of sparsity of the unknown image. The reconstructed image is made of the same number of shapes, each a C^2 normal deformation of a shape appearing in the sougt-after image.

Part III. This third part is dedicated to the numerical resolution of variational problems associated to total variation regularization. We propose a grid-free approach based on the Frank-Wolfe (or conditional gradient) algorithm. A complete implementation of the proposed method is available at https://github.com/rpetit/pycheeger and https://github.com/rpetit/tvsfw.

- **Frank-Wolfe approach.** In this first section, we describe the theoretical algorithm we propose to solve our variational problem of interest. We explain how the linear minimization step, which is at the core of Frank-Wolfe algorithm, is closely linked to a shape optimization problem called the (generalized) *Cheeger problem*. We then turn to convergence results, before discussing the interest of a special final update, called the *sliding step*.
- **Polygonal approximation of generalized Cheeger sets.** This section is devoted to the numerical approximation of solutions of the (generalized) Cheeger problem. We propose to optimize its objective over simple polgons with a fixed number of sides, and prove the existence of optimizers for this new problem. We then discuss a two-step numerical method to approach them. Finally, we discuss how polygonal maximizers compare to their continuous counterpart in the specific case of radial weight functions.
- **Numerical results.** We begin this last section by discussing the implementation of the sliding step. We describe the iterative method we propose to locally optimize the objective, and also mention the issue of *topology changes*, which might appear over the course of this evolution. We conclude this part by assessing the performance of our algorithm on a few recovery examples. We provide comparisons with two standard fixed-grid methods, and discuss parameter choices.

Part IV. We conclude this manuscript by discussing interesting avenues for future research.

Publications. This thesis gave rise to two articles, which are listed below.

- **[De Castro et al., in preparation] Exact recovery of the support of piecewise constant images via total variation regularization**, Y. De Castro, V. Duval and R. Petit, *in preparation*.
- **[De Castro et al., 2022] Towards off-the-grid algorithms for total variation regularized inverse problems**, Y. De Castro, V. Duval and R. Petit, *Journal of Mathematical Imaging and Vision*, 2022

A short version of the second article appeared in the proceedings of the 8th international conference on scale space and variational methods in computer vision (SSVM 2021) [De Castro et al., 2021].

Notations

Functions spaces. Let $m, n \in \mathbb{N}^*$ and $\Omega \subset \mathbb{R}^m$ be an open set. For any $k \in \mathbb{N}$, we denote by $\mathrm{C}^k(\Omega, \mathbb{R}^n)$ ($\mathrm{C}^k(\Omega)$ if $n = 1$) the set of $\mathbb{R}^n$-valued functions of class C^k on Ω, and by $\mathrm{C}^k_b(\Omega, \mathbb{R}^n)$ ($\mathrm{C}^k_b(\Omega)$ if $n = 1$) the set of functions $u \in \mathrm{C}^k(\Omega, \mathbb{R}^n)$ that satisfy $\|u\|_{\mathrm{C}^k(\Omega)} < +\infty$ with

$$\|u\|_{\mathrm{C}^k(\Omega)} \overset{\text{def.}}{=} \sum_{i=1}^{k} \|D^i u\|_\infty \,,$$

where $D^i u$ denotes the i-th derivative of u. If a function u is bounded and uniformly continuous on Ω, it has a unique continuous extension to $\overline{\Omega}$. We denote by $\mathrm{C}^k(\overline{\Omega})$ the space of functions in $\mathrm{C}^k(\Omega)$ with bounded and uniformly continuous derivatives up to order k.

Hausdorff measure. We denote by $\mathcal{H}^1$ the 1-dimensional Hausdorff measure on $\mathbb{R}^2$, and, for every Borel set $A \subset \mathbb{R}^2$, by $\mathcal{H}^1 \llcorner A$ the measure $\mathcal{H}^1$ restricted to A, i.e. such that for every Borel set E we have

$$\left(\mathcal{H}^1 \llcorner A\right)(E) = \mathcal{H}^1(A \cap E) \,.$$

Part 1

Introduction

1. Inverse problems and variational regularization

Broadly speaking, an inverse problem consists in reconstructing an unknown object from partial and often indirect observations of it. Many problems fall in this category. To name only a few, let us mention the reconstruction of a 3D object from a few 2D photographs, the location of an earthquake's epicenter from measurements of the resulting seismic waves, or the determination of the composition of a material from the way it interacts with light (or the way other kinds of waves travel through it).

This thesis is dedicated to a particular family of inverse problems, which belongs to the class of imaging inverse problems, in which the unknown object to reconstruct is an image. In this section, we first provide a broad introduction to this domain[1], before presenting the problem we are specifically interested in. Then, we introduce a family of methods for solving inverse problems, namely variational regularization techniques, that are at the core of our work. Finally, we present the main aim of this thesis, which is to adapt recent advances in the field to our specific setting.

1.1. Inverse problems in imaging

As mentioned above, an imaging inverse problem consists in reconstructing an image (in the general sense) from a set of measurements. Such problems are ubiquitous in biological, medical, and astronomical imaging, as well as in computational photography. Before discussing these applications in greater detail, let us introduce the mathematical framework commonly used to model these problems.

Framework and assumptions. The unknown image is modeled as an element f_0 of some vector space E. We assume that the measurement operator Φ, which accounts for the measurements acquisition process, is a linear[2] mapping from E to some Hilbert space $\mathcal{H}$ (a common scenario is to have access to m real-valued measurements, in which case $\mathcal{H} = \mathbb{R}^m$). Finally, the noiseless and noisy observations are respectively defined by $y_0 = \Phi f_0$ and $y = y_0 + w$ where $w \in \mathcal{H}$ is an additive noise. Solving the inverse problem means (approximately) recovering f from the knowledge of y_0 (noiseless case) or y (noisy case).

Modeling images. The choice of the vector space E in which the sought-after image lies depends on the application. When dealing with digital images (say grayscale images for simplicity), which are arrays of real values, it is natural to take $E = \mathbb{R}^{m \times n}$ for some pair of integers $m, n \in \mathbb{N}^*$, or simply $E = \mathbb{R}^n$ for some $n \in \mathbb{N}^*$. On the other hand, if one wishes to model "continuous images", choosing E to be an infinite dimensional space of functions or measures is relevant. The

[1] Our presentation is greatly inspired by [Peyré, 2021, Chapter 8]. We refer the reader to this book draft and the companion website for more details on imaging. Let us also mention the reference book [Scherzer et al., 2008].

[2] In many situations of interest, the assumption that Φ is linear is unrealistic. Still, the class of linear inverse problems encompass a large variety of practically relevant applications. We stress that the techniques used to study nonlinear inverse problems significantly differ from those we present in this manuscript.

main examples of interest for our purpose are $E = \mathcal{M}(\Omega)$, the set of bounded Radon measures on Ω, and $E = \mathrm{L}^2(\Omega)$, where Ω is the image domain.

Structural assumptions on the sought-after image. In most cases, a major obstacle towards the resolution of an inverse problem is the non-injectivity of the measurement operator Φ. To put it another way, given a set of observations y_0, multiple images f in E could satisfy $\Phi f = y_0$, and the difficulty is to select the right one among them. Even in situations where Φ is injective, the linear system $\Phi f = y_0$ is often ill-conditioned. Inverting Φ hence yields reconstructions that are highly sensitive to changes of y_0, and which are therefore not robust to noise. The main ingredient usually used to overcome these issues is to make structural assumptions on the sought-after image f_0. If we know a priori that f_0 has a special structure, then this information can be exploited to recover it among all images f satisfying $\Phi f = y_0$. In the following subsection, we give examples of interesting structural assumptions on f_0, which are all closely linked to the concept of sparsity.

1.1.1. Sparse signals

A class of signals that has been extensively studied since the 90's is the class of signals having some kind of sparsity. Given a set of atoms called a dictionary, a signal is said to be sparse if it can be decomposed as the weighted sum of a few atoms.

Sparse vectors. The first example of sparse objects that were extensively studied are sparse vectors. The sparsity of a vector x of $\mathbb{R}^n$ is defined as its number of non-zero coordinates, and a vector is said to be sparse if its sparsity is significantly smaller than n (it can hence be written as a weighted sum of a few 1-sparse vectors). The problem of recovering a sparse vector from a set of linear measurements has a tremendous number of applications. It has been studied empirically from the 70's in different domains, including seismic imaging [Claerbout and Muir, 1973]. Its mathematical analysis started in the 90's with the work of David Donoho and his co-authors (see for instance [Donoho, 1992, Chen et al., 1998]). In some situations, it may happen than one is interested in recovering images directly represented by sparse vectors (for example, if one wishes to reconstruct a superposition of a few point light sources). However, it often happens that the sought-after image is not directly sparse, but admits a sparse decomposition in some dictionary. In other terms, there exists a sparse vector x_0 and a dictionary Ψ such that $f_0 = \Psi x_0$. In this case, in order to exploit the sparsity of x_0 for solving the considered inverse problem, one can define a new measurement operator $A = \Phi\Psi$, and solve the inverse problem associated to x_0 and A[1]. This ultimately allows to obtain an approximation of f_0. We refer the reader to [Mallat, 2008, Chapter 12 and 13] for more details on this subject.

Discrete measures. If one wishes to reconstruct a superposition of point light sources, it is often interesting, as explained in Section 1.3, to model the sought-after image as the weighted sum of a few Dirac masses. In this case, one choses $E = \mathcal{M}(\mathcal{X})$ the space of bounded Radon measures on a domain $\mathcal{X}$ (usual choices of $\mathcal{X}$ include the n-dimension torus $\mathbb{T}^n$ or an open subset

[1] This is the so-called *synthesis* approach. Another one is the *analysis* approach (see for instance [Elad et al., 2007] and [März et al., 2022, Section 1.2.2]).

of $\mathbb{R}^n$), and assumes that the unknown signal is a measure μ_0 of the form

$$\mu_0 = \sum_{i=1}^{s} a_i \delta_{x_i} \,,$$

with $a \in \mathbb{R}^s$ and $x \in \mathcal{X}^s$. The recovery of such signals has been the subject of many works in recent years (see for instance [De Castro and Gamboa, 2012, Bredies and Pikkarainen, 2013, Candès and Fernandez-Granda, 2014]). We refer the reader to Section 1.3.2 below and the review article [Laville et al., 2021] for more details. A part of the work presented in this thesis is an attempt to adapt these results to the reconstruction of piecewise constant functions.

Time-dependant discrete measures. One could also consider the *dynamic* inverse problem which consists in recovering the evolution of a superposition of point light sources from a set of measurements. In this case the unknown signal is a time-dependent measure of the form

$$\mu_0 : t \mapsto \sum_{i=1}^{s} a_i(t) \delta_{x_i(t)}$$

where a and x take values in $\mathbb{R}^s$ and $\mathcal{X}^s$. This task was notably studied in the fixed mass setting in [Alberti et al., 2019, Bredies and Fanzon, 2020, Bredies et al., 2021], and then in the varying mass setting in [Bredies et al., 2022a] .

Measures supported on 1D curves. Another class of signals presenting some kind of sparsity are measures supported on a few 1D curves. Such measures could be relevant to model structures of interest in biological imaging. We provide in Figure 1 two images of microtubules to support this claim. In this context, one would chose $E = \mathcal{M}(\mathcal{X})$ but assume the sought-after signal is a measure μ_0 of the form

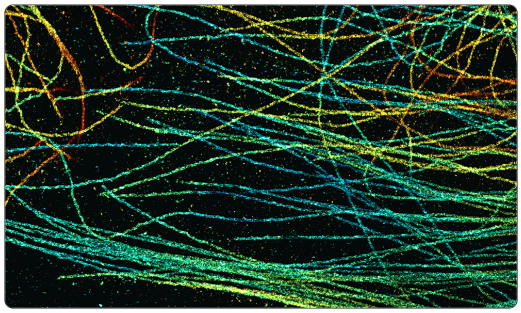
$$\mu_0 = \sum_{i=1}^{s} a_i \, \mathcal{H}^1 \llcorner \Gamma_i \,,$$

with $a \in \mathbb{R}^s$ and Γ_i a connected 1D curve for all $i \in \{1, ..., s\}$. Let us stress that a related task is considered in [Laville et al., 2022]. This work however deals with vector-valued measures, which are hence associated (just as time-dependant measures) to *oriented* curves.

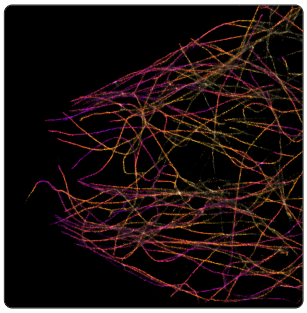
Piecewise constant functions. In this thesis, the signals we focus on are models for piecewise constant images (also called "cartoon images"). The recovery of such objects could have potential applications in e.g. cell imaging. We provide in Figure 2 a set of images which could be considered piecewise constant as a first approximation. In this setting, one could choose $E = \mathrm{L}^2(\mathbb{R}^2)$ and assume the sought-after signal is an element u_0 of E of the form

$$u_0 = \sum_{i=1}^{s} a_i \mathbf{1}_{E_i} \,,$$

with $a \in \mathbb{R}^s$ and $(E_i)_{i=1}^s$ a collection of sufficiently regular subsets of the plane. We give a precise definition of the class of piecewise constant functions we deal with in Section 1 of Part 2.

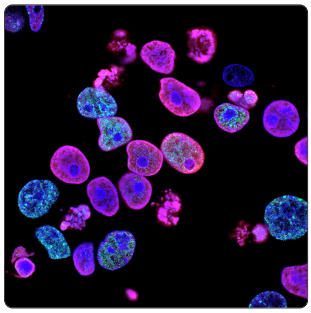


Microtubules labeled with antibodies (Jonas Ries)

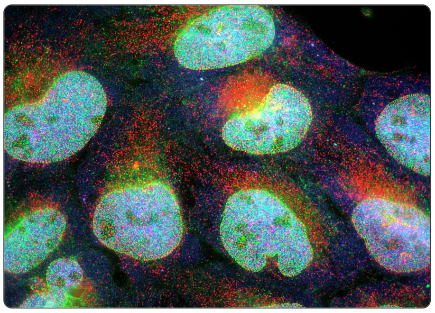


3D dSTORM reconstruction of microtubules (UMIF)

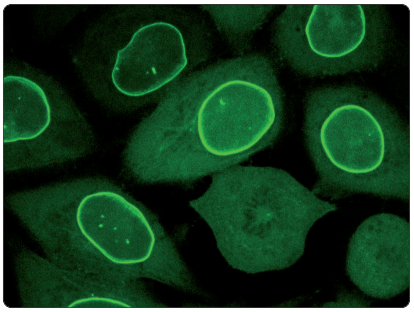
Figure 1 – Examples of microtubule images.



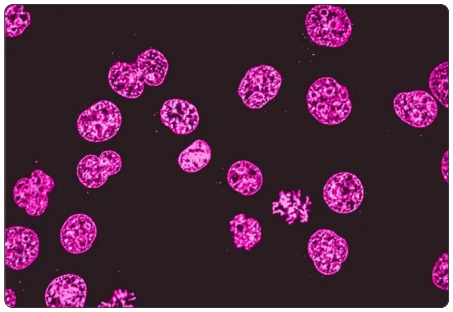
Human colorectal cancer cells marked with a chromosomal protein (Rozenn Jossé, Yves Pommier)



Immunofluorescence light micrograph of human osteosarcoma cells (Nancy Kedersha)



Cytoskeleton imaging using chromobodies (Bjoern Traenkle, Ulrich, Rothbauer)



Confocal microscopy of human acute leukemia cells (Volha Vshyukova)

Figure 2 – Examples of cell images.

1.1.2. Examples of inverse problems

We now give a few examples of measurement operators of interest.

Denoising. The choice of $\Phi = Id$ corresponds to the denoising setting, where one wishes to recover an image corrupted by some (additive) noise.

Deconvolution. Another setting of interest is when one has access to a blurry version of the sought-after image. This is modeled by defining $\Phi f = \varphi \star f$ with φ a convolution kernel. The "$\star$" operation denotes a discrete or continuous convolution depending on the nature of f. The kernel φ could for example model the point spread function of an imaging device, or a motion blur.

Fourier transform sampling. The reconstruction of an image from a sampling of its Fourier transform has many practical applications. We describe two of them below. The first is the detection of point light sources using radio interferometry in astronomical imaging. The second is medical imaging, and more precisely X-ray tomography and magnetic resonance imaging.

- **Radio interferometry:** if one wishes to observe point light sources in the sky, a natural model for the sought-after image is a measure μ_0 that is a combination of a few Dirac masses. In radio interferometry, one uses an array of antennas to record the light emitted by celestial bodies. Computing the cross-correlation between these recordings gives access to a sampling of the Fourier transform of μ_0 (see e.g. [Pan et al., 2017]). The measurement operator can hence be modeled as

$$\begin{aligned} \Phi : \mathcal{M}(\mathbb{R}^n) &\to \mathbb{C}^{|\Omega|} \\ \mu &\mapsto \left[\int_{\mathbb{R}^d} \exp(-i\langle \omega, x \rangle)\, d\mu(x) \right]_{\omega \in \Omega}, \end{aligned}$$

  where $\Omega \subset \mathbb{R}^n$ is a set of observed frequencies, and d is typically equal to two (e.g. in the context of astronomical imaging).

- **X-ray tomography, magnetic resonance imaging:** other situations in which one wishes to recover an image from a sampling of its Fourier transform include X-ray tomographgy and magnetic resonance imaging (MRI). Here, the aim is to to reconstruct a general digital ($E = \mathbb{R}^n$) or continuous (e.g. $E = \mathrm{L}^2(\mathbb{R}^2)$) image. The measurement operator is given by

$$\Phi f = \left[\hat{f}(\omega) \right]_{\omega \in \Omega},$$

  where $\hat{f}$ denotes the (discrete or continuous) Fourier transform of f and Ω is a set of frequencies. The main difference between these two imaging techniques is the type of Fourier samples they give access to. For X-ray tomography Ω is a set of points located on radial lines in the Fourier plane. In the case of MRI, these points are located on smooth curves.

Laplace transform sampling (fluorescence microscopy). A major challenge in modern biology is to understand mechanisms at the sub-cellular level. This requires the development of so-called super-resolution microscopy techniques, able to overcome the diffraction limit (see Figure 2 for an example of images obtained using such techniques). A popular method to achieve this is single-molecule localization microscopy (SMLM) [Betzig et al., 2006, Rust et al., 2006, Shroff et al., 2008]. It consists in introducing fluorophores in a biological sample, and recording the light they emit after illuminating a random subset of them. In this context, acquiring depth information on the sample is a major challenge that can be adressed using several techniques [Huang et al., 2008, Pavani et al., 2009]. Among them, the multi-angle total internal reflection fluorescence method (MA-TIRF) [Boulanger et al., 2014, Santos et al., 2014, Soubies et al., 2014] consists in illumating the biological sample along different directions. The resulting measurements correspond to a sampling of the Laplace transform (in depth) of the sought-after image, which is modeled as a discrete measure. In other words, neglecting the horizontal blur, this corresponds to the choice of $E = \mathcal{M}(\mathbb{R}_+)$ and

$$\begin{aligned} \Phi : \mathcal{M}(\mathbb{R}_+) &\to \mathbb{R}^K \\ \mu &\mapsto \left[\int_{\mathbb{R}_+} \exp(-s_k x)\, d\mu(x) \right]_{k=1,\dots,K}, \end{aligned}$$

where $s_k \in \mathbb{R}_+$ for every $k \in \{1, ..., K\}$. Let us stress that this model is a gross simplification of the real models used in fluorescence microscopy, and we refer the reader to [Denoyelle, 2018, Chapter 5] for more information on this subject.

1.2. Variational regularization

1.2.1. Presentation

A fruitful idea for leveraging a priori knowledge on the sought-after signal f_0 is that of variational regularization. It consists, in the case of noiseless observations, in finding an approximation of f_0 by solving

$$\inf_{f \in E} R(f) \;\text{ s.t. }\; \Phi f = y_0\,, \tag{$\mathcal{P}_0(y_0)$}$$

where $R : E \to \mathbb{R} \cup \{+\infty\}$ is called the regularizer. This mapping is chosen in order to promote solutions having the same structure as f_0. Loosely speaking, the value of $R(f)$ should be small if f has the desired structure, and large otherwise. The choice of R given a priori assumptions on the sought-after signal is discussed in Section 1.2.2.

If one has access to noisy observations, i.e. to $y = \Phi f_0 + w$, then minimizing R among functions f satisfying $\Phi f = y$ makes little sense (even f_0 would not belong to the admissible set). One hence solves instead

$$\inf_{f \in E} R(f) + \frac{1}{2\lambda} \|\Phi f - y\|_{\mathcal{H}}^2\,, \tag{$\mathcal{P}_\lambda(y)$}$$

with λ a positive real number called the regularization parameter. Typically, λ should be chosen according to an estimate of $\|w\|_{\mathcal{H}}$.

Having defined these two variational problems, it is natural to investigate if solving them allows to solve the considered inverse problems. Three questions therefore arise.

1. Is f_0 the unique solution of $(\mathcal{P}_0(y_0))$?
2. If w is small and λ well chosen, are solution of $(\mathcal{P}_\lambda(y_0+w))$ close to solutions of $(\mathcal{P}_0(y_0))$?
3. Can one numerically solve $(\mathcal{P}_0(y_0))$ and $(\mathcal{P}_\lambda(y))$?

1.2.2. Choice of the regularizer and a priori information

Discrete ℓ_1 regularization. The most important (and first historical) example of regularizer is arguably the ℓ_1-norm in the Euclidean space. From the 70's, ℓ_1 regularization has been used empirically in various fields (notably in seismic imaging) to recover sparse vectors. It was later studied mathematically by David Donoho and its co-authors (among others) from the 90's (see e.g. [Donoho, 1992, Chen et al., 1998] and also [Tibshirani, 1996] for applications in statistics).

Convex regularizers. In order to obtain convex problems $(\mathcal{P}_0(y_0))$ and $(\mathcal{P}_\lambda(y))$, it is natural to choose R among convex functions. Although non-convex regularizers could also be considered, being able to leverage the theory of convex optimization is really useful for analyzing variation regulatization techniques.

Representer theorems and convex regularization. The following series of recent works [Chandrasekaran et al., 2012, Boyer et al., 2019, Bredies and Carioni, 2019] has recently shed a new light on convex regularization. In these articles, the link between the choice of a regularizer and the structural properties of some solutions of $(\mathcal{P}_0(y_0))$ and $(\mathcal{P}_\lambda(y))$ is investigated. Their main finding is that the atoms promoted by a convex regularizer are the extreme points of its associated unit ball: loosely speaking, if $\mathcal{H}$ has finite dimension, then some solutions of $(\mathcal{P}_0(y_0))$ and $(\mathcal{P}_\lambda(y))$ are linear combinations of extreme points of $\{R \leq 1\}$. We provide below a few examples of regularizers and describe the atoms they promote.

- If $E = \mathbb{R}^n$ and $R = \|.\|_1$ is the ℓ_1-norm, then the extreme points of $\{R \leq 1\}$ are the vectors $\pm e_i$ with $i \in \{1, ..., d\}$, where $(e_i)_{i=1,...,n}$ is the canonical basis of $\mathbb{R}^n$.
- If $E = \mathbb{R}^{n\times m}$ and $R = \|.\|_*$ is the nuclear norm of matrices then the extreme points of $\{R \leq 1\}$ are the matrices $u^T v$ with $u \in \mathbb{R}^n$, $v \in \mathbb{R}^m$ and $\|u\|_2 = \|v\|_2 = 1$.
- If $E = \mathcal{M}(\mathcal{X})$ with $\mathcal{X}$ and $R = |.|(\mathcal{X})$ is the total variation (of measures) then the extreme points of $\{R \leq 1\}$ are the measures $\pm\delta_x$ with $x \in \mathcal{X}$.
- If $E = \mathrm{L}^2(\mathbb{R}^2)$ and $R = |\mathrm{D}.|(\mathbb{R}^2)$ is the total (gradient) variation then the extreme points of $\{R \leq 1\}$ are the functions $\pm\mathbf{1}_E/P(E)$ with E a simple set with positive finite measure[1].

Besides extreme points, the results of [Boyer et al., 2019] show that the structure of the solution set of $(\mathcal{P}_0(y_0))$ or $(\mathcal{P}_\lambda(y))$ is linked to the structure of the faces of $\{R \leq 1\}$. The importance of closely studying these faces is highlighted in Part 2. If one has access to m observations (i.e. $\mathcal{H} = \mathbb{R}^m$), then some solutions lie in a face of dimension at most $m-1$ (or m in degenerate cases). In particular, these solutions can be written as the sum of at most m (or $m+1$) extreme points of $\{R \leq 1\}$.

[1] See Section 2.1.1 for a precise definition of the total variation and simple sets.

1.2.3. Identifiability

If the first question stated in Section 1.2.1 has a positive answer, i.e. if f_0 is the unique solution of $(\mathcal{P}_0(y_0))$, then it is said to be *identifiable*. Even if we rather focus on noise robustness results in this thesis, let us briefly discuss this question.

Discrete ℓ_1 regularization. We begin by considering the recovery of sparse vectors using the ℓ_1 norm as a regularizer (i.e. $E = \mathbb{R}^n$, $R = \|\cdot\|_1$ and $\mathcal{H} = \mathbb{R}^m$). The first historical example of identifiability condition is the *nullspace property*. This name was first used in [Cohen et al., 2009], although the condition also appeared in [Gribonval and Nielsen, 2003] and implicitly in previous works. If Φ is the measurement matrix, then every s-sparse vector is identifiable if and only if the nullspace property of order s is satisfied. This condition is difficult to check in practice. To circumvent this issue, the stronger *restricted isometry property* was introduced in [Candes and Tao, 2005]. Even if computing the restricted isoperimetry constants of a matrix is known to be NP-hard, large classes of random matrices satisfy the property with high probability (see for instance [Candes and Tao, 2006]). We refer the reader to the reference book [Foucart and Rauhut, 2013] for more details.

Extensions. Sufficient identifiability conditions also exist for low-rank matrix recovery, i.e. when $E = \mathbb{R}^{n\times m}$ and $R = \|\cdot\|_*$ (see for instance [Foucart and Rauhut, 2013, Section 4.6]). Likewise, perfect recovery results were proven in the sparse spikes setting, i.e. with $E = \mathcal{M}(\mathcal{X})$ and $R = |\cdot|(\mathcal{X})$. We provide further details on this case in Section 1.3.2.

1.2.4. Noise robustness

As mentioned above, the second question stated in Section 1.2.1 is closely related to noise robustness. Indeed, if a signal f_0 is identifiable and one proves that solutions of $(\mathcal{P}_\lambda(y))$ converge to a solution of $(\mathcal{P}_0(y_0))$ when w is small and λ well chosen, then solving $(\mathcal{P}_\lambda(y))$ allows to approximately recover f_0 in a low noise regime.

Stability and model stability. Noise robustness results typically fall into two categories. The first ones are concerned with the convergence of a solution $f_{\lambda,w}$ of $(\mathcal{P}_\lambda(y_0 + w))$ towards f_0 with respect to some distance on the signal space E. This distance is typically the ℓ_2 distance if $E = \mathbb{R}^n$, the L^2 distance if $E = \mathrm{L}^2(\mathbb{R}^2)$, or an optimal transport-based distance if $E = \mathcal{M}(\mathcal{X})$. This notion of stability is however quite weak. Indeed, if f_0 is a combination of a few atoms, a convergence result of this kind does not tell anything about the structure of $f_{\lambda,w}$. To be more precise, it does not allow to answer the following question: is $f_{\lambda,w}$ made of the same number of atoms as f_0, and are these atoms related to those appearing in the decomposition of f_0? This shows the necessity of a second category of noise robustness results, allowing to answer these questions. They are sometimes called *model stability* results, and are the ones we focus on in this thesis.

Discrete ℓ_1 regularization. Given a sparse vector, several works investigated whether solving $(\mathcal{P}_\lambda(y_0 + w))$ (with $E = \mathbb{R}^n$ and $R = \|.\|_1$) in a low noise regime allows to exactly recover its support. To name only a few, let us cite [Fuchs, 2004, Dossal and Mallat, 2005, Tropp, 2006], in which sufficient conditions are derived. Loosely speaking and using a modern terminology,

exact recovery is possible if some *minimal norm dual certificate* interpolates the sign of the signal coordinates.

Extensions. In [Vaiter et al., 2015, Vaiter et al., 2018], the analysis of ℓ_1 regularization was extended to the family of so-called *partly smooth* regularizers, which promote signals having some notion of low complexity. Numerous popular regularizers fall in this category. To name only a few, let us mention the ℓ_1 norm, the $\ell_1 - \ell_2$ norm (used to enforce group sparsity), and the nuclear norm of matrices. An extension to so-called *mirror-stratifiable* regularizers was also studied in [Fadili et al., 2018, Fadili et al., 2019]. This last series of works has similarities with the dual-based approach of [Duval and Peyré, 2015], in which an exact support recovery result is proved for discrete measures. We discuss this topic in greater details in Section 1.3.2

1.3. Grid-free sparse recovery

At first glance, solving ($\mathcal{P}_\lambda(y)$) or ($\mathcal{P}_0(y_0)$) with E an infinite dimensional space seems intractable. Considering such problems hence only seems to be of theoretical interest. However, a recent line of works demonstrated that it is in fact possible to solve ($\mathcal{P}_\lambda(y)$) and ($\mathcal{P}_0(y_0)$) efficiently when $E = \mathcal{M}(\mathcal{X})$ and $R = |.|(\mathcal{X})$ is the total variation of measures. Doing so even has numerous advantages, both from a numerical point of view and with respect to recovery guarantees. This section provides a review of these advances, and situates the subject of this thesis within this context.

1.3.1. Sparse spikes super-resolution on thin grids

Sparse spikes recovery and the BLASSO. As explained in Sections 1.1.1 and 1.1.2, the recovery of a discrete measure $\mu_0 = \sum_{i=1}^s a_{0,i}\,\delta_{x_{0,i}}$ from noisy linear measurements $y = \Phi\mu_0 + w$ has numerous applications. We also mentioned that solving ($\mathcal{P}_\lambda(y_0 + w)$) where $y_0 = \Phi\mu_0$ with $E = \mathcal{M}(\mathcal{X})$ and $R = |.|(\mathcal{X})$ is a good way to approximate μ_0. The resulting problem, which writes

$$\inf_{\mu\in\mathcal{M}(\mathcal{X})} |\mu|(\mathcal{X}) + \frac{1}{2\lambda}\|\Phi\mu - y\|_{\mathcal{H}}^2 \qquad (\mathcal{P}_\lambda(y))$$

and is known as the BLASSO[1], is however infinite-dimensional. How to numerically solve it is hence unclear.

Restriction to a finite grid. A first approach one can consider is to look for an approximate solution of ($\mathcal{P}_\lambda(y)$) whose support lies in some discrete grid $\mathcal{G} = \{x_i\}_{i=0}^{n-1} \subset \mathcal{X}$, and therefore to solve ($\mathcal{P}_\lambda(y)$) with the additional constraint $\mathrm{Supp}(\mu) \subset \mathcal{G}$. This yields the finite dimensional problem

$$\inf_{a\in\mathbb{R}^n} \|a\|_1 + \frac{1}{2\lambda}\|\Phi_{\mathcal{G}}a - y\|_{\mathcal{H}}^2\,, \qquad (\mathcal{P}_\lambda^{\mathcal{G}}(y))$$

[1]The name BLASSO stands for Beurling LASSO, and was coined in [De Castro and Gamboa, 2012]. We provide more information on this problem in Section 1.3.2 below.

which is the celebrated basis pursuit, also known as the LASSO (see for instance [Chen et al., 1998, Tibshirani, 1996]), where $\Phi_{\mathcal{G}}$ is defined by

$$\begin{aligned} \Phi_{\mathcal{G}} : \mathbb{R}^n &\to \mathcal{H} \\ a &\mapsto \Phi\left(\sum_{i=1}^{n} a_i \, \delta_{x_i}\right). \end{aligned}$$

Thin grids. Since no information on the support of μ_0 is known a priori, it is natural to consider increasingly "thinner" grids $\mathcal{G}$, in order to cover the whole domain $\mathcal{X}$. However, the analysis conducted in [Duval and Peyré, 2017] shows this approach yields reconstructed measures whose structure is different from that of μ_0. To be more precise, each spike in μ_0 is estimated by a pair of neighbooring spikes. This demonstrates the impossibility to exactly recover the support of μ_0 by relying on ℓ_1 regularization techniques on (even arbitrarily thin) finite grids.

1.3.2. Grid-free sparse spikes recovery

Considering the negative result presented above, it is natural to wonder if, despite its infinite-dimensional nature, numerically solving the BLASSO problem ($\mathcal{P}_\lambda(y)$) is possible. Over the last ten years, numerous methods have in fact been developed to achieve this. We briefly review them here, before discussing available results on noise robustness. We refer the reader to the lecture notes [Poon, 2019] for a more detailed introduction to sparse spikes recovery using the BLASSO.

Numerical resolution of the BLASSO. The first method allowing to solve the BLASSO was introduced in [Tang et al., 2013, Candès and Fernandez-Granda, 2014]. It relies on the equivalence, in the specific case of 1D Fourier measurements, of ($\mathcal{P}_\lambda(y)$) with a finite dimensional semi-definite program, for which efficient solvers exist. Extensions to higher dimensions rely on the so-called Lasserre hierarchy (see e.g. [Catala et al., 2019], [Catala, 2020, Chapter 2] and also [De Castro et al., 2017] for the case of polynomial measurements). For general measurements, approaches based on the conditional gradient or Frank-Wolfe algorithm, which does not require any Hilbertian structure (and is hence particularly well suited for optimization on the space of measures), were developed in [Bredies and Pikkarainen, 2013, Boyd et al., 2017, Denoyelle et al., 2019]. Finally, let us mention the work of [Chizat, 2022], which is based on conic particle gradient descent.

Identifiability. The groundbreaking work [Candès and Fernandez-Granda, 2014] shows that a discrete measure can be exactly recovered from $2f_c + 1$ Fourier samples, provided the minimal distance Δ between its atoms satisfies $\Delta \geq 2/f_c$. In [Tang et al., 2013], it has also been shown that this result remains valid with high probability if a small number of Fourier samples are randomly selected. Let us stress that, for positive measures, the minimal separation condition can be lifted (see for instance [De Castro and Gamboa, 2012, Schiebinger et al., 2018]).

Noise robustness. The convergence of solutions $\mu_{\lambda,w}$ of ($\mathcal{P}_\lambda(y_0 + w)$) in a low noise regime has been investigated in several works. The weak-* convergence of $\mu_{\lambda,w}$ towards a solution μ_0 of the noiseless problem is proved in [Bredies and Pikkarainen, 2013]. In a similar spirit, [Candès and Fernandez-Granda, 2013, Fernandez-Granda, 2013, Azaïs et al., 2015] provide error

bounds (based on [Burger and Osher, 2004]) which show that $\mu_{\lambda,w}$ concentrates around the support of μ_0. However, these results do not address the question of support recovery: if μ_0 is discrete and identifiable, is $\mu_{\lambda,w}$ discrete, and does it have the same number of spikes as μ_0?

Exact support recovery. In [Duval and Peyré, 2015], an answer to this last question is provided. The authors prove that, if a so-called non-degenerate source condition holds, $\mu_{\lambda,w}$ has exactly the same number of spikes as μ_0 in a low noise regime, and the locations of these spikes (as well as the associated amplitudes) smoothly converge to the limit ones. This is in sharp contrast with the negative result presented for finite grids in Section 1.3.1.

1.3.3. Perspectives

In Section 1.3.2 above, we have seen that, in the context of sparse spikes recovery, grid-free approaches both have theoretical and numerical benefits. In this subsection, we mention recent attempts to develop similar methods for recovering other kinds of signals. We finally introduce the problem this thesis is dedicated to, and advocate for studying noise robustness properties and developing numerical methods using the fruitful ideas presented in Section 1.3.2.

Dynamic sparse spikes recovery. As mentioned in Section 1.1.1, the recovery of time-dependant discrete measures has recently been investigated. Several regularizers were introduced to tackle the fixed and varying mass settings, and extreme points results were derived. Although we are not aware of any identifiability or noise robustness result, numerical methods received a lot of attention, and approaches based on the Frank-Wolfe (or conditional gradient) algorithm were very recently proposed in [Bredies et al., 2022b, Duval and Tovey, 2021].

Curve recovery. We also discussed in Section 1.1.1, the reconstruction of images which are the superposition of a few 1D curves. To our knowledge, the only work in which this task is investigated is the recent work [Laville et al., 2022]. Its authors prove an extreme point result which we briefly state below. Let $E = \mathcal{M}(\mathcal{X})^2$ be the dual of $\mathcal{C}_0(\mathcal{X}, \mathbb{R}^2)$ (equipped with the supremum norm), and let R be defined by

$$\begin{aligned} R : \mathcal{M}(\mathcal{X})^2 &\to \mathbb{R} \cup \{+\infty\} \\ \mu &\mapsto |\mu|(\mathcal{X}) + |\mathrm{div}(\mu)|(\mathcal{X}) \,, \end{aligned}$$

with $\mathrm{div}(\mu)$ the divergence of μ (defined in the sense of distributions). Then the extreme points of $\{R \leq 1\}$ are the measures $\mu_\gamma / R(\mu_\gamma)$ where μ_γ is the measure canonically associated to a simple oriented Lipschitz curve γ (see [Laville et al., 2022, Definition 7] for a precise definition). An interesting avenue for future research is to investigate the performance of this regularizer for solving inverse problems (i.e. to derive identifiability and noise robustness results), and also to develop numerical methods allowing to solve ($\mathcal{P}_0(y_0)$) and ($\mathcal{P}_\lambda(y)$) efficiently. Let us also stress that the measures this work deals with are associated to *oriented* curves, and that they also carry tangential information. Recovering measures supported on *unoriented* curves has, to our knowledge, not be considered, although they seem a natural model for images such as those of Figure 1.

Reconstruction of piecewise constant images. In this thesis, we are concerned with the recovery of piecewise constant images, i.e. which are the superposition of a few simple shapes. To recover such objects, the natural choice of regularizer is the total (gradient) variation. The first appearance of this functional for imaging applications dates back to the pioneering work of Rudin Osher and Fatemi [Rudin et al., 1992], which focuses on the denoising problem. Since then, the performance of this regularizer has been the subject of a large number of works. It has notably been shown in [Chambolle et al., 2016, Iglesias et al., 2018] that, in a small noise regime, solutions of $(\mathcal{P}_\lambda(y_0 + w))$ converge to a solution of $(\mathcal{P}_0(y_0))$ (which is unique and equal to the image to recover in the case of denoising). The type of convergence proved in this result is already quite strong. Namely, it is shown that the level sets of solutions of $(\mathcal{P}_\lambda(y_0 + w))$ converge to the level sets of solutions of $(\mathcal{P}_0(y_0))$ (see Proposition 1.12 for a precise statement). This convergence result for level sets is of particular interest in the context of imaging. Even if the analysis presented in Section 1.2.2 suggests that the total variation is particularly well suited to recover piecewise constant functions, the above mentioned works do not focus on unknown images having this structure. If we make this assumption, we argue that the emphasis should be on structure-preserving convergence guarantees, in the spirit of [Duval and Peyré, 2015]. To be more specific, one should study whether the reconstructed image is composed of the same number of shapes than the sought-after image, and whether these shapes are close to the unknown ones. Finally, let us mention that, in this setting, the problem of deriving sufficient identifiability conditions is to our knowledge mostly open. Still, let us point out that, in [Bredies and Vicente, 2019], a perfect reconstruction result is obtained in the case where the measurement associated to a function is its image under a linear partial differential operator with unknown boundary conditions.

The main aim of this thesis is to study total variation regularization from a sparse recovery viewpoint.

2. Total variation regularization

The use of the total variation as a regularizer started with the pioneering work of Rudin Osher and Fatemi [Rudin et al., 1992] in the context of image denoising. This functional penalizes oscillations while allowing discontinuities. Reconstruction methods based on its minimization therefore have the striking feature of preserving image edges (see Figure 3 for a few examples). Although state of the art algorithms now exhibit much better performance, the total variation is still routinely used in specific applications, and remains an import baseline. In this section, we provide an introduction to total variation regularization. We review existing theoretical results and numerical methods, and also mention the open problems that we attempt to address in this thesis.

Figure 3 – Reconstructions obtained with [Getreuer, 2012a, Getreuer, 2012b] in the case of deconvolution (first and third rows) and inpainting (second and fourth rows).
Left: observations $y = \Phi u_0 + w$, middle: unknown image u_0, right: approximate solution of $(\mathcal{P}_\lambda(y_0 + w))$ (source images: bazzier, Le Pays Malouin).

2.1. Presentation and preliminaries

2.1.1. The total variation

We begin this section by defining the total variation functional. Then, we introduce a few notions to be able to state the extreme point result mentioned in Section 1.2.2, which characterizes the atoms promoted by this regularizer. Finally, we give several results about the subdifferential of the total variation, which are useful for the analysis of the variational problems ($\mathcal{P}_0(y_0)$) and ($\mathcal{P}_\lambda(y)$).

Definition. If $u \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^2)$, its total variation is defined by

$$\mathrm{TV}(u) \stackrel{\text{def.}}{=} \sup \left\{ -\int_{\mathbb{R}^2} u \operatorname{div}\phi \;\middle|\; \phi \in \mathrm{C}_c^\infty(\mathbb{R}^2, \mathbb{R}^2),\ \|\phi\|_\infty \leq 1 \right\}. \tag{1}$$

A function u has finite total variation if and only if its distributional gradient, denoted $\mathrm{D}u$, is a finite Radon measure. In this case one has $\mathrm{TV}(u) = |\mathrm{D}u|(\mathbb{R}^2)$. Let us now provide more intuition on this functional by considering two classical examples. If u has integrable gradient one has $\mathrm{TV}(u) = \int_{\mathbb{R}^2} |\nabla u|$. If E is a set of class C^1 (see Appendix B for a precise definition) by the Gauss-Green theorem $\int_E \operatorname{div} \phi = \int_{\partial E} \phi \cdot \nu_E \, d\mathcal{H}^1$, which yields $\mathrm{TV}(\mathbf{1}_E) = \mathcal{H}^1(\partial E)$. Further details about the total variation can be found in [Ambrosio et al., 2000, Chapter 3] and [Evans and Gariepy, 2015, Chapter 5]. Let us also mention [Chambolle et al., 2010a] for more information on its use in imaging.

Remark 1.1

In this thesis, we extensively work with square integrable functions with finite total variation. These do not belong to the more classical space $\mathrm{BV}(\mathbb{R}^2)$ *of integrable functions with finite total variation. The main interest of this choice is that, in* $\mathbb{R}^2$*, the total variation dominates the* L^2 *norm (and more generally the* $\mathrm{L}^{N/(N-1)}$ *norm in* $\mathbb{R}^N$*), as shown by the isoperimetric inequality* (3) *presented below. As a consequence, the total variation unit ball is weakly compact in* $\mathrm{L}^2(\mathbb{R}^2)$*, which is hence the natural space in which we can obtain existence of solutions for variational problems with a total variation term.*

In all the following we consider TV as a mapping from $\mathrm{L}^2(\mathbb{R}^2)$ to $\mathbb{R} \cup \{+\infty\}$. This function is convex, proper and lower semi-continuous.

Sets of finite perimeter. If a measurable set $E \subset \mathbb{R}^2$ is such that $P(E) \stackrel{\text{def.}}{=} \mathrm{TV}(\mathbf{1}_E)$ is finite, it is said to have finite perimeter. Such sets have similar properties to those of smooth sets, in a measure-theoretic sense. In particular, they satisfy a generalized Gauss-Green formula, and have an approximate normal and an approximate tangent space. They are central in the study of geometric variational problems, mainly thanks to useful compactness results. We refer the reader to [Maggi, 2012, Part II] for further details on this subject. We collect notations and a few properties of sets of finite perimeter in Appendix A.

The isoperimetric inequality

We now state the useful isoperimetric inequality, and take this opportunity to make a digression on its quantitative version. This allows us to introduce several topics of interest for our purpose, namely shape calculus and stability in geometric variational problems.

For every set of finite perimeter $E \subset \mathbb{R}^2$, the isoperimetric inequality states that

$$\sqrt{\min(|E|, |E^c|)} \leq c_2 \, P(E) \,, \tag{2}$$

with equality if and only if E is a ball, and where $c_2 \overset{\text{def.}}{=} 1/\sqrt{4\pi}$ is the isoperimetric constant (see e.g. [Maggi, 2012, Chapter 14]). In particular, if E is a set of finite perimeter, either E or E^c has finite measure. The functional version[1] of (2) is:

$$\forall u \in \mathrm{L}^2(\mathbb{R}^2), \ \|u\|_{\mathrm{L}^2(\mathbb{R}^2)} \leq c_2 \, \mathrm{TV}(u) \,. \tag{3}$$

A sharp quantitative version of (2) was first proved in [Fusco et al., 2008] using symmetrization techniques (see also the review article [Fusco, 2015]). Denoting $B = B(0,1)$ the unit ball, for every measurable set $E \subset \mathbb{R}^2$ with $|E| = |B|$, it writes:

$$\min_{x \in \mathbb{R}^2} |(x+E) \triangle B|^2 \leq C[P(E) - P(B)] \,. \tag{4}$$

Another proof, relying on optimal transport, was given in [Figalli et al., 2010]. A third approach, which is the most relevant for our purpose, was proposed in [Cicalese and Leonardi, 2012]. It relies on two main ingredients. The first is a so-called selection principle, which allows to reduce the proof to nearly spherical sets, i.e. sets whose boundary is the normal graph of a sufficiently small Lipschitz function on the sphere. The second is the proof of (4) for such sets, which dates back to [Fuglede, 1989].

The proof of Fuglede relies on shape calculus. Since nearly spherical sets are parametrized by a function φ on the sphere, one can look at the mapping $\varphi \mapsto P(B_\varphi)$, where B_φ is the nearly spherical set associated to φ, and use calculus to study the stability of $B = B_0$ as a minimizer. To be more precise, one can show that the second order derivative of this mapping at zero, which is called the second order shape derivative of the perimeter, is coercive, which ultimately yields the result.

The derivation of stability results using second order shape derivatives has been generalized to other geometric functionals in several works (see [Dambrine and Lamboley, 2019] and the references therein). The noise robustness analysis we conduct in Part 2 heavily relies on these ideas, which we apply to the prescribed curvature problem presented below.

Indecomposable and simple sets. Now, we introduce the notion of indecomposable and simple sets, which are the measure-theoretic analogues of connected and simply connected sets (we refer the reader to [Ambrosio et al., 2001] for more details). A set of finite perimeter $E \subset \mathbb{R}^2$ is said to be decomposable if there exists a partition of E in two sets of positive measure A and B

[1] This inequality is for example proved in [Ambrosio et al., 2000, Theorem 3.47] by using (2) and the coarea formula (see e.g. [Maggi, 2012, Chapter 13]). In this reference u is assumed to be integrable, but the constant m appearing therein is also zero for any p-integrable function with $1 \leq p < +\infty$.

with $P(E) = P(A) + P(B)$. We say that E is indecomposable if it is not decomposable[1]. An indecomposable set E is called simple if $E = \mathbb{R}^2$ or $|E| < +\infty$ and $\mathbb{R}^2 \setminus E$ is indecomposable.

Extreme points of the total variation unit ball. We are now able to state the extreme point result mentioned Section 1.2.2.

Proposition 1.2 (*[Fleming, 1957, Ambrosio et al., 2001]*)

The extreme points of the convex set

$$\{\mathrm{TV} \leq 1\} \overset{\text{def.}}{=} \left\{u \in \mathrm{L}^2(\mathbb{R}^2) \,\middle|\, \mathrm{TV}(u) \leq 1\right\}$$

are the functions of the form $\pm \mathbf{1}_E / P(E)$, where E is a simple set with $0 < |E| < +\infty$.

This result, in conjunction with the representer theorems mentioned in Section 1.2.2, justifies the affirmation that using the total variation as a regularizer promotes piecewise constant functions (i.e. that are the weighted sum of a few indicator functions of simple sets).

Subdifferential. Let us now collect several results on the subdifferential of TV, which are useful to derive and analyze the dual problems of ($\mathcal{P}_0(y_0)$) and ($\mathcal{P}_\lambda(y)$). Since TV is the support function of the convex set

$$C \overset{\text{def.}}{=} \left\{\operatorname{div} z \,\middle|\, z \in \mathrm{C}_c^\infty(\mathbb{R}^2, \mathbb{R}^2),\ \|z\|_\infty \leq 1\right\},$$

its subdifferential at 0 is the closure of C in $\mathrm{L}^2(\mathbb{R}^2)$, that is

$$\partial \mathrm{TV}(0) = \overline{C} = \left\{\operatorname{div} z \,\middle|\, z \in \mathrm{L}^\infty(\mathbb{R}^2, \mathbb{R}^2),\ \operatorname{div} z \in \mathrm{L}^2(\mathbb{R}^2),\ ||z||_\infty \leq 1\right\}. \tag{5}$$

We also have the following useful identity:

$$\partial \mathrm{TV}(0) = \left\{\eta \in \mathrm{L}^2(\mathbb{R}^2) \,\middle|\, \forall u \in \mathrm{L}^2(\mathbb{R}^2),\ \left|\int_{\mathbb{R}^2} \eta\, u\right| \leq \mathrm{TV}(u)\right\}.$$

Finally, the sudifferential of TV at some $u \in \mathrm{L}^2(\mathbb{R}^2)$ is given by:

$$\partial \mathrm{TV}(u) = \left\{\eta \in \partial \mathrm{TV}(0) \,\middle|\, \int_{\mathbb{R}^2} \eta\, u = \mathrm{TV}(u)\right\}.$$

Hence, if $\eta \in \partial \mathrm{TV}(u)$, then η is an element of $\partial \mathrm{TV}(0) = \overline{C}$ for which the supremum in the definition of the total variation is attained.

From the subdifferential to the level sets. Now, there is an important sufficient and necessary condition for $\eta \in \partial \mathrm{TV}(u)$ to hold. In the rest of the article, given $u : \mathbb{R}^2 \to \mathbb{R}$ and $t \in \mathbb{R}$, we use the following notation:

$$U^{(t)} \overset{\text{def.}}{=} \begin{cases} \{x \in \mathbb{R}^2 \,|\, u(x) \geq t\} & \text{if } t \geq 0\,, \\ \{x \in \mathbb{R}^2 \,|\, u(x) \leq t\} & \text{otherwise.} \end{cases}$$

It is worth noting that, if $u \in \mathrm{L}^2(\mathbb{R}^2)$, then $|U^{(t)}| < +\infty$ for all $t \neq 0$. With this notation, we may state the following result.

[1] This notion is known under the name of inseparable set in the literature on submodular functions (see e.g. [Bach, 2013]).

Proposition 1.3 (*see e.g. [Chambolle et al., 2016]*)

Let $u \in \mathrm{L}^2(\mathbb{R}^2)$ be such that $\mathrm{TV}(u) < +\infty$, and let $\eta \in \mathrm{L}^2(\mathbb{R}^2)$. Then the following conditions are equivalent.

(i) $\eta \in \partial \mathrm{TV}(u)$.

(ii) $\eta \in \partial \mathrm{TV}(0)$ *and the level sets of* u *satisfy*

$$\forall t > 0,\ P(U^{(t)}) = \int_{U^{(t)}} \eta \quad \text{and} \quad \forall t < 0,\ P(U^{(t)}) = -\int_{U^{(t)}} \eta \,.$$

(iii) The level sets of u *satisfy*

$$\forall t > 0,\ U^{(t)} \in \operatorname*{Argmin}_{E \subset \mathbb{R}^2,\, |E| < +\infty} \left(P(E) - \int_E \eta \right),$$
$$\forall t < 0,\ U^{(t)} \in \operatorname*{Argmin}_{E \subset \mathbb{R}^2,\, |E| < +\infty} \left(P(E) + \int_E \eta \right).$$

The geometric variational problem appearing in (iii), which is

$$\inf_{E \subset \mathbb{R}^2,\ |E| < +\infty} P(E) - \int_E \eta \,, \tag{6}$$

is called the *prescribed curvature problem* associated to η. This terminology stems from the fact that, if η is sufficiently regular, every solution of (6) has a (scalar) distributional curvature (defined in Appendix A) equal to η. This problem plays a crucial role in the analysis of total variation regularization, as explained below. If $\eta \in \partial \mathrm{TV}(0)$, one should note that a set of finite measure E solves (6) if and only if $P(E) = \int_E \eta$, which is why (ii) and (iii) are equivalent.

2.1.2. Primal and dual problems, dual certificates

We can now specialize the variational problems introduced in Section 1.2.1 to our setting, which corresponds to the choice of $E = \mathrm{L}^2(\mathbb{R}^2)$ and $R = \mathrm{TV}$.

Assumptions. In all the following, we assume that $\Phi : \mathrm{L}^2(\mathbb{R}^2) \to \mathcal{H}$ is a continuous linear operator, which is equivalent to the existence of $\varphi \in \mathrm{L}^2(\mathbb{R}^2, \mathcal{H})$ such that

$$\begin{aligned} \Phi : \mathrm{L}^2(\mathbb{R}^2) &\to \mathcal{H} \\ u &\mapsto \int_{\mathbb{R}^2} u(x)\, \varphi(x)\, dx \,. \end{aligned}$$

We also assume that $y_0 = \Phi u_0$ with $\mathrm{TV}(u_0) < +\infty$.

Primal problems. As mentioned above, in our setting, the two variational problems introduced in Section 1.2.1 are

$$\inf_{u \in \mathrm{L}^2(\mathbb{R}^2)} \mathrm{TV}(u) \ \text{ s.t. } \ \Phi u = y_0 \,, \tag{$\mathcal{P}_0(y_0)$}$$

$$\inf_{u\in \mathrm{L}^2(\mathbb{R}^2)} \mathrm{TV}(u) + \frac{1}{2\lambda}\|\Phi u - y\|_{\mathcal{H}}^2 \,. \tag{$\mathcal{P}_\lambda(y)$}$$

Existence of solutions for ($\mathcal{P}_0(y_0)$) and ($\mathcal{P}_\lambda(y)$) can be shown using the direct method of the calculus of variations and the isoperimetric inequality (3).

Dual problems. In the remaining of this section, we collect useful results following from standard duality arguments. We refer the reader to [Iglesias et al., 2018, Section 2] for their proof (which had been established in the particular case of denoising in [Chambolle et al., 2016]). The Fenchel-Rockafellar dual problems of ($\mathcal{P}_0(y_0)$) and ($\mathcal{P}_\lambda(y)$) are:

$$\sup_{p\in\mathcal{H}} \langle p, y_0\rangle_{\mathcal{H}} \quad \text{s.t. } \Phi^* p \in \partial \mathrm{TV}(0)\,, \tag{$\mathcal{D}_0(y_0)$}$$

$$\sup_{p\in\mathcal{H}} \langle p, y\rangle_{\mathcal{H}} - \frac{\lambda}{2}||p||_{\mathcal{H}}^2 \quad \text{s.t. } \Phi^* p \in \partial \mathrm{TV}(0)\,. \tag{$\mathcal{D}_\lambda(y)$}$$

The existence of a solution of ($\mathcal{D}_0(y_0)$) is not always guaranteed. Requiring that there indeed exists a maximizer is called the *source condition* in the literature. On the contrary, ($\mathcal{D}_\lambda(y)$) can be reformulated as the projection problem of y/λ on the closed convex set $\{p \in \mathcal{H} \,|\, \Phi^* p \in \partial \mathrm{TV}(0)\}$. It hence has a unique solution. Strong duality holds between ($\mathcal{P}_0(y_0)$) and ($\mathcal{D}_0(y_0)$) (respectively ($\mathcal{P}_\lambda(y)$) and ($\mathcal{D}_\lambda(y)$)), as stated in the following two propositions.

Proposition 1.4

Strong duality holds between ($\mathcal{P}_0(y_0)$) and ($\mathcal{D}_0(y_0)$). Moreover, if there exists a solution p to ($\mathcal{D}_0(y_0)$), then for every solution u of ($\mathcal{P}_0(y_0)$) we have

$$\Phi^* p \in \partial \mathrm{TV}(u)\,. \tag{7}$$

Conversely, if $(u, p) \in \mathrm{L}^2(\mathbb{R}^2) \times \mathcal{H}$ with $\Phi u = y_0$ and (7) holds, then u and p respectively solve ($\mathcal{P}_0(y_0)$) and ($\mathcal{D}_0(y_0)$).

Proposition 1.5

Strong duality holds between ($\mathcal{P}_\lambda(y)$) and ($\mathcal{D}_\lambda(y)$). Moreover, denoting p the unique solution of ($\mathcal{D}_\lambda(y)$), for every solution u of ($\mathcal{P}_\lambda(y)$) we have

$$\begin{cases} \Phi u = y - \lambda\, p\,, \\ \Phi^* p \in \partial \mathrm{TV}(u)\,. \end{cases} \tag{8}$$

Conversely, if (8) holds, then u and p respectively solve ($\mathcal{P}_\lambda(y)$) and ($\mathcal{D}_\lambda(y)$).

Remark 1.6

Although there might not be a unique solution to ($\mathcal{P}_\lambda(y)$), Proposition 1.5 shows that all of them have the same image by Φ and the same total variation.

Dual certificates. If $\eta = \Phi^* p$ and $\eta \in \partial \mathrm{TV}(u)$, we call η a dual certificate for u in $(\mathcal{P}_0(y_0))$, as its existence proves the optimality of u for $(\mathcal{P}_0(y_0))$ (assuming u is admissible). Similarly, if $\eta = -\Phi^*(\Phi u - y)/\lambda$ and $\eta \in \partial \mathrm{TV}(u)$, we call η a dual certificate for u in $(\mathcal{P}_\lambda(y))$. There could be multiple dual certificates associated to $(\mathcal{P}_0(y_0))$. One of them, the minimal norm certificate, plays a crucial role in our analysis. A quick look at the objective of $(\mathcal{D}_\lambda(y))$ indeed suggests that, as λ goes to 0, its solution converges to the solution of the limit problem $(\mathcal{D}_0(y_0))$ with minimal norm. This is Proposition 1.8 below.

Definition 1.7

If there exists a solution to $(\mathcal{D}_0(y_0))$, the minimal norm dual certificate associated to $(\mathcal{P}_0(y_0))$, denoted η_0, is defined as

$$\eta_0 = \Phi^* p_0 \quad \textit{with} \quad p_0 = \operatorname{argmin} \|p\|_{\mathcal{H}} \quad \textit{s.t.} \quad p \textit{ solves } (\mathcal{D}_0(y_0)) \,.$$

If $\lambda > 0$, we denote $p_{\lambda,w}$ the unique solution of $(\mathcal{D}_\lambda(y_0 + w))$, and $\eta_{\lambda,w} = \Phi^* p_{\lambda,w}$ the associated dual certificate. Noise robustness results extensively rely on the behaviour of $\eta_{\lambda,w}$ as λ and w go to zero. This behaviour is described by the following results.

Proposition 1.8

If there exists a solution to $(\mathcal{D}_0(y_0))$, then $p_{\lambda,0}$ converges strongly to p_0 as $\lambda \to 0$.

Since $p_{\lambda,w}$ is the projection of $(y_0+w)/\lambda$ onto the closed convex set $\{p \in \mathcal{H} \mid \Phi^* p \in \partial \mathrm{TV}(0)\}$, the non-expansiveness of the projection mapping yields

$$\forall (\lambda, w) \in \mathbb{R}_+^* \times \mathcal{H}, \ \|p_{\lambda,w} - p_{\lambda,0}\|_{\mathcal{H}} \leq \frac{\|w\|_{\mathcal{H}}}{\lambda} \,, \tag{9}$$

and hence

$$\forall (\lambda, w) \in \mathbb{R}_+^* \times \mathcal{H}, \ \|\eta_{\lambda,w} - \eta_{\lambda,0}\|_{\mathrm{L}^2(\mathbb{R}^2)} \leq \frac{\|\Phi^*\| \, \|w\|_{\mathcal{H}}}{\lambda} \,.$$

As a result, if $\lambda \to 0$ and $\|w\|_{\mathcal{H}}/\lambda \to 0$, the dual certificate $\eta_{\lambda,w}$ converges in $\mathrm{L}^2(\mathbb{R}^2)$ to the minimal norm certificate η_0.

2.2. Noise robustness results

In this section, we review existing results concerning noise robustness (i.e. question 2 in Section 1.2.1). The main references on this topic are [Chambolle et al., 2016, Iglesias et al., 2018].

Setting. In all the following we consider two sequences $(w_n)_{n\geq 0} \in \mathcal{H}^N$ and $(\lambda_n)_{n\geq 0}$ of noises and regularization parameters, and we assume that $(u_n)_{n\geq 0}$ is such that u_n solves $(\mathcal{P}_{\lambda_n}(y_0 + w_n))$ for every $n \geq 0$.

Set convergence. Let us define a useful notion of set convergence, known as *Painlevé-Kuratowski* set convergence. We refer the reader to [Rockafellar and Wets, 1998, Chapter 4] for more details.

Definition 1.9

If $(E_n)_{n\geq 0}$ is a sequence of sets in $\mathbb{R}^d$, its inner and outer limits are defined by

$$\liminf_{n\to+\infty} E_n \overset{\text{def.}}{=} \left\{ x \in \mathbb{R}^d \,\middle|\, \limsup_{n\to+\infty} \operatorname{dist}(x, E_n) = 0 \right\},$$
$$\limsup_{n\to+\infty} E_n \overset{\text{def.}}{=} \left\{ x \in \mathbb{R}^d \,\middle|\, \liminf_{n\to+\infty} \operatorname{dist}(x, E_n) = 0 \right\}.$$

If there exists a set E such that

$$\liminf_{n\to+\infty} E_n = \limsup_{n\to+\infty} E_n = E\,,$$

we say that $(E_n)_{n\geq 0}$ converges to E and write $\lim_{n\to+\infty} E_n = E$.

If a sequence of closed sets $(E_n)_{n\geq 0}$ is bounded (i.e. there exists $R > 0$ such that $E_n \subset B(0, R)$ for every $n \geq 0$), then the above notion of set convergence is equivalent to convergence in the Hausdorff sense, that is to

$$\lim_{n\to+\infty} \|\operatorname{dist}(\cdot, E_n) - \operatorname{dist}(\cdot, E)\|_\infty = 0\,.$$

First convergence result. General convergence results given in [Hofmann et al., 2007] apply to our setting. This yields the following proposition.

Proposition 1.10 (*[Hofmann et al., 2007, Theorem 3.5]*)

If $\lambda_n \to 0$ and $\|w_n\|^2_{\mathcal{H}} = o(\lambda_n)$ then, up to the extraction of a subsequence (not relabeled), we have that $(u_n)_{n\geq 0}$ converges strongly in $\mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^2)$ and weakly in $\mathrm{L}^2(\mathbb{R}^2)$ to a solution u_* of ($\mathcal{P}_0(y_0)$). Moreover, we have $\mathrm{TV}(u_n) \to \mathrm{TV}(u_*)$.

We stress that the strong $\mathrm{L}^1_{\mathrm{loc}}$ convergence of $(u_n)_{n\geq 0}$ towards u_* and the fact $\mathrm{TV}(u_n) \to \mathrm{TV}(u_*)$ imply $\mathrm{D}u_n \overset{*}{\rightharpoonup} \mathrm{D}u_*$, which in turn implies[1]

$$\operatorname{Supp}(\mathrm{D}u_*) \subseteq \liminf_{n\to+\infty} \operatorname{Supp}(\mathrm{D}u_n)\,. \tag{10}$$

Properties of the level sets in the low noise regime. Stronger convergence results can be obtained by having a closer look at the level sets of u_n. Indeed, from Propositions 1.3 and 1.5, we know they are solutions of the prescribed curvature problem associated to η_{λ_n, w_n}. This information can be exploited to obtain uniform properties of the level sets in the low noise regime. The main ingredient to achieve this is the following lemma, that we use in different parts of this manuscript.

[1] This is for instance proved in [Chambolle et al., 2016, proof of Proposition 8].

Lemma 1.11 (*[Chambolle et al., 2016, Section 5]*)

Let $(\eta_n)_{n\geq 0} \subset \partial \mathrm{TV}(0)$ be a sequence converging strongly in $\mathrm{L}^2(\mathbb{R}^2)$ to η_∞, and let $\mathcal{E}$ be defined by

$$\mathcal{E} \overset{\text{def.}}{=} \left\{ E \subset \mathbb{R}^2,\ 0 < |E| < +\infty \,\middle|\, \exists n \in \mathbb{N} \cup \{\infty\},\ P(E) = \left| \int_E \eta_n \right| \right\}.$$

Then the following holds:

1. $\inf_{E\in\mathcal{E}} P(E) > 0$ *and* $\sup_{E\in\mathcal{E}} P(E) < +\infty$,
2. $\inf_{E\in\mathcal{E}} |E| > 0$ *and* $\sup_{E\in\mathcal{E}} |E| < +\infty$,
3. *there exists $R > 0$ such that, for every $E \in \mathcal{E}$, it holds $E \subset B(0, R)$,*
4. *there exists $r_0 > 0$ and $C \in (0, 1/2)$ such that for every $r \in (0, r_0]$ and $E \in \mathcal{E}$:*

$$\forall x \in \partial E,\ C \leq \frac{|E \cap B(x, r)|}{|B(x, r)|} \leq 1 - C\,.$$

Point 4 in Lemma 1.11 is a weak regularity property. In particular, it ensures that E does not have cusps.

Convergence of level sets. Before stating Proposition 1.12 below,

Proposition 1.12 (*[Chambolle et al., 2016, Iglesias et al., 2018]*)

If $(\mathcal{D}_0(y_0))$ has a solution, $\lambda_n \to 0$ and

$$\frac{\|w_n\|_{\mathcal{H}}}{\lambda_n} \leq \frac{1}{4\, c_2\, \|\Phi^*\|}\,,$$

then $(\mathrm{Supp}(u_n))_{n\geq 0}$ is bounded and, up to the extraction of a subsequence (not relabeled), we have that $(u_n)_{n\geq 0}$ converges strictly in $\mathrm{BV}(\mathbb{R}^2)$ to a solution u_* of $(\mathcal{P}_0(y_0))$. Moreover, for almost every $t \in \mathbb{R}$, we have:

$$\left| U_n^{(t)} \triangle U_*^{(t)} \right| \longrightarrow 0 \quad \text{and} \quad \partial U_n^{(t)} \longrightarrow \partial U_*^{(t)}\,.$$

Remark 1.13

The result of Proposition 1.12 slightly differs from the one proved in [Chambolle et al., 2016, Iglesias et al., 2018]. Indeed, the latter guarantees the convergence of level sets up to the extraction of a further subsequence. We argue that this can be avoided by using Proposition D.1, that we prove in Appendix D.

The extended support. The notion of *extended support* associated to u_0[1] was introduced in [Chambolle et al., 2016, Section 6]. If $(\mathcal{D}_0(y_0))$ has a solution and η_0 is the minimal norm certificate defined in Appendix B, this set, denoted $\mathrm{Ext}(\mathrm{D}u_0)$, is defined as follows:

$$\mathrm{Ext}(\mathrm{D}u_0) \overset{\text{def.}}{=} \overline{\bigcup \left\{ \mathrm{Supp}(\mathrm{D}u) \,\middle|\, \eta_0 \in \partial \mathrm{TV}(u) \right\}}\,.$$

From Proposition 1.3, the extended support is also characterized by

$$\mathrm{Ext}(\mathrm{D}u_0) = \overline{\bigcup \left\{ \partial^* E \,\middle|\, |E| < +\infty \text{ and } P(E) = \left| \int_E \eta_0 \right| \right\}},$$

where $\partial^* E$ denotes the reduced boundary of E (see Appendix A for a definition). In other words, $\mathrm{Ext}(\mathrm{D}u_0)$ contains the reduced boundary of all solutions of the prescribed curvature problems associated to the minimum norm dual certificate and its opposite. In the above-mentioned work, the following result was proved in the case of denoising (i.e. $\Phi = Id$). One may check that it still holds in the case of general measurement operators.

Proposition 1.14

If $(\mathcal{D}_0(y_0))$ has a solution, $\lambda_n \to 0$ and $w_n = o(\lambda_n)$ then

$$\limsup_{n \to +\infty} \mathrm{Supp}(\mathrm{D}u_n) \subseteq \mathrm{Ext}(\mathrm{D}u_0)\,,$$

or equivalently, for every $r > 0$ there exists $n_0 \in \mathbb{N}$ such that

$$\forall n \geq n_0,\ \mathrm{Supp}(\mathrm{D}u_n) \subseteq \{x \in \mathbb{R}^2 \,|\, \mathrm{dist}(x,\ \mathrm{Ext}(\mathrm{D}u_0)) \leq r\}\,.$$

Combining Proposition 1.14 and (10) yields

$$\mathrm{Supp}(\mathrm{D}u_*) \subseteq \liminf_{n \to +\infty} \mathrm{Supp}(\mathrm{D}u_n) \subseteq \limsup_{n \to +\infty} \mathrm{Supp}(\mathrm{D}u_n) \subseteq \mathrm{Ext}(\mathrm{D}u_0)\,.$$

Structure-preserving convergence results. Considering the exact support recovery results mentioned for other reconstruction problems, one could wonder if stronger convergence guarantees can be derived. Namely, if u_0 is identifiable and the sum of a few indicator functions of simple sets, are solutions of $(\mathcal{P}_\lambda(y_0 + w))$ made of the same number of atoms? And if so, are these atoms related to those appearing in the decomposition of u_0? In Part 2, we provide an answer to these questions by adapting the tools introduced in [Duval and Peyré, 2015].

2.3. Numerical methods

In this section, we provide a quick overview of existing numerical methods for solving $(\mathcal{P}_\lambda(y))$[2]. We refer the reader to the recent survey [Chambolle and Pock, 2021a] for a complete review of

[1] We stress that this set only depends on η_0, and hence on y_0, rather than on u_0.

[2] One could also wonder how to numerically solve $(\mathcal{P}_0(y_0))$. As this is less relevant for practical applications, we choose to not cover this question here.

the topic. In the following, as we are interested in the practical resolution of ($\mathcal{P}_\lambda(y)$), we only consider the case of finitely many measurements, i.e. $\mathcal{H} = \mathbb{R}^m$ from some $m \in \mathbb{N}^*$.

The main idea behind most numerical methods is to introduce a fixed spatial discretization and a discrete version of the total variation. To be more precise, one defines a discretization parameter h, a finite dimensional space E^h, and discrete versions of the measurement operator and the total variation $\Phi^h : E^h \to \mathcal{H}$ and $\mathrm{TV}^h : E^h \to \mathbb{R}$, and looks for a solution of

$$\inf_{u \in E^h} \frac{1}{2\lambda} \|\Phi^h u - y\|^2 + \mathrm{TV}^h(u) \,. \tag{11}$$

In most situations (11) is a non-smooth convex problem and can be solved efficiently using standard algorithms. The main difficulty usually lies in the choice of a relevant discretization TV^h of the total variation. In the remaining of this section, we briefly present a few standard choices for TV^h. Our aim is to give an idea of why discretizing the total variation is challenging, and to underline the drawbacks which are common to most existing approaches.

Spatial discretization. As all solutions of ($\mathcal{P}_\lambda(y)$) have their support included in some common ball[1], one can equivalently solve ($\mathcal{P}_\lambda(y)$) in $[-R, R]^2$ (with Dirichlet boundary conditions) for a sufficiently large $R > 0$. The most standard choice of spatial discretization is then to take a positive integer N, define $h \overset{\text{def.}}{=} 2R/N$, and choose E^h to be the space of N by N matrices. Every element $u = (u_{i,j})_{(i,j)\in[1,N]^2}$ of E^h encodes the values of a piecewise constant function on the partition of $[-R, R]^2$ composed of squares of equal size. The resulting approximate solution of ($\mathcal{P}_\lambda(y)$) is hence piecewise constant on this pixel grid.

2.3.1. Finite differences discretizations

In order to simplify the exposition, we only review the two most basic discretizations, and refer to [Chambolle and Pock, 2021a, Section 2] for more details.

For every matrix $u = (u_{i,j})_{(i,j)\in[1,N]^2} \in E^h$ we define

$$\partial_x^h u_{i,j} \overset{\text{def.}}{=} u_{i+1,j} - u_{i,j} \,, \qquad \partial_y^h u_{i,j} \overset{\text{def.}}{=} u_{i,j+1} - u_{i,j} \,,$$

for all $(i, j) \in [0, N]^2$, with the convention $u_{i,j} = 0$ if either i or j is in $\{0, N + 1\}$. We also define the discrete gradient

$$\nabla^h u_{i,j} \overset{\text{def.}}{=} \left(\partial_x^h u_{i,j}, \partial_y^h u_{i,j}\right)$$

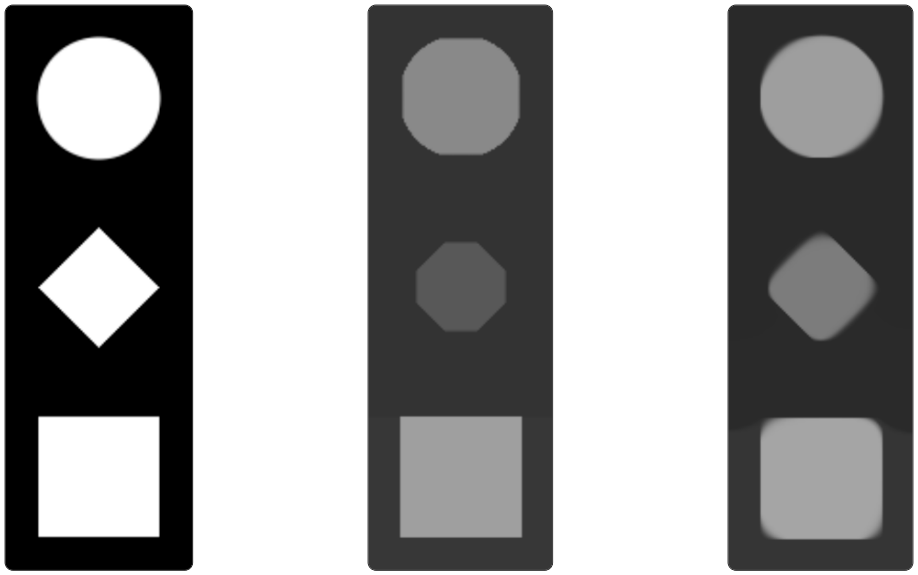
Anisotropic total variation. The most basic idea to discretize the total variation is to define TV^h as follows:

$$\mathrm{TV}^h(u) \overset{\text{def.}}{=} h \sum_{i=0}^{N} \sum_{j=0}^{N} |\nabla^h u_{i,j}|_1 = h \|\nabla^h u\|_{1,1} \,.$$

With this choice $\mathrm{TV}^h(u) = \mathrm{TV}(\tilde{u})$ with $\tilde{u}$ the piecewise constant function naturally associated to u. The main drawback of this choice is that the functional TV^h is strongly "anisotropic": it over-estimates the total variation of oblique discontinuities, and its minimization hence favors

[1] To see this, one simply needs to use Proposition 1.3 with Lemma 1.11.

horizontal and vertical discontinuities. Figure 4 shows how this anisotropy induces a strong bias on the reconstructed images. In fact, one can prove this discretization is not even consistent, as it Γ-converges towards the perimeter associated with a crystalline total variation (see e.g. [Chambolle et al., 2010b]).



Original image Anistropic TV Isotropic TV

Figure 4 – Solutions of the discrete Rudin-Osher-Fatemi denoising problem using the anisotropic and isotropic total variations (figure taken from [Tabti et al., 2018]).

Isotropic total variation. To avoid the anisotropy of the above-mentioned discretization, it is standard to consider the so-called isotropic discrete total variation, defined by

$$\mathrm{TV}^h(u) \overset{\text{def.}}{=} h \sum_{i=0}^{N} \sum_{j=0}^{N} |\nabla^h u_{i,j}|_2 = h \|\nabla^h u\|_{2,1} \,.$$

It is arguably the most widely used discretization. As explained in [Chambolle and Pock, 2021a, Sections 2.1 and 2.2], it still has some anisotropy, since it is not invariation by a rotation of $90°$ of the input image. Reducing this anisotropy is however possible with slightly more sophisticated discretizations. Its main drawback (which is also shared by more complex discretizations) is that its minimization produces blurry images (see Figure 4 for an illustration of this phenomenon), while the (continuous) total variation is precisely used to recover sharp edges.

2.3.2. Dual-based discretizations

Another idea to define TV^h is to consider the dual formulation of TV, given in (1), and to "discretize" the constraint $\|\phi\|_\infty \leq 1$ imposed on the dual variable. To be more precise, this amounts to defining TV^h as follows:

$$\mathrm{TV}^h(u) \overset{\text{def.}}{=} \sup \left\{ h \left\langle p, \nabla^h u \right\rangle \,\middle|\, \|Lp\|_{2,\infty} \leq 1 \right\} , \tag{12}$$

where L is a linear operator encoding the constraints to be imposed on the dual variable p.

Hintermüller, Rautenberg, Hahn and Condat's discretization. A discretization of the form (12) was introduced in two independant works [Hintermüller et al., 2014, Condat, 2017]. The operator L is chosen in this case to enforce the constraint $|p| \leq 1$ both at pixel centers and edges. The performance of this discrete total variation is acknowledged in [Chambolle and Pock, 2021a, Section 4.2]. The experiments presented in [Condat, 2017] indeed show reconstructions that are significantly less blurry than those obtained with the isotropic discretization, although recovered edges are never exactly sharp (i.e. the resulting images are not binary).

Learned dual discretization. In [Chambolle and Pock, 2021b], a learning approach is proposed to find the best choice of operator L on a particular task. The consistency of the resulting discrete total variations is guaranteed: they all Γ-converge towards the continuous total variation[1]. Interestingly, discretizations learned on different tasks significantly differ, which, according to the authors, show that it is hopeless to look for a universal best discrete total variation.

2.3.3. Motivation for grid-free approaches

We now motivate the numerical approach introduced in Part 3.

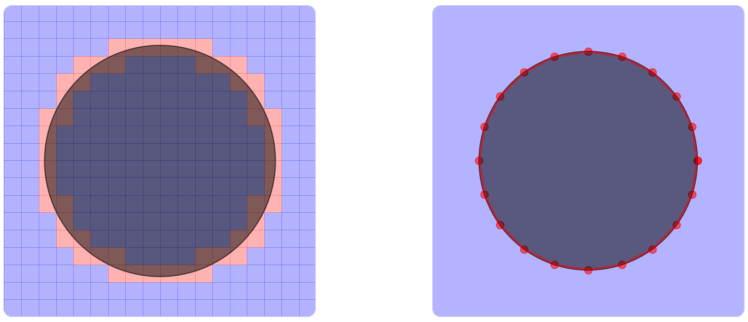
Common drawbacks of fixed grid approaches. As mentioned above, most existing numerical methods for solving ($\mathcal{P}_\lambda(y)$) rely on the introduction of a fixed spatial discretization. This often yield reconstruction artifacts, such as anisotropy or blur. Most importantly, they often fail to preserve the structure exhibited by (some) solutions of ($\mathcal{P}_\lambda(y)$), which are piecewise constant.

Adaptative discretizations. To circumvent these issues, mesh adaptation techniques were introduced (see e.g. [Viola et al., 2012, Bartels et al., 2021]). The spatial discretization is hence adapted to the reconstructed image during the reconstruction process. The refinement rules these works propose are, however, either heuristic or too restrictive to faithfully recover edges. In any case, they still rely on a discretization of the whole domain, and hence do not provide a compact representation of the reconstruced image, which is highly desirable when working with simple images (i.e. that are the superposition of a few simple shapes).

A first grid-free approach. In [Ongie and Jacob, 2016], a method for recovering piecewise constant images from few Fourier samples is introduced. Its orginiality is to produce a continuous domain representation of the image, assuming its edge set is a trigonometric curve. However, this approach heavily relies on relations satisfied by the Fourier coefficients of the image. As such, it does not seem possible to adapt it to handle other types of measurements. In some sense, this method can be seen as the counterpart of Prony-type methods used for the recovery of sparse spikes (see e.g. [Catala, 2020, Chapter 1]), which are specific to certain type of measurements. On the other hand, greedy approaches (that we try to adapt to our setting in Part 3) are agnostic to the choice of measurement operator.

[1] As underlined by the authors, this consistency is however a bit weak, as it does not provide convergence rates or error bounds. This property is for example also satisfied by the isotropic discretization.

Towards general grid-free numerical methods. Our goal is to design an algorithm which does not suffer from some grid bias, while providing a continuous domain representation of solutions. To this aim, we propose to construct an approximate solution built from the atoms promoted by the total variation, namely indicator functions of simple sets. To obtain numerically tractable algorithms, we choose the simple sets we deal with to be simple polygons, although other choices could be considered. A schematic comparison of this approach and grid-based approaches is provided in Figure 5.



Grid-based discretization Boundary discretization

Figure 5 – Comparison of grid-based and grid-free approaches for representing simple images.

Part 2

Noise robustness

This part is devoted to the noise robustness analysis of total variation regularization. It is based on the following forthcoming publication.

- **[De Castro et al., in preparation] Exact recovery of the support of piecewise constant images via total variation regularization**, Y. De Castro, V. Duval and R. Petit, *in preparation*.

We begin by collecting results about the total variation unit ball, and more specifically about the structure of its (exposed) faces. The importance of such an analysis, hinted in Section 1.2.2, is showcased at then end, where we state that elements of a given face can be decomposed using a common set of atoms. Then, we turn to the analysis of the prescribed curvature problem. Its solutions are closely linked with elements of exposed faces, and our noise robustness analysis is intimately related to the stability of its minimizers. Finally, we introduce a so-called non-degenerate source condition, under which exact support recovery is obtained in a low noise regime. We end this part by a discussion on this condition, investigating whether it holds in a simple case.

Contents

1. Piecewise constant functions

Piecewise constant functions are the sparse objects associated to the total variation. They play the same role as sparse vectors for the ℓ_1 norm, and as discrete measures for the total variation of measures. However, these two kinds of objects are arguably easier to deal with, mainly because there is a canonical way to write them as the sum of a few atoms. Finding a good way to decompose piecewise constant functions is somewhat more intricate. This increased complexity is closely linked to properties of the total variation unit ball, whose faces have a non-trivial structure. In this section, we discuss these issues in detail, and precisely define the class of piecewise constant functions we aim at recovering, which we call M-simple.

1.1. Exposed faces of the total variation unit ball

Motivation. As a result of Proposition 1.5 (and recalling Remark 1.6), given $(\lambda, w) \in \mathbb{R}_+^* \times \mathcal{H}$, we have that every nonzero solution u of ($\mathcal{P}_\lambda(y_0 + w)$) satisfies

$$\int_{\mathbb{R}^2} \eta_{\lambda,w} \frac{u}{\mathrm{TV}(u)} = 1 \,,$$

with $\eta_{\lambda,w}$ the dual certificate associated to this problem (defined in Section 2.1.2). The solution set of ($\mathcal{P}_\lambda(y_0 + w)$) is hence included (up to a normalization) in the solution set of

$$\sup_{u \in \mathrm{L}^2(\mathbb{R}^2)} \int_{\mathbb{R}^2} \eta_{\lambda,w}\, u \;\text{ s.t. }\; \mathrm{TV}(u) \leq 1 \,, \tag{13}$$

which is in fact the face of $\{\mathrm{TV} \leq 1\}$ exposed by $\eta_{\lambda,w}$ (see [Rockafellar, 1970, Section 18] for more details on the notion of exposed face). This observation plays a crucial role in our analysis. In the following, we describe the structure of the exposed faces of the total variation unit ball, and derive a useful result allowing to represent its elements using a common set of atoms. This representation is at the core of the proof of our support recovery result. The following is based on the habilitation thesis of Vincent Duval [Duval, 2022] and on a forthcoming work [Boyer et al., in preparation][1]. The results we present here are a subset of those given in [Duval, 2022]. In this last work, general linearly closed faces of the total variation unit ball are considered. We only treat the case of exposed faces, which is enough for our purpose. This allows slightly different (and often shorter) proofs.

Setting. Let us fix $\eta \in \mathrm{L}^2(\mathbb{R}^2)$, and denote $\mathcal{F}$ the face of $\{\mathrm{TV} \leq 1\}$ exposed by η, which is the solution set of

$$\begin{aligned} \sup_{u \in \mathrm{L}^2(\mathbb{R}^2)} \quad & \int_{\mathbb{R}^2} \eta u \\ \text{s.t.} \quad & \mathrm{TV}(u) \leq 1 \,. \end{aligned} \tag{14}$$

The weak compactness of $\{\mathrm{TV} \leq 1\}$ yields that $\mathcal{F}$ is non-empty (i.e. the supremum in (14) is attained). Moreover, if η is not equal to zero almost everywhere, it has a Lebesgue point x such

[1] We also stress, as in [Duval, 2022, Section 2.6.2], that the total variation is a particular case of submodular function. The faces of the unit ball defined by such functions have been studied in the monographs [Fujishige, 2005, Bach, 2013].

that $|\eta(x)| > 0$. There hence exists $r > 0$ such that $|\int_{B(x,r)} \eta| > 0$, which shows the value of (14) is strictly positive, say $1/\alpha$ for some $\alpha \in \mathbb{R}_+^*$. Hence, we have that

$$\forall u \in \mathrm{L}^2(\mathbb{R}^2),\ \mathrm{TV}(u) - \int_{\mathbb{R}^2} (\alpha\eta)u \geq 0\,,$$

with equality if and only if u solves (14). For convenience reasons, we assume in the following that $\alpha = 1$ (all the results below remain valid if $\alpha \neq 1$ by replacing η with $\eta' \stackrel{\text{def.}}{=} \alpha\eta$), and consequently obtain

$$\forall u \in \mathrm{L}^2(\mathbb{R}^2),\ \mathrm{TV}(u) - \int_{\mathbb{R}^2} \eta u \geq 0\,. \tag{15}$$

With this assumption we have the following characterization of $\mathcal{F}$:

$$\mathcal{F} = \left\{ u \in \mathrm{L}^2(\mathbb{R}^2) \;\middle|\; \mathrm{TV}(u) \leq 1 \text{ and } \int_{\mathbb{R}^2} \eta\, u = 1 \right\},$$

and every $u \in \mathcal{F}$ satisfies $\int_{\mathbb{R}^2} \eta\, u = \mathrm{TV}(u)$.

Remark 2.1

Let us stress that Problem 14 has value 1 (i.e. $\alpha = 1$) if and only if their exists a nonzero function $u \in \mathrm{L}^2(\mathbb{R}^2)$ which satisfies $\int_{\mathbb{R}^2} \eta\, u = \mathrm{TV}(u)$ (i.e. $\eta \in \partial\mathrm{TV}(u)$). Dual certificates introduced above all fall in this category, unless the unique solution of ($\mathcal{P}_0(y_0)$) (respectively ($\mathcal{P}_\lambda(y)$)) is zero.

1.1.1. Indicator functions

From Proposition 1.2, we know that the extreme points of $\{\mathrm{TV} \leq 1\}$ are the functions of the form $\pm\mathbf{1}_E/P(E)$, with E a simple set such that $0 < |E| < +\infty$. Elements of $\mathcal{F}$ that are proportional to indicator functions hence play a special role in the analysis of its structure. Let us therefore define

$$\begin{aligned}
\mathcal{E} &\stackrel{\text{def.}}{=} \mathcal{E}^+ \cup \mathcal{E}^- \cup \{\emptyset, \mathbb{R}^2\}\,,\\
\mathcal{E}^+ &\stackrel{\text{def.}}{=} \left\{ E \subset \mathbb{R}^2 \;\middle|\; |E| < +\infty,\ 0 < P(E) < +\infty,\ \frac{\mathbf{1}_E}{P(E)} \in \mathcal{F} \right\},\\
\mathcal{E}^- &\stackrel{\text{def.}}{=} \left\{ E \subset \mathbb{R}^2 \;\middle|\; |E^c| < +\infty,\ 0 < P(E^c) < +\infty,\ \frac{-\mathbf{1}_{E^c}}{P(E^c)} \in \mathcal{F} \right\}.
\end{aligned} \tag{16}$$

Remark 2.2

We stress that an extreme point of $\{\mathrm{TV} \leq 1\}$ is not necessarily exposed. In fact, $E \in \mathcal{E}^+$ if and only if $P(E) = \int_E \eta$ (and $E \in \mathcal{E}^-$ if and only if $P(E^c) = -\int_{E^c} \eta$). This is equivalent to require that E (respectively E^c) is a solution of the prescribed curvature problem associated to η (respectively $-\eta$) studied in Section 2. In particular, this implies that E (respectively E^c) is bounded, that the number of Jordan curves in the decomposition of $\partial^M E$ is finite, and that the following weak regularity property holds:

$$\exists r_0 > 0,\ \forall r \in (0, r_0],\ \forall x \in \partial E,\ \frac{1}{16} \leq \frac{|E \cap B(x,r)|}{|B(x,r)|} \leq 1 - \frac{1}{16}\,.$$

We refer the reader to [Chambolle et al., 2016, Section 5] for more details.

The set $\mathcal{E}$ has the remarkable property of being closed under several operations that we describe below. We first consider the intersection and union operations, before turning to M-connected components and holes.

Proposition 2.3

The set $\mathcal{E}$ is closed under countable unions and countable intersections.

Proof : If $E \in \mathcal{E}^+$ and $F \in \mathcal{E}^+$ the submodularity of the perimeter (see e.g. [Ambrosio et al., 2001, Proposition 1]) yields:

$$P(E \cap F) + P(E \cup F) \leq P(E) + P(F) = \int_E \eta + \int_F \eta = \int_{E \cap F} \eta + \int_{E \cup F} \eta \,.$$

We hence obtain:

$$\left(P(E \cap F) - \int_{E \cap F} \eta \right) + \left(P(E \cup F) - \int_{E \cup F} \eta \right) \leq 0 \,.$$

Using (15), we get that the two terms above are nonnegative, which yields $E \cap F \in \mathcal{E}^+$ (unless $|E \cap F| = 0$) and $E \cup F \in \mathcal{E}^+$. The same reasoning shows that the result also holds when replacing $\mathcal{E}^+$ by $\mathcal{E}^-$.

If $E \in \mathcal{E}^+$ and $F \in \mathcal{F}^-$ we have:

$$\begin{aligned} P(E \cap F) + P((E \cup F)^c) &= P(E \cap F) + P(E \cup F) \\ &\leq P(E) + P(F) \\ &= P(E) + P(F^c) \\ &= \int_E \eta - \int_{F^c} \eta \\ &= \int_{E \cap F} \eta - \int_{(E \cup F)^c} \eta \end{aligned}$$

Reasoning as above, we obtain $E \cap F \in \mathcal{E}^+$ (unless $|E \cap F| = 0$) and $E \cup F \in \mathcal{E}^-$ (unless $|(E \cup F)^c| = 0$), which shows $E \cap F \in \mathcal{E}$.

Now, we have proved that $\mathcal{E}$ is stable under finite unions and finite intersections. The countable case follows from the fact exposed faces are closed in $\mathrm{L}^2(\mathbb{R}^2)$. ∎

Proposition 2.4

Let $E, F \in \mathcal{E}$ with $F \subset E$. If there exists $C_1, C_2 \subset \mathbb{R}^2$ such that

1. *$|C_1| > 0$ and $|C_2| > 0$*
2. *$E \setminus F = C_1 \cup C_2$*
3. *$P(E \setminus F) = P(C_1) + P(C_2)$*

then $F \cup C_1 \in \mathcal{E}$ and $F \cup C_2 \in \mathcal{E}$.

Proof : We first note that $C_1 \subsetneq \mathbb{R}^2$ and $C_2 \subsetneq \mathbb{R}^2$ (otherwise we would have $E = \mathbb{R}^2$ and $F = \emptyset$, which would contradict the indecomposability of $\mathbb{R}^2$). Using 2. and 3., from [Ambrosio et al., 2001, Proposition 3] we hence obtain that $\mathcal{H}^1(\partial^* C_1 \cap \partial^* C_2) = 0$. Now, using Proposition 1 of the same

reference, we obtain:

$$\begin{aligned} P(E) + P(F) &= 2\,P(F) + P(E \setminus F) - 2\,\mathcal{H}^1(\partial^* F \cap \partial^*(E \setminus F)) \\ &= 2\,P(F) + P(E \setminus F) - 2\,\mathcal{H}^1(\partial^* F \cap \partial^*(C_1 \cup C_2)) \\ &= (P(F) + P(C_1) - 2\,\mathcal{H}^1(\partial^* F \cap \partial^* C_1)) + (P(F) + P(C_2) - 2\,\mathcal{H}^1(\partial^* F \cap \partial^* C_2)) \\ &= P(G_1) + P(G_2)\,, \end{aligned}$$

with $G_1 \overset{\text{def.}}{=} F \cup C_1$ and $G_2 \overset{\text{def.}}{=} F \cup C_2$. We can then conclude as in the proof of Proposition 2.3. ■

As a consequence of Proposition 2.4, we obtain the following result. We remind the reader that M-connected components and holes are defined in Appendix A.

Corollary 2.5

If $E \in \mathcal{E}$, any M-connected component of E belongs to $\mathcal{E}$. Moreover, for any hole Y of E, we have that $E \cup Y \in \mathcal{E}$.

1.1.2. Chains and maximal chains

Considering exposed faces (instead of general linearly closed faces) does not allow for shorter proofs of the results given in this section. We hence do not include them and refer the reader to [Duval, 2022, Section 2.3].

We say that a collection $\mathcal{C}$ of subsets of $\mathbb{R}^2$ is a chain if for every $E, F \in \mathcal{C}$ we have $E \subseteq F$ or $F \subseteq E$, i.e. if $\mathcal{C}$ is totally ordered for the inclusion relation. We call its cardinal (denoted $|\mathcal{C}|$) the length of $\mathcal{C}$. Denoting $\dim(\mathcal{F})$ the dimension of the affine hull of $\mathcal{F}$, we may state the following result.

Proposition 2.6

If $\mathcal{C} \subseteq \mathcal{E} \setminus \{\emptyset, \mathbb{R}^2\}$ is a chain and $\dim(\mathcal{F}) = d$, then $\mathcal{C}$ has length at most $d + 1$.

We say that a chain $\mathcal{C}$ in $\mathcal{E}$ is maximal if there is no $E \in \mathcal{E} \setminus \mathcal{C}$ such that $\mathcal{C} \cup \{E\}$ is a chain. Such a chain must therefore contain $\emptyset$ and $\mathbb{R}^2$. As a result, if $\mathcal{C} = \{E_i\}_{i=0}^m$ is a maximal chain, then the collection of its increments $\{E_i \setminus E_{i-1}\}_{i=1}^m$ is a partition of $\mathbb{R}^2$. The following proposition shows that every set in $\mathcal{E}$ can be written as the union of elements of this partition.

Proposition 2.7

Assume $\mathcal{C} = \{E_i\}_{i=0}^m$ is a maximal chain in $\mathcal{E}$ with

$$\emptyset = E_0 \subset E_1 \subset \ldots \subset E_m = \mathbb{R}^2\,.$$

Then the following holds.

1. *For every $i \in \{1, \ldots, m\}$, the set $E_i \setminus E_{i-1}$ is indecomposable.*
2. *For every $F \in \mathcal{E}$, there exists $I \subseteq \{1, \ldots, m\}$ such that $F = \bigcup_{i \in I} (E_i \setminus E_{i-1})$.*

Even if there could be several maximal chains in $\mathcal{F}$, they all have the same length and the same collection of increments, as the following proposition shows. In particular, we see that each

exposed face $\mathcal{F}$ with finite dimension is naturally associated to a partition of $\mathbb{R}^2$ (which is the collection of increments of any maximal chain in $\mathcal{F}$).

Proposition 2.8

If $\mathcal{C} = \{E_i\}_{i=0}^m$ and $\mathcal{C}' = \{E'_i\}_{i=0}^{m'}$ are two maximal chains in $\mathcal{E}$, we have:

1. $|\mathcal{C}| = |\mathcal{C}'| = \dim(\mathcal{F}) + 3$,
2. $\{E_i \setminus E_{i-1}\}_{i=1}^m = \{E'_i \setminus E'_{i-1}\}_{i=1}^{m'}$.

1.2. Chains associated to a piecewise constant function

1.2.1. Definition and properties

Let $u \in \mathrm{L}^2(\mathbb{R}^2)$ be a function with finite total variation, taking a finite number of values[1]

$$t_1 > ... > t_{i_0-1} > t_{i_0} = 0 > t_{i_0+1} > ... > t_m .$$

Let $E_i \overset{\text{def.}}{=} \{u \geq t_i\}$ if $1 \leq i \leq m$ and $E_0 \overset{\text{def.}}{=} \emptyset$. We hence have $\emptyset = E_0 \subset E_1 \subset ... \subset E_m = \mathbb{R}^2$, i.e. $\{E_i\}_{i=0}^m$ is a chain.

Now if $F_i \overset{\text{def.}}{=} E_i \setminus E_{i-1} = \{u = t_i\}$ for every $1 \leq i \leq m$, we have $\mathcal{CC}^M(F_i) = \{F_{i,j}\}_{j \in J_i}$ with J_i at most countable. Defining $J \overset{\text{def.}}{=} \{(i,j) \,|\, 1 \leq i \leq m,\ 1 \leq j \leq |J_i|\}$, for every $(i,j) \in J$ we introduce

$$E_{i,j} \overset{\text{def.}}{=} E_{i-1} \bigcup \left(\bigcup_{1 \leq k \leq j} F_{i,j} \right) .$$

With these notations, the lexicographic order on J yields a total order on $\{E_{i,j}\}_{(i,j)\in J}$, which is hence a chain. We say that it is a chain associated to the function u (see Figure 6 for an illustration). By construction, u is constant on each of its increments $(F_{i,j})_{(i,j)\in J}$, which are all indecomposable.

Remark 2.9

Since the order we choose on $\mathcal{CC}^M(\{u = t_i\})$ is arbitrary, multiple chains can be associated to u (see Figure 6 for an illustration). These chains however have the same collection of increments $\mathcal{I}$, which is given by:

$$\mathcal{I} = \bigcup_{i=1}^m \mathcal{CC}^M(\{u = t_i\}) .$$

Considering the construction above, we now define the sparse objects naturally associated to the total variation, which we call *M-simple functions*. This is the class of piecewise constant functions we work with in the rest of this manuscript.

[1] Meaning that $|\{u = t_i\}| > 0$ for all $i \in \{1, ..., m\}$ and, for almost every $x \in \mathbb{R}^2$, we have $u(x) \in \{t_i\}_{i=1}^m$.

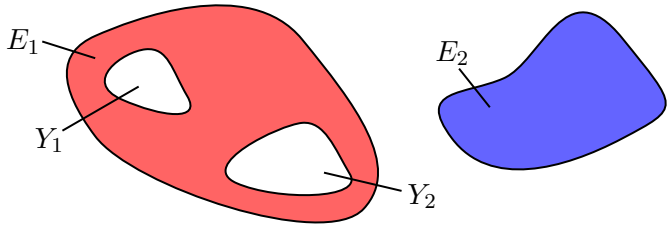


Figure 6 – A function $u = \mathbf{1}_{E_1} - \mathbf{1}_{E_2}$.

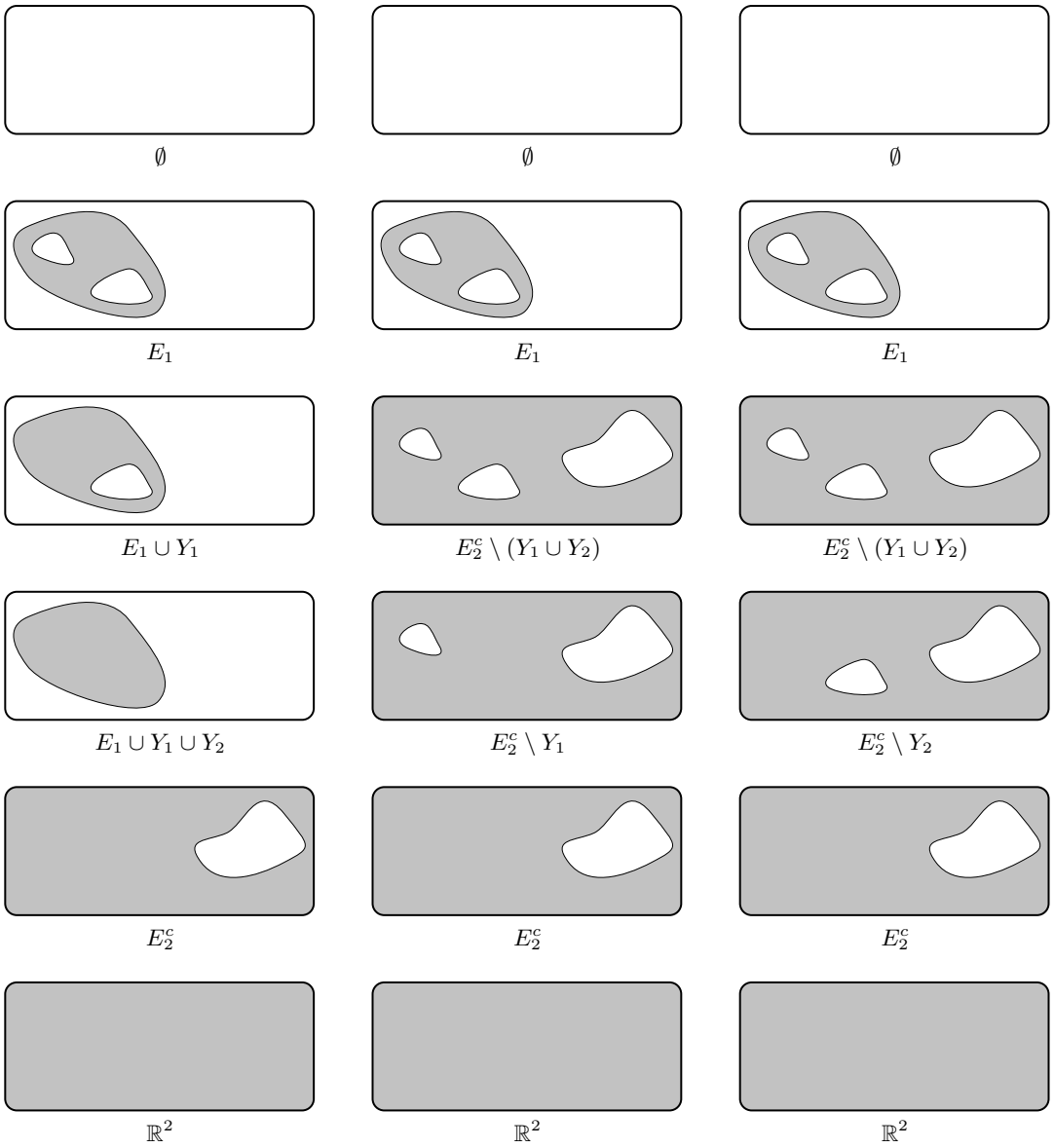


Figure 7 – Three chains associated to u (see Figure 6). Each column corresponds to a chain, whose elements are ordered from top to bottom.

Definition 2.10 (*M-simple functions*)

We say that a function $u : \mathbb{R}^2 \to \mathbb{R}$ is M-simple if there exists a finite set I, a collection $\{E_i\}_{i\in I}$ of indecomposable sets of finite measure and $t \in \mathbb{R}^I$ such that

$$u = \sum_{i\in I} t_i \mathbf{1}_{E_i} .$$

We stress that, with this definition, any M-simple function u belongs to $\mathrm{L}^2(\mathbb{R}^2)$ and has finite total variation. We also have that a function $u \in \mathrm{L}^2(\mathbb{R}^2)$ with finite total variation is M-simple if and only if its associated chains have finite length. Moreover, the length of these chains is exactly the sum, over all values t taken by u, of the number of bounded M-connected components of $\{u = t\}$.

Sparsity of a piecewise constant function. The "right" measure of sparsity for a piecewise constant function u is the sum, over all values t taken by u, of the number of bounded M-connected components of $\{u = t\}$.

We conclude this section by stating the following useful result, whose importance is explained below.

Proposition 2.11

Assume that $\dim(\mathcal{F}) = d$ and $u \in \mathcal{F}$. Then u is M-simple. Moreover, any chain associated to u is a chain in $\mathcal{E}$, and hence has at most $d + 1$ elements (not counting $\emptyset$ and $\mathbb{R}^2$). Finally, u is constant on the increments of any maximal chain in $\mathcal{E}$.

Proof : Using Carathéodory's theorem, we have that u can be written as a convex combination of $d + 1$ extreme points of $\mathcal{F}$, which are of the form $\epsilon\, \mathbf{1}_E/P(E)$ with $\epsilon \in \{-1, 1\}$ and E a simple set such that $0 < |E| < +\infty$. As a consequence, u takes a finite number of values. Using the above-defined notations we have:

$$u = \sum_{i=1}^{i_0-1} \theta_i \mathbf{1}_{E_i} + \sum_{i=i_0}^{m-1} \theta_i \left(-\mathbf{1}_{E_i^c}\right) \quad \text{and} \quad \int_{\mathbb{R}^2} \eta\, u = \sum_{i=1}^{i_0-1} \theta_i \int_{E_i} \eta - \sum_{i=i_0}^{m-1} \theta_i \int_{E_i^c} \eta ,$$

with $\theta_i > 0$. By the coarea formula (see e.g. [Maggi, 2012, Chapter 13]), we also have:

$$\sum_{i=1}^{i_0-1} \theta_i\, P(E_i) + \sum_{i=i_0}^{m-1} \theta_i\, P(E_i^c) = \mathrm{TV}(u) .$$

Using $\int_{\mathbb{R}^2} \eta\, u = \mathrm{TV}(u)$ we hence obtain $P(E_i) = \int_{E_i} \eta$ for $1 \le i \le i_0 - 1$ and $P(E_i^c) = -\int_{E_i^c} \eta$ for $i_0 \le i \le m - 1$, which shows $\{E_i\}_{1\le i\le m-1}$ is a chain in $\mathcal{E}$. We conclude that $\{E_{i,j}\}_{(i,j)\in J}$ is a chain in $\mathcal{E}$ by using Proposition 2.4. The last claim is a straightforward consequence of Proposition 2.7, since the level sets of u all belong to $\mathcal{E}$. ∎

The main interest of this proposition is that, denoting $\{F_j\}_{1\le j\le m}$ the collection of increments of any maximal chain in $\mathcal{F}$, every function $u \in \mathcal{F}$ is of the form

$$u = \sum_{j=1}^{m} t_j\, \mathbf{1}_{F_j} ,$$

with $t \in \mathbb{R}^m$. To put it another way, the elements of $\mathcal{F}$ can be decomposed using a common set of atoms, which is the collection $\{\mathbf{1}_{F_j}\}_{1\leq j\leq m}$.

1.2.2. A sufficient condition for identifiability

In this subsection, we derive a sufficient identifiability condition using the above results. To this aim, we first recall a condition, introduced in [Burger and Osher, 2004] as the source condition, which ensures that a given function u_0 solves $(\mathcal{P}_0(y_0))$.

Definition 2.12

A function u_0 satisfies the source condition if there exists $\eta \in \operatorname{Im}\Phi^*$ such that $\eta \in \partial \mathrm{TV}(u_0)$.

If u_0 is an M-simple function, the source condition can be strengthened to ensure it is the unique solution of $(\mathcal{P}_0(y_0))$ (i.e. u_0 is identifiable). To this aim, for any M-simple function u and any chain $\mathcal{C}$ associated to u, whose ordered increments are denoted $(F_j)_{1\leq j\leq m}$, we define

$$\begin{aligned}\Phi_{\mathcal{C}} : \mathbb{R}^m &\to \mathcal{H} \\ t &\mapsto \Phi\left(\sum_{j=1}^m t_j \, \mathbf{1}_{F_j}\right).\end{aligned}$$

We stress that this mapping only mildly depends on the choice of a chain $\mathcal{C}$, in the sense that if $\mathcal{C}'$ is another chain associated to u, then $\mathcal{C}$ and $\mathcal{C}'$ have the same collection of increments. There hence exists a permutation of the coordinates σ such that $\Phi_{\mathcal{C}} = \Phi_{\mathcal{C}'} \circ \sigma$.

We may now state the following result, which is a natural analog of [Duval and Peyré, 2015, Proposition 5].

Proposition 2.13

Let u_0 be an M-simple function. Assume there exists $\eta \in \operatorname{Im}\Phi^* \cap \partial \mathrm{TV}(u_0)$. If a chain $\mathcal{C}_0$ associated to u_0 is maximal in the face exposed by η, and if $\Phi_{\mathcal{C}_0}$ is injective, then u_0 is the unique solution of $(\mathcal{P}_0(y_0))$.

Proof : Since $\eta \in \operatorname{Im}\Phi^*$ and $\eta \in \partial \mathrm{TV}(u_0)$, by Proposition 1.4 we have that every solution u of $(\mathcal{P}_0(y_0))$ satisfies $\eta \in \partial \mathrm{TV}(u)$, and hence that $u/\mathrm{TV}(u)$ is in the face exposed by η. Since $\mathcal{C}_0$ is a maximal chain in this face, denoting $(F_j)_{1\leq j\leq m}$ its increments, we obtain that $u/\mathrm{TV}(u)$ is constant on each F_j, which yields the existence of $t \in \mathbb{R}^m$ such that

$$u = \sum_{j=1}^m t_j \mathbf{1}_{F_j}.$$

The injectivity of $\Phi_{\mathcal{C}_0}$ finally allows to conclude. ∎

Remark 2.14

We stress that a chain associated to an M-simple function u is maximal in an exposed face of $\{\mathrm{TV} \leq 1\}$ if and only if all chains associated to u are maximal in this face. This directly follows from the fact all these chains have the same length. As a consequence, the result stated in Proposition 2.13 does not depend on the choice of a particular chain.

The main question which remains unanswered is the following: if u_0 is M-simple and identifiable, with w and λ small enough, are solutions of $(\mathcal{P}_\lambda(y_0 + w))$ M-simple, what is the length of their associated chains, and how are their elements related to those of $\mathcal{C}_0$? To answer an analog question for the sparse spikes problem, a non-degenerate version of the source condition was introduced in [Duval and Peyré, 2015, Definition 5].

From Proposition 1.3 and (8), the elements of the chains associated to solutions of $(\mathcal{P}_\lambda(y_0+w))$ are all solutions of the prescribed curvature problem associated to $\eta_{\lambda,w}$. In Section 2.1.2, we have also seen that, under a few assumptions, $\eta_{\lambda,w}$ converges to the minimal norm certificate η_0 when w and λ go to zero. It is therefore natural to investigate how solutions of the prescribed curvature problem behave under variations of the curvature functional, which is the topic of the following section. The results therein allow us to state a natural analog of the non-degenerate source condition in Section 3, and to finally answer the above question.

2. The prescribed curvature problem

As mentioned above, this section is dedicated to the study of the prescribed curvature problem associated to some function $\eta \in \partial \mathrm{TV}(0)$:

$$\inf_{E\subset\mathbb{R}^2,\,|E|<+\infty} J(E) \overset{\text{def.}}{=} P(E) - \int_E \eta\,. \tag{$\mathcal{PC}(\eta)$}$$

Our main aim is to investigate how the solution set of $(\mathcal{PC}(\eta))$ behaves when η varies. To be more specific, given two sufficiently close curvature functionals η and η' we are interested in answering the following two questions.

(i) Are solutions of $(\mathcal{PC}(\eta'))$ close to solutions of $(\mathcal{PC}(\eta))$?

(ii) How many solutions of $(\mathcal{PC}(\eta'))$ are there in a neighborhood of a given solution of $(\mathcal{PC}(\eta))$?

In the following subsection, we answer the first question using the notion of quasi-minimizers of the perimeter, as well as first order optimality conditions for $(\mathcal{PC}(\eta))$. Then, under a strict stability assumption on solutions of $(\mathcal{PC}(\eta))$, we answer the second question using the notion of second order shape derivatives.

Tools. As discussed in the following subsection, if η is sufficiently regular, the solutions of $(\mathcal{PC}(\eta))$ also enjoy some regularity. In this section, we hence mainly work with smooth sets and extensively use related notions. Relevant definitions and properties are collected in Appendix B. We also use the notion of curvature (defined in Appendix A) and of second fundamental form (see e.g. [Maggi, 2012, Section 17.6]). Given a sufficiently smooth set E, we often consider sets that are normal deformations of E, i.e. whose boundary is a normal graph over ∂E. Such sets are parametrized by real-valued functions on ∂E, which leads us to use the notion of tangential gradient, tangential Jacobian, and the spaces $\mathrm{C}^k(\partial E)$, $\mathrm{L}^p(\partial E)$ and $\mathrm{H}^1(\partial E)$ (along with their associated norms). We refer to the reader to [Henrot and Pierre, 2018, Sections 5.4.1, 5.4.3 and 5.9.1] for a precise definition of these objects.

2.1. Generalities and first convergence result

Existence of minimizers. We first stress that existence of solutions is guaranteed for $(\mathcal{PC}(\eta))$. Indeed, since $\eta \in \partial \mathrm{TV}(0)$, the objective J is always nonnegative, and equal to zero when evaluated at the empty set, which is admissible. From Proposition 1.3, we also know that it has a non trivial solution as soon as there exists a nonzero function $u \in \mathrm{L}^2(\mathbb{R}^2)$ such that $\eta \in \partial \mathrm{TV}(u)$.

Boundedness. As already mentioned, by Lemma 1.11 all solutions of $(\mathcal{PC}(\eta))$ are included in some common ball, i.e. there exists $R > 0$ such that, for every solution E of $(\mathcal{PC}(\eta))$, we have $E \subset B(0, R)$.

Regularity of the solutions. Let us now discuss the regularity of solutions of $(\mathcal{PC}(\eta))$. If we have $\eta \in \mathrm{L}^\infty(\mathbb{R}^2)$, then any solution of $(\mathcal{PC}(\eta))$ is a strong quasi-minimizer of the perimeter, and, consequently, is of class $\mathrm{C}^{1,1}$ (see [Ambrosio, 2010, Definition 4.7.3 and Theorem 4.7.4]). If η is moreover continuous, then the boundary of any solution is locally the graph of a function u which solves (in the sense of distributions) the Euler-Lagrange equation associated to $(\mathcal{PC}(\eta))$, which is (up to a translation and a rotation):

$$\left(\frac{u'}{\sqrt{1+u'^2}}\right)'(z) = \frac{u''(z)}{\left(1+u'(z)^2\right)^{3/2}} = \eta(z, u(z))\,. \tag{17}$$

This in turn implies that u is C^2 ($\mathrm{C}^{k+2,\alpha}$ if $\eta \in \mathrm{C}^{k,\alpha}(\mathbb{R}^2)$) and solves (17) in the classical sense. As a result, assuming η is bounded and of class C^1, every solution of $(\mathcal{PC}(\eta))$ is of class C^3.

Convergence result. Now, we state a result that, loosely speaking, tells that any neighborhood (in terms of C^2-normal deformations) of the solution set of $(\mathcal{PC}(\eta_0))$ contains the solution set of $(\mathcal{PC}(\eta))$ provided η is sufficiently close to η_0 in $\mathrm{C}^1(\mathbb{R}^2)$ and $\mathrm{L}^2(\mathbb{R}^2)$. This provides an answer to question (i).

Proposition 2.15

Let $\eta_0 \in \partial \mathrm{TV}(0) \cap \mathrm{C}^1_b(\mathbb{R}^2)$. For every $\epsilon > 0$ there exists $r > 0$ such that for every $\eta \in \partial \mathrm{TV}(0)$ with $\|\eta - \eta_0\|_{\mathrm{L}^2(\mathbb{R}^2)} + \|\eta - \eta_0\|_{\mathrm{C}^1(\mathbb{R}^2)} \leq r$, the following holds: every solution F of $(\mathcal{PC}(\eta))$ is a C^2-normal deformation of size at most ϵ of a solution E of $(\mathcal{PC}(\eta_0))$ (i.e., with the notation of Proposition B.5, $F = E_\varphi$ with $\|\varphi\|_{\mathrm{C}^2(\partial E)} \leq \epsilon$).

Proof : We argue by contradiction and assume the existence of two sequences $(\eta_n)_{n\in\mathbb{N}^*}$ and $(F_n)_{n\in\mathbb{N}^*}$ such that

- for all $n \in \mathbb{N}^*$, $\eta_n \in \partial \mathrm{TV}(0) \cap \mathrm{C}^1_b(\mathbb{R}^2)$,
- the sequence $(\eta_n)_{n\in\mathbb{N}^*}$ converges to η_0 in $\mathrm{L}^2(\mathbb{R}^2)$ and $\mathrm{C}^1_b(\mathbb{R}^2)$,
- for all $n \in \mathbb{N}^*$, the set F_n solves $(\mathcal{PC}(\eta_n))$ and cannot be written as a C^2-normal deformation of size at most ϵ of a solution of $(\mathcal{PC}(\eta_0))$.

We hence have that $(F_n)_{n\in\mathbb{N}^*}$ is bounded and that F_n is a strong (Λ, r_0)-quasi minimizer of the perimeter (in short form $F_n \in \mathcal{QM}(\Lambda, r_0)$, see [Maggi, 2012, Section 21] and [Ambrosio, 2010, Definition 4.7.3]) for all $n \in \mathbb{N}^*$, with $\Lambda = \sup\{\|\eta_n\|_\infty,\ n \in \mathbb{N}^*\}$ and r_0 any positive real number.

Taking r_0 small enough to have $\Lambda\, r_0 \leq 1$, from [Maggi, 2012, Propositions 21.13 and 21.14] we get that, up to the extraction of a subsequence (not relabeled), $(F_n)_{n\geq 0}$ converges in measure to a bounded set $E \in \mathcal{QM}(\Lambda, r_0)$, and that $(\partial F_n)_{n\geq 0}$ converges to ∂E. From $|F_n \triangle E| \to 0$ we obtain that E is a solution of ($\mathcal{PC}(\eta_0)$), and the convergence of $(\partial F_n)_{n\geq 0}$ towards E yields

$$\forall r > 0,\ \exists n_0 \in \mathbb{N},\ \forall n \geq n_0,\ \partial F_n \subset \bigcup_{x\in\partial E} \mathrm{C}(x, r, \nu_E(x))\,,$$

where $\mathrm{C}(x, r, \nu_E(x))$ denotes the square of axis $\nu_E(x)$ and side r centered at x, defined in (60). From [Ambrosio, 2010, 4.7.4], and arguing as in the proof of [Maggi, 2012, Theorem 26.6], for every $x \in \partial E$ we obtain the existence of $r > 0$, of $n_0 \in \mathbb{N}$, of $u \in \mathrm{C}^{1,1}([-r, r])$ and of a sequence $(u_n)_{n\geq n_0}$ which is uniformly bounded in $\mathrm{C}^{1,1}([-r, r])$, such that, in $\mathrm{C}(x, r, \nu_E(x))$, the set E is the hypograph of u and, for every $n \geq n_0$, the set F_n is the hypograph of u_n. Moreover, we have that $\|u_n - u\|_{\mathrm{C}^1([-r,r])} \to 0$.

Now, we also have that u and u_n (for $n \geq n_0$) respectively solve (in the sense of distributions) the following equations in $(-r, r)$:

$$\begin{aligned} \frac{u''(z)}{(1+u'(z)^2)^{3/2}} &= H(z, u(z))\,, \qquad \text{with } H(z,t) \overset{\text{def.}}{=} \eta_0(x + R_{\nu_E(x)}(z,t))\,, \\ \frac{u_n''(z)}{(1+u_n'(z)^2)^{3/2}} &= H_n(z, u_n(z))\,, \quad \text{with } H_n(z,t) \overset{\text{def.}}{=} \eta_n(x + R_{\nu_E(x)}(z,t))\,. \end{aligned} \tag{18}$$

We hence immediately obtain that u and u_n belong to $\mathrm{C}^2([-r, r])$. Moreover, for every $z \in (-r, r)$ we have:

$$\begin{aligned} |u_n''(z) - u''(z)| &= \left| H_n(z, u_n(z)) \left(1 + u_n'(z)^2\right)^{3/2} - H(z, u(z)) \left(1 + u'(z)^2\right)^{3/2} \right| \\ &\leq \quad (\|H_n - H\|_\infty + |H(z, u_n(z)) - H(z, u(z))|) \left(1 + u_n'(z)^2\right)^{3/2} \\ &\quad + \|H\|_\infty \left[\left(1 + u_n'(z)^2\right)^{3/2} - \left(1 + u'(z)^2\right)^{3/2} \right], \end{aligned}$$

from which we obtain that $\|u_n'' - u''\|_\infty \to 0$.

Using these new results in combination with (18), we get that u and u_n belong to $\mathrm{C}^3([-r, r])$. Differentiating (18), we obtain, for every $z \in (-r, r)$:

$$\begin{aligned} u^{(3)}(z) = &\quad \left[\partial_1 H(z, u(z)) + u'(z)\, \partial_2 H(z, u(z))\right] (1 + u'(z)^2)^{3/2} \\ &+ 3\, H(z, u(z))\, u''(z)\, u'(z)\, (1 + u'(z)^2)^{3/2}\,, \\ u_n^{(3)}(z) = &\quad \left[\partial_1 H_n(z, u_n(z)) + u_n'(z)\, \partial_2 H_n(z, u_n(z))\right] (1 + u_n'(z)^2)^{3/2} \\ &+ 3\, H_n(z, u_n(z))\, u_n''(z)\, u_n'(z)\, (1 + u_n'(z)^2)^{3/2}\,, \end{aligned}$$

from which we can finally show $\|u_n^{(3)} - u^{(3)}\|_\infty \to 0$.

Finally, using the compactness of ∂E, we obtain that $(F_n)_{n\geq 0}$ converges in C^3 towards E, and Proposition B.3 allows to eventually write F_n as a C^2-normal deformation of E, whose norm converges to zero. This yields a contradiction. ∎

2.2. Stability result

Question (ii) is closely linked to the stability of minimizers of ($\mathcal{PC}(\eta)$), that is to the behaviour of the objective J in a neighborhood of a solution. To analyze this behaviour, we use the general

framework presented in [Dambrine and Lamboley, 2019], which relies on the notion of second order shape derivative.

Approach. The natural path to obtain our main stability result, which is Proposition 2.23, is to prove that J is in some sense of class C^2, i.e. that its second order shape derivative is continuous at zero (see Proposition 2.21 for a precise statement). We could not find this result in the literature. A large part of this section is hence dedicated to its proof. To the best of our knowledge, deriving such a result is necessary to obtain Proposition 2.23. In particular, we had to use a stronger condition than the "improved continuity condition" ($\mathbf{IC}_{\mathrm{H}^1,\mathrm{W}^{2,\infty}}$) of [Dambrine and Lamboley, 2019], which is satisfied by our functional. This condition only requires some uniform control of second order directional derivatives at zero, which is weaker than the result of Proposition 2.21.

Structure of shape derivatives. Given an open solution E of ($\mathcal{PC}(\eta)$), we introduce the following mapping, where E_φ denotes the normal deformation of E associated to φ, defined in Proposition B.5:

$$\begin{aligned} \mathrm{j}_E : \mathrm{C}^1(\partial E) &\to \mathbb{R} \\ \varphi &\mapsto J(E_\varphi)\,. \end{aligned}$$

With this notation, the following result holds.

Proposition 2.16 (*See e.g. [Henrot and Pierre, 2018, Chapter 5]*)

If E is a bounded open set of class C^2 and $\eta \in \mathrm{C}^1(\mathbb{R}^2)$, then j_E is twice Fréchet differentiable at 0 and, for every $\psi \in \mathrm{C}^1(\partial E)$, we have:

$$\begin{aligned} \mathrm{j}'_E(0).(\psi) &= \int_{\partial E} [\eta - H]\,\psi\, d\mathcal{H}^1 \\ \mathrm{j}''_E(0).(\psi,\psi) &= \int_{\partial E} \left[|\nabla_\tau \psi|^2 - \left(H\,\eta + \frac{\partial \eta}{\partial \nu} \right) \psi^2 \right] d\mathcal{H}^1 \end{aligned}$$

where H denotes the curvature of E and $\nabla_\tau \psi \overset{\text{def.}}{=} \nabla \psi - (\nabla \psi \cdot \nu)\,\nu$ is the tangential gradient of ψ with respect to E.

From the expression of $\mathrm{j}'_E(0)$ and $\mathrm{j}''_E(0)$ given above, we immediately notice that $\mathrm{j}'_E(0)$ can be extended to a continuous linear form on $\mathrm{L}^1(\partial E)$, and $\mathrm{j}''_E(0)$ to a continuous bilinear form on $\mathrm{H}^1(\partial E)$.

Strict stability. Following [Dambrine and Lamboley, 2019], we say that a solution E of ($\mathcal{PC}(\eta)$) is strictly stable if $\mathrm{j}''_E(0)$ is coercive in $\mathrm{H}^1(\partial E)$, i.e. if the following property holds:

$$\exists \alpha > 0,\ \forall \psi \in \mathrm{H}^1(\partial E),\ \mathrm{j}''_E(0).(\psi,\psi) \geq \alpha\, \|\psi\|^2_{\mathrm{H}^1(\partial E)}\,.$$

Under a few assumptions (that are satisfied by our functional), this strict stability condition ensures that E is a strict local minimizer of J (see Theorem 1.1 in the above-mentioned reference), and is hence the only minimizer (modulo Lebesgue negligible sets) among sets E_φ with φ in a neighborhood of 0. It plays a crucial role in our answer to question (ii).

Continuity results. Now, we prove a few results concerning the convergence of j''_E towards $\mathrm{j}''_{0,E}$ and the continuity of $\varphi \mapsto \mathrm{j}''_E(\varphi)$, where j_E and $\mathrm{j}_{0,E}$ are the functionals respectively associated to η and η_0. To achieve this, we need to compute $\mathrm{j}''_E(\varphi)$ for $\varphi \in \mathrm{C}^1(\partial E)$ in a neighborhood of 0. This may be done using Lemma 2.17 below. To state it, given a bounded open set E of class C^2 and φ in a neighborhood of 0 in $\mathrm{C}^1(\partial E)$, we introduce the mapping $f_\varphi = Id + \xi_\varphi$, with ξ_φ defined as in Lemma B.4. If $\|\varphi\|_{\mathrm{C}^1(\partial E)}$ is sufficiently small then f_φ is a C^1-diffeomorphism, and we denote its inverse by g_φ.

Lemma 2.17

Let E be a bounded open set of class C^2. Then for every φ in a neighborhood of 0 in $\mathrm{C}^1(\partial E)$, and for every $\psi \in \mathrm{H}^1(\partial E)$, we have:

$$\mathrm{j}''_E(\varphi).(\psi,\psi) = j''_{E_\varphi}(0).(\xi_\psi \circ g_\varphi \cdot \nu_\varphi,\, \xi_\psi \circ g_\varphi \cdot \nu_\varphi) + \mathrm{j}'_{E_\varphi}(0).(Z_{\varphi,\psi}) \tag{19}$$

where ν_φ is the unit outward normal to E_φ and

$$Z_{\varphi,\psi} = B_\varphi((\xi_\psi \circ g_\varphi)_{\tau_\varphi}, (\xi_\psi \circ g_\varphi)_{\tau_\varphi}) - 2(\nabla_{\tau_\varphi}(\xi_\psi \circ g_\varphi \cdot \nu_\varphi)) \cdot (\xi_\psi \circ g_\varphi)_{\tau_\varphi}\,,$$

with ζ_{τ_φ} and $\nabla_{\tau_\varphi}\zeta$ the tangential part and the tangential gradient of ζ with respect to E_φ, and B_φ the second fundamental form of E_φ.

Proof : To prove this result, we need to introduce $\mathcal{J}_E$[1] defined by

$$\begin{aligned} \mathcal{J}_E : \mathrm{C}^1_b(\mathbb{R}^2,\mathbb{R}^2) &\to \mathbb{R} \\ \xi &\mapsto J((Id+\xi)(E))\,. \end{aligned}$$

We denote ν the outward unit normal to E and B its second fundamental form. We also denote ζ_τ and $\nabla_\tau\zeta$ the tangential part and the tangential gradient of ζ with respect to E. The structure theorem (see e.g. [Henrot and Pierre, 2018, Theorem 5.9.2] or [Dambrine and Lamboley, 2019, Theorem 2.1]) then yields, for every sufficiently smooth vector field ζ:

$$\begin{aligned} \mathcal{J}'_E(0).(\zeta) &= \mathrm{j}'_E(0).(\zeta|_{\partial E}\cdot\nu)\,, \\ \mathcal{J}''_E(0).(\zeta,\zeta) &= \mathrm{j}''_E(0).(\zeta|_{\partial E}\cdot\nu, \zeta|_{\partial E}\cdot\nu) + \mathrm{j}'_E(0).(Z_\zeta)\,, \end{aligned}$$

where

$$Z_\zeta \overset{\text{def.}}{=} B(\zeta_\tau,\zeta_\tau) - 2\,(\nabla_\tau(\zeta\cdot\nu))\cdot\zeta_\tau\,.$$

Now, we first notice that, for every pair of vector fields ξ,ζ such that $Id+\xi$ is invertible, we have:

$$(Id+\xi+\zeta)(E) = (Id+\zeta\circ(Id+\xi)^{-1})((Id+\xi)(E))\,.$$

Defining $F \overset{\text{def.}}{=} (Id+\xi)(E)$ we hence obtain $\mathcal{J}_E(\xi+\zeta) = \mathcal{J}_F(\zeta\circ(Id+\xi)^{-1}, \zeta\circ(Id+\xi)^{-1})$, which yields

$$\mathcal{J}''_E(\xi).(\zeta,\zeta) = \mathcal{J}''_F(0).(\zeta\circ(Id+\xi)^{-1})\,.$$

Using this with $\xi=\xi_\varphi$ and $\zeta=\xi_\psi$, we get:

$$\mathrm{j}''_E(\varphi).(\psi,\psi) = \mathcal{J}''_E(\xi_\varphi)(\xi_\psi,\xi_\psi) = \mathcal{J}''_{E_\varphi}(0).(\xi_\psi\circ g_\varphi, \xi_\psi\circ g_\varphi)\,,$$

and we finally obtain (19) by applying the structure theorem. ■

[1] This mapping allows to study the behaviour of the objective in a neighborhood of E with respect to general deformations, while j_E is only related to normal deformations.

Most of the results below rely on the following technical lemma, whose first part is contained in [Dambrine and Lamboley, 2019, Lemma 4.7].

Lemma 2.18

Let E be a bounded C^2 set. If $\|\varphi\|_{\mathrm{C}^1(\partial E)} \to 0$ we have:

$$(i)\quad \|f_\varphi - Id\|_{\mathrm{C}^1(\partial E)} \to 0\,, \quad \|\nu_\varphi \circ f_\varphi - \nu\|_{\mathrm{C}^0(\partial E)} \to 0\,, \quad (iii)$$
$$(ii)\quad \|g_\varphi - Id\|_{\mathrm{C}^1(\partial E_\varphi)} \to 0\,, \quad \|\mathrm{Jac}_\tau f_\varphi - 1\|_{\mathrm{C}^0(\partial E)} \to 0\,. \quad (iv)$$

If $\|\varphi\|_{\mathrm{C}^2(\partial E)} \to 0$ then we also have:

$$(v)\quad \|H_\varphi \circ f_\varphi - H\|_{\mathrm{C}^0(\partial E)} \to 0\,, \quad \|B_\varphi \circ f_\varphi - B\|_{\mathrm{C}^0(\partial E)} \to 0\,. \quad (vi)$$

Moreover, the following holds:

$$(a)\quad \lim_{\|\varphi\|_{\mathrm{C}^1(\partial E)}\to 0} \sup_{\psi\in \mathrm{L}^2(\partial E)\setminus\{0\}} \frac{\|(\xi_\psi \circ g_\varphi)_{\tau_\varphi}\|_{\mathrm{L}^2(\partial E_\varphi)}}{\|\psi\|_{\mathrm{L}^2(\partial E)}} = 0\,,$$
$$(b)\quad \lim_{\|\varphi\|_{\mathrm{C}^1(\partial E)}\to 0} \sup_{\psi\in \mathrm{H}^1(\partial E)\setminus\{0\}} \frac{\left|\|\nabla_{\tau_\varphi}(\xi_\psi \circ g_\varphi \cdot \nu_\varphi)\|_{\mathrm{L}^2(\partial E_\varphi)} - \|\nabla_\tau \psi\|_{\mathrm{L}^2(\partial E)}\right|}{\|\psi\|_{\mathrm{H}^1(\partial E)}} = 0\,,$$
$$(c)\quad \lim_{\|\varphi\|_{\mathrm{C}^2(\partial E)}\to 0} \sup_{\psi\in \mathrm{H}^1(\partial E)\setminus\{0\}} \frac{\|Z_{\varphi,\psi}\|_{\mathrm{L}^1(\partial E_\varphi)}}{\|\psi\|^2_{\mathrm{H}^1(\partial E)}} = 0\,.$$

Proof : See [Dambrine and Lamboley, 2019, Lemma 4.7] for a proof of the results stated in the first part of the lemma. To prove (a) we use the fact that

$$\begin{aligned}\|(\xi_\psi \circ g_\varphi)_{\tau_\varphi}\|^2_{\mathrm{L}^2(\partial E_\varphi)} &= \int_{\partial E} (\nu \circ g_\varphi)^2_{\tau_\varphi} \circ f_\varphi \,\mathrm{Jac}_\tau f_\varphi \,\psi^2 \, d\mathcal{H}^1 \\ &\le \|\mathrm{Jac}_\tau f_\varphi\|_{\mathrm{C}^0(\partial E)} \,\|(\nu \circ g_\varphi)_{\tau_\varphi} \circ f_\varphi\|^2_{\mathrm{C}^0(\partial E)} \,\|\psi\|^2_{\mathrm{L}^2(\partial E)} \\ &= \|\mathrm{Jac}_\tau f_\varphi\|_{\mathrm{C}^0(\partial E)} \,\|\nu - (\nu \cdot \nu_\varphi \circ f_\varphi)\,\nu_\varphi \circ f_\varphi\|^2_{\mathrm{C}^0(\partial E)} \,\|\psi\|^2_{\mathrm{L}^2(\partial E)}\,.\end{aligned}$$

which, using (i), (iii) and (iv), yields the result.

To prove (b), we notice that:

$$\nabla_{\tau_\varphi}(\xi_\psi \circ g_\varphi \cdot \nu_\varphi) = \left[c^1_\varphi \,\psi \circ g_\varphi + c^2_\varphi \cdot \nabla_\tau \psi \circ g_\varphi\right] \tau_\varphi\,,$$

with $\tau = \nu^\perp$, $\tau_\varphi = \nu_\varphi^{\perp}$[1] and

$$c^1_\varphi \overset{\text{def.}}{=} \tau \circ g_\varphi \cdot \nu_\varphi \,(J_{g_\varphi}\,\tau_\varphi) \cdot \tau \circ g_\varphi + \tau_\varphi \cdot \nu \circ g_\varphi\,, \qquad c^2_\varphi \overset{\text{def.}}{=} \nu \circ g_\varphi \cdot \nu_\varphi \,(J_{g_\varphi}\,\tau_\varphi)\,.$$

We hence obtain

$$|\nabla_{\tau_\varphi}(\xi_\psi \circ g_\varphi \cdot \nu_\varphi) \circ f_\varphi \,\mathrm{Jac}_\tau f_\varphi - \nabla_\tau \psi| \le c_\varphi \,(|\psi| + |\nabla_\tau \psi|)$$

with c_φ independant of ψ. Moreover, using (ii) and (iii), we have:

$$\lim_{\|\varphi\|_{\mathrm{C}^1(\partial E)}\to 0} \|c_\varphi\|_{\mathrm{C}^0(\partial E)} \to 0\,.$$

[1] These two vectors are defined as the application of the rotation of angle $\pi/2$ to ν and ν_φ.

Denoting $\mathcal{A} \overset{\text{def.}}{=} \left| \|\nabla_{\tau_\varphi}(\xi_\psi \circ g_\varphi \cdot \nu_\varphi)\|_{\mathrm{L}^2(\partial E_\varphi)} - \|\nabla_\tau \psi\|_{\mathrm{L}^2(\partial E)} \right|$, this finally yields

$$\begin{aligned}\mathcal{A} &\leq \|\nabla_{\tau_\varphi}(\xi_\psi \circ g_\varphi \cdot \nu_\varphi) \circ f_\varphi \operatorname{Jac}_\tau f_\varphi - \nabla_\tau \psi\|_{\mathrm{L}^2(\partial E)} \\ &\leq \sqrt{2}\, \|c_\varphi\|_{\mathrm{C}^0(\partial E)}\, \|\psi\|_{\mathrm{H}^1(\partial E)}\,.\end{aligned}$$

We now prove (c). Since

$$\|B_\varphi((\xi_\psi \circ g_\varphi)_{\tau_\varphi}, (\xi_\psi \circ g_\varphi)_{\tau_\varphi})\|_{\mathrm{L}^1(\partial E_\varphi)} \leq \|B_\varphi\|_{\mathrm{C}^0(\partial E_\varphi)}\, \|(\xi_\psi \circ g_\varphi)_{\tau_\varphi}\|^2_{\mathrm{L}^2(\partial E_\varphi)}$$

and

$$\mathcal{B} \leq \|\nabla_{\tau_\varphi}(\xi_\psi \circ g_\varphi \cdot \nu_\varphi))\|_{\mathrm{L}^2(\partial E_\varphi)}\, \|(\xi_\psi \circ g_\varphi)_{\tau_\varphi}\|_{\mathrm{L}^2(\partial E_\varphi)}$$

with $\mathcal{B} \overset{\text{def.}}{=} \|(\nabla_{\tau_\varphi}(\xi_\psi \circ g_\varphi \cdot \nu_\varphi)) \cdot (\xi_\psi \circ g_\varphi)_{\tau_\varphi}\|_{\mathrm{L}^1(\partial E_\varphi)}$, we get the result. ■

Using the above result, we now prove the continuity of $\varphi \mapsto \mathrm{j}''_E(\varphi)$ by proving the continuity of the two terms appearing in its expression. In all the following, if E is a (real) vector space, we denote $\mathcal{Q}(E)$ the set of quadratic forms over E.

Proposition 2.19

If E is a bounded C^2 set and $\mathrm{p}_E : \varphi \mapsto P(E_\varphi)$, the mapping

$$\begin{aligned}\mathrm{p}''_E : \mathrm{C}^2(\partial E) &\to \mathcal{Q}(\mathrm{H}^1(\partial E)) \\ \varphi &\mapsto \mathrm{p}''_E(\varphi)\end{aligned}$$

is continuous at 0.

Proof : Using Lemma 2.18, for every $\varphi \in \mathrm{C}^2(\partial E)$ in a neighborhood of 0 and $\psi \in \mathrm{H}^1(\partial E)$, we obtain:

$$\mathrm{p}''_E(\varphi).(\psi,\psi) - \mathrm{p}''_E(0).(\psi,\psi) = \mathcal{A} + \mathrm{p}'_{E_\varphi}(0).(Z_{\varphi,\psi})\,,$$

with $\mathcal{A} \overset{\text{def.}}{=} \mathrm{p}''_{E_\varphi}(0).((\xi_\psi \circ g_\varphi) \cdot \nu_\varphi, (\xi_\psi \circ g_\varphi) \cdot \nu_\varphi) - \mathrm{p}''_E(0).(\psi,\psi)$. Now, we also have:

$$\mathcal{A} = \|\nabla_{\tau_\varphi}(\xi_\psi \circ g_\varphi \cdot \nu_\varphi)\|^2_{\mathrm{L}^2(\partial E_\varphi)} - \|\nabla_\tau \psi\|^2_{\mathrm{L}^2(\partial E)}\,,$$

and using Lemma 2.18 we obtain

$$\lim_{\|\varphi\|_{\mathrm{C}^2(\partial E)} \to 0} \sup_{\psi \in \mathrm{H}^1(\partial E)\setminus\{0\}} \frac{\left|\mathrm{p}''_{E_\varphi}(0).((\xi_\psi \circ g_\varphi) \cdot \nu_{E_\varphi}, (\xi_\psi \circ g_\varphi) \cdot \nu_{E_\varphi}) - \mathrm{p}''_E(0).(\psi,\psi)\right|}{\|\psi\|^2_{\mathrm{H}^1(\partial E)}} = 0\,.$$

Moreover

$$|\mathrm{p}'_E(0).(Z_{\varphi,\psi})| \leq \|H_\varphi\|_{\mathrm{L}^\infty(\partial E_\varphi)}\, \|Z_{\varphi,\psi}\|_{\mathrm{L}^1(\partial E_\varphi)}\,,$$

and Lemma 2.18 allows to conclude. ■

Proposition 2.20

If E is a bounded C^2 set, $\eta \in \mathrm{C}^1_b(\mathbb{R}^2)$ and $\mathrm{g}_E : \varphi \mapsto \int_{E_\varphi} \eta$, the mapping

$$\begin{aligned}\mathrm{g}''_E : \mathrm{C}^2(\partial E) &\to \mathcal{Q}(\mathrm{H}^1(\partial E)) \\ \varphi &\mapsto \mathrm{g}''_E(\varphi)\end{aligned}$$

is continuous at 0.

Proof : We proceed as in Proposition 2.19. Defining

$$\mathcal{A} \overset{\text{def.}}{=} \mathrm{g}''_{E_\varphi}(0).((\xi_\psi \circ g_\varphi)\cdot\nu_{E_\varphi}, (\xi_\psi \circ g_\varphi)\cdot\nu_{E_\varphi})$$

we have:

$$\begin{aligned}\mathcal{A} &= \int_{\partial E_\varphi}\left[H_\varphi\,\eta + \frac{\partial\eta}{\partial\nu_\varphi}\right]((\psi\,\nu)\circ g_\varphi\cdot\nu_{E_\varphi})^2\,d\mathcal{H}^1\\ &= \int_{\partial E}\left[H_\varphi\,\eta + \frac{\partial\eta}{\partial\nu_\varphi}\right]\circ f_\varphi\,(\nu\cdot\nu_\varphi\circ f_\varphi)^2\,\mathrm{Jac}_\tau f_\varphi\,\psi^2\,d\mathcal{H}^1.\end{aligned}$$

This yields:

$$\frac{\left|\mathrm{g}''_{E_\varphi}(0).((\xi_\psi\circ g_\varphi)\cdot\nu_\varphi, (\xi_\psi\circ g_\varphi)\cdot\nu_\varphi) - \mathrm{g}''_E(0).(\psi,\psi)\right|}{\|\psi\|^2_{\mathrm{L}^2(\partial E)}} \le c_\varphi\,,$$

with

$$c_\varphi \overset{\text{def.}}{=} \left\|\left[H_\varphi\,\eta + \frac{\partial\eta}{\partial\nu_\varphi}\right]\circ f_\varphi\,(\nu\cdot\nu_\varphi\circ f_\varphi)^2\,\mathrm{Jac}_\tau f_\varphi - \left[H\,\eta + \frac{\partial\eta}{\partial\nu}\right]\right\|_\infty.$$

Using Lemma 2.18 we obtain

$$\lim_{\|\varphi\|_{\mathrm{C}^2(\partial E)}\to 0}\ \sup_{\psi\in\mathrm{H}^1(\partial E)\setminus\{0\}} \frac{\left|\mathrm{g}''_{E_\varphi}(0).((\xi_\psi\circ g_\varphi)\cdot\nu_\varphi, (\xi_\psi\circ g_\varphi)\cdot\nu_\varphi) - \mathrm{g}''_E(0).(\psi,\psi)\right|}{\|\psi\|^2_{\mathrm{H}^1(\partial E)}} = 0\,.$$

Moreover

$$|\mathrm{g}'_{E_\varphi}(0).(Z_{\varphi,\psi})| \le \|\eta\|_\infty\,\|Z_{\varphi,\psi}\|_{\mathrm{L}^1(\partial E_\varphi)}\,,$$

and using again Lemma 2.18 we finally obtain the result. ■

As a consequence of the two propositions above, we obtain:

Proposition 2.21

If E is a bounded C^2 set and $\eta\in\mathrm{C}^1_b(\mathbb{R}^2)$, the mapping

$$\begin{aligned}\mathrm{j}''_E : \mathrm{C}^2(\partial E) &\to \mathcal{Q}(\mathrm{H}^1(\partial E))\\ \varphi &\mapsto \mathrm{j}''_E(\varphi)\end{aligned}$$

is continuous at 0.

Now, we prove that for φ in a neighborhood of 0 in $\mathrm{C}^1(\partial E)$, the mapping $\mathrm{j}''_E(\varphi)$ is uniformly close to $\mathrm{j}''_{0,E}(\varphi)$ provided $\|\eta-\eta_0\|_{\mathrm{C}^1(\mathbb{R}^2)}$ is small engouh.

Proposition 2.22

Let E be a bounded C^2 set and $\eta_0\in\mathrm{C}^1_b(\mathbb{R}^2)$. There exists $\epsilon>0$ such that

$$\lim_{\|\eta-\eta_0\|_{\mathrm{C}^1(\mathbb{R}^2)}\to 0}\ \sup_{\|\varphi\|_{\mathrm{C}^2(\partial E)}\le\epsilon} \left\|\mathrm{j}''_E(\varphi) - \mathrm{j}''_{0,E}(\varphi)\right\|_{\mathcal{Q}(\mathrm{H}^1(\partial E))} = 0\,.$$

Proof : Since $|(\mathrm{j}_E - \mathrm{j}_{0,E})''(\varphi).(\psi,\psi)| \leq c^1_\varphi + c^2_\varphi$ with

$$c^1_\varphi \overset{\text{def.}}{=} \left| \int_{\partial E_\varphi} \left(H_\varphi \left(\eta - \eta_0\right) + \frac{\partial \left(\eta - \eta_0\right)}{\partial \nu_\varphi} \right) \left(\xi_\psi \circ g_\varphi \cdot \nu_\varphi\right)^2 d\mathcal{H}^1 \right|,$$

$$c^2_\varphi \overset{\text{def.}}{=} \left| \int_{\partial E_\varphi} \left(\eta - \eta_0\right) Z_{\varphi,\psi}\, d\mathcal{H}^1 \right|,$$

the result readily follows from Lemma 2.18. ■

Stability result. We are now able to state the final result of this section, which, loosely speaking, states that if E is a strictly stable solution of ($\mathcal{PC}(\eta_0)$), there is at most one φ in a neighborhood of 0 such that E_φ is a solution of ($\mathcal{PC}(\eta)$), provided $\|\eta - \eta_0\|_{\mathrm{C}^1(\mathbb{R}^2)}$ is small engouh.

Proposition 2.23

Let $\eta_0 \in \partial \mathrm{TV}(0) \cap \mathrm{C}^1_b(\mathbb{R}^2)$ and E be a strictly stable solution of ($\mathcal{PC}(\eta_0)$). Then there exists $\epsilon > 0$ and $r > 0$ such that for every $\eta \in \partial \mathrm{TV}(0)$ with $\|\eta - \eta_0\|_{\mathrm{C}^1(\mathbb{R}^2)} \leq r$ there is at most one $\varphi \in \mathrm{C}^2(\partial E)$ such that $\|\varphi\|_{\mathrm{C}^2(\partial E)} \leq \epsilon$ and E_φ solves ($\mathcal{PC}(\eta)$).

Proof : The fact E is a strictly stable solution of ($\mathcal{PC}(\eta_0)$) and Propositions 2.21 and 2.22 give the existence of $\epsilon > 0$, $r > 0$ and $\alpha > 0$ such that, for every $(\varphi, \eta) \in \mathrm{C}^2(\partial E) \times \mathrm{C}^1_b(\mathbb{R}^2)$ with $\|\varphi\|_{\mathrm{C}^2(\partial E)} \leq \epsilon$ and $\|\eta - \eta_0\|_{\mathrm{C}^1(\mathbb{R}^2)} \leq r$, we have:

$$\sup_{\psi \in \mathrm{H}^1(\partial E)\setminus\{0\}} \frac{\mathrm{j}''_E(\varphi).(\psi,\psi)}{\|\psi\|^2_{\mathrm{H}^1(\partial E)}} \geq \alpha$$

As a result, $\mathrm{j}''_E(\varphi)$ is coercive (and hence positive definite) for every φ such that $\|\varphi\|_{\mathrm{C}^2(\partial E)} \leq \epsilon$. We therefore obtain that j_E is strictly convex on this set and the result follows. ■

Summary. Combining the results of Propositions 2.15 and 2.23, we have proved that, provided η is sufficiently close to η_0 in $\mathrm{C}^1(\mathbb{R}^2)$ and $\mathrm{L}^2(\mathbb{R}^2)$, every solution of ($\mathcal{PC}(\eta)$) belongs to a neighborhood (in terms of C^2-normal deformations) of a solution of ($\mathcal{PC}(\eta_0)$), and that, under a strict stability assumption, each of these neighborhoods contain at most one solution of ($\mathcal{PC}(\eta)$). In Theorem 2.27 below, we prove (under suitable assumptions) that, if $\eta = \eta_{\lambda,w}$ is the dual certificate associated to ($\mathcal{P}_\lambda(y_0 + w)$) and η_0 the minimal norml dual certificate associated to ($\mathcal{P}_0(y_0)$), then each neighborhood of a solution of ($\mathcal{PC}(\eta_0)$) contains exactly one solution of ($\mathcal{PC}(\eta_{\lambda,w})$).

3. Exact support recovery

In this section, we provide an answer to the following question: if u_0 is M-simple and identifiable, are solutions of ($\mathcal{P}_\lambda(y_0 + w)$) M-simple, what is the length of their associated chains, and how are they related to the chains associated to u_0? To answer it, we first use the results proved in Section 2 to show that, under a few assumptions, the dimension of the faces of the total variation unit ball is in some sense stable. Then, we introduce a non-degenerate version of the source condition, and finally prove our exact support recovery result.

3.1. Stability of the dimension of the faces of the total variation unit ball

If $\eta \in \partial \mathrm{TV}(0) \cap \mathrm{C}^1_b(\mathbb{R}^2)$ and $\mathcal{F}$ is the face exposed by η, defining $\mathcal{E}$ as in (16), we say that $E \subset \mathbb{R}^2$ is a strictly stable element of $\mathcal{E}$ if E is a strictly stable solution of ($\mathcal{PC}(\eta)$) or E^c is a strictly stable solution of ($\mathcal{PC}(-\eta)$).

Theorem 2.24

Let $\eta_0 \in \partial \mathrm{TV}(0) \cap \mathrm{C}^1_b(\mathbb{R}^2)$. Assume the face $\mathcal{F}_0$ exposed by η_0 has a maximal chain $\mathcal{C}_0$ whose elements are all strictly stable. Then for every $\epsilon > 0$ there exists $r > 0$ such that, for every $\eta \in \partial \mathrm{TV}(0) \cap \mathrm{C}^1_b(\mathbb{R}^2)$ with

$$\|\eta - \eta_0\|_{\mathrm{L}^2(\mathbb{R}^2)} + \|\eta - \eta_0\|_{\mathrm{C}^1(\mathbb{R}^2)} \leq r\,,$$

the face $\mathcal{F}$ exposed by η has a maximal chain whose elements are C^2-normal deformations of size at most ϵ of elements of $\mathcal{C}_0$. In particular $\dim(\mathcal{F}) \leq \dim(\mathcal{F}_0)$.

Remark 2.25

To be more precise, Theorem 2.24 states that, if $\mathcal{C}_0 = \{E_i\}_{i=0}^m$, then $\mathcal{F}$ has a maximal chain $\mathcal{C}$ of length $n \leq m$, say $\mathcal{C} = \{F_j\}_{j=0}^n$, with

$$\forall j \in \{1, ..., n-1\},\ F_j = \left(E_{\theta(j)}\right)_{\varphi_j}\,,$$

where $\theta : \{1, ..., n-1\} \to \{1, ..., m-1\}$ is a strictly increasing function and

$$\forall j \in \{1, ..., n-1\},\ \|\varphi_j\|_{\mathrm{C}^2(\partial E_{\theta(j)})} \leq \epsilon\,.$$

Proof : We argue by contradiction and assume the existence of a sequence $(\eta_k)_{k\in\mathbb{N}^*}$ converging in $\mathrm{L}^2(\mathbb{R}^2)$ and $\mathrm{C}^1(\mathbb{R}^2)$ to η_0. We denote by $\mathcal{E}_0$ the set associated to $\mathcal{F}_0$ defined as in (16) (and similarly by $\mathcal{E}_k$ the set associated to the face $\mathcal{F}_k$ exposed by η_k). For simplicity, we first treat the case where, for every $i \in \{1, ..., m-1\}$, the set E_i belongs to $\mathcal{E}_0^+$ (i.e. E_i is a solution of ($\mathcal{PC}(\eta_0)$)). The general case, where E_i belongs to $\mathcal{E}_0^-$ (i.e. E_i^c is a solution of ($\mathcal{PC}(-\eta_0)$)) for every $i \geq i_0 + 1$ with i_0 possibly strictly smaller than $m-1$, is discussed at the end of the proof.

From Propositions 2.15 and 2.23 we know that there exists $\epsilon > 0$ and $k_1 \in \mathbb{N}^*$ such that, for every $k \geq k_1$, every solution of ($\mathcal{PC}(\eta_k)$) is in a neighborhood of size ϵ (in terms of C^2-normal deformations) of a solution of ($\mathcal{PC}(\eta_0)$), and that each of these neighborhoods contains at most one of these solutions. We hence have that, for every $i \in \{1, ..., m-1\}$, there exists infinitely many k such that the neighborhood of size ϵ of E_j contains exactly one solution of ($\mathcal{PC}(\eta_k)$), or infinitely many k such that this neighborhood does not contain any solution of ($\mathcal{PC}(\eta_k)$). We therefore obtain the existence of a partition $(\mathcal{I}, \mathcal{J})$ of $\{1, ..., m-1\}$ such that, up to extraction and possibly increasing k_1, for every $k \geq k_1$ the following holds:

- if $i \in \mathcal{I}$ then there is no solution of ($\mathcal{PC}(\eta_k)$) in the neighborhood of E_i of size ϵ,
- if $i \in \mathcal{J}$ there is exactly one solution of ($\mathcal{PC}(\eta_k)$) in the neighborhood of E_i of size ϵ.

There hence exists $n = |\mathcal{J}| + 1 \leq m$ and an increasing bijection $\theta : \{1, ..., n-1\} \to \mathcal{J}$ with the following property: for every $k \geq k_1$ and $j \in \{1, ..., n-1\}$, there exists $\varphi_{k,j}$ such that $F_{k,j} \overset{\text{def.}}{=} \left(E_{\theta(j)}\right)_{\varphi_{k,j}}$ is the unique solution of ($\mathcal{PC}(\eta_k)$) in the neighborhood of $E_{\theta(j)}$of size ϵ.

Now, let us show that there exists $k_2 \in \mathbb{N}$ such that, for every $k \geq k_2$, the collection

$$\mathcal{C}_k \overset{\text{def.}}{=} \{F_{k,j}\}_{0\leq j\leq n} \text{ with } F_{k,0} \overset{\text{def.}}{=} \emptyset \text{ and } F_{k,n} \overset{\text{def.}}{=} \mathbb{R}^2$$

is a chain. For every $j \in \{1, ..., n-2\}$, we have that $(F_{k,j} \cap F_{k,j+1})_{k\geq k_1}$ solves ($\mathcal{PC}(\eta_k)$) and converges in measure to $E_{\theta(j)} \cap E_{\theta(j+1)} = E_{\theta(j)}$. Arguing as in the proof of Proposition 2.15, we obtain the existence of $k_{2,j}$ such that $F_{k,j} \cap F_{k,j+1}$ is in the neighborhood of $E_{\theta(j)}$ of size ϵ for every $k \geq k_{2,j}$. Using Proposition 2.23 then yields $F_{k,j} \cap F_{k,j+1} = F_{k,j}$ for every $k \geq k_{2,j}$. Repeatedly applying this argument we obtain the existence of $k_{2,1}, ..., k_{2,n-2}$ and defining k_2 as the maximum of these integers we get that $\mathcal{C}_k$ is a chain for every $k \geq k_2$.

Now, let us show the existence of $k_3 \in \mathbb{N}^*$ such that, for every $k \geq k_3$, the chain $\mathcal{C}_k$ is maximal in the face $\mathcal{F}_k$ exposed by η_k. Arguing by contradiction, we assume the existence of $(G_k)_{k\geq k_2}$ such that $G_k \in \mathcal{E}_k \setminus \mathcal{C}_k$ and $\mathcal{C}_k \cup \{G_k\}$ is a chain for every $k \geq k_2$[1]. Arguing as in the proof of Proposition 2.15, we get the existence of $G \in \mathcal{E}_0$ such that, up to the extraction of a subsequence that we do not relabel, $G_k = G_{\psi_k}$ for k large enough, with $\|\psi_k\|_{C^2(\partial G)} \to 0$. Now, since $\mathcal{C}_k \cup \{G_k\}$ is a chain for every $k \geq k_2$, we have that, for every $j \in \{1, n-1\}$, there exists infinitely many k such that $F_{k,j} \subset G_k$ or infinitely many k such that $G_k \subset F_{k,j}$. Up to another extraction, we hence get the existence of $j \in \{0, ..., n-1\}$ and $k_3 \in \mathbb{N}^*$ such that $F_{k,j} \subset G_k \subset F_{k,j+1}$ for every $k \geq k_3$. As a result, we obtain

$$\begin{cases} \emptyset \subset G \subset E_{\theta(1)} & \text{if } j = 0\,, \\ E_{\theta(n-1)} \subset G \subset \mathbb{R}^2 & \text{if } j = n-1\,, \\ E_{\theta(j)} \subset G \subset E_{\theta(j+1)} & \text{otherwise.} \end{cases}$$

Using the maximality of $\mathcal{C}_0$, we get $G = E_i$ with $i \in \mathcal{I}$. Since $G_k = G_{\psi_k}$ with $\|\psi_k\|_{C^2(\partial G)} \to 0$ and, for every $k \geq k_1$, sufficiently small neighborhoods of E_i do not contain any solution of ($\mathcal{PC}(\eta_k)$), we get a contradiction.

Finally, let us comment on the general case, where there exists $i_0 \in \{0, ..., m-1\}$ such that, for every $i \in \{1, ..., m-1\}$, $E_i \in \mathcal{E}_0^+$ if $i \leq i_0$ and $E_i \in \mathcal{E}_0^-$ if $i \geq i_0 + 1$. The same arguments can be applied to obtain the result, bearing in mind that if $i \geq i_0 + 1$, then E_i^c is a solution of ($\mathcal{PC}(-\eta_0)$) (instead of E_i being a solution of ($\mathcal{PC}(\eta_0)$)). ■

3.2. Main result

We are now able to introduce a non-degenerate version of the source condition, which ultimately allows us to state our main result.

[1] Here $\mathcal{E}_k$ denotes the set $\mathcal{E}$ defined in (16) associated to $\mathcal{F}_k$.

Definition 2.26 (*Non-degenerate source condition*)

Let u_0 be an M-simple function and $\mathcal{C}_0$ a chain associated to u_0. We say that u_0 satisfies the non-degenerate source condition if

1. *$\Phi_{\mathcal{C}_0}$ has full rank*
2. *$\eta_0 \in \partial \mathrm{TV}(u_0)$,*
3. *$\mathcal{C}_0$ is maximal in the face exposed by η_0,*
4. *every element of $\mathcal{C}_0$ is strictly stable,*

with η_0 the minimal norm certificate. In that case, we say that η_0 is non-degenerate.

Let us stress that, as pointed out in Remark 2.14, Definition 2.26 does not depend on the choice of a chain $\mathcal{C}_0$ associated to u_0, as one is maximal in an exposed face if and only if they all are.

Theorem 2.27

Let u_0 be an M-simple function and $\mathcal{C}_0 = \{E_i\}_{i=0}^m$ an associated chain. Assume Φ^* is continuous from $\mathcal{H}$ to $\mathrm{C}_b^1(\mathbb{R}^2)$ and u_0 satisfies the non-degenerate source condition. Then there exists constants $\alpha, \lambda_0 \in \mathbb{R}_+^*$ such that for every $(\lambda, w) \in \mathbb{R}_+^* \times \mathcal{H}$ with $\lambda \leq \lambda_0$ and $\|w\|_{\mathcal{H}}/\lambda \leq \alpha$, any solution $u_{\lambda,w}$ of ($\mathcal{P}_\lambda(y)$) is M-simple and its associated chains have the same length as $\mathcal{C}_0$.

Moreover, writing

$$u_0 = \sum_{i=1}^m t_i \, \mathbf{1}_{E_i \setminus E_{i-1}} \,,$$

we have

$$u_{\lambda,w} = \sum_{i=1}^m t_i^{\lambda,w} \, \mathbf{1}_{E_i^{\lambda,w} \setminus E_{i-1}^{\lambda,w}} \,, \tag{20}$$

with $E_0^{\lambda,w} \overset{\text{def.}}{=} \emptyset$, $E_m^{\lambda,w} \overset{\text{def.}}{=} \mathbb{R}^2$, and

$$\forall i \in \{1, ..., m-1\}, \; E_i^{\lambda,w} = (E_i)_{\varphi_i^{\lambda,w}} \text{ with } \varphi_i^{\lambda,w} \in \mathrm{C}^2(\partial E_i) \,. \tag{21}$$

Finally, if $\lambda = \|w\|_{\mathcal{H}}/\alpha$, we have:

$$\forall i \in \{1, ..., m\}, \; \lim_{w \to 0} t_i^{\lambda,w} = t_i \,,$$
$$\forall i \in \{1, ..., m-1\}, \; \lim_{w \to 0} \left\| \varphi_i^{\lambda,w} \right\|_{\mathrm{C}^2(\partial E_i)} = 0 \,.$$

Proof : Let ϵ and r be such that the results of Theorem 2.24 hold. Since Φ^* is continuous from $\mathcal{H}$ to $\mathrm{C}_b^1(\mathbb{R}^2)$, using Proposition 1.8 and (9) we obtain the existence of α and λ_0 such that for every $(\lambda, w) \in \mathbb{R}_+^* \times \mathcal{H}$ with $\lambda \leq \lambda_0$ and $\|w\|_{\mathcal{H}}/\lambda \leq \alpha_0$, we have:

$$\|\eta_{\lambda,w} - \eta_0\|_{\mathrm{L}^2(\mathbb{R}^2)} + \|\eta_{\lambda,w} - \eta_0\|_{\mathrm{C}^1(\mathbb{R}^2)} \leq r \,.$$

Since $\mathcal{C}_0$ is a maximal chain in the face exposed by η_0, Theorem 2.24 ensures the existence of a maximal chain $\mathcal{C}_{\lambda,w}$ in the face exposed by $\eta_{\lambda,w}$, whose length is not greater than the length of $\mathcal{C}_0$. As a result, $u_{\lambda,w}$ is M-simple. We also know the elements of $\mathcal{C}_{\lambda,w}$ are C^2-normal deformations

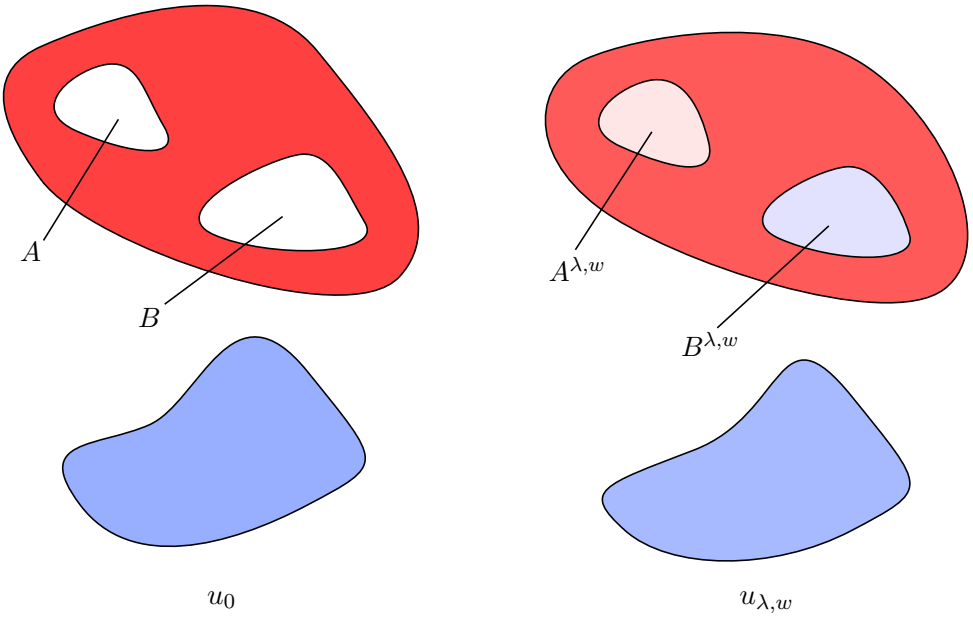


Figure 8 – Illustration of the result stated in Theorem 2.27. Here u_0 is equal to 0 in A and B. The values taken by $u_{\lambda,w}$ in $A^{\lambda,w}$ and $B^{\lambda,w}$ are close (but not necessarily equal) to 0.

of elements of $\mathcal{C}_0$. Using the fact $u_{\lambda,w}$ is constant on each increment of $\mathcal{C}_{\lambda,w}$ and converges in $\mathrm{L}^1(\mathbb{R}^2)$ towards u_0, we obtain that $\mathcal{C}_0$ and $\mathcal{C}_{\lambda,w}$ have equal length, and that (20) and (21) hold.

Finally, the convergence of $\varphi_i^{\lambda,w}$ towards 0 in $\mathrm{C}^2(\partial E_i)$ follows from Theorem 2.24, and the convergence of $t_i^{\lambda,w}$ from the convergence of $u_{\lambda,w}$ towards u_0 in $\mathrm{L}^1(\mathbb{R}^2)$. ■

3.3. A sufficient condition for strict stability

The core assumption in Theorem 2.27 is the strict stability of the elements of the chains associated to u_0, as minimizers of the prescribed curvature functional associated to the minimal norm certificate η_0. In this section, we study this strict stability assumption and derive a natural sufficient condition for it to hold. We then discuss to what extent this condition is necessary.

Setting. We fix $\eta \in \partial\mathrm{TV}(0) \cap \mathrm{C}_b^1(\mathbb{R}^2)$ and E a (non-trivial) solution of the prescribed curvature problem ($\mathcal{PC}(\eta)$) associated to η. We recall that E is a strictly stable solution of ($\mathcal{PC}(\eta)$) if $\mathrm{j}''_E(0)$ is coercive in $\mathrm{H}^1(\partial E)$, with

$$\forall \psi \in \mathrm{H}^1(\partial E),\ \mathrm{j}''_E(0).(\psi,\psi) = \int_{\partial E} \left[|\nabla_{\tau_E} \psi|^2 - \left(H_E\, \eta + \frac{\partial \eta}{\partial \nu_E} \right) \psi^2 \right] d\mathcal{H}^1 .$$

Equivalence of coercivity and positive definiteness. As explained (in a more general context) in [Dambrine and Lamboley, 2019], under a few assumptions, the bilinear form $\mathrm{j}''_E(0)$ is in

fact coercive if and only if it is positive definite. Instead of proving that our functional J fits the assumptions of Lemma 3.1 in the above reference, we use their arguments to provide a direct proof in our specific setting.

Lemma 2.28 (*[Dambrine and Lamboley, 2019, Lemma 3.1]*)

Let $\eta \in \mathrm{C}^1(\mathbb{R}^2)$ and E be a bounded open set of class C^2. Then the following propositions are equivalent:

(i) $\mathrm{j}''_E(0)$ *is positive definite, i.e.*

$$\forall \psi \in \mathrm{H}^1(\partial E) \setminus \{0\},\ \mathrm{j}''_E(0).(\psi,\psi) > 0\,,$$

(ii) $\mathrm{j}''_E(0)$ *is coercive, i.e.*

$$\exists \alpha > 0,\ \forall \psi \in \mathrm{H}^1(\partial E),\ \mathrm{j}''_E(0).(\psi,\psi) \geq \alpha \|\psi\|^2_{\mathrm{H}^1(\partial E)}\,.$$

Proof : The fact (ii) implies (i) is trivial. We assume (i) and let $(\psi_n)_{n\geq 0}$ be a minimizing sequence for

$$\inf_{\psi \in \mathrm{H}^1(\partial E)} \mathrm{j}''_E(0).(\psi,\psi) \ \text{ s.t. } \ \|\psi\|_{\mathrm{H}^1(\partial E)} = 1\,. \tag{22}$$

Up to the extraction of a subsequence (not relabeled), we have that $(\psi_n)_{n\geq 0}$ converges weakly in $\mathrm{H}^1(\partial E)$ and (by the compactness of the embedding $\mathrm{H}^1(\partial E) \subset\subset \mathrm{L}^2(\partial E)$) strongly in $\mathrm{L}^2(\partial E)$ to some $\psi \in \mathrm{H}^1(\partial E)$. We hence obtain

$$\lim_{n\to+\infty} \int_{\partial E} \left(H_E\,\eta + \frac{\partial \eta}{\partial \nu_E}\right) \psi_n^2\, d\mathcal{H}^1 = \int_{\partial E} \left(H_E\,\eta + \frac{\partial \eta}{\partial \nu_E}\right) \psi^2\, d\mathcal{H}^1\,.$$

Let us now distinguish two cases.

1. If $\psi \neq 0$, then we use

$$\int_{\partial E} |\nabla_{\tau_E} \psi|^2\, d\mathcal{H}^1 \leq \liminf_{n\to+\infty} \int_{\partial E} |\nabla_{\tau_E} \psi_n|^2\, d\mathcal{H}^1\,,$$

which, using (i), yields $\liminf_{n\to+\infty} \mathrm{j}''_E(0).(\psi_n,\psi_n) \geq \mathrm{j}''_E(0).(\psi,\psi) > 0$.

2. If $\psi = 0$, using that $\|\psi_n\|_{\mathrm{H}^1(\partial E)} = 1$ for all n, we obtain

$$\liminf_{n\to+\infty} \mathrm{j}''_E(0).(\psi_n,\psi_n) = \liminf_{n\to+\infty} \int_{\partial E} |\nabla_{\tau_E} \psi_n|^2\, d\mathcal{H}^1 = 1 > 0\,.$$

As a result, we obtain that, in both cases, the infimum in (22) is strictly positive, which shows that $\mathrm{j}''_E(0)$ is coercive. ∎

A sufficient condition for coercivity. Using the expression of $\mathrm{j}''_E(0)$ and the equivalence between coercivity and positive definiteness, the following result can be directly obtained.

Proposition 2.29

Let $\eta \in \mathrm{C}^1(\mathbb{R}^2)$ and E be a bounded open set of class C^2 such that $\mathrm{j}'_E(0) = 0$. If

$$\sup_{x \in \partial E} \left[H_E(x)^2 + \frac{\partial \eta}{\partial \nu_E}(x) \right] < 0 \,,$$

then $\mathrm{j}''_E(0)$ is coercive.

Necessity of the condition? A natural question to investigate is whether the condition given in Proposition 2.29 is necessary. A first result in this direction is given in Proposition 2.30 below. Before stating it, let us define $\lambda_1(\Gamma)$, the first Dirichlet eigenvalue of the Laplacian associated to some simple C^1 curve Γ with finite length (see for instance [Kuttler and Sigillito, 1984] for the more classical case of open bounded sets):

$$\lambda_1(\Gamma) \stackrel{\text{def.}}{=} \inf_{\psi \in \mathrm{H}^1_0(\Gamma) \setminus \{0\}} \frac{\|\nabla_{\tau_\Gamma} \psi\|^2_{\mathrm{L}^2(\Gamma)}}{\|\psi\|^2_{\mathrm{L}^2(\Gamma)}} \,. \tag{23}$$

The infimum in (23) is attained and is actually equal to the Dirichlet eigenvalue of the interval $I = (0, \mathcal{H}^1(\Gamma)) \subset \mathbb{R}$, which is $1/(\pi \mathcal{H}^1(\Gamma)^2)$. To see this, one simply needs to consider an arc-length parameterization γ of Γ, and notice that, for every $\psi \in \mathrm{H}^1_0(\Gamma) \setminus \{0\}$, we have:

$$\frac{\|\nabla_{\tau_\Gamma} \psi\|^2_{\mathrm{L}^2(\Gamma)}}{\|\psi\|^2_{\mathrm{L}^2(\Gamma)}} = \frac{\int_I (\psi \circ \gamma)'(t)^2 \, dt}{\int_I (\psi \circ \gamma)(t)^2 \, dt} \,.$$

We can now state the following proposition.

Proposition 2.30

If there exists $\alpha > 0$ such that $H_E^2 + \frac{\partial \eta}{\partial \nu_E} \geq \alpha$ on a connected subset Γ of ∂E with

$$\alpha \, \mathcal{H}^1(\Gamma)^2 \geq 1/\pi \,,$$

then $\mathrm{j}''_E(0)$ is not coercive.

Proof : Since the infimum in the definition of $\lambda_1(\Gamma)$ is attained, we have the existence of a nonzero function $\varphi \in \mathrm{H}^1_0(\Gamma)$ such that

$$\frac{\|\nabla_{\tau_\Gamma} \varphi\|^2_{\mathrm{L}^2(\Gamma)}}{\|\varphi\|^2_{\mathrm{L}^2(\Gamma)}} = \lambda_1(\Gamma) = \frac{1}{\pi \, \mathcal{H}^1(\Gamma)^2} \leq \alpha \,.$$

We hence obtain

$$\int_\Gamma \left[|\nabla_{\tau_\Gamma} \varphi|^2 - \left(H_E^2 + \frac{\partial \eta}{\partial \nu_E} \right) \varphi^2 \right] d\mathcal{H}^1 \leq \int_\Gamma \left[|\nabla_{\tau_\Gamma} \varphi|^2 - \alpha \, \varphi^2 \right] d\mathcal{H}^1 \leq 0 \,.$$

We can then extend φ to $\psi \in \mathrm{H}^1(\partial E)$ whose support is compactly included in Γ, which yields

$$\int_{\partial E} \left[|\nabla_{\tau_E} \psi|^2 - \left(H_E^2 + \frac{\partial \eta}{\partial \nu_E} \right) \psi^2 \right] d\mathcal{H}^1 \leq 0 \,.$$

We can therefore conclude that $\mathrm{j}''_E(0)$ is not coercive. ∎

We stress that Proposition 2.30 does not show that the condition in Proposition 2.29 is necessary, since it does not cover the case where $H_E^2 + \frac{\partial \eta}{\partial \nu_E}$ is greater than a "small" positive constant on a "small" portion of the boundary.

3.4. Computing the minimal norm certificate

In this subsection, we describe two ways to construct candidates for the minimal norm dual certificate defined in Definition 1.7. The first one relies on an attempt to adapt the notion of calibrability, which is exploited for the denoising case in [Chambolle et al., 2016], to general measurement operators. The second is based on vanishing derivatives pre-certificates, introduced in [Duval and Peyré, 2015, Section 4].

As computing the minimal norm certificate in full generality is challenging, we provide a further study of a particular setting, where Φ is assumed to be a convolution operator with radial kernel h, and the unknown image u_0 is the indicator of a ball of radius $R_0 > 0$. Our aim is to investigate, in this simplified scenario, when the non-degenerate source condition is satisfied.

3.4.1. Φ-calibrability

In [Chambolle et al., 2016, Section 4.3], the minimal norm certificate is computed when Φ is the identity (and $(\mathcal{P}_\lambda(y))$ is hence the Rudin Osher Fatemi denoising problem) for a special class of sets called calibrable sets (see e.g. [Bellettini et al., 2002, Alter et al., 2005]). It is hence tempting to adapt this strategy to our setting. An extension of the notion of calibrability for a set E with respect to Φ is to require that $\mathbf{1}_E$ is a singular vector of the total variation as defined in [Benning and Burger, 2013, Definition 4]. This leads to the following definition.

Definition 2.31

A set $E \subset \mathbb{R}^2$ is said to be Φ-calibrable if $\lambda_E \Phi^* \Phi \mathbf{1}_E \in \partial \mathrm{TV}(\mathbf{1}_E)$ with $\lambda_E \overset{\text{def.}}{=} \frac{P(E)}{\|\Phi \mathbf{1}_E\|^2}$.

Remark 2.32

If $\lambda \Phi^* \Phi \mathbf{1}_E \in \partial \mathrm{TV}(\mathbf{1}_E)$ for some $\lambda \in \mathbb{R}$, taking the inner product with $\mathbf{1}_E$ shows $\lambda = \lambda_E$.

In [Chambolle et al., 2016, Proposition 6], it is proven that, if $\Phi = Id$ and E is calibrable, then $\lambda_E \mathbf{1}_E$ is the minimal norm certificate. With Definition 2.31, the following analog of this result is straightforward to obtain.

Proposition 2.33

If E is Φ-calibrable, then the minimal norm dual certificate is $\lambda_E \Phi^* \Phi \mathbf{1}_E$.

Proof : Let $p \in \mathcal{H}$ such that $\Phi^* p \in \partial \mathrm{TV}(\mathbf{1}_E)$. Then, we have:

$$P(E) = \langle \Phi^* p, \mathbf{1}_E \rangle = \langle p, \Phi \mathbf{1}_E \rangle \leq \|p\|_{\mathcal{H}} \|\Phi \mathbf{1}_E\|_{\mathcal{H}} .$$

Since $\|\lambda_E \Phi \mathbf{1}_E\| = P(E)/\|\Phi \mathbf{1}_E\|_{\mathcal{H}}$, we obtain that $\|p\|_{\mathcal{H}} \leq \|\lambda_E \Phi \mathbf{1}_E\|$. ■

Now, from Proposition 1.3, we can readily obtain the following useful characterization of Φ-calibrable sets.

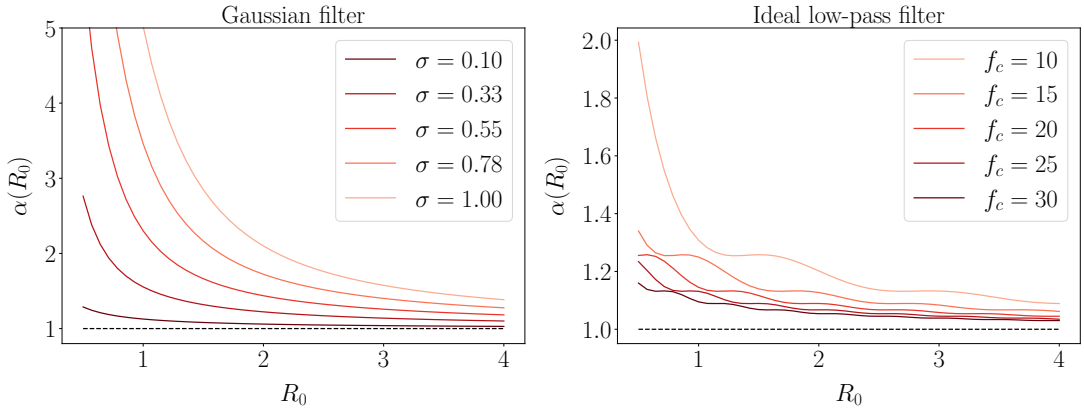


Figure 9 – Graphs of $R_0 \mapsto \alpha(R_0)$ for the Gaussian and ideal low-pass filters, for several values of the variance and cut-off frequency.

Proposition 2.34

A set E is Φ-calibrable if and only if E maximizes

$$F \mapsto K(E,F) \overset{\text{def.}}{=} \left\langle \Phi \frac{\mathbf{1}_E}{P(E)}, \Phi \frac{\mathbf{1}_F}{P(F)} \right\rangle$$

among sets of finite perimeter with positive finite measure.

When $\Phi = Id$, a complete characterization of calibrable sets is available [Alter et al., 2005]. We are not aware of any result outside this specific setting. In fact, we provide numerical evidence below suggesting that even disks are not Φ-calibrable for natural choices of Φ (namely the convolution with the Gaussian and ideal low-pass kernels).

Deconvolution of a disk. Let us define

$$\alpha(R_0) \overset{\text{def.}}{=} \sup \left\{ \frac{K(R,R_0)}{K(R_0,R_0)},\ R>0 \right\} \quad \text{with} \quad K(R,R') = K\big(\mathbf{1}_{B(0,R)}, \mathbf{1}_{B(0,R')}\big)\,.$$

With this notation we have that, if $\mathbf{1}_{B(0,R_0)}$ is Φ-calibrable, then $\alpha(R_0) \leq 1$. Figure 9 shows the graph of $R_0 \mapsto \alpha(R_0)$ when Φ is the convolution with the Gaussian or ideal low-pass kernel, for several values of the variance and cut-off frequency. In every situation, these plots seem to indicate that $\alpha(R_0) > 1$ for every $R_0 > 0$, and hence that $\mathbf{1}_{B(0,R_0)}$ is never Φ-calibrable. This questions the relevance of our definition of Φ-calibrability. In fact, it could even be that there are no Φ-calibrable sets, even for reasonable choices of measurement operator Φ.

3.4.2. Pre-certificates

In [Duval and Peyré, 2015, Section 4], the notion of *dual pre-certificate* is introduced. Its authors define it as any "good candidate" $p \in \mathcal{H}$ for solving $\Phi^* p \in \partial \mathrm{TV}(u)$, where u is an admissible function for ($\mathcal{P}_0(y_0)$) whose optimality is to be proven. In this subsection, we define an analog of the so-called *vanishing derivative pre-certificate*, in order to study the optimality of $\mathbf{1}_E$ with E a simple set. We assume as in Theorem 2.27 that Φ^* is continuous from $\mathcal{H}$ to $\mathrm{C}^1_b(\mathbb{R}^2)$. We

also assume that E is of class C^3, as any set E satisfying $\Phi^* p \in \partial \mathrm{TV}(\mathbf{1}_E)$ for some $p \in \mathcal{H}$ has this regularity.

Necessary optimality conditions. We know that $\Phi^* p$ is a dual certificate associated to $\mathbf{1}_E$ if and only if $\Phi^* p \in \partial \mathrm{TV}(0)$ and E minimizes $J(E) \overset{\text{def.}}{=} P(E) - \int_E \Phi^* p$. Using the expression of the first order shape derivative of J, we obtain the following necessary optimality conditions:

$$\int_E \Phi^* p = P(E)\,, \tag{24}$$

$$\Phi^* p|_{\partial E} = H_E\,. \tag{25}$$

Now, provided E is simple, (25) in particular implies that

$$\int_{\partial E} \Phi^* p = \int_{\partial E} H_E = 2\pi\,. \tag{26}$$

One could then define the vanishing derivative pre-certificate as the solution of (24) and (26) with minimal norm.

Definition 2.35

We call vanishing derivative pre-certificate associated to a simple set E the function $\eta_v = \Phi^* p_v$ with p_v the unique solution of

$$\min_{p \in \mathcal{H}} \|p\|_{\mathcal{H}} \quad \text{s.t.} \quad \left\langle p, \int_E \varphi \right\rangle = P(E) \ \text{ and } \ \left\langle p, \int_{\partial E} \varphi \right\rangle = 2\pi\,. \tag{27}$$

If the source condition holds (i.e. there exists $\eta \in \mathrm{Im}\Phi^* \cap \partial \mathrm{TV}(0)$ such that $\eta \in \partial \mathrm{TV}(\mathbf{1}_E)$), then p_v is well-defined as Problem (27) is feasible. Since any dual certificate satisfies (24) and (26), we have the following result.

Proposition 2.36

If $\eta_v \in \partial \mathrm{TV}(0)$, then η_v is the minimal norm dual certificate.

The constraints in (27) can be rewritten as follows:

$$\begin{cases} \langle p, f_E \rangle = 1\,, \\ \langle p, g_E \rangle = 1\,, \end{cases} \qquad \text{with} \qquad \begin{cases} f_E \overset{\text{def.}}{=} \dfrac{1}{P(E)} \displaystyle\int_E \varphi\,, \\ g_E \overset{\text{def.}}{=} \dfrac{1}{2\pi} \displaystyle\int_{\partial E} \varphi\,. \end{cases} \tag{28}$$

We hence have that p_v satisfies

$$p_v = \alpha f_E + \beta g_E \quad \text{with} \quad \alpha \overset{\text{def.}}{=} \frac{\|g_E\|^2 - \langle f_E, g_E \rangle}{\|f_E\|^2 \|g_E\|^2 - \langle f_E, g_E \rangle^2} \ \text{ and } \ \beta \overset{\text{def.}}{=} \frac{\|f_E\|^2 - \langle f_E, g_E \rangle}{\|f_E\|^2 \|g_E\|^2 - \langle f_E, g_E \rangle^2}\,.$$

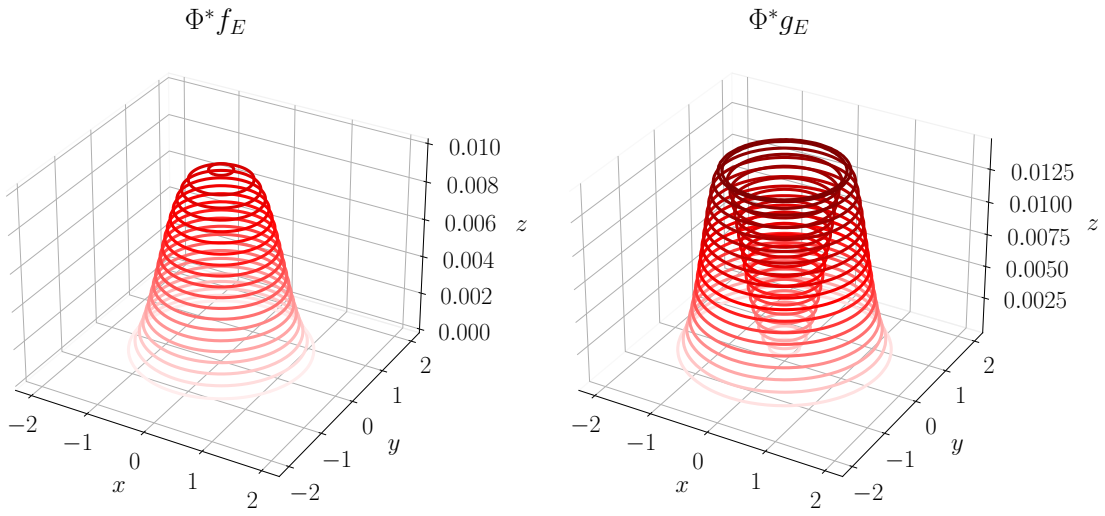


Figure 10 – Plots of $\Phi^* f_E$ and $\Phi^* g_E$ defined in (27) for $E = \mathbf{1}_{B(0,1)}$ and Φ the convolution with the Gaussian kernel with variance $\sigma = 0.2$.

Remark 2.37

In some sense, the function $\lambda_E \Phi^* \Phi \mathbf{1}_E$ considered above can be seen as the pre-certificate associated to the single constraint (24). Imposing the second constraint (26) allows to build a pre-certificate which is a combination of the two functions f_E and g_E (instead of f_E only for $\lambda_E \Phi^* \Phi \mathbf{1}_E$). Other constraints could also be considered, as every certificate satisfies (25), which is only exploited "in average" in (26). Considering [Duval and Peyré, 2015, Propositions 7 and 8], it would be interesting to investigate whether the non-degenerate source condition holds if and only if $\eta_v = \eta_0$ and η_0 is non-degenerate, and also whether $\eta_v = \eta_0$ is a necessary condition for support recovery.

Deconvolution of the unit disk. We now focus on the case where $E = \mathbf{1}_{B(0,1)}$ and Φ is the convolution with the Gaussian kernel with variance σ. We provide in Figure 10 a plot of $\Phi^* f_E$ and $\Phi^* g_E$, which are the two "basis functions" from which η_v is built. From (5), we know that a way to show $\eta_v \in \partial \mathrm{TV}(0)$ is to find a vector field $z \in \mathrm{L}^\infty(\mathbb{R}^2, \mathbb{R}^2)$ such that $\operatorname{div} z = \eta_v$. Since f_E and g_E are radial, so is η_v. It is hence natural to look for a radial vector field z (i.e. such that there exists $z_r : \mathbb{R}_+ \to \mathbb{R}$ with $z(x) = z_r(\|x\|)\, x/\|x\|$ for almost every $x \in \mathbb{R}^2$). In this case we have $\operatorname{div} z = \eta_v$ if and only if, for every $r > 0$:

$$\begin{aligned} \tilde{\eta}_v(r) = \frac{1}{r}\frac{\partial}{\partial r}(r\, z_r)(r) &\iff r\, \tilde{\eta}_v(r) = \frac{\partial}{\partial r}(r\, z_r)(r) \\ &\iff z_r(r) = \frac{1}{r}\int_0^r \tilde{\eta}_v(s)\, s\, ds\,. \end{aligned}$$

This shows one only needs to ensure the mapping f_v defined by

$$\begin{aligned} f_v : \mathbb{R}_+ &\to \mathbb{R} \\ r &\mapsto \frac{1}{r}\int_0^r \tilde{\eta}_v(s)\, s\, ds \end{aligned} \tag{29}$$

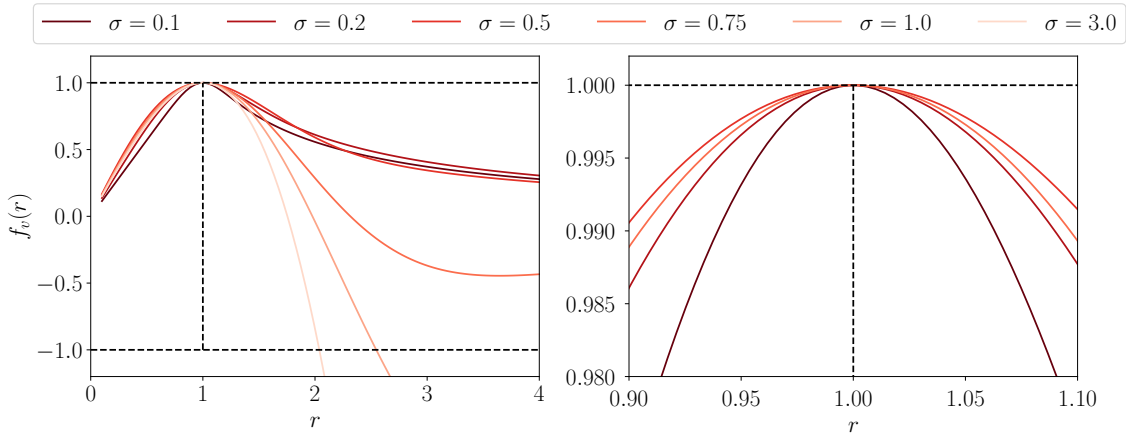


Figure 11 – Graph of f_v defined in (29) (left: global graph, right: zoom around 1).

satisfies $\|f_v\|_\infty \leq 1$ to show $\eta_v \in \partial \mathrm{TV}(0)$. Figure 11 shows the graph of f_v for several values of σ. This suggests that there exists $\sigma_0 > 0$ such that η_v is a dual certificate (and hence the one with minimal norm) for every $\sigma < \sigma_0$. It even seems that $\sigma_0 \geq 0.75$.

Remark 2.38

One could wonder whether looking for a radial vector field is not restrictive. In fact, if a vector field z is suitable, then so is the radial vector field $\tilde{z}$ defined by

$$\tilde{z}(x) \overset{\text{def.}}{=} \tilde{z}_r(\|x\|)\frac{x}{\|x\|} \quad \text{with} \quad \tilde{z}_r(r) \overset{\text{def.}}{=} \frac{1}{2\pi}\int_0^{2\pi} z_r(r,\theta)\,d\theta\,,$$

where z_r denotes the radial component of z. Indeed, we have $|\tilde{z}(r)| \leq 1$ for all r with equality if and only if $z_r(r,\theta) = 1$ for almost every θ or $z_r(r,\theta) = -1$ for almost every θ. Moreover

$$\begin{aligned}\eta_v(r) = \frac{1}{2\pi}\int_0^{2\pi}\eta_v(r)\,dr &= \frac{1}{r}\frac{\partial}{\partial r}\left(r\,\frac{1}{2\pi}\int_0^{2\pi} z_r(r,\theta)\,d\theta\right) + \frac{1}{r}\,\frac{1}{2\pi}\int_0^{2\pi}\frac{\partial z_\theta}{\partial\theta}(r,\theta)\,d\theta \\ &= \frac{1}{r}\frac{\partial}{\partial r}\left(r\,\tilde{z}_r\right) = \operatorname{div}\tilde{z}.\end{aligned}$$

Finally, we can investigate if the non-degenerate source condition holds. As explained in Section 3.3, it is sufficient to show that

$$\sup_{x\in\partial E}\left[H_E^2(x) + \frac{\partial\eta_0}{\partial\nu_E}(x)\right] < 0\,.$$

In our case H_E is constant equal to one, and, since η_0 is radial, $\frac{\partial\eta_0}{\partial\nu_E}$ is constant on ∂E. Proving that

$$\frac{\partial\eta_v}{\partial r}(1) < -1$$

is hence sufficient. In Figure 12, we numerically compute this quantity and notice it is always the case, even when $\eta_v \notin \partial\mathrm{TV}(0)$. This suggests that there exists $\sigma_0 > 0$ such that, for every $\sigma > 0$, the non-degenerate source condition holds (and, from our experiments, it seems that $\sigma_0 \geq 0.75$). Surprisingly, $\sigma \mapsto \frac{\partial\eta_v}{\partial r}(1)$ does not seem to be monotonous, even on $[0,\sigma_0)$.

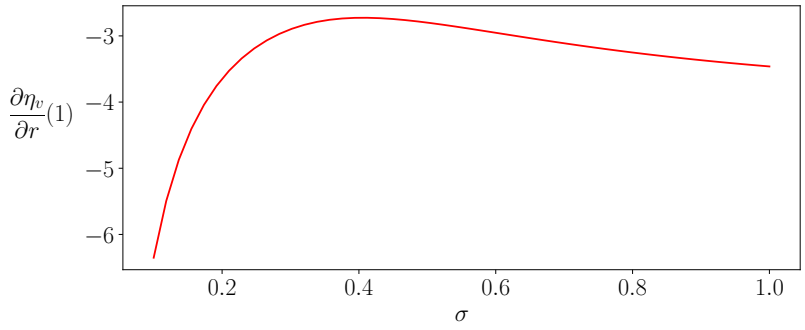


Figure 12 – Graph of $\frac{\partial \eta_v}{\partial r}(1)$ as a function of σ.

Beyond the radial case. A way to ensure $\eta_v \in \partial \mathrm{TV}(0)$ in the general case is to solve the (generalized) Cheeger problem associated to η_v. This can be done using the numerical method presented in Section 2. Now, to ensure the non-degenerate source condition holds, one must also show that E is the unique (non-trivial) solution of the prescribed curvature problem associated to η_v. In [Buttazzo et al., 2007, Carlier et al., 2009], a method is proposed to compute a "maximal" solution of the Cheeger problem, which is closely related to the prescribed curvature problem. It relies on the introduction of a small strictly convex penalization. We argue that these ideas could be adapted to numerically compute a "maximal" element of the face exposed by η_v (i.e. a function whose chains are maximal in this face). If this element is proportional to $\mathbf{1}_E$, then $\mathbf{1}_E$ is identifiable and η_v is the minimal norm certificate.

4. Conclusion

Summary. In this part, we collected properties of the total variation unit ball and its (exposed) faces. The main interest of this analysis is a useful result, stating that elements of a given face can be decomposed using a common set of atoms. We also defined the class of M-simple functions, which are the sparse objects naturally associated to the total variation. Then, we investigated the behaviour, under variations of the associated curvature functional, of solutions of the precribed curvature problem. This allowed us to prove that the dimension of the faces of the total variation unit ball is in some sense stable. Using this last result, we finally managed to prove that, under a so-called non-degenerate source condition, the jump set of an unknown M-simple function can be exactly recovered in a low noise regime.

Towards grid-free numerical methods. Our support recovery result is yet another motivation for designing grid-free numerical methods allowing to solve ($\mathcal{P}_\lambda(y)$). Indeed, it shows that, under suitable assumptions, if $y = \Phi u_0 + w$ with u_0 an M-simple function, then solutions of ($\mathcal{P}_\lambda(y)$) are themselves M-simple. Being able to compute approximate solutions with this structure is hence highly desirable. As explained in Section 2.3, traditional grid-based method are not designed with that goal in mind. They do not produce a continuous domain representation of the reconstructed images, which are often blurry. The aim of Part 3 is to propose a numerical solver which does not suffer from these drawbacks, and accurately estimates the jump set of solutions.

Part 3

Grid-free numerical resolution

In this part, we focus on the numerical resolution of $(\mathcal{P}_\lambda(y))$. We present a so-called *grid-free* algorithm, which does not rely on the introduction of a fixed spatial discretization. It iteratively constructs an approximate solution built from the atoms promoted by the total variation, namely indicator functions of simple sets. We choose to numerically represent these sets using simple polygons. We provide a complete implementation of the proposed method in the following online repositories.

- https://github.com/rpetit/pycheeger
- https://github.com/rpetit/tvsfw

Its performance is investigated in the last section, and comparisons with grid-based approaches are provided. This part is based on the following publication.

- **[De Castro et al., 2022] Towards off-the-grid algorithms for total variation regularized inverse problems**, Y. De Castro, V. Duval and R. Petit, *Journal of Mathematical Imaging and Vision*, 2022.

Contents

1. Frank-Wolfe approach

In [Bredies and Pikkarainen, 2013, Boyd et al., 2017, Denoyelle et al., 2019], variants of the conditional gradient algorithm, also known as Frank-Wolfe algorithm, were introduced to perform continuous domain sparse spikes recovery. In this section, we adapt this fruitful approach to our setting, and derive a modified Frank-Wolfe algorithm allowing to solve ($\mathcal{P}_\lambda(y)$) in a grid-free manner.

In all the following, as we are interested in the numerical resolution of ($\mathcal{P}_\lambda(y)$), we only consider the case of finitely many measurements, i.e. $\mathcal{H} = \mathbb{R}^m$ from some $m \in \mathbb{N}^*$.

1.1. Frank-Wolfe algorithm

Frank-Wolfe algorithm (see Algorithm 1) allows to minimize a function f over a subset C of a Banach space. The objective f is assumed to be convex and differentiable, and the admissible set C convex and weakly compact. Each step of the algorithm consists in minimizing the first order expansion of f on C, and building the next iterate as a convex combination of the obtained point and the current iterate.

Algorithm 1: Frank-Wolfe algorithm

Data: objective f, domain C, starting point $x^{[0]} \in C$
Result: point x^*

1 **while** *true* **do**
2 find $s^{[k]} \in \operatorname{Argmin}_{s \in C} f(x^{[k]}) + df(x^{[k]}).(s - x^{[k]})$
3 **if** $df(x^{[k]}).(s^{[k]} - x^{[k]}) = 0$ **then**
4 output $x^* \leftarrow x^{[k]}$, which is optimal
5 **else**
6 $\gamma^{[k]} \leftarrow \operatorname{Argmin}_{\gamma \in [0,1]} f(x^{[k]} + \gamma(s^{[k]} - x^{[k]}))$ *// line search*
7 $\tilde{x}^{[k+1]} \leftarrow x^{[k]} + \gamma^{[k]}(s^{[k]} - x^{[k]})$ *// tentative update*
8 choose any $x^{[k+1]}$ such that $f(x^{[k+1]}) \leq f(\tilde{x}^{[k+1]})$ *// final update*
9 **end**
10 **end**

Sparse greedy updates. Our main interest for Frank-Wolfe algorithm lies in the following classical observation: since the linear minimization step (Line 2) consists in minimizing a linear form over the weakly compact convex set C, at least one of its solutions is an extreme point of C. Choosing such a point at every iteration, we obtain that $x^{[k]}$ is a convex combination of at most k extreme points of C, provided $x^{[0]} = 0$. This is particularly interesting in situations where the extreme points of C are "atoms" and where one is looking for a solution which is a sparse combination of those.

Choice of the final update. An important feature of the algorithm is that, while the classical update (Line 8) is to take $x^{[k+1]}$ to be equal to $\tilde{x}^{[k+1]}$, all convergence guarantees are preserved if one chooses instead any $x^{[k+1]} \in C$ such that $f(x^{[k+1]}) \leq f(\tilde{x}^{[k+1]})$. As suggested by [Bredies and Pikkarainen, 2013, Boyd et al., 2017, Denoyelle et al., 2019] in the context of sparse spikes recovery, on can extensively make use of this flexibility to produce iterates that are as sparse as possible.

1.2. Proposed algorithm

Frank-Wolfe algorithm can not be directly applied to ($\mathcal{P}_\lambda(y)$), as its objective

$$F(u) \overset{\text{def.}}{=} \mathrm{TV}(u) + \frac{1}{2\lambda}\|\Phi u - y\|^2$$

is not differentiable and its admissible set is unbounded. In this subsection, we derive an equivalent formulation of ($\mathcal{P}_\lambda(y)$) which fits into this framework, and then describe the resulting algorithm.

Epigraphical lift. To obtain a problem equivalent to ($\mathcal{P}_\lambda(y)$) with a differentiable objective and a weakly compact admissible set, we perform an epigraphical lift. This strategy is used in [Denoyelle et al., 2019] (following an idea of [Harchaoui et al., 2015, Section 2, Paragraph "penalized norm minimization"]).

Proposition 3.1

Problem ($\mathcal{P}_\lambda(y)$) is equivalent to

$$\begin{aligned} \inf_{(u,t)\in \mathrm{L}^2(\mathbb{R}^2)\times\mathbb{R}} \quad & G(u,t) \overset{\text{def.}}{=} \frac{1}{2\lambda}\|\Phi u - y\|^2 + t \\ \text{s.t.} \quad & \mathrm{TV}(u) \leq t \leq \frac{1}{2\lambda}\|y\|^2 , \end{aligned} \tag{$\mathcal{Q}_\lambda(y)$}$$

i.e. both problems have the same value and:

1. *if (u,t) is a solution of ($\mathcal{Q}_\lambda(y)$), then $t = \mathrm{TV}(u)$ and (u,v) is a solution of ($\mathcal{P}_\lambda(y)$),*
2. *if u is a solution of ($\mathcal{P}_\lambda(y)$), then $(u, \mathrm{TV}(u))$ is a solution of ($\mathcal{Q}_\lambda(y)$).*

Proof : This is a straightforward consequence of the following fact: if u is a solution of ($\mathcal{P}_\lambda(y)$), then we have that $F(u) \leq F(0)$, which yields $\mathrm{TV}(u) \leq \frac{1}{2\lambda}\|y\|^2$. The admissible set of ($\mathcal{P}_\lambda(y)$) can hence be restricted to functions u satisfying $\mathrm{TV}(u) \leq \frac{1}{2\lambda}\|y\|^2$ without modifying its solutions. ■

The admissible set of ($\mathcal{Q}_\lambda(y)$) is now weakly compact and convex. Moreover, its objective G is differentiable with, for all $(u,t) \in \mathrm{L}^2(\mathbb{R}^2) \times \mathbb{R}$:

$$\begin{aligned} dG(u,t) \colon \mathrm{L}^2(\mathbb{R}^2) \times \mathbb{R} &\to \mathbb{R} \\ (v,s) &\mapsto \left[\int_{\mathbb{R}^2} \frac{1}{\lambda}\Phi^*(\Phi u - y)\, v\right] + s\,. \end{aligned}$$

Linear minimization step. Applying Algorithm 1 to $(\mathcal{Q}_\lambda(y))$, the linear minimization step at iteration k (Line 2) writes

$$\begin{aligned} \inf_{(u,t)\in \mathrm{L}^2(\mathbb{R}^2)\times\mathbb{R}} \quad & \left[\int_{\mathbb{R}^2} \frac{1}{\lambda}\Phi^*(\Phi u^{[k]} - y)\, u\right] + t \\ \text{s.t.} \quad & \mathrm{TV}(u) \leq t \leq \frac{1}{2\lambda}\|y\|^2 \,. \end{aligned} \tag{30}$$

As explained above, there always exists a solution of (30) which is an extreme point of the admissible set. These points are described by the following lemma, whose proof is given in Appendix E.

Lemma 3.2

If $t^* > 0$, the extreme points of $C \overset{\text{def.}}{=} \{u \in \mathrm{L}^2(\mathbb{R}^2) \,|\, \mathrm{TV}(u) \leq t \leq t^*\}$ are:

- $(0,0)$,
- $(\epsilon t^* \mathbf{1}_E / P(E), t^*)$ *with* $\epsilon \in \{-1,1\}$, $E \subset \mathbb{R}^2$ *simple and* $0 < |E| < +\infty$.

With this result, we obtain that finding a solution of (30) among the extreme points of the admissible set is equivalent to solving

$$\begin{aligned} \sup_{\substack{E\subset\mathbb{R}^2 \\ \epsilon\in\{-1,1\}}} \quad & \frac{\epsilon}{P(E)} \int_E \eta^{[k]} \\ \text{s.t.} \quad & E \text{ simple},\ 0 < |E| < +\infty \,, \end{aligned} \tag{31}$$

with $\eta^{[k]} \overset{\text{def.}}{=} (-1/\lambda)\Phi^*(\Phi u^{[k]} - y)$. Indeed, a simple computation shows that, if the value of (31) is smaller than 1, then $(0,0)$ is optimal. Otherwise, defining $M \overset{\text{def.}}{=} \|y\|^2/(2\lambda)$, for any solution (E,ϵ) of (31), we get that $(\epsilon M \mathbf{1}_E / P(E), M)$ is optimal. Properties of (31), which is reminiscent of the Cheeger problem (see the surveys [Parini, 2011, Leonardi, 2015]), are discussed in Section 2.1.

Form of the iterates. Applying Algorithm 1 to $(\mathcal{Q}_\lambda(y))$ with (30) solved as described above, one obtains a sequence of iterates $(u^{[k]}, t^{[k]})_{k\geq 0}$ whose form is described by the following lemma.

Lemma 3.3

If $(u^{[0]}, t^{[0]}) = (0,0)$, for every $k \in \mathbb{N}$, we either have that $(u^{[k]}, t^{[k]}) = (0,0)$ or that there exists $N^{[k]} \in \mathbb{N}^*$, $\epsilon^{[k]} \in \{-1,1\}^{N^{[k]}}$, $a^{[k]} \in (\mathbb{R}_+^*)^{N^{[k]}}$ and a collection $E^{[k]} = \left(E_1^{[k]}, \ldots, E_{N^{[k]}}^{[k]}\right)$ of simple sets of positive finite measure such that

$$(u^{[k]}, t^{[k]}) = \left(\sum_{i=1}^{N^{[k]}} \epsilon_i^{[k]} a_i^{[k]} \mathbf{1}_{E_i^{[k]}}, \sum_{i=1}^{N^{[k]}} a_i^{[k]} P(E_i^{[k]}) \right). \tag{32}$$

Proof : We argue by induction. The result is already known for $k = 0$. Let us fix $k \in \mathbb{N}$ and assume that $(u^{[k]}, t^{[k]})$ has the right form.

1. If $(u^{[k]}, t^{[k]}) = (0, 0)$ then:
   - if $(0, 0)$ solves (30) then $(u^{[k+1]}, t^{[k+1]}) = (0, 0)$,
   - otherwise there exists $\epsilon \in \{-1, 1\}$ and a simple set E with $0 < |E| < +\infty$ such that the couple $(\epsilon M \mathbf{1}_E / P(E), M)$ solves (30). We therefore obtain that $(u^{[k]}, t^{[k]})$ is as in (32) with
   $$N^{[k+1]} = 1,\ \epsilon^{[k+1]} = (\epsilon),\ a^{[k+1]} = (\gamma^{[k]} M / P(E)) \text{ and } E^{[k+1]} = (E)\,.$$
2. If $(u^{[k]}, t^{[k]})$ is as in (32) then:
   - if $(0, 0)$ solves (30), we get that $(u^{[k+1]}, t^{[k+1]})$ is as in (32) with
   $$N^{[k+1]} = N^{[k]},\ \epsilon^{[k+1]} = \epsilon^{[k]},\ a^{[k+1]} = (1 - \gamma^{[k]})\, a^{[k]} \text{ and } E^{[k+1]} = E^{[k]}\,.$$
   - Otherwise as above denoting $(\epsilon M \mathbf{1}_E / P(E), M)$ a solution of (30) we obtain the result with
   $$N^{[k+1]} = N^{[k]} + 1,\ \epsilon^{[k+1]} = (\epsilon^{[k]}, \epsilon),\ a^{[k+1]} = ((1 - \gamma^{[k]})\, a^{[k]}, \gamma^{[k]} M / P(E))$$
   and $E^{[k+1]} = (E^{[k]}, E)$. ■

"Fully-corrective" variant. Lemma 3.3 allows us to use a so-called *fully corrective* variant of Frank-Wolfe, meaning that instead of obtaining the next iterate as a convex combination of the new atom and the previous iterate (as in Line 7 of Algorithm 1), we find $a^{[k+1]}$ by minimizing $G_N(\epsilon, a, E)$ defined by

$$G_N(\epsilon, a, E) \overset{\text{def.}}{=} G\left(\sum_{i=1}^{N} \epsilon_i a_i \mathbf{1}_{E_i}, \sum_{i=1}^{N} a_i P(E_i)\right)$$

with $N = N^{[k]} + 1$ and $(\epsilon, E) = (\epsilon^{[k+1]}, E^{[k+1]})$ fixed. This minimization amounts to solving a LASSO problem with a positivity constraint and a weighted ℓ_1 penalty (the weights being the perimeters of the sets in $E^{[k+1]}$). Indeed, given $N \in \mathbb{N}^*$, a collection of sets $E_1, ..., E_N$, and two vectors $a \in (\mathbb{R}_+)^N$, $\epsilon \in \{-1, 1\}^N$ we have:

$$G_N(\epsilon, a, E) = \frac{1}{2\lambda} \|\Phi_E^\epsilon a - y\|^2 + \sum_{i=1}^{N} P(E_i)\, a_i \quad \text{with} \quad \Phi_E^\epsilon \overset{\text{def.}}{=} \left[\left(\epsilon_i \int_{E_i} \varphi_j\right)_{\substack{1 \le i \le N \\ 1 \le j \le m}}\right]^T \in \mathbb{R}^{m \times N}\,.$$

The main interest of Lemma 3.3 is that, considering (32), this choice of $a^{[k+1]}$ yields a value of the objective no higher than that obtained with the standard update, and hence does not break convergence guarantees.

Modified Frank-Wolfe algorithm applied to ($\mathcal{Q}_\lambda(y)$). Dropping the dependence on the auxiliary variable $t \in \mathbb{R}$ we obtain Algorithm 2. As a result of the equivalence between ($\mathcal{P}_\lambda(y)$) and ($\mathcal{Q}_\lambda(y)$), this algorithm is a valid application of Algorithm 1 to ($\mathcal{P}_\lambda(y)$), in the sense that Proposition 3.5 below holds. Before studying convergence results, let us make a comment on the stopping condition.

Algorithm 2: modified Frank-Wolfe algorithm applied to ($\mathcal{P}_\lambda(y)$)

Data: measurement operator Φ, observations y, regularization parameter λ
Result: function u^*

1 $u^{[0]} \leftarrow 0$
2 $N^{[0]} \leftarrow 0$
3 **while** *true* **do**
4 $\quad \eta^{[k]} \leftarrow -\frac{1}{\lambda}\Phi^*(\Phi u^{[k]} - y)$
5 $\quad (E_*, \epsilon_*) \leftarrow \underset{\substack{E\subset\mathbb{R}^2 \\ \epsilon\in\{-1,1\}}}{\text{Argmax}}\ \epsilon \int_E \eta^{[k]}/P(E)$ s.t. E simple, $0 < |E| < +\infty$
6 $\quad$ **if** $\left|\int_{E_*} \eta^{[k]}\right| \leq P(E_*)$ **then**
7 $\quad\quad$ output $u^* \leftarrow u^{[k]}$, which is optimal
8 $\quad$ **else**
9 $\quad\quad E^{[k+1]} \leftarrow (E_1^{[k]}, ..., E_{N^{[k]}}^{[k]}, E_*)$
10 $\quad\quad \epsilon^{[k+1]} \leftarrow (\epsilon_1^{[k]}, ..., \epsilon_{N^{[k]}}^{[k]}, \epsilon_*)$
11 $\quad\quad a^{[k+1]} \leftarrow \underset{a\in(\mathbb{R}_+)^{N^{[k]}+1}}{\text{Argmin}}\ \frac{1}{2\lambda}\|\Phi^{\epsilon^{[k+1]}}_{E^{[k+1]}} a - y\|^2 + \sum_{i=1}^{N^{[k]}+1} P(E_i^{[k+1]})\, a_i$
12 $\quad\quad$ remove atoms with zero amplitude
13 $\quad\quad N^{[k+1]} \leftarrow$ number of atoms in $E^{[k+1]}$
14 $\quad\quad u^{[k+1]} \leftarrow \sum_{i=1}^{N^{[k+1]}} a_i^{[k+1]} \epsilon_i^{[k+1]} \mathbf{1}_{E_i^{[k+1]}}$
15 $\quad$ **end**
16 **end**

Stopping condition. The stopping condition (Line 3 of Algorithm 1) is replaced in Algorithm 2 by

$$\sup\left\{\frac{\epsilon}{P(E)}\int_E \eta^{[k]},\ E\subset\mathbb{R}^2 \text{ simple},\ 0<|E|<+\infty,\ \epsilon\in\{-1,1\}\right\} \leq 1\,, \tag{33}$$

which is equivalent to $\eta^{[k]} \in \partial \mathrm{TV}(0)$. Since the optimality of $a^{[k]}$ at Line 11 always ensures

$$\int_{\mathbb{R}^2} \eta^{[k]} u^{[k]} = \sum_{i=1}^{N^{[k]}} a_i^{[k]} P(E_i^{[k]}) \geq \mathrm{TV}(u^{[k]})\,,$$

we obtain that $u^{[k]}$ solves ($\mathcal{P}_\lambda(y)$) as soon as (33) holds.

1.3. Convergence results

1.3.1. Convergence in objective value

Curvature constant. As pointed out in [Jaggi, 2013], in the convergence analysis of Frank-Wolfe algorithm applied to a function f, a measure of the "nonlinearity" of f, called the *curvature*

constant, naturally appears. Let f be a convex differentiable function and C a weakly compact convex set. The curvature constant C_f of f with respect to C is defined by:

$$C_f \overset{\text{def.}}{=} \sup_{\substack{x,s\in C,\\ \gamma\in[0,1],\\ y=x+\gamma(s-x)}} \frac{2}{\gamma^2}\left[f(y)-f(x)-df(x).(y-x)\right] .$$

Proposition 3.4

The curvature constant C_G of G with respect to the admissible set C of $(\mathcal{Q}_\lambda(y))$ satisfies:

$$C_G \leq \frac{1}{\lambda}\left(c_2\,\frac{\|\Phi\|\,\|y\|^2}{\lambda}\right)^2 ,$$

where $c_2 = 1/\sqrt{4\pi}$ is the isoperimetric constant.

Proof : For any $u, v \in \mathrm{L}^2(\mathbb{R}^2)$ and $t, s \in \mathbb{R}$, we have:

$$G(u,t) - G(v,s) - dG(u,t).(v,s) = \frac{1}{2\lambda}\|\Phi(u-v)\|^2 .$$

We hence obtain that

$$\begin{aligned} C_G = \sup_{u,v\in\mathrm{L}^2(\mathbb{R}^2)} \quad & \frac{1}{\lambda}\|\Phi(u-v)\|^2 \\ \text{s.t.} \quad & \mathrm{TV}(u) \leq \frac{1}{2\lambda}\|y\|^2 , \\ & \mathrm{TV}(v) \leq \frac{1}{2\lambda}\|y\|^2 . \end{aligned} \tag{34}$$

Now, if u, v are admissible for (34), using the isoperimetric inequality (3), we finally get:

$$\begin{aligned} \|\Phi(u-v)\| &\leq \|\Phi\|\|u-v\|_{\mathrm{L}^2(\mathbb{R}^2)} \\ &\leq \|\Phi\|\, c_2\, \mathrm{TV}(u-v) \leq c_2\,\frac{\|\Phi\|\,\|y\|^2}{\lambda} . \end{aligned}$$

■

Convergence in objective value. As already mentioned, Algorithm 2 is a valid application of Algorithm 1 to $(\mathcal{P}_\lambda(y))$, in the sense that the following property holds (see [Jaggi, 2013]):

Proposition 3.5

Let $(u^{[k]})_{k\geq 0}$ be a sequence produced by Algorithm 2. Then for any solution u^* of $(\mathcal{P}_\lambda(y))$, we have:

$$\forall k \in \mathbb{N},\ F(u^{[k]}) - F(u^*) \leq \frac{2\,C_G}{k+2} .$$

Approximate linear minimization step. As discussed in [Jaggi, 2013], the linear minimization step, which consists in solving (30) or equivalently (31), can be carried out approximately. In fact if there exists $\delta > 0$ such that, for every k, the couple (E_*, ϵ_*) computed at Line 5 of Algorithm 2 is an $\epsilon^{[k]}$- minimizer of (31) with $\epsilon^{[k]} = \delta\, C_G/(k+2)$, then

$$\forall k \in \mathbb{N}^*,\ F(u^{[k]}) - F(u^*) \leq \frac{2\,C_G}{k+2}(1+\delta) .$$

1.3.2. Convergence of the iterates

Convergence of minimizing sequences. Now, let us state a general property of minimizing sequences of $(\mathcal{P}_\lambda(y))$, which hence applies to the sequence of iterates produced by Algorithm 2.

Proposition 3.6

Let $(u^{[k]})_{k\geq 0}$ be a minimizing sequence for $(\mathcal{P}_\lambda(y))$. Then there exists a subsequence which converges strongly in $\mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^2)$ and weakly in $\mathrm{L}^2(\mathbb{R}^2)$ to a solution u^* of $(\mathcal{P}_\lambda(y))$. Moreover, we have $\mathrm{TV}(u^{[k]}) \to \mathrm{TV}(u^*)$.

Proof : We first notice that since $(u^{[k]})_{k\geq 0}$ is a minimizing sequence then $(\mathrm{TV}(u^{[k]}))_{k\geq 0}$ is bounded, and (from the isoperimetric inequality) $(u^{[k]})_{k\geq 0}$ is therefore bounded in $\mathrm{L}^2(\mathbb{R}^2)$. There hence exists a subsequence (not relabeled) which converges weakly in $\mathrm{L}^2(\mathbb{R}^2)$ to $u^* \in \mathrm{L}^2(\mathbb{R}^2)$. Since this subsequence satisfies

$$\sup_{k\in\mathbb{N}} \left[\|u^{[k]}\|_{\mathrm{L}^2(\mathbb{R}^2)} + \mathrm{TV}(u^{[k]}) \right] < +\infty \,,$$

from [Ambrosio et al., 2000, Theorem 3.23] we obtain that (up to the extraction of a further subsequence, still not relabeled) $(u^{[k]})_{k\geq 0}$ converges strongly in $\mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^2)$ (to a limit which is necessarily u^*). The fact u^* solves $(\mathcal{P}_\lambda(y))$ then follows from the lower semicontinuity of the data fidelity term with respect to the weak $\mathrm{L}^2(\mathbb{R}^2)$ topology, and the lower semicontinuity of the total variation with respect to the strong $\mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^2)$ topology. Finally, since

$$\begin{aligned}
\frac{1}{2\lambda}\|\Phi u^* - y\|^2 &\leq \liminf_{k\to+\infty} \frac{1}{2\lambda}\|\Phi u^{[k]} - y\|^2 \,, \\
\mathrm{TV}(u^*) &\leq \liminf_{k\to+\infty} \mathrm{TV}(u^{[k]}) \,, \\
\frac{1}{2\lambda}\|\Phi u^* - y\|^2 + \mathrm{TV}(u^*) &= \lim_{k\to+\infty} \frac{1}{2\lambda}\|\Phi u^{[k]} - y\|^2 + \mathrm{TV}(u^{[k]}) \,,
\end{aligned}$$

we obtain $\mathrm{TV}(u^{[k]}) \to \mathrm{TV}(u^*)$. ∎

As explained earlier, the fact $\mathrm{TV}(u^{[k]}) \to \mathrm{TV}(u^*)$ and the strong $\mathrm{L}^1_{\mathrm{loc}}$ convergence of $(u^{[k]})_{k\geq 0}$ towards u^* also imply $\mathrm{D}u^{[k]} \overset{*}{\rightharpoonup} \mathrm{D}u_*$.

Properties of sets appearing in Algorithm 2. An important observation is that, for every $k \in \mathbb{N}$, the set E_* introduced at Line 5 of Algorithm 2 is a minimizer of the prescribed curvature problem associated to $\lambda^{[k]}\eta^{[k]}$ with $1/\lambda^{[k]}$ the value of (31)[1]. Moreover, if $(u^{[k]})$ is a subsequence as in Proposition 3.6, we have that $\Phi u^{[k]} \to \Phi u^*$ and hence $\eta^{[k]} \to \eta^*$ strongly in $\mathrm{L}^2(\mathbb{R}^2)$ with $\eta^* = (-1/\lambda)\Phi^*(\Phi u - y)$, which also yields $\lambda^{[k]} \to 1$[2]. This can be exploited by using Lemma 1.11 to obtain "uniform" properties of sets appearing in Algorithm 2, which ultimately yields the stronger convergence results presented below.

[1] The value of (31) is nonzero unless $\eta^{[k]} = 0$, which occurs if and only if $y = 0$ (in which case $u^* = 0$ is the unique solution of $(\mathcal{P}_\lambda(y))$). To see this, one simply needs to consider a Lebesgue point of $\eta^{[k]}$ at which the function is nonzero, and to show the objective is strictly positive for sufficiently small balls.

[2] To see this, it is sufficient to note that $\left| \frac{\int_E \eta_1}{P(E)} - \frac{\int_E \eta_2}{P(E)} \right| \leq c_2 \, \|\eta_1 - \eta_2\|_{\mathrm{L}^2(\mathbb{R}^2)}$ for every E with $0 < |E| < +\infty$ and $P(E) < +\infty$.

Additional properties of sequences produced by Algorithm 2. If $(u^{[k]})_{k\geq 0}$ is a subsequence as in Proposition 3.6 then $(\lambda^{[k]}\eta^{[k]})_{k\geq 0}$ converges strongly in $\mathrm{L}^2(\mathbb{R}^2)$, and its terms all belong to $\partial\mathrm{TV}(0)$. Lemma 1.11 hence yields the existence of $R>0$ such that $\mathrm{Supp}(u^{[k]})\subset B(0,R)$ for every $k\in\mathbb{N}$. We consequently obtain the following strict BV convergence result.

Proposition 3.7

Let $(u^{[k]})_{k\geq 0}$ be a sequence produced by Algorithm 2. Then, up to the extraction of a subsequence, we have that $(u^{[k]})_{k\geq 0}$ converges strictly in $\mathrm{BV}(\mathbb{R}^2)$ to a solution u^* of $(\mathcal{P}_\lambda(y))$.

This in turn yields a convergence result for the level sets of the iterates, which is given below (we remind the reader that the inner limit of a sequence of sets is defined in Definition 1.9).

Corollary 3.8

Let $(u^{[k]})_{n\geq 0}$ be a subsequence as in Proposition 3.7. Then for almost every $t\in\mathbb{R}$, we have

$$\lim_{n\to+\infty} |U_k^{(t)}\triangle U_*^{(t)}| = 0 \;\text{ and }\; \partial U_*^{(t)} \subseteq \liminf_{n\to+\infty} \partial U_k^{(t)}\,.$$

Proof : The strong convergence of $(u^{[k]})_{k\geq 0}$ towards u^* in $\mathrm{L}^1(\mathbb{R}^2)$ and Proposition D.1 yield:

$$\lim_{n\to+\infty} |U_k^{(t)}\triangle U_*^{(t)}| = 0 \text{ for almost every } t\in\mathbb{R}\,.$$

We now fix such $t\in\mathbb{R}$ and let $x\in\partial U_*^{(t)}$. We want to show that $x\in\liminf_{n\to+\infty}\partial U_k^{(t)}$, which is equivalent to

$$\limsup_{k\to+\infty} \mathrm{dist}\big(x,\,\partial U_k^{(t)}\big) = 0\,.$$

By contradiction, if the last identity does not hold, we have the existence of $r>0$ and of φ such that

$$\forall k\in\mathbb{N},\;\; B(x,r)\cap\partial U_{\varphi(k)}^{(t)} = \emptyset\,.$$

Hence for all k, we either have

$$B(x,r)\subset U_{\varphi(k)}^{(t)} \;\text{ or }\; B(x,r)\subset\mathbb{R}^2\setminus U_{\varphi(k)}^{(t)}\,.$$

If $B(x,r)\subset U_{\varphi(k)}^{(t)}$ for a given k, using that $\eta^*\in\partial\mathrm{TV}(u^*)$ and hence that $U_*^{(t)}$ satisfies point 4 of Lemma 1.11 we obtain

$$\left|U_{\varphi(k)}^{(t)}\triangle U_*^{(t)}\right| \geq \left|U_{\varphi(k)}^{(t)}\setminus U_*^{(t)}\right| \geq \left|B(x,r)\setminus U_*^{(t)}\right| \geq C\,|B(x,r)|\,.$$

We can in the same way show

$$\left|U_{\varphi(k)}^{(t)}\triangle U_*^{(t)}\right| \geq C\,|B(x,r)|$$

if $B(x,r)\subset\mathbb{R}^2\setminus U_{\varphi(k)}^{(t)}$, and hence get the inequality for all k. Using that $|U_k^{(t)}\triangle U_*^{(t)}|\to 0$, we get a contradiction. ∎

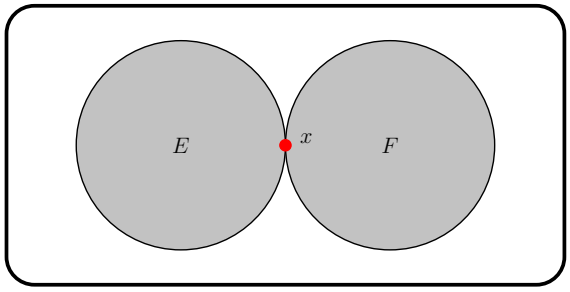


Figure 13 – The sets E and F have the weak regularity property (point 4 of Lemma 1.11), but $E \cup F$ does not, since the density ratio of x with respect to $E \cup F$ converges to 1.

Convergence of the boundaries of the level sets ? Considering Corollary 3.8, it is natural to wonder whether

$$\limsup_{n \to +\infty} \partial U_k^{(t)} \subseteq \partial U_*^{(t)}$$

holds, which would yield the convergence of $(\partial U_k^{(t)})_{k \geq 0}$ towards $\partial U_*^{(t)}$. However, it is unclear whether $U_k^{(t)}$ satisfies point 4 of Lemma 1.11 (a property that one would crucially need to mimic the proof of Corollary 3.8). The level sets $U_k^{(t)}$ are obtained as a finite number of unions and intersections of the sets in $E^{[k]}$ and their complements. Although the latter sets have the desired property, it is not stable by such operations, as Figure 13 shows.

1.4. Sliding step

1.4.1. Presentation

Several works [Bredies and Pikkarainen, 2013, Boyd et al., 2017, Denoyelle et al., 2019] have advocated for the use of a special final update, which helps identify the sparse structure of the sought-after signal. Loosely speaking, it would amount in our case to running, at the end of iteration $k-1$, the gradient flow of the mapping

$$(a, E) \mapsto G(\epsilon^{[k]}, a, E) \tag{35}$$

initialized with $(a^{[k]}, E^{[k]})$, so as to find a set of parameters at which the objective is smaller. Formally, this would correspond[1] to finding a curve $t \mapsto (a_i(t), E_i(t))_{i=1,\dots,N^{[k]}}$ such that for all t:

$$\forall i \in \{1, \dots, N^{[k]}\}, \begin{cases} a_i'(t) = -\left(P(E_i(t)) - \epsilon_i^{[k]} \displaystyle\int_{E_i(t)} \eta(t) \right), \\ V_i(t) = -a_i(t) \left(H_{E_i(t)} - \epsilon_i^{[k]} \eta(t) \right), \end{cases} \tag{36}$$

[1] The formulas in (36) can be formally obtained by using the notion of shape derivative, see [Henrot and Pierre, 2018, Chapter 5].

where $V_i(t)$ denotes the normal velocity of the boundary of $E_i(t)$ and

$$\eta(t) = -\frac{1}{\lambda}\Phi^*\left(\Phi u(t) - y\right), \; u(t) = \sum_{i=1}^{N^{[k]}} a_i(t)\,\epsilon_i^{[k]}\,\mathbf{1}_{E_i(t)}\,.$$

The study of this gradient flow (existence, uniqueness) seems to be challenging, and we did not investigate these questions.

For our purpose, it is enough to introduce the sliding step as a local descent on the mapping defined in (35) initialized with $(a^{[k]}, E^{[k]})$. We however need to ensure this step does not increase the value of the objective, and we hence ask to find $(a, E) = (a_i, E_i)_{1\leq i\leq N^{[k]}}$ such that E_i is simple for all i and

$$G_{N^{[k]}}\big(\epsilon^{[k]}, a, E\big) \leq G_{N^{[k]}}\big(\epsilon^{[k]}, a^{[k]}, E^{[k]}\big). \tag{37}$$

The resulting algorithm is Algorithm 3. The introduction of Line 15, which is discussed in the next paragraph, ensures that all convergence guarantees derived for Algorithm 2 remain valid.

Algorithm 3: modified Frank-Wolfe algorithm applied to ($\mathcal{P}_\lambda(y)$) (with sliding)

Data: measurement operator Φ, observations y, regularization parameter λ
Result: function u^*

1 $u^{[0]} \leftarrow 0$
2 $N^{[0]} \leftarrow 0$
3 **while** *true* **do**
4 $\quad \eta^{[k]} \leftarrow -\frac{1}{\lambda}\Phi^*\left(\Phi u^{[k]} - y\right)$
5 $\quad (E_*, \epsilon_*) \leftarrow \underset{\substack{E\subset\mathbb{R}^2\\ \epsilon\in\{-1,1\}}}{\text{Argmax}}\ \epsilon \int_E \eta^{[k]}/P(E)$ s.t. E simple, $0 < |E| < +\infty$
6 $\quad$ **if** $\left|\int_{E_*} \eta^{[k]}\right| \leq P(E_*)$ **then**
7 $\quad\quad$ output $u^* \leftarrow u^{[k]}$, which is optimal
8 $\quad$ **else**
9 $\quad\quad E^{[k+1]} \leftarrow (E_1^{[k]}, ..., E_{N^{[k]}}^{[k]}, E_*)$
10 $\quad\quad \epsilon^{[k+1]} \leftarrow (\epsilon_1^{[k]}, ..., \epsilon_{N^{[k]}}^{[k]}, \epsilon_*)$
11 $\quad\quad a^{[k+1]} \leftarrow \underset{a\in(\mathbb{R}_+)^{N^{[k]}+1}}{\text{Argmin}} \frac{1}{2\lambda}\|\Phi_{E^{[k+1]}}^{\epsilon^{[k+1]}} a - y\|_{\mathcal{H}}^2 + \sum_{i=1}^{N^{[k]}+1} P(E_i^{[k+1]})\, a_i$
12 $\quad\quad$ remove atoms with zero amplitude
13 $\quad\quad N^{[k+1]} \leftarrow$ number of atoms in $E^{[k+1]}$
14 $\quad\quad$ perform a local descent on $(a, E) \mapsto G_{N^{[k+1]}}(\epsilon^{[k+1]}, a, E)$ initialized with $(a^{[k+1]}, E^{[k+1]})$
15 $\quad\quad$ repeat the operations of Lines 11-13
16 $\quad\quad u^{[k+1]} \leftarrow \sum_{i=1}^{N^{[k+1]}} a_i^{[k+1]}\,\mathbf{1}_{E_i^{[k+1]}}$
17 $\quad$ **end**
18 **end**

1.4.2. Influence on convergence

The version of the sliding step we use was introduced in [Denoyelle et al., 2019]. In practice, it allows to considerably improve the convergence speed of the algorithm. It also produces sparser solutions: if the solution is a combination of a few indicator functions, removing the sliding step typically produces iterates made of a much larger number of atoms. In fact, it seems that the majority of these atoms correct the crude approximations of the support of the solution made during the first iterations.

In [Denoyelle et al., 2019], the introduction of the sliding step allowed the authors to derive improved convergence guarantees (i.e. finite time convergence) in the context of sparse spikes recovery. Their proof relies on the fact that, at the end of every sliding step, a "critical point" of the objective is reached. However, the existence issues mentioned in Section 1.4.1 make the adaptation of these results to our setting difficult.

A reasonable definition of "critical point" in our context could be as follows.

Definition 3.9 (*Critical point of G_N*)

Let $N \in \mathbb{N}^*$, $\epsilon \in \{-1,1\}^N$, $a \in (\mathbb{R}_+^*)^N$ and $E_1, ..., E_N$ be simple subsets of $\mathbb{R}^2$ with positive and finite measure. We say that $(\epsilon_i, a_i, E_i)_{i=1,...,N}$ is a critical point of the mapping G_N if

$$\forall i \in \{1, ..., N\}, \quad \begin{cases} P(E_i) = \epsilon_i \displaystyle\int_{E_i} \eta\,, \\ H_{E_i} = \epsilon_i\, \eta\,, \end{cases} \tag{38}$$

where

$$\eta \overset{\text{def.}}{=} -\frac{1}{\lambda}\Phi^*\left(\Phi u - y\right) \quad \text{and} \quad u \overset{\text{def.}}{=} \sum_{i=1}^{N} a_i\, \epsilon_i\, \mathbf{1}_{E_i}\,.$$

With this definition, reaching a critical point at the end of the sliding step implies all sets in $E^{[k]}$ have the same distributional curvature (up to a sign). On the contrary, without sliding, sets are never modified and their curvature depends on the iteration during which they were introduced.

We stress that if, for a given iteration, a critical point is reached at the end of the sliding step, then Line 15 can be skipped, since the first equality in (38) ensure $a^{[k+1]}$ minimizes

$$a \mapsto G_{N^{[k+1]}}(\epsilon^{[k+1]}, a, E^{[k+1]})\,.$$

If a critical point is reached, the fact the sets in $E^{[k]}$ have the same distributional curvature can be exploited to obtain "uniform" density estimates for the level sets of $u^{[k]}$, in the spirit of [Maggi, 2012, Corollary 17.18]. It is then natural to wonder whether this could be used to prove

$$\limsup_{n\to+\infty} \partial U_n^{(t)} \subseteq \partial U_*^{(t)}\,, \tag{39}$$

and hence the convergence of $(\partial U_k^{(t)})_{k\geq 0}$ towards $\partial U_*^{(t)}$. A major obstacle towards this result is that, although Lemma 1.11 provides a uniform upper bound on the perimeter of sets appearing in the algorithm, to our knowledge, it does not seem possible to derive such a bound for the perimeter of the level sets, which prevents one from using the potential weak-* convergence of $\mathrm{D}\mathbf{1}_{U_k^{(t)}}$ towards $\mathrm{D}\mathbf{1}_{U_*^{(t)}}$.

Link with conic particle gradient descent. Assuming for simplicity that $\epsilon^{[k]} = (1, ..., 1)$ and, for every $i \neq j$, that $\mathcal{H}^1(\partial^* E_i \cap \partial^* E_j) = 0$, we obtain

$$F\left(\sum_{i=1}^{N^{[k]}} a_i \mathbf{1}_{E_i}\right) = G_{N^{[k]}}(\epsilon^{[k]}, a, E)\,,$$

and hence see that the sliding step amounts to performing a local descent on the mapping

$$(a, E) \mapsto F\left(\sum_{i=1}^{N^{[k]}} a_i \mathbf{1}_{E_i}\right).$$

In [Chizat, 2021, Chizat, 2022], a seemingly related task is considered. Given Θ a parameter space, the author investigates the minimization of

$$\begin{aligned} J : \mathcal{M}_+(\Theta) &\to \mathbb{R} \\ \mu &\mapsto \frac{1}{2\lambda}\|\Phi\mu - y\|^2 + \mu(\Theta) \end{aligned}$$

using conic particle gradient algorithms, which consist in performing gradient-based optimization on

$$(a, x) \mapsto J\left(\sum_{i=1}^{N} h(a_i)\delta_{\theta_i}\right), \tag{40}$$

with $h : \mathbb{R}_+ \to \mathbb{R}_+$ a smooth increasing bijection. The main finding of these works is that, for $h(a) = a^2$ and for a specific choice of metric on $\mathbb{R}_+ \times \Theta$ (plus a few assumptions), the gradient flow of (40) converges to a minimizer of J. An interesting avenue for future research could be to study how this analysis translates to our setting. On the practical side, one could investigate whether performing multiplicative (rather than additive) updates on the amplitudes a improves the convergence speed of the local descent, as suggested in [Chizat, 2021].

2. Polygonal approximation of generalized Cheeger sets

To implement Algorithm 3, one needs two oracles to carry out the operations of Lines 5 and 14[1]. In this section, we focus on the first task (the second is the subject of Section 3.1). Our aim is hence to design a numerical method to approximately solve

$$\begin{aligned} \sup_{E \subset \mathbb{R}^2} \quad & J(E) \overset{\text{def.}}{=} \frac{1}{P(E)}\left|\int_E \eta\right| \\ \text{s.t.} \quad & E \text{ simple},\ 0 < |E| < +\infty\,, \end{aligned} \tag{41}$$

which is called a generalized Cheeger problem[2]. In section 2.1, we explain its connection with the classical Cheeger problem and provide properties of its solutions, which are called generalized

[1] We recall that Line 11 amounts to solve a LASSO-type problem for which efficient solvers exist.

[2] Indeed, (31) and (41) are equivalent in the sense that (E, ϵ) solves (31) if and only if E solves (41) and ϵ is the opposite of the sign of $\int_E \eta$.

Cheeger sets. We also review existing numerical methods for solving it, and motivate our choice to introduce a slightly different procedure.

Our approach consists in choosing a (possibly large) integer $n \geq 3$, and looking for an optimal set among simple polygons with at most n sides. The existence of such sets is proved in Section 2.2. To approximate this optimal polygon, we optimize the vertices of an initial polygon using a first order algorithm. The initialization is performed by solving a relaxation of (41) on a fixed grid. The details of this numerical method are presented in Sections 2.3 and 2.4. Finally, in Section 2.5, we study how these optimal polygons compare to solutions of (41) in a specific setting, where η is assumed to be radial.

2.1. A generalized Cheeger problem

Related problems. We first stress that (41) is reminiscent of the Cheeger problem, which, given an open bounded set $\Omega \subset \mathbb{R}^d$ with Lipschitz boundary, consists in finding a subset E of Ω minimizing $P(E)/|E|$ (see the surveys [Parini, 2011, Leonardi, 2015]). Many generalizations of the Cheeger problem have been considered. Let us mention [Ionescu and Lachand-Robert, 2005], in which the usual perimeter and volume are replaced by weighted versions (the integral on $\partial^* E$ and E of some nonnegative weight functions), and [Caselles et al., 2009, Brasco et al., 2014], in which the case of anisotropic and fractional perimeters are respectively investigated.

Our setting. The problem we consider slightly differs from the ones mentioned above. First, the emphasis is not on the domain Ω anymore ((41) is solved on the whole plane), but on the weight function η. To put it another way, maximizing J on the whole plane or inside some domain Ω is equivalent as soon as η^2 concentrates[1] in a compact subset of Ω. Moreover, contrary to [Ionescu and Lachand-Robert, 2005], the function η is not assumed to be nonnegative, and is taken in $\mathrm{L}^2(\mathbb{R}^2)$ (instead of $\mathrm{L}^\infty(\Omega)$ in this last work).

Existence of solutions. Two strategies can be considered to prove the existence of maximizers in (41). The first is the one hinted in Section 1.2. It consists in proving the existence of maximizers for

$$\begin{aligned} \sup_{u \in \mathrm{L}^2(\mathbb{R}^2)} \quad & \int_{\mathbb{R}^2} \eta u \\ \text{s.t.} \quad & \mathrm{TV}(u) \leq 1\,, \end{aligned} \tag{42}$$

and then showing at least one solution is the indicator function of a simple set (using the form of the extreme points of $\{\mathrm{TV} \leq 1\}$ and the Krein-Milman theorem, or a result like [Carlier and Comte, 2007, Theorem 2]). The second strategy relies on purely geometric arguments, in the spirit of [Maggi, 2012, Section 12]. One can indeed prove a uniform lower bound on the perimeter of minimizing sequences, and obtain local convergence (up to extraction) towards some limit set, whose optimality can subsequently be shown.

Generic uniqueness. As explained in [Buttazzo et al., 2007, Proposition 4.1], the solution of (41) is in some sense generically unique. Indeed, denoting $V(\eta)$ the value of (42), we obtain

[1] To be more precise, we mean that there exists a compact set $K \subset \Omega$ such that $\int_{\mathbb{R}^2 \setminus K} \eta^2$ is sufficiently small.

that V is a continuous convex function on $\mathrm{L}^2(\mathbb{R}^2)$. A theorem of Mazur [Phelps, 1993, Theorem 1.20] then shows V is Gâteaux differentiable on a dense G_δ subset of $\mathrm{L}^2(\mathbb{R}^2)$. Since $\partial V(\eta)$ is precisely the solution set of (41), generic uniqueness follows.

Existing numerical methods. In [Carlier et al., 2009] a method is proposed to approximate the so-called maximal Cheeger set. It covers the case of a weighted area term and of weighted and anisotropic perimeters. The proposed approach is to solve a discretized version of (42), where the maximization is performed over the set of piecewise constant functions on some fixed grid. A method which is very similar to ours is introduced in [Ionescu and Lupaşcu-Stamate, 2019]. It consists in iteratively deforming an initial domain by computing a shape gradient at each iteration. We discovered the existence of this work during the writing of this manuscript.

2.2. Existence of polygonal generalized Cheeger sets

The aim of this section is to prove the existence of polygonal generalized Cheeger sets, i.e. maximizers of J among simple polygons with a given number of sides. In fact, we prove a slightly stronger result, namely the existence of maximizers of a relaxed energy which coincides with J on simple polygons, and the existence of a simple polygon among these maximizers.

Existence proofs in polygonal shape optimization problems. To prove the existence of optimizers for shape optimization problems in the class of polygons, a typical strategy is to use a compactness result for the Hausdorff convergence of open sets. One constructs a minimizing sequence whose terms are open polygons included in a common compact set, and obtain its convergence (up to extraction) in the Hausdorff sense to a "generalized polygon" (see for example [Henrot, 2006, Theorem 3.3.1] for the minimization of the first Dirichlet eigenvalue of the Laplacian, or [Bucur and Fragalà, 2016, Proposition 9] for the minimization of the Cheeger constant). At the time we considered the problem we are discussing here, we were not aware of this literature, and followed a slightly different (but essentially similar) strategy. As explained below, our approach amounts to considering the objective as a function of the vertices of the input polygon.

Notations. In the following, we fix an integer $n \geq 3$ and denote

$$\mathcal{X}_n = \left\{ x \in \mathbb{R}^{n\times 2} \,\middle|\, [x_1, x_2], ..., [x_n, x_1] \text{ is simple} \right\}.$$

We recall that a polygonal curve is said to be simple if non-adjacent sides do not intersect. If $x \in \mathcal{X}_n$, then[1] $\cup_{i=1}^n [x_i, x_{i+1}]$ is a Jordan curve. It hence divides the plane in two regions, one of which is bounded. We denote this region by E_x (it is hence a simple polygon). With this notation the set $\mathcal{P}_n$ of simple polygons with at most n sides is given by

$$\mathcal{P}_n = \{E_x,\ x \in \mathcal{X}_n\}.$$

[1] If $i > n$ we define $x_i \overset{\text{def.}}{=} x_{i \bmod n}$, i.e. $x_{n+1} = x_1$.

Relaxed perimeter and weighted area. We first begin by defining relaxed versions of the perimeter and the (weighted) area. To be able to deal with polygons with a number of vertices smaller than n, which will be useful in the following, we define for all $m \geq 2$ and $x \in \mathbb{R}^{m\times 2}$ the following quantities:

$$P(x) \overset{\text{def.}}{=} \sum_{i=1}^{m} \|x_{i+1} - x_i\| \text{ and } A(x) \overset{\text{def.}}{=} \int_{\mathbb{R}^2} \eta\, \chi_x\,,$$

where $\chi_x(y)$ denotes the index (or winding number) of any parametrization of the oriented polygonal curve $[x_1, x_2], ..., [x_m, x_1]$ around $y \in \mathbb{R}^2$ (see for instance [Rudin, 1986, Theorem 10.10]). In particular, for every $x \in \mathcal{X}_m$ (i.e. for every $x \in \mathbb{R}^{m\times 2}$ defining a simple polygon), we have $\chi_x = \pm \mathbf{1}_{E_x}$ and

$$P(x) = P(E_x) \text{ and } |A(x)| = \left|\int_{E_x} \eta\right|,$$

and hence, as soon as $P(x) > 0$:

$$J(E_x) = \frac{|A(x)|}{P(x)}\,.$$

This naturally leads us to define

$$\mathcal{Y}_m \overset{\text{def.}}{=} \left\{x \in \mathbb{R}^{m\times 2} \,\middle|\, P(x) > 0\right\}$$

and to define, abusing notation, $J(x) \overset{\text{def.}}{=} |A(x)|/P(x)$ for every $x \in \mathcal{Y}_m$.

Properties of the index. As mentioned above, if E_x is simple then $\chi_x = \pm \mathbf{1}_{E_x}$. In the general case χ_x is constant on each connected component of $\mathbb{R}^2 \setminus \Gamma_x$ with $\Gamma_x \overset{\text{def.}}{=} \cup_{i=1}^{m} [x_i, x_{i+1}]$. It takes values in $\{-m, ..., m\}$ and is equal to zero on the only unbounded connected component of $\mathbb{R}^2 \setminus \Gamma_x$. We also have $\partial\, \text{supp}(\chi_x) \subset \Gamma_x$. Moreover χ_x has bounded variation and, for $\mathcal{H}^1$-almost every $y \in \Gamma_x$, there exists $u_\Gamma^+(y), u_\Gamma^-(y)$ in $\{-m, ..., m\}$ such that

$$\mathrm{D}\chi_x = (u^+_{\Gamma_x} - u^-_{\Gamma_x})\, \nu_{\Gamma_x} \mathcal{H}^1 \llcorner \Gamma_x\,.$$

Positivity of the problem value. Now, we define $\alpha \overset{\text{def.}}{=} \sup\{J(x),\ x \in \mathcal{Y}_n\}$. If $\eta = 0$, then the existence of maximizers is trivial. Otherwise, there exists a Lebesgue point x_0 of η at which η is non-zero. Now the family of regular n-gons inscribed in any circle centered at x_0 has bounded eccentricity. Hence, if $x_{n,r}$ defines a regular n-gon inscribed in a circle of radius r centered at x_0, the Lebesgue differentiation theorem[1] ensures that

$$\lim_{r\to 0^+} \frac{\left|\int_{E_{x_{n,r}}} \eta\right|}{|E_{x_{n,r}}|} > 0\,,$$

and the fact that $\alpha > 0$ follows.

[1] For details on the Lebesgue differentiation theorem where balls are replaced by a family of sets with bounded eccentricity, see [Stein, 1993, Chapter 2, Paragraph 3.1].

Lemma 3.10

Let $C > 0$. There exists $R > 0$ and $c > 0$ such that, for every $x \in \mathcal{Y}_n$

$$J(x) \geq C \implies P(x) \geq c \text{ and } \|x_i\| \leq R \text{ for all } i.$$

Proof : The proof is similar to that of Lemma 1.11.

Upper bound on the perimeter: the integrability of η^2 yields that for every $\epsilon > 0$ there exists $R_1 > 0$ such that

$$\int_{\mathbb{R}^2 \setminus B(0,R_1)} \eta^2 \leq \epsilon^2 \,. \tag{43}$$

Let $\epsilon > 0$ and $R_1 > 0$ such that (43) holds. We have

$$\begin{aligned}
P(x) &\leq \frac{1}{C} |A(x)| \\
&\leq \frac{1}{C} \left[\left| \int_{\mathbb{R}^2 \cap B(0,R)} \eta \, \chi_x \right| + \left| \int_{\mathbb{R}^2 \setminus B(0,R)} \eta \, \chi_x \right| \right] \\
&\leq \frac{1}{C} \left[\|\eta\|_{\mathrm{L}^2} \, \|\chi_x\|_{\mathrm{L}^\infty} \, \sqrt{|B(0,R)|} + \epsilon \, \|\chi_x\|_{\mathrm{L}^2} \right] \\
&\leq \frac{1}{C} \left[\|\eta\|_{\mathrm{L}^2} \, n \, \sqrt{|B(0,R)|} + \epsilon \, \frac{1}{\sqrt{c_2}} |\mathrm{D}\chi_x|(\mathbb{R}^2) \right] \\
&\leq \frac{1}{C} \left[\|\eta\|_{\mathrm{L}^2} \, n \, \sqrt{|B(0,R)|} + \epsilon \, \frac{2n}{\sqrt{c_2}} P(x) \right] .
\end{aligned}$$

Now, taking

$$\epsilon \overset{\text{def.}}{=} \frac{C}{4 c_2 n} \text{ and } c' = \frac{2n}{C} \|\eta\|_{\mathrm{L}^2} \sqrt{|B(0,R)|} \,,$$

we finally get that $P(x) \leq c'$.

Inclusion in a ball: we take $\epsilon = \frac{1}{4c_2 n}$ and fix $R_2 > 0$ such that $\int_{\mathbb{R}^2 \setminus B(0,R_2)} \eta^2 \leq \epsilon^2$. Let us show that

$$\operatorname{supp}(\chi_x) \cap B(0,R_2) \neq \emptyset \,.$$

By contradiction, if $\operatorname{supp}(\chi_x) \cap B(0,R_2) = \emptyset$, we would have:

$$\begin{aligned}
P(x) \leq \frac{1}{C} |A(x)| = \frac{1}{C} \left| \int_{\mathbb{R}^2 \setminus B(0,R_2)} \eta \, \chi_x \right| &\leq \sqrt{\int_{\mathbb{R}^2 \setminus B(0,R_2)} \eta^2} \, \|\chi_x\|_{\mathrm{L}^2} \\
&\leq \epsilon \, c_2 \, |D\chi_x|(\mathbb{R}^2) \leq 2n \, \epsilon \, c_2 \, P(x) \,.
\end{aligned}$$

Dividing by $P(x) > 0$ yields a contradiction. Now since

$$\partial \operatorname{supp}(\chi_x) \subset \Gamma_x \,,$$

we have $\operatorname{diam}(\operatorname{supp}(\chi_x)) \leq P(x) \leq c'$ which shows

$$\operatorname{supp}(\chi_x) \subset B(0,R) \text{ with } R \overset{\text{def.}}{=} c' + R_2 \,.$$

This in turn implies that $\|x_i\| \leq R$ for all i.

Lower bound on the perimeter: the integrability of η^2 shows that, for every $\epsilon > 0$, there exists $\delta > 0$ such that

$$\forall E \subset \mathbb{R}^2, \; |E| \leq \delta \implies \left| \int_E \eta^2 \right| \leq \epsilon^2 \,.$$

Taking $\epsilon \overset{\text{def.}}{=} C/(2\,c_2)$, we obtain that if $|\mathrm{supp}(\chi_x)| \leq \delta$

$$P(x) \leq \frac{1}{C}\,|A(x)| = \frac{1}{C}\left|\int_{\mathrm{supp}(\chi_x)} \eta\right| \leq \frac{1}{C}\sqrt{\int_{\mathrm{supp}(\chi_x)} \eta^2}\,\sqrt{|\mathrm{supp}(\chi_x)|} \leq \frac{\epsilon\,c_2}{C}\,P(\mathrm{supp}(\chi_x)) \leq \frac{\epsilon\,c_2}{C}\,P(x)\,,$$

the last inequality holding because $\partial\,\mathrm{supp}(\chi_x) \subset \Gamma_x$. We get a contradiction since $P(x)$ is positive. ■

Applying Lemma 3.10 with e.g. $C = \alpha/2$, and defining

$$\mathcal{Y}_n' \overset{\text{def.}}{=} \left\{x \in \mathbb{R}^{n\times 2} \;\middle|\; P(x) \geq c \text{ and } \|x_i\| \leq R \text{ for all } i\right\},$$

we see that any maximizer of J over $\mathcal{Y}_n'$ (if it exists) is also a maximizer of J over $\mathcal{Y}_n$, and conversely.

Lemma 3.11

Let $x \in \mathbb{R}^{n\times 2}$. Then, for every $a \in \mathbb{R}^2$, denoting $(c_1\ c_2)$ the 2×2 matrix whose columns are $c_1, c_2 \in \mathbb{R}^2$, we have:

$$\begin{aligned} A(x) &= \sum_{i=1}^{n} \mathrm{sign}(\det(x_i - a\ x_{i+1} - a)) \int_{a x_i x_{i+1}} \eta \\ &= \sum_{i=1}^{n} \det(x_i - a\ x_{i+1} - a) \int_{T_1} \eta((x_i - a\ x_{i+1} - a)\,y)\,dy\,, \end{aligned}$$

where $a x_i x_{i+1}$ denotes the triangle with vertices a, x_i, x_{i+1} and

$$T_1 \overset{\text{def.}}{=} \left\{(\alpha,\beta) \in (\mathbb{R}_+)^2 \;\middle|\; \alpha + \beta \leq 1\right\}$$

is the unit triangle.

Proof : Let us show that for all $a \in \mathbb{R}^2$ we have

$$\chi_x = \sum_{i=1}^{n} \mathrm{sign}(\det(x_i - a\ x_{i+1} - a))\,\mathbf{1}_{a x_i x_{i+1}} \tag{44}$$

almost everywhere. We have that $y \in \mathbb{R}^2$ is in the (open) triangle $a x_i x_{i+1}$ if and only if the ray issued from y directed by $y - a$ intersects $]x_i, x_{i+1}[$. Moreover, if y is in this triangle, then $\det(x_i - a\ x_{i+1} - a))$ is positive if and only if the triangle $y x_i x_{i+1}$ is oriented counterclockwise. This shows that, if $\mathbb{R}^2 \setminus \cup_{i=1}^n [x_i, x_{i+1}]$ does not belong to any of the segments $[a, x_i]$, evaluating the right hand side of (44) at y amounts to computing the winding number $\chi_x(y)$ by applying the ray-crossing algorithm described in [Hormann and Agathos, 2001]. This in particular means that (44) holds almost everywhere, and the result follows. ■

From Lemma 3.11, we get that A is continuous on $\mathbb{R}^{n\times 2}$. This is also the case of P. Now $\mathcal{Y}_n'$ is compact and included in $\mathcal{Y}_n$, hence the existence of maximizers of J over $\mathcal{Y}_n'$, which in turn implies the existence of maximizers of J over $\mathcal{Y}_n$.

Existence of a maximizing simple n-gon. Let us now show there exists a maximizer which belongs to $\mathcal{X}_n$, i.e. that defines a simple polygon. To do so, we rely on the following lemma. Loosely speaking, it states that starting from a non-simple polygon, one can always construct another polygon with strictly fewer vertices achieving a higher objective value.

Lemma 3.12

Let $m \geq 3$ and $x \in \mathcal{Y}_m \setminus \mathcal{X}_m$. Then there exists m' with $2 \leq m' < m$ and $y \in \mathcal{Y}_{m'}$ such that

$$J(x) \leq J(y)\,.$$

Proof : If $x \in \mathcal{Y}_m \setminus \mathcal{X}_m$ then $[x_1, x_2], ..., [x_m, x_1]$ is not simple. If there exists i with $x_i = x_{i+1}$ then

$$y = (x_1, ..., x_i, x_{i+2}, ..., x_m)$$

is suitable, and likewise if $x_1 = x_m$ then

$$y = (x_1, ..., x_{m-1})$$

is suitable. Otherwise we distinguish the following cases:

If there exists $i < j$ with $x_i = x_j$: we define

$$\begin{aligned} y &= (x_1, ..., x_i, x_{j+1}, ..., x_m) \in \mathbb{R}^{m-(j-i)}\,, \\ z &= (x_i, x_{i+1}, ..., x_{j-1}) \in \mathbb{R}^{j-i}\,. \end{aligned}$$

We notice that $2 \leq j - i < m$ and $2 \leq m - (j - i) < m$.

If there exists $i < j$ with $x_i \in]x_j, x_{j+1}[$: we necessarily have $(i, j) \neq (1, m)$. We define

$$\begin{aligned} y &= (x_1, ..., x_i, x_{j+1}, ..., x_m) \in \mathbb{R}^{m-(j-i)}\,, \\ z &= (x_i, x_{i+1}, ..., x_j) \in \mathbb{R}^{j-i+1}\,. \end{aligned}$$

We again have $2 \leq m - (j - i) < m$, and since $(i, j) \neq (1, m)$, we have $j - i < m - 1$ which shows that $2 \leq j - i + 1 < m$.

If there exists $i < j$ with $x_j \in]x_i, x_{i+1}[$: we necessarily have $j > i + 1$. We define

$$\begin{aligned} y &= (x_1, ..., x_i, x_j, ..., x_m) \in \mathbb{R}^{m-(j-i)+1}\,, \\ z &= (x_{i+1}, ..., x_j) \in \mathbb{R}^{j-i}\,. \end{aligned}$$

We again have $2 \leq j - i < m$, and since $j > i + 1$ we obtain that $2 \leq m - (j - i) + 1 < m$.

If there exists $i < j$ with $x' \in]x_i, x_{i+1}[\cap]x_j, x_{j+1}[$: if we have $j = i + 1$ then either $x_{i+2} \in]x_i, x_{i+1}[$ or $x_i \in]x_{i+1}, x_{i+2}[$ and in both cases we fall back on the previously treated cases. The same holds if $(i, j) = (1, m)$. Otherwise, we define

$$\begin{aligned} y &= (x_1, ..., x_i, x', x_{j+1}, ..., x_m) \in \mathbb{R}^{m-(j-i)+1}\,, \\ z &= (x', x_{i+1}, ..., x_j) \in \mathbb{R}^{j-i+1}\,. \end{aligned}$$

Since $j > i + 1$ and $(i, j) \neq (1, m)$ we get $2 \leq m - (j - i) + 1 < m$ and $2 \leq j - i + 1 < m$.

Now, one can see that in each case we have $P(x) = P(y) + P(z)$ and $\chi_x = \chi_y + \chi_z$ almost everywhere, which in turn gives that $A(x) = A(y) + A(z)$. We hence get that $P(y) = 0$ or $P(z) = 0$, and hence $J(x) = J(y)$ or $J(x) = J(z)$, or that $P(y) > 0$ and $P(z) > 0$, which yields

$$\frac{|A(x)|}{P(x)} \leq \frac{|A(y)| + |A(z)|}{P(y) + P(z)} = \frac{P(y)}{P(y) + P(z)} \frac{|A(y)|}{P(y)} + \frac{P(z)}{P(y) + P(z)} \frac{|A(z)|}{P(z)} .$$

Hence $J(x)$ is smaller than a convex combination of $J(y)$ and $J(z)$, which gives that it is smaller than $J(y)$ or $J(z)$. This shows that y or z is suitable. ■

We can now prove our final result, i.e. that there exists $x_* \in \mathcal{X}_n$ such that

$$\forall x \in \mathcal{Y}_n,\ J(x_*) \geq J(x) .$$

Indeed, repeatedly applying the above lemma starting with a maximizer x_* of J over $\mathcal{Y}_n$, we either have that there exists m with $3 \leq m \leq n$ and $x'_* \in \mathcal{X}_m$ such that $J(x_*) = J(x'_*)$, or that there exists $y \in \mathcal{Y}_2$ such that $J(x_*) \leq J(y)$, which is impossible since in that case $J(y) = 0$ and $J(x_*) = \alpha > 0$. We hence have $x'_* \in \mathcal{X}_m$ such that

$$\forall x \in \mathcal{Y}_n,\ J(x'_*) = J(x_*) \geq J(x) .$$

We can finally build $x''_* \in \mathcal{X}_n$ such that $J(x''_*) = J(x'_*)$ by adding dummy vertices to x'_*, which allows to conclude.

2.3. Fixed grid initialization

In this section, we detail the initialization of our first order method. We proceed almost exactly as in [Carlier et al., 2009], i.e. we solve a discrete version of (42), in which the minimization is performed over the set of piecewise constant functions on a fixed grid. The only difference with this last work is that we do not look for some maximal solution of (42), and hence do not need the strictly concave penalization introduced in [Buttazzo et al., 2007][1].

Discrete problem. Since every solution of (42) has its support included in some ball[2], we can solve (42) in $[-R, R]^2$ (with Dirichlet boundary conditions) for a sufficiently large $R > 0$. Let N be a positive integer and $h \overset{\text{def.}}{=} 2R/N$. We denote E^h the set of N by N matrices. For every matrix $u = (u_{i,j})_{(i,j)\in[1,N]^2} \in E^h$ we define

$$\partial_x^h u_{i,j} \overset{\text{def.}}{=} u_{i+1,j} - u_{i,j} , \qquad \partial_y^h u_{i,j} \overset{\text{def.}}{=} u_{i,j+1} - u_{i,j} , \tag{45}$$

for all $(i, j) \in [0, N]^2$, with the convention $u_{i,j} = 0$ if either i or j is in $\{0, N + 1\}$. The isotropic discrete total variation is then defined by

$$\mathrm{TV}^h(u) \overset{\text{def.}}{=} h \sum_{i=0}^{N} \sum_{j=0}^{N} ||\nabla^h u_{i,j}||_2 = h \, \|\nabla^h u\|_{2,1} \text{ with } \nabla^h u_{i,j} \overset{\text{def.}}{=} \left(\partial_x^h u_{i,j}, \partial_y^h u_{i,j} \right) .$$

[1] We stress that, although we do not follow this approach, finding some maximal solution of (42) and adding several atoms at each iteration could potentially be interesting. This would lead to what could be called a "polyatomic" Frank-Wolfe algorithm (see [Jarret et al., 2022] and the related approach of [Flinth et al., 2021]).

[2] Indeed, if u solves (42), then there exists α such that $\alpha\eta \in \partial \mathrm{TV}(u)$, and the result follows from Proposition 1.3 and Lemma 1.11

We then solve the following discretized version of (42) for increasingly small values of $h > 0$

$$\min_{u \in E^h} h^2 \langle \overline{\eta}^h, u \rangle \;\text{ s.t. }\; \mathrm{TV}^h(u) \leq 1\,, \tag{46}$$

where $\overline{\eta}^h = \left(\frac{1}{h^2}\int_{C^h_{i,j}} \eta\right)_{(i,j)\in[1,N]^2}$ and $(C^h_{i,j})_{(i,j)\in[1,N]^2}$ is a partition of $[-R,R]^2$ composed of squares of equal size, i.e.

$$C^h_{i,j} \overset{\text{def.}}{=} [-R+(i-1)h, -R+ih] \times [-R+(j-1)h, -R+jh]\,.$$

For convenience reasons, we also use the above expression to define $C^h_{i,j}$ if i or j belongs to $\{0, N+1\}$.

Optimization algorithm. To solve (46), we propose to use the primal-dual algorithm introduced in [Chambolle and Pock, 2011]: we take (τ,σ) such that $\tau\sigma\|D\|_2^2 < 1$ with $D \overset{\text{def.}}{=} h\nabla^h$ and define

$$\begin{cases} \phi^{n+1} = \mathrm{prox}_{\sigma\|\cdot\|_{2,\infty}}(\phi^n + \sigma\, D\bar{u}^n)\,, \\ u^{n+1} = (u^n - \tau\, D^*\phi^{n+1}) - \tau\, h^2\, \bar{\eta}^h\,, \\ \bar{u}^{n+1} = 2\,u^{n+1} - u^n\,, \end{cases} \tag{47}$$

where $\mathrm{prox}_{\sigma\|\cdot\|_{2,\infty}}$ is given by:

$$\mathrm{prox}_{\sigma\|\cdot\|_{2,\infty}}(\phi) = \phi - \sigma\,\mathrm{proj}_{\{\|\cdot\|_{2,1}\leq 1\}}\left(\frac{\phi}{\sigma}\right).$$

The projection onto the $(2,1)$-unit ball can be computed efficiently (see [Condat, 2016]).

Convergence as $h \to 0$. The following proposition shows that, when the grid becomes finer, solutions of (46) converge to a solution of (42). Its proof is almost the same as the one of [Carlier et al., 2009, Theorem 4.1]. Since the latter however gives a slightly different result about the minimization of a quadratic objective (linear in our case) on the total variation unit ball, we decided to include it for the sake of completeness.

Proposition 3.13

Let u^h be the piecewise constant function on $(C^h_{i,j})_{(i,j)\in[1,N]^2}$, extended to 0 outside $[-R,R]^2$ associated to a solution of (46). Then there exists a (not relabeled) subsequence converging strongly in $\mathrm{L}^1(\mathbb{R}^2)$ and weakly in $\mathrm{L}^2(\mathbb{R}^2)$ to a solution u^* of (42) when $h \to 0$. Moreover, we have $\mathrm{D}u^h \overset{*}{\rightharpoonup} \mathrm{D}u$.

Proof: First, let us stress that for any function v that is piecewise constant on $(C_{i,j})_{(i,j)\in[1,N]^2}$ and that is equal to 0 outside $[-R,R]^2$, we have $\mathrm{TV}(v) = h\,\|\nabla^h v\|_{1,1}$ where by abuse of notation $\nabla^h v$ is given by (45) with $v_{i,j}$ the value of v in $C_{i,j}$. Hence $\mathrm{TV}^h(u^h) \leq 1$ for all h implies that $\mathrm{TV}(u^h)$ (and hence $\|u^h\|_{\mathrm{L}^2}$) is uniformly bounded in h. There hence exists a (not relabeled) subsequence that converges strongly in $\mathrm{L}^1_{loc}(\mathbb{R}^2)$ and weakly in $\mathrm{L}^2(\mathbb{R}^2)$ to a function u, with moreover $\mathrm{D}u^h \overset{*}{\rightharpoonup} \mathrm{D}u$.

Let us now take $\phi = (\phi^{(1)}, \phi^{(2)}) \in C_c^\infty(\mathbb{R}^2, \mathbb{R}^2)$ such that $||\phi||_\infty \leq 1$. The weak-* convergence of the gradients give us that

$$\begin{aligned}\int_{\mathbb{R}^2} \phi \cdot dDu &= \lim_{h\to 0} \int_{\mathbb{R}^2} \phi \cdot dDu^h \\ &= \lim_{h\to 0} \sum_{i=0}^{N}\sum_{j=0}^{N} \left(\textstyle\int_{C_{i,j}^h \cap C_{i+1,j}^h} \phi^{(1)}\, d\mathcal{H}^1 \int_{C_{i,j}^h \cap C_{i,j+1}^h} \phi^{(2)}\, d\mathcal{H}^1 \right) \cdot \nabla^h u_{i,j}^h \,.\end{aligned}$$

One can moreover show there exists $C > 0$ such that for h small enough and all (i, j) we have:

$$\begin{aligned}\left| \left[\int_{C_{i,j}^h \cap C_{i+1,j}^h} \phi^{(1)}\, d\mathcal{H}^1 \right] - h\, \phi^{(1)}(x_{i+1,j+1}^h) \right| &\leq Ch^2 \,, \\ \left| \left[\int_{C_{i,j}^h \cap C_{i,j+1}^h} \phi^{(2)}\, d\mathcal{H}^1 \right] - h\, \phi^{(2)}(x_{i+1,j+1}^h) \right| &\leq Ch^2 \,,\end{aligned}$$

with $x_{i,j} \overset{\text{def.}}{=} (-R + i\,h, -R + j\,h)$. We use the above inequalities and the fact $\|\phi(x)\| \leq 1$ for all x to obtain the existence of $C' > 0$ such that for h small enough and for all (i, j) we have:

$$\left\| \begin{pmatrix} \int_{C_{i,j}^h \cap C_{i+1,j}^h} \phi^{(1)}\, d\mathcal{H}^1 \\ \int_{C_{i,j}^h \cap C_{i,j+1}^h} \phi^{(2)}\, d\mathcal{H}^1 \end{pmatrix} \right\|_2 \leq h\,\sqrt{1 + C'h}\,.$$

This finally yields

$$\begin{aligned}\sum_{i=0}^{N}\sum_{j=0}^{N} \begin{pmatrix} \int_{C_{i,j}^h \cap C_{i+1,j}^h} \phi^{(1)}\, d\mathcal{H}^1 \\ \int_{C_{i,j}^h \cap C_{i,j+1}^h} \phi^{(2)}\, d\mathcal{H}^1 \end{pmatrix} \cdot \nabla^h u_{i,j}^h &\leq \sum_{i=0}^{N}\sum_{j=0}^{N} h\,\sqrt{1 + C'h}\, \|\nabla^h u_{i,j}^h\| \\ &= \sqrt{1 + C'h}\, \mathrm{TV}^h(u^h)\,,\end{aligned}$$

which gives

$$\int_{\mathbb{R}^2} \phi \cdot dDu \leq \limsup_{h\to 0} \sqrt{1 + C'h}\, \mathrm{TV}^h(u^h) \leq 1\,.$$

We now have to show that

$$\forall v \in \mathrm{L}^2(\mathbb{R}^2),\ \mathrm{TV}(v) \leq 1 \implies \int_{\mathbb{R}^2} \eta\, u \leq \int_{\mathbb{R}^2} \eta\, v\,.$$

Let $v \in C^\infty([-R, R]^2)$ be such that $\mathrm{TV}(v) \leq 1$. We define

$$v^h \overset{\text{def.}}{=} \left(v\left(\left(i + \frac{1}{2} \right) h,\ \left(j + \frac{1}{2} \right) h \right) \right)_{(i,j)\in[1,N]^2}\,.$$

One can then show that

$$\lim_{h\to 0} \mathrm{TV}^h(v^h) = \mathrm{TV}(v) = 1\,,$$

so that for every $\delta > 0$ we have $\mathrm{TV}^h\left(\frac{v^h}{1+\delta}\right) \leq 1$ for h small enough. Now this yields

$$\int_{[-R,R]^2} \eta\, u = \lim_{h\to 0} \int_{[-R,R]^2} \eta\, u^h \leq \lim_{h\to 0} \int_{[-R,R]^2} \eta\, \frac{v^h}{1+\delta} = \int_{[-R,R]^2} \eta\, \frac{v}{1+\delta}\,.$$

Since this holds for all $\delta > 0$ we get that

$$\int_{[-R,R]^2} \eta\, u \leq \int_{[-R,R]^2} \eta\, v\,. \tag{48}$$

Finally, if $v \in \mathrm{L}^2(\mathbb{R}^2)$ is such that $v = 0$ outside $[-R, R]^2$ and $\mathrm{TV}(v) \leq 1$, by standard approximation results (see [Ambrosio et al., 2000, remark 3.22]) we also have that (48) holds, and hence u solves (42). Finally, since u solves (42), its support is included in $[-R, R]^2$, which shows the strong $\mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^2)$ convergence of (u^h) towards u^* in fact implies its strong $\mathrm{L}^1(\mathbb{R}^2)$ convergence. ■

Extraction of a level set. Since we are interested in finding a simple set E that approximately solves (41), and now have a good way of approximating solutions of (42), we make use of the following result, which is a direct consequence of Proposition 1.3.

Proposition 3.14

Let u be a solution of (42). Then the level sets of u are such that for all $t \in \mathbb{R}^*$ with $|U^{(t)}| > 0$, the set $U^{(t)}$ solves (41).

If we have v^h converging strongly in $\mathrm{L}^1(\mathbb{R}^2)$ to a solution v^* of (42), then for almost every $t \in \mathbb{R}$ we have that

$$\lim_{h \to 0} \left| V_h^{(t)} \triangle V_*^{(t)} \right| = 0 \,.$$

The above results hence show we can construct a sequence of sets $(E_k)_{k \geq 0}$ such that $|E_k \triangle E_*|$ converges to 0, with E_* a solution of (41). However, this convergence only implies that

$$\limsup_{k \to \infty} J(E_k) \leq J(E_*) \,,$$

and given that E_k is a union of squares this inequality is likely to be strict, with the perimeter of E_k not converging to the perimeter of E_*. From the last paragraph of Section 1.3.1, we know we have to design a numerical method that allows to find a set at which the value of J is arbitrarily close to $J(E_*)$. This hence motivates the introduction of the refinement step described in the next subsection.

Obtaining a simple polygon. As a final remark, we note that, even for k large enough, E_k could be non-simple. However, using the notations of (59) and **??**, since for every set of finite perimeter E, we have that $J(E)$ is a convex combination of the

$$\left(J(\mathrm{int}(\gamma_i^+))\right)_{i \in I} \,, \; \left(J(\mathrm{int}(\gamma_{i,j}^-))\right)_{i \in I, j \in J_i} \,,$$

there is a simple set F in the decomposition of E such that $J(F) \geq J(E)$. In practice, such a set can be found by extracting all the contours of the binary image $\mathbf{1}_E$, and finding the one with highest objective value. This procedure guarantees that the output of the fixed grid step is a simple polygon. We stress that in all our experiments, v^h is close to being (proportional to) the indicator of a simple set for h large enough, so that its non-trivial level sets are all simple.

2.4. Optimization of the vertices

As explained at the beginning of Section 2, our approach for approximating polygonal Cheeger sets essentially consists in optimizing the vertices of an initial polygon with a first order method. This is in spirit very similar to so-called shape gradient algorithms (see e.g. [Allaire et al., 2021]).

We iteratively construct a sequence of polygons by finding at each step a displacement of steepest ascent for J, along which the vertices of the previous polygon are moved. This displacement is found by exploiting Lemmas 3.15 and 3.16 below.

Fixed sign assumption. Given E_{x^0} the polygon produced by the fixed grid initization, we only aim at *locally* optimizing the objective J around x_0. We can therefore assume the sign of the weighted area term A remains constant and equal to $\epsilon \overset{\text{def.}}{=} \operatorname{sign}(A(x^0))$ during the optimization. In the following, in order to simplify the exposition, we only consider the case $\epsilon = 1$, and consequently seek to maximize

$$J(x) = \frac{A(x)}{P(x)}\,.$$

Lemma 3.15

If $x \in \mathcal{X}_n$ we have:

$$P(E_{x+h}) = P(E_x) - \sum_{i=1}^{n} \left\langle h_i, \tau_i^+ + \tau_i^- \right\rangle + o\left(\|h\|\right), \tag{49}$$

where $\tau_i^+ \overset{\text{def.}}{=} \frac{x_{i+1}-x_i}{\|x_{i+1}-x_i\|}$ and $\tau_i^- \overset{\text{def.}}{=} \frac{x_{i-1}-x_i}{\|x_{i-1}-x_i\|}$.

Proof : If $\|h\|$ is small enough we have:

$$\begin{aligned}
P(E_{x+h}) &= \sum_{i=1}^{n} \|x_{i+1} - x_i + h_{i+1} - h_i\| \\
&= \sum_{i=1}^{n} \sqrt{\|x_{i+1} - x_i + h_{i+1} - h_i\|^2} \\
&= \sum_{i=1}^{n} \|x_{i+1} - x_i\| \left(1 + \frac{\langle x_{i+1} - x_i, h_{i+1} - h_i\rangle}{\|x_{i+1} - x_i\|^2} + o\left(\|h\|\right)\right) \\
&= P(E_x) + \sum_{i=1}^{n} \langle \tau_i^+, h_{i+1} - h_i\rangle + o\left(\|h\|\right),
\end{aligned}$$

and the result follows by re-arranging the terms in the sum. ■

Lemma 3.16

Let $x \in \mathcal{X}_n$ and $\eta \in \mathrm{C}^0(\mathbb{R}^2)$. Then we have

$$\int_{E_{x+h}} \eta = \int_{E_x} \eta + \sum_{i=1}^{n} \left\langle h_i, w_i^- \nu_i^- + w_i^+ \nu_i^+ \right\rangle + o\left(\|h\|\right) \tag{50}$$

where ν_i^+, ν_i^- are respectively the outward unit normal to E_x on $[x_i, x_{i+1}]$, $[x_{i-1}, x_i]$ and

$$w_i^+ \overset{\text{def.}}{=} \|x_{i+1} - x_i\| \int_0^1 \eta((1-t)x_i + t x_{i+1})(1-t)dt\,,$$
$$w_i^- \overset{\text{def.}}{=} \|x_{i-1} - x_i\| \int_0^1 \eta((1-t)x_i + t x_{i-1})(1-t)dt\,.$$

Proof : Our proof relies on the following identity (see Lemma 3.11 for a proof of a closely related formula):

$$\int_{E_x} \eta = \operatorname{sign}\left(\sum_{i=1}^{n} \det(x_i\ x_{i+1})\right) \sum_{i=1}^{n} \omega(x_i, x_{i+1})\,,$$

where $\omega(a_1, a_2) \overset{\text{def.}}{=} \det(a_1\ a_2) \int_{T_1} \eta((a_1\ a_2)\, y)\, dy$ and $T_1 \overset{\text{def.}}{=} \left\{(\alpha, \beta) \in (\mathbb{R}_+)^2 \;\middle|\; \alpha + \beta \leq 1\right\}$ is the unit triangle. Assuming $\eta \in \mathrm{C}^1(\mathbb{R}^2)$ and denoting $\operatorname{adj}(A)$ the adjugate of a matrix A we get:

$$\begin{aligned}
\omega(a_1 + h_1, a_2 + h_2) = \quad & \omega(a_1, a_2) + \det(a_1\ a_2) \int_{T_1} \nabla\eta((a_1\ a_2)\, y) \cdot ((h_1\ h_2) y)\, dy \\
& + \operatorname{tr}\left(\operatorname{adj}(a_1\ a_2)^T (h_1\ h_2)\right) \int_{T_1} \eta((a_1\ a_2)\, y)\, dy + o(\|h\|) \\
= \quad & \omega(a_1, a_2) \\
& + \operatorname{sign}(\det(a_1\ a_2)) \int_{O a_1 a_2} \nabla\eta(y) \cdot ((h_1\ h_2)\, (a_1\ a_2)^{-1})\, y)\, dy \\
& + \frac{\operatorname{tr}\left(\operatorname{adj}(a_1\ a_2)^T (h_1\ h_2)\right)}{|\det(a_1\ a_2)|} \int_{O a_1 a_2} \eta(y)\, dy + o(\|h\|)\,.
\end{aligned}$$

Denoting $g(y) \overset{\text{def.}}{=} (h_1\ h_2)(a_1, a_2)^{-1}\, y$, we obtain:

$$\begin{aligned}
\omega(a_1 + h_1, a_2 + h_2) &= \omega(a_1, a_2) + \operatorname{sign}(\det(a_1\ a_2)) \int_{O a_1 a_2} [\nabla\eta \cdot g + \eta\, \mathrm{div} g] + o(\|h\|) \\
&= \omega(a_1, a_2) + \operatorname{sign}(\det(a_1\ a_2)) \int_{\partial(O a_1 a_2)} \eta\, (g \cdot \nu_{O a_1 a_2})\, d\mathcal{H}^1 + o(\|h\|)\,,
\end{aligned}$$

where we used Gauss-Green theorem to obtain the last equality. Now if $\|h\|$ is small enough then

$$\sum_{i=1}^{n} \det(x_i + h_i\ \ x_{i+1} + h_{i+1}) \text{ and } \sum_{i=1}^{n} \det(x_i\ x_{i+1})$$

have the same sign, so that, defining $g_i : y \mapsto ((h_i\ h_{i+1})(x_i\ x_{i+1})^{-1}\, y)$ we get

$$d\left(\int_{E_\bullet} \eta\right)(x)\,.\,h = \epsilon \sum_{i=1}^{n} \operatorname{sign}(\det(x_i\ x_{i+1}))\, \omega_i\,,$$

with

$$\epsilon \overset{\text{def.}}{=} \operatorname{sign}\left(\sum_{i=1}^{n} \det(x_i\ x_{i+1})\right) \text{ and } \omega_i \overset{\text{def.}}{=} \int_{\partial^*(Ox_ix_{i+1})} \eta\,(g_i \cdot \nu_{Ox_ix_{i+1}})\,d\mathcal{H}^1\,.$$

Then one can decompose each integral in the sum and show the integrals over $[0, x_i]$ cancel out each other, which allows to obtain

$$d\left(\int_{E_\bullet} \eta\right)(x)\,.\,h = \sum_{i=1}^{n} \int_{[x_i,x_{i+1}]} \eta\,(g_i \cdot \nu_i)\,d\mathcal{H}^1\,.$$

But now if $y \in [x_i, x_{i+1}]$ then

$$(x_i\ x_{i+1})^{-1}y = \frac{1}{\|x_{i+1} - x_i\|}\begin{pmatrix}\|y - x_{i+1}\| \\ \|y - x_i\|\end{pmatrix},$$

and the result follows by re-arranging the terms in the sum. One can then use an approximation argument as in [Maggi, 2012, Proposition 17.8] to show it also holds when η is only continuous. ■

Remark 3.17

The results given in Lemmas 3.15 and 3.16 are in fact closely linked to the notion of shape derivative[1]*. If E_x is a simple polygon with n vertices, there exists a constant $C > 0$ such that, given any displacement h of the vertices of E_x, one can extend h to $\theta \in \mathrm{W}^{1,\infty}(\mathbb{R}^2, \mathbb{R}^2)$ with $\|\theta\|_{1,\infty} \leq C\|h\|$ and such that $E_{x+h} = (Id + \theta)(E_x)$ and, for every $i \in \{1, ..., n\}$*

$$\forall t \in [0,1],\ \theta((1-t)x_i + tx_{i+1}) = (1-t)\theta(x_i) + t\theta(x_{i+1}) = (1-t)h_i + th_{i+1}\,. \quad (51)$$

The expressions of the shape derivatives of the perimeter and area (see [Henrot and Pierre, 2018, Chapter 5] for more details, including precise assumptions on the regularity of E and η) yield

$$P((Id+\theta)(E)) = P(E) + \int_{\partial^* E} \mathrm{H}_E \cdot \theta\,d\mathcal{H}^1 + o(\|\theta\|_{1,\infty})\,,$$
$$\int_{(Id+\theta)(E)} \eta = \int_E \eta + \int_{\partial^* E} \eta\,\theta \cdot \nu_E\,d\mathcal{H}^1 + o(\|\theta\|_{1,\infty})\,,$$

where H_E denotes the curvature vector of E. To recover the formulas given in the above lemmas, one simply needs to use (51) and the fact

$$\mathrm{H}_{E_x} = -\sum_{i=1}^{n}(\tau_i^+ + \tau_i^-)\delta_{x_i} \text{ with } \tau_i^+ \overset{\text{def.}}{=} \frac{x_{i+1} - x_i}{\|x_{i+1} - x_i\|} \text{ and } \tau_i^- \overset{\text{def.}}{=} \frac{x_{i-1} - x_i}{\|x_{i-1} - x_i\|}\,.$$

Considering Lemmas 3.15 and 3.16, we construct our sequence of polygons as follows:

[1] We already introduced this tool in Part 2, where we only considered normal deformations (in contrast with the general deformations we use here).

given $x^t \in \mathcal{X}_n$ and a step size α^t, we define the next iterate by $x^{t+1} = x^t + \alpha^t \theta^t$ with

$$\begin{aligned} \theta_j^t &\overset{\text{def.}}{=} \frac{1}{P(E_{x^t})} \left(\theta_{\text{area},j}^t - \frac{\int_{E_{x^t}} \eta}{P(E_{x^t})} \theta_{\text{per},j}^t \right), \\ \theta_{\text{area},j}^t &\overset{\text{def.}}{=} w_j^{t-} \nu_j^{t-} + w_j^{t+} \nu_j^{t+}, \\ \theta_{\text{per},j}^t &\overset{\text{def.}}{=} -(\tau_j^{t+} + \tau_j^{t-}), \end{aligned} \tag{52}$$

where, for all j, ν_j^{t+} and ν_j^{t-} are the outward unit normals of the two sides stemming from x_j^t, and

$$\begin{aligned} \tau_j^{t+} &\overset{\text{def.}}{=} \frac{x_{i+1} - x_i}{\|x_{i+1} - x_i\|}, & w_j^{t+} &\overset{\text{def.}}{=} \|x_{i+1} - x_i\| \int_0^1 \eta((1-t)x_i + t x_{i+1})(1-t)dt, \\ \tau_j^{t-} &\overset{\text{def.}}{=} \frac{x_i - x_{i-1}}{\|x_i - x_{i-1}\|}, & w_j^{t+} &\overset{\text{def.}}{=} \|x_{i-1} - x_i\| \int_0^1 \eta((1-t)x_i + t x_{i-1})(1-t)dt. \end{aligned}$$

Lemmas 3.15 and 3.16 in fact show that the displacement θ^t we apply to the vertices of E_{x^t} is such that

$$\lim_{\alpha \to 0^+} \frac{J(E_{x^t + \alpha\theta}) - J(E_{x^t})}{\alpha} = \langle \theta^t, \theta \rangle, \tag{53}$$

i.e. that it is the displacement of steepest ascent for J at E_{x^t}.

Numerical computation of θ^t. To compute the integral of η over E_{x^t}, we integrate η on each triangle of a sufficiently fine triangulation of E_{x^t} (this triangulation must be updated at each iteration, and sometimes re-computed from scratch to avoid the presence of ill-shaped triangles). The integral of η on a triangle and w_j^{t+}, w_j^{t-} are computed using standard numerical integration schemes for triangles and line segments. If $|\mathcal{T}|$ denotes the number of triangles in the triangulation of E_{x^t}, $|\mathcal{S}_T|$ (resp. $|\mathcal{S}_L|$) the number of points used in the numerical integration scheme for triangles (resp. line segments), the complexity of each iteration is of order $\mathcal{O}\left(m\left(|\mathcal{T}|\,|\mathcal{S}_T| + n\,|\mathcal{S}_L|\right)\right)$.

Comments. Two potential concerns about the above procedure are whether the iterates remain simple polygons (i.e. $x^t \in \mathcal{X}_n$ for all t) and whether they converge to a global maximizer of J over $\mathcal{P}_n$. We could not prove that the iterates remain simple polygons along the process, but since the initial polygon can be taken arbitrarily close to a simple set solving (41) (in terms of the Lebesgue measure of the symmetric difference), we do not expect nor observe in practice any change of topology during the optimization. Moreover, even if J could have non-optimal critical points[1], the above initialization allows us to start our local descent with a polygon that hopefully lies in the basin of attraction of a global maximizer.

Critical points. If the sequence of polygons defined above converges to a simple polygon E_x, then E_x is such that

$$\forall j \in \{1, ..., n\},\ w_j^+ \nu_j^+ + w_j^- \nu_j^- = \frac{\int_{E_x} \eta}{P(E_x)} (-\tau_i^+ - \tau_i^-). \tag{54}$$

[1]Here critical point is to be understood in the sense that the limit appearing in (53) is equal to zero for every θ.

This can be seen as a discrete version of the following first order optimality condition for solutions of (41):

$$\eta = \frac{\int_E \eta}{P(E)} H_E \text{ on } \partial^* E \,. \tag{55}$$

Note that (55) is similar to the optimality condition for the classical Cheeger problem (i.e. with $\eta = 1$ and the additional constraint $E \subseteq \Omega$), namely $H_E = P(E)/|E|$ in the free boundary of E (see [Parini, 2011, Proposition 2.4]).

2.5. The case of radial weight functions

In this section, we study how generalized Cheeger sets and their polygonal approximations compare in a particular setting, where the weight function η is radial. To be more precise, we assume that there exists a strictly decreasing nonnegative function $\tilde{\eta} : [0, +\infty) \to \mathbb{R}_+$ such that $\eta(x) = \tilde{\eta}(\|x\|)$ for almost every $x \in \mathbb{R}^2$[1].

We first prove that, in this setting, (generalized) Cheeger sets associated to η are all disks centered at the origin. Then, we prove that the maximizers of J over $\mathcal{P}_n$ are regular and inscribed in a circle centered at the origin when $n \in \{3, 4\}$, and conjecture that this result holds for every $n \geq 3$.

To enforce the uniqueness of the generalized Cheeger set, we may also invoke the following assumption, which is for example satisfied by $\tilde{\eta} : r \mapsto \exp(-r^2/(2\sigma^2))$ for all $\sigma > 0$.

Assumption 1. *The function $\tilde{\eta}$ is C^1. Moreover, defining $f : r \mapsto r\,\tilde{\eta}(r)$, we have $f(r) \xrightarrow{r \to +\infty} 0$. Finally, there exists $r_0 > 0$ such that $f'(r) > 0$ for all $r < r_0$ and $f'(r) < 0$ for all $r > r_0$.*

2.5.1. Description of generalized Cheeger sets

To prove the above-mentioned result, we rely on Steiner symmetrization. Let us briefly recall its definition and main property. If E is a set of finite perimeter with finite measure, $\nu \in \mathbb{S}^1$ and $z \in \mathbb{R}$, we denote

$$E_{\nu,z} \overset{\text{def.}}{=} \{t \in \mathbb{R} \,|\, z\,\nu + t\,\nu^\perp \in E\} \,.$$

The Steiner symmetrization of E with respect to the line through the origin and directed by ν, denoted E^s_ν, is then defined by

$$E^s_\nu \overset{\text{def.}}{=} \{x \in \mathbb{R}^2 \,|\, |\langle x, \nu^\perp \rangle| \leq |E_{\nu,\langle x,\nu\rangle}|/2\} \,.$$

The fundamental property of Steiner symmetrization is that it preserves volume and does not increase perimeter (see [Maggi, 2012, section 14.1] for more details).

To prove our result, we first prove two useful lemmas.

[1] In particular, this implies that η is essentially bounded.

Lemma 3.18

Let $f : \mathbb{R} \to \mathbb{R}_+$ be even and strictly decreasing on $\mathbb{R}_+$. Then for every measurable set A such that $|A| < +\infty$ we have:

$$\int_A f \leq \int_{A^s} f\,,$$

where $A^s \overset{\text{def.}}{=} (-|A|/2, |A|/2)$. Moreover, equality holds if and only if $|A \triangle A^s| = 0$.

Proof : We have

$$\int_A f = \int_0^{+\infty} |\{f\,\mathbf{1}_A \geq t\}|\,dt = \int_0^{+\infty} |\{f \geq t\} \cap A|\,dt\,.$$

Since f is even, for all $t > 0$, there exists α such that $\{f \geq t\} = [-\alpha, \alpha]$, so that we have

$$\begin{aligned} |\{f \geq t\} \cap A| &= |[-\alpha, \alpha] \cap A| \\ &\leq \min(2\alpha, |A|) \\ &= |[-\alpha, \alpha] \cap [-|A|/2, |A|/2]| \\ &= |\{f \geq t\} \cap A^s|\,. \end{aligned}$$

Hence

$$\int_A f \leq \int_0^{+\infty} \left|\{f \geq t\} \cap A^s\right| dt = \int_{A^s} f\,.$$

Now if $|A \triangle A^s| > 0$, since $|A| = |A^s|$ then $|A \setminus A^s| = |A^s \setminus A| > 0$ and we have

$$\begin{aligned} \int_{A^s} f &= \int_{A \cap A^s} f + \int_{A^s \setminus A} f \\ &> \int_{A \cap A^s} f + f(|A|/2)\,|A^s \setminus A| \\ &\geq \int_{A \cap A^s} f + \int_{A \setminus A^s} f = \int_A f\,, \end{aligned}$$

which proves the second part of the result. ∎

Lemma 3.19

Let $E \subset \mathbb{R}^2$ be a set of finite perimeter with $0 < |E| < +\infty$. Then for any $\nu \in \mathbb{S}^1$ we have

$$\frac{\int_{E_\nu^s} \eta}{P(E_\nu^s)} \geq \frac{\int_E \eta}{P(E)}\,,$$

with equality if and only if $|E \triangle E_\nu^s| = 0$.

Proof : From [Maggi, 2012, theorem 14.4] we know that we have $P(E_\nu^s) \leq P(E)$. We now perform a change of coordinates in order to have $E_\nu^s = \{(x_1, x_2) \in \mathbb{R}^2 \,|\, |x_2| \leq |E_{x_1}|/2\}$ with

$$E_{x_1} \overset{\text{def.}}{=} \{x_2 \in \mathbb{R} \,|\, (x_1, x_2) \in E\}\,.$$

Now, we have

$$\begin{aligned} \int_E \eta &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} \eta(x_1, x_2)\,\mathbf{1}_E(x_1, x_2)\,dx_2 \right) dx_1 \\ &= \int_{-\infty}^{+\infty} \left(\int_{E_{x_1}} \eta(x_1, \cdot) \right) dx_1\,, \end{aligned}$$

with $E_{x_1} = \{x_2 \in \mathbb{R} \,|\, (x_1, x_2) \in E\}$. For almost every $x_1 \in \mathbb{R}$ we have that E_{x_1} is measurable, has finite measure, and that $\eta(x_1, \cdot)$ is nonnegative, even and strictly decreasing on $\mathbb{R}_+$. We can hence apply Lemma 3.18 and get that

$$\int_E \eta \geq \int_{-\infty}^{+\infty} \left(\int_{(E_{x_1})^s} \eta(x_1, \cdot) \right) d_{x_1} = \int_{E_\nu^s} \eta \,. \tag{56}$$

Moreover, if $|E \triangle E_\nu^s| > 0$, then since

$$\begin{aligned} |E \triangle E_\nu^s| &= \int_0^{+\infty} \left(\int_0^{+\infty} |\mathbf{1}_E(x_1, x_2) - \mathbf{1}_{E_\nu^s}(x_1, x_2)| \, dx_2 \right) dx_1 \\ &= \int_0^{+\infty} \left(\int_0^{+\infty} \left| \mathbf{1}_{E_{x_1}}(x_2) - \mathbf{1}_{(E_{x_1})^s}(x_2) \right| \, dx_2 \right) dx_1 \\ &= \int_0^{+\infty} \left| E_{x_1} \triangle \left(E_{x_1}^s \right) \right| \, dx_1 \,, \end{aligned}$$

we get that $\mathcal{L}^1 \left(\{x_1 \in \mathbb{R} \,|\, |E_{x_1} \triangle (E_{x_1})^s| > 0\} \right) > 0$ and hence that (56) is strict. ■

Using the above lemmas, we may now state the following result.

Proposition 3.20

The generalized Cheeger sets associated to η are disks centered at the origin. Moreover, under Assumption 1, the optimal set is unique and its radius is

$$R^* = \operatorname*{argmax}_{R>0} \frac{1}{R} \int_0^R r\tilde{\eta}(r) \, dr \,.$$

Proof : If $E \subset \mathbb{R}^2$ is a set of finite perimeter such that $0 < |E| < +\infty$ and $\nu \in \mathbb{S}^1$ we have:

$$\frac{\int_E \eta}{P(E)} \leq \frac{\int_{E_\nu^s} \eta}{P(E_\nu^s)} \,,$$

with equality if and only if $|E \triangle E_\nu^s| = 0$. Hence if E^* solves (41), arguing as in [Maggi, 2012, Section 14.2], we get that E^* is a convex set which is invariant by reflection with respect to any line through the origin, and hence that E^* is a disk centered at the origin.

Let us define

$$G(R) \overset{\text{def.}}{=} 2\pi J(B(0,R)) = \frac{1}{R} \int_0^R r\tilde{\eta}(r) \, dr \,.$$

Under Assumption 1 G can be extended to a C^1 function on $\mathbb{R}_+$ with $G(0) = 0$, $G'(0) = \tilde{\eta}(0)/2$ and (denoting $f : r \mapsto r\tilde{\eta}(r)$ as in Assumption 1):

$$\forall R > 0, \; G'(R) = \frac{Rf(R) - \int_0^R f(r) \, dr}{R^2} \,.$$

Now defining $F : R \mapsto Rf(R) - \int_0^R f(r) \, dr$ we see that $F'(R) = Rf'(R)$ for all $R > 0$, so that F and f have the same variations. Using Assumption 1, we hence have that F is strictly increasing on $[0, R_0]$ and strictly decreasing on $[R_0, +\infty)$. Moreover, $F(0) = 0$ and, since $f(R) \xrightarrow{R \to +\infty} 0$, we also get $F(R) < 0$ for R large enough. Therefore, there exists $R^* > 0$ such that F (and hence G') is strictly positive on $(0, R^*)$ and strictly negative on $(R^*, +\infty)$. We hence obtain that R^* is the unique maximizer (and the unique critical point) of G. ■

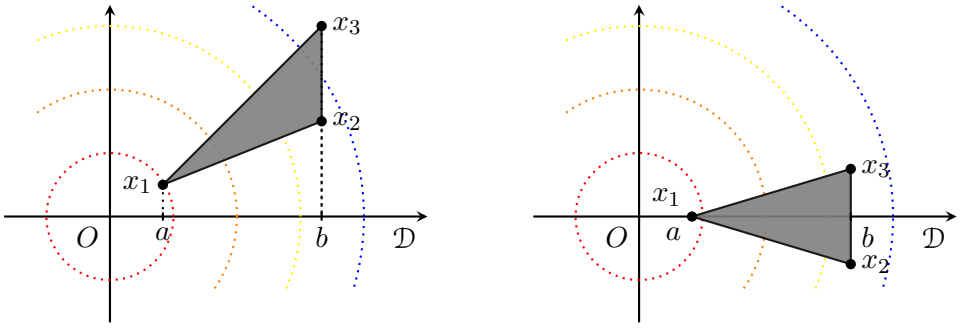


Figure 14 – Steiner symmetrization of triangles

2.5.2. Description of generalized polygonal Cheeger sets

In practice, instead of solving (41), we look for an element of $\mathcal{P}_n$ (a simple polygon with at most n sides) maximizing J, for some given integer $n \geq 3$. It is hence natural to investigate the proximity of these optimal polygons with the "continuous" Cheeger sets. Solving classical geometric variational problems restricted to the set of n-gons is involved, as the Steiner symmetrization procedure might increase the number of sides [Polya and Szegö, 1951, Sec. 7.4]. However, using a trick from Pólya and Szegö, one may prove:

Proposition 3.21

Let $n \in \{3, 4\}$. Then all the maximizers of J over $\mathcal{P}_n$ are regular and inscribed in a circle centered at the origin.

Proof :

Triangles ($n = 3$): let E^* be a maximizer of J among triangles. Then the Steiner symmetrization of E^* with respect to any of its heights through the origin (see Figure 14) is still a triangle, and Lemma 3.19 ensures it has a higher energy except if this operation leaves E^* unchanged. As a consequence, E^* must be symmetric with respect to all its heights through the origin. This shows E^* is equilateral and inscribed in a circle centered at the origin.

Quadrilaterals ($n = 4$): we notice that if E is a simple quadrilateral, then its Steiner symmetrization with respect to any line perpendicular to one of its diagonals (see Figure 15) is still a simple quadrilateral. We can then proceed exactly as for triangles to prove any maximizer E^* of J over $\mathcal{P}_4$ is symmetric with respect to every line through the origin and perpendicular to one of its diagonals. This shows E^* is a rhombus centered at the origin. We can now symmetrize with respect to any line through the origin perpendicular to one of its sides to finally obtain that E^* must be a square centered at the origin. ■

Our proof does not extend to $n \geq 5$, but the following conjecture is natural:

Conjecture 1

The result stated in Proposition 3.21 holds for all $n \geq 3$.

For $n \in \{3, 4\}$ or if Conjecture 1 holds, it remains to compare the radius of optimal polygons with the one of "continuous" Cheeger sets. Let us fix a regular n-gon inscribed in the circle of radius R centered at 0, and denote it $B_n(0, R)$. Abusing notation, we define $J(R) \overset{\text{def.}}{=} J(B(0, R))$ and $J_n(R) \overset{\text{def.}}{=} J(B_n(0, R))$, and can now state the following result.

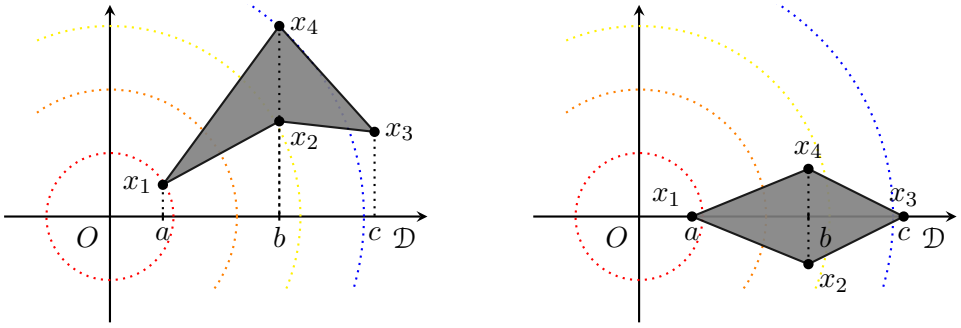


Figure 15 – Steiner symmetrization of quadrilaterals

Proposition 3.22

It holds

$$\|J_n - J\|_\infty = O\left(\frac{1}{n^2}\right).$$

Moreover, under Assumption 1 and assuming f is C^2 with $f''(r_0) < 0$, then for n large enough we have that J_n has a unique maximizer R_n^* and

$$|R_n^* - R^*| = O\left(\frac{1}{n}\right).$$

Proof : Recalling that $P(B_n(0,R)) = 2\pi R \frac{\sin(\pi/n)}{\pi/n}$ we have:

$$\begin{aligned}
|J(R) - J_n(R)| &= \left| \frac{1}{2\pi R} \int_{B(0,R)} \eta - \frac{\pi/n}{2\pi R \sin(\pi/n)} \int_{B_n(0,R)} \eta \right| \\
&= \left| \left(1 - \frac{\pi/n}{\sin(\pi/n)}\right) \frac{1}{2\pi R} \int_{B(0,R)} \eta + \frac{\pi/n}{2\pi R \sin(\pi/n)} \int_{B(0,R)\setminus B_n(0,R)} \eta \right| \\
&\leq \left| 1 - \frac{\pi/n}{\sin(\pi/n)} \right| J(R) + \frac{\pi/n}{2\pi R \sin(\pi/n)} \|\eta\|_\infty |B(0,R) \setminus B_n(0,R)| .
\end{aligned}$$

Now, we define

$$R_n(\theta) \overset{\text{def.}}{=} R \frac{\cos(\pi/n)}{\cos\left((\theta \bmod 2\pi/n) - \pi/n\right)},$$

so that in polar coordinates an equation of $\partial B_n(0,R)$ is given by $r(\theta) = R_n(\theta)$. We obtain:

$$\begin{aligned}
|B(0,R) \setminus B_n(0,R)| &= \int_0^{2\pi} \int_{R_n(\theta)}^{R} r \, dr \, d\theta \\
&\leq 2\pi \sup_{\theta \in [0,2\pi]} \int_{R_n(\theta)}^{R} r \, dr \\
&= \pi \sup_{\theta \in [0,2\pi]} R^2 - R_n^2(\theta) \\
&= \pi(R^2 - R^2 \cos^2(\pi/n)) .
\end{aligned}$$

We hence obtain

$$\|J_n - J\|_\infty \leq \|J\|_\infty \left| 1 - \frac{\pi/n}{\sin(\pi/n)} \right| + R\|\eta\|_\infty \frac{\pi/n}{\sin(\pi/n)} (1 - \cos(\pi/n))$$

and we can finally conclude regarding the first statement.

Now, defining $\alpha_n(s) \overset{\text{def.}}{=} \cos(\pi/n)/\cos(\pi s/n)$, we have:

$$\begin{aligned}
J_n(R) &= \frac{\pi/n}{2\pi R\sin(\pi/n)} \int_0^{2\pi} \int_0^{R_n(\theta)} r\tilde{\eta}(r)\, dr\, d\theta \\
&= \frac{\pi/n}{2\pi R\sin(\pi/n)} n \int_0^{2\pi/n} \int_0^{R\frac{\cos(\pi/n)}{\cos(\theta-\pi/n)}} r\tilde{\eta}(r)\, dr\, d\theta \\
&= \frac{\pi/n}{2\pi R\sin(\pi/n)} 2n \int_0^{\pi/n} \int_0^{R\frac{\cos(\pi/n)}{\cos(\theta)}} r\tilde{\eta}(r)\, dr\, d\theta \\
&= \frac{\pi/n}{R\sin(\pi/n)} \int_0^1 \int_0^{R\alpha_n(s)} r\tilde{\eta}(r)\, dr\, ds \\
&= \frac{\pi/n}{\sin(\pi/n)} \frac{1}{R} \int_0^R r \left[\int_0^1 (\alpha_n(s))^2\, \tilde{\eta}(r\alpha_n(s))\, ds\right] dr\,.
\end{aligned}$$

We can now define $f_n : r \mapsto r\left[\int_0^1 (\alpha_n(s))^2\, \tilde{\eta}(r\alpha_n(s))\, ds\right]$ and proceed as in Proposition 3.20: showing f_n' is strictly positive on $(0, r_1)$ and strictly negative on $(r_1, +\infty)$ for some $r_1 > 0$ is sufficient to prove J_n has a unique maximizer (its unique critical point). Now, we have:

$$\forall r \in \mathbb{R}_+,\ f_n'(r) = \int_0^1 (\alpha_n(s))^2(\tilde{\eta}(r\alpha_n(s)) + r\alpha_n(s)\tilde{\eta}'(r\alpha_n(s)))\, ds\,.$$

The image of $[0,1]$ by $s \mapsto r\alpha_n(s)$ is $[r\cos(\pi/n), r]$. Since $r \mapsto \tilde{\eta}(r) + r\tilde{\eta}'(r) = (r\tilde{\eta})'(r)$ is strictly positive on $(0, r_0)$ and strictly negative on $(r_0, +\infty)$, we get that f_n' is strictly positive on $(0, r_0)$ and strictly negative on $(r_0/\cos(\pi/n), +\infty)$. It hence remains to investigate the sign of f_n' on the interval $[r_0, r_0/\cos(\pi/n)]$. But since f is C^2 and $f''(r_0) < 0$ there exists $\epsilon > 0$ such that $f''(r) < 0$ for every $r \in (r_0 - \epsilon, r_0 + \epsilon)$. For n large enough, we hence have:

$$[r_0\cos(\pi/n), r_0/\cos(\pi/n)] \subset (r_0 - \epsilon, r_0 + \epsilon)\,.$$

This implies that for all $r \in [r_0, r_0/\cos(\pi/n)]$ we have $r\alpha_n(s) \in (r_0 - \epsilon, r_0 + \epsilon)$, and hence that $f_n''(r) < 0$. This finally shows there exists $r_1 > 0$ such that f_n' is strictly positive on $(0, r_1)$ and strictly negative on $(r_1, +\infty)$, and the result follows as in Proposition 3.20.

Now R^* and R_n^* are respectively the unique solutions of $F(0, R) = 0$ and $F(\pi/n, R) = 0$ with

$$\begin{aligned}
F(t, R) &\overset{\text{def.}}{=} \left[\int_0^R f_t(r)\, dr\right] - Rf_t(R)\,, \\
f_t(r) &\overset{\text{def.}}{=} r\int_0^1 \alpha(t,s)^2\, \tilde{\eta}(r\alpha(t,s))ds\,, \\
\alpha(t,s) &\overset{\text{def.}}{=} \frac{\cos(t)}{\cos(ts)}\,.
\end{aligned}$$

Moreover, F is C^1 in a neighborhood of $(0, R^*)$ and $\partial_R F(0, R^*) = -R^* f'(R^*)$. From Assumption 1 we know the unique $R > 0$ such that $f'(R) = 0$ is R_0, which (from the proof of Proposition 3.20) is different from R^*. We can hence apply the implicit function theorem and obtain that

$$|R_n^* - R^*| = \frac{|\partial_t F(0, R^*)|}{|\partial_R F(0, R^*)|} \frac{1}{n} + o\left(\frac{1}{n}\right),$$

which yields the result. ∎

3. Numerical results

Before assessing the practical performance of Algorithm 3, we briefly explain how we implement the sliding step (Line 14).

3.1. Local minimization of the objective

This section is dedicated to the implementation of the sliding step (Line 14 of Algorithm 3). Our approach is exactly the same as in Section 2.4, that is we optimize the vertices of a set of polygons and their corresponding amplitudes using a first order algorithm. To be more precise, we perform a gradient descent on the mapping

$$(a, x) \mapsto G_N(\epsilon, a, E_x)\,, \tag{57}$$

where $E_x = (E_{x_1}, ..., E_{x_N})$. Given a step size α^t, a vector $a^t \in \mathbb{R}^N$ and $x_1^t, ..., x_N^t$ in $\mathcal{X}_n$, we define

$$u^t \overset{\text{def.}}{=} \sum_{i=1}^{N} a_i^t \, \mathbf{1}_{E_{x_i^t}}\,, \quad \eta^t \overset{\text{def.}}{=} -\frac{1}{\lambda}\Phi^*(\Phi u^t - y)\,,$$

and perform the following update:

$$\begin{aligned}
a_i^{t+1} &\overset{\text{def.}}{=} a_i^t - \alpha^t \, h_i^t\,,\\
h_i^t &\overset{\text{def.}}{=} P\big(E_{x_i^t}\big) - \epsilon_i \int_{E_{x_i^t}} \eta^t,\\
x_{i,j}^{t+1} &\overset{\text{def.}}{=} x_{i,j}^t - \alpha^t \, \theta_{i,j}^t\,,\\
\theta_{i,j}^t &\overset{\text{def.}}{=} a_i^t \left[-(\tau_{i,j}^{t+} + \tau_{i,j}^{t-}) + \epsilon_i \, \theta_{\text{data},i,j}^t\right],\\
\theta_{\text{data},i,j}^t &\overset{\text{def.}}{=} \langle \Phi u^t - y,\, w_{i,j}^{t+} \rangle \, \nu_{i,j}^{t+} + \langle \Phi u^t - y,\, w_{i,j}^{t-} \rangle \, \nu_{i,j}^{t-}\,,
\end{aligned}$$

where $\tau_{i,j}^{t\pm}$, $\nu_{i,j}^{t\pm}$ are defined as in Section 2.4 and

$$\begin{aligned}
w_{i,j}^{t+} &\overset{\text{def.}}{=} \|x_{i,j+1}^t - x_{i,j}^t\| \int_0^1 \varphi((1-t)\, x_{i,j}^t + t\, x_{i,j+1}^t)\,(1-t)\, dt\,,\\
w_{i,j}^{t-} &\overset{\text{def.}}{=} \|x_{i,j-1}^t - x_{i,j}^t\| \int_0^1 \varphi((1-t)\, x_{i,j}^t + t\, x_{i,j-1}^t)\,(1-t)\, dt\,.
\end{aligned}$$

Using the notations of Section 2.4, the complexity of each iteration is of order

$$\mathcal{O}\left(N\, m\, (|\mathcal{T}|\, |\mathcal{S}_T| + n\, |\mathcal{S}_L|)\right).$$

Comments. We first stress that the above update is similar to the evolution formally described in (36). Now, unlike the local optimization we perform to approximate Cheeger sets, the sliding step may tend to induce topology changes (see Section 3.3 for an example). This is of course linked to the possible appearance of singularities mentioned in Section 1.4.1. Typically, a simple set may

tend to split in two simple sets over the course of the descent. This is a major difference (and challenge) compared to the sliding steps used in sparse spikes recovery (where the optimization is carried out over the space of Radon measures) [Bredies and Pikkarainen, 2013, Boyd et al., 2017, Denoyelle et al., 2019]. This phenomenon is closely linked to topological properties of the faces of the total (gradient) variation unit ball: its extreme points do not form a closed set for any reasonable topology (e.g. the weak $\mathrm{L}^2(\mathbb{R}^2)$ topology), nor do its faces of dimension $d \leq k$ for any $k \in \mathbb{N}$. As a result, when moving continuously on the set of faces of dimension $d = k$, it is possible to "stumble upon" a point which only belongs to a face of dimension $d > k$.

Handling (or not) topology changes. Our current implementation does not allow to handle these topology changes in a consistent way, and finding a way to deal with them "off-the-grid" is an interesting avenue for future research. It is important to note that not allowing topological changes during the sliding step is not an issue, since all convergence guarantees hold as soon as the output of the sliding step decreases the energy more than the standard update. One can hence stop the local descent at any point before any change of topology occurs, which avoids having to treat them. Still, in order to yield iterates that are as sparse as possible (and probably to decrease the objective as quickly as possible), it seems preferable to allow topological changes.

3.2. Recovery examples

Here, we investigate the practical performance of Algorithm 3. We focus on the case where Φ is a sampled Gaussian convolution operator, i.e.

$$\forall x \in \mathbb{R}^2,\ \varphi(x) = \left(\exp\left(-\frac{||x - x_i||^2}{2\sigma^2} \right) \right)_{i=1}^{m}$$

for a given $\sigma > 0$ and a sampling grid $(x_i)_{i=1}^m$. The noise is drawn from a multivariate Gaussian with zero mean and isotropic covariance matrix $\tau^2\, I_m$. We take λ of the order of $\sqrt{2\log(m)\,\tau^2}$.

Numerically certifying that a given function is an approximate solution of ($\mathcal{P}_\lambda(y)$) is difficult. However, as the sampling grid becomes finer, Φ tends to the convolution with the Gaussian kernel, which is injective. Relying on a Γ-convergence argument, one may expect that if u_0 is a piecewise constant image and w is some small additive noise, the solutions of ($\mathcal{P}_\lambda(y)$) with $y = \Phi u_0 + w$ are all close to u_0, modulo the regularization effects of the total variation.

Comparison with other algorithms. We assess the performance of our algorithm by comparing its output to that of a primal dual algorithm minimizing a discretized version of ($\mathcal{P}_\lambda(y)$) on a pixel grid, where the total variation term is replaced by the discrete isotropic total variation or the dual-based discretization of [Hintermüller et al., 2014, Condat, 2017]. To minimize discretization artifacts, we artificially introduce a downsampling in the forward operator, so that the reconstruction is performed on a grid four times larger than the sampling one.

Experiments. Our first experiment consists in recovering a function u_0 that is a linear combination of three indicator functions (see Figures 16 and 17). During each of the three iterations required to obtain a good approximation of u_0, a new atom is added to its support. One can see the sliding step is crucial: the large atom on the left, added during the second iteration, is

Figure 16 – From left to right: observations, unknown function, output of Algorithm 3, outputs of the fixed grid method using the isotropic and Condat's total variation

significantly refined during the sliding step of the third iteration, when enough atoms have been introduced. The second experiment (see Figure 18) consists in recovering the indicator function of a set with a hole (which can also be seen as the sum of two indicator functions of simple sets). The support of u_0 and its gradient are accurately estimated. Still, the typical effects of total variation regularization are noticeable: corners are slightly rounded, and there is a "loss of contrast" in the eye of the pacman. The third experiment (Figure 19) also showcases the rounding of corners, and highlights the influence of the regularization parameter: as λ decreases, the curvature of the edge set increases. Finally, we provide in Fig. Figure 20 the results of an experiment on a more challenging task, which consists in reconstructing a natural grayscale image.

Choice of parameters. The number of observations in the first experiment is 60×60, 75×75 in the second and third ones, and 64×64 in the last one. In all experiments, we solved (46) on a grid of size 80×80. In both local descent steps (for approximating Cheeger sets and for the sliding step), the simple polygons have a number of vertices of order 30 times the length of their boundary (100 for the last experiment), and the maximum area of triangles in their inner mesh is 10^{-2} (the domain being a square of side 1). The inner triangulation of a simple polygon is obtained by using Richard Shewchuk's Triangle library. The boundary of the polygons are resampled every 30 iterations. Line integrals are computed using the Gauss-Patterson scheme of order 3 (15 points) and integrals on triangles using the Hammer-Marlowe-Stroud scheme of order 5 (7 points).

3.3. Topology changes during the sliding step

Here, we illustrate the changes of topology that may occur during the sliding step (Line 14 of Algorithm 3). All relevant plots are given in Figure 21. The unknown function (see (a)) is the sum of two indicator functions:

$$u_0 = \mathbf{1}_{B((-1,0),0.6)} + \mathbf{1}_{B((1,0),0.6)}\,,$$

and observations are shown in (b). The Cheeger set computed at Line 5 of the first iteration covers the two disks (see (c)).

In this setting, our implementation of the sliding step converges to a function similar to (f)[1], and we obtain a valid update that decreases the objective more than the standard Frank-Wolfe update. The next iteration of the algorithm will then consist in adding a new atom to the

[1]This only occurs when λ is small enough. For higher values of λ, the output is similar to (d) or (e).

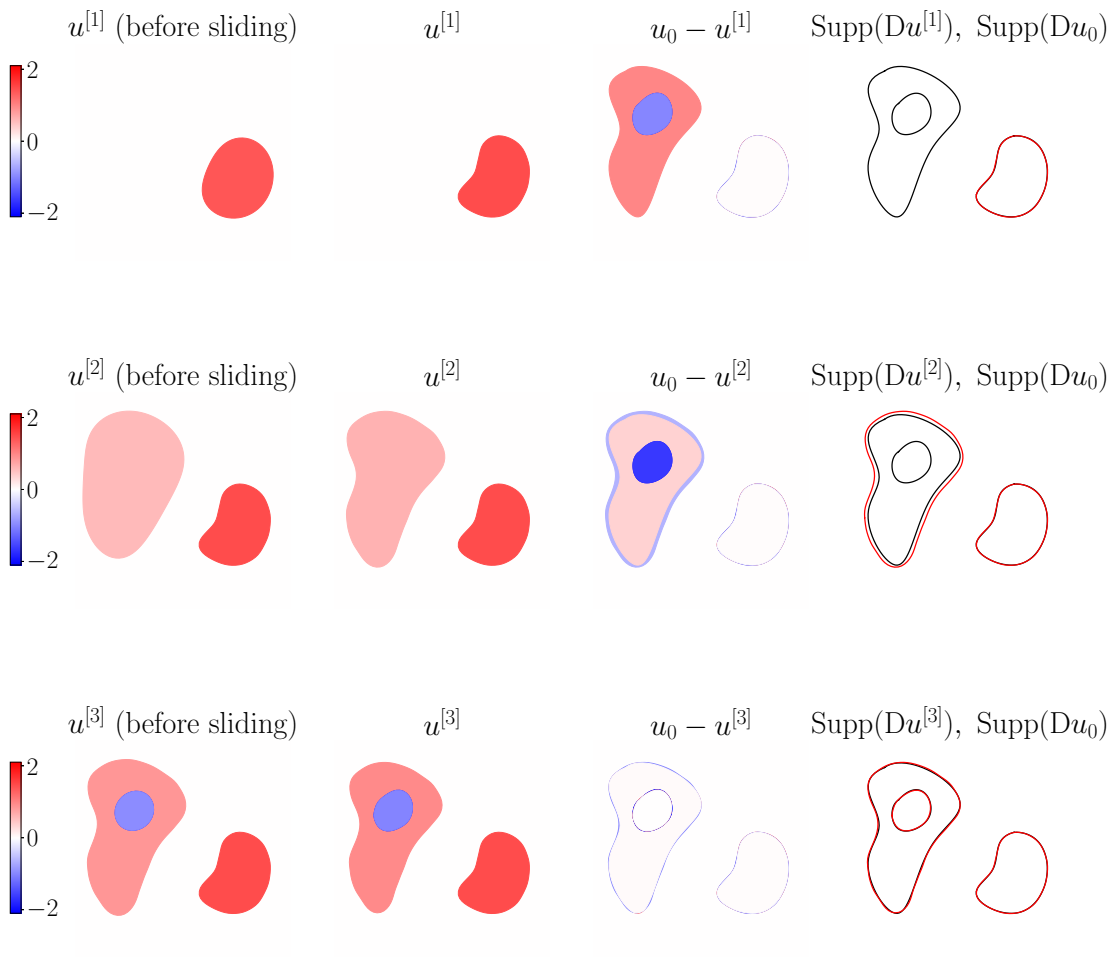


Figure 17 – Unfolding of Algorithm 3 for the first experiment ($u^{[k]}$ denotes the k-th iterate)

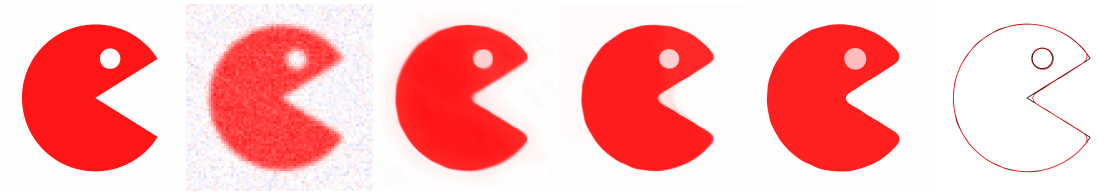


Figure 18 – From left to right: unknown function, observations, outputs of the fixed grid method using the isotropic and Condat's total variation, output of Algorithm 3, gradients support (red: output of Algorithm 3, black: unknown)

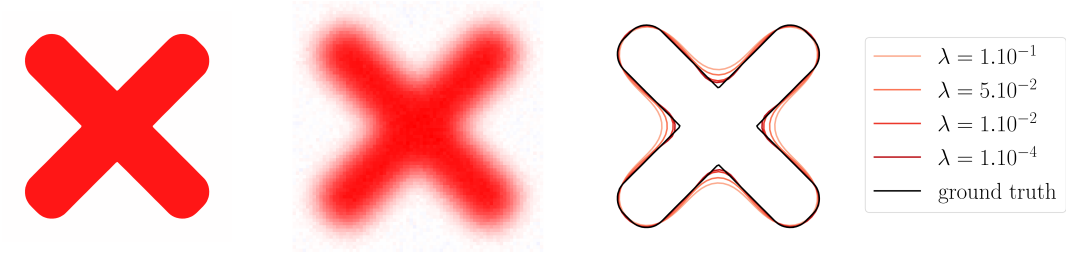


Figure 19 – Left: unknown function, middle: observations, right: output of Algorithm 3 for different values of λ

Figure 20 – From left to right: original image, observations, iterates $u^{[k]}$ $(k = 1, 4)$ produced by Algorithm 3, outputs of the fixed grid method using the isotropic and Condat's total variation

approximation, with negative amplitude, so as to compensate for the presence of the small bottleneck.

However, it seems natural that the support of (f) should split into two disjoint simple sets, which is not possible with our current implementation. To investigate what would happen in this case, we manually split the two sets (see (g)) and let them evolve independently. The support of the approximation converges to the union of the two disks, which produces an update that decreases the objective even more than (f).

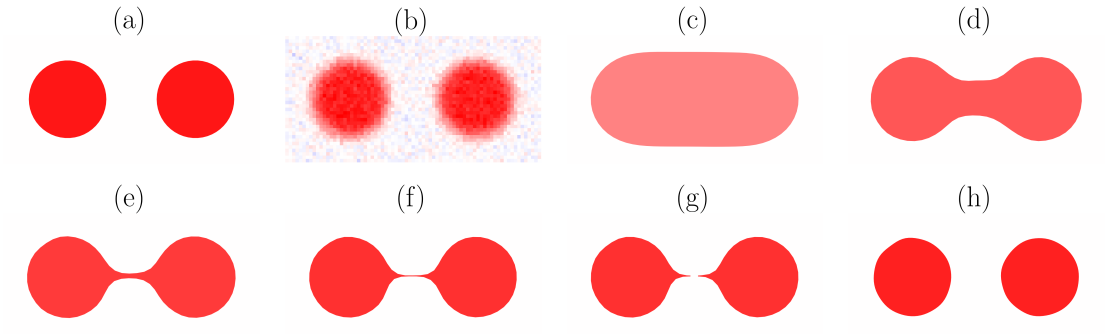


Figure 21 – Topology change experiment. (a): unknown signal, (b): observations, (c): weighted Cheeger set, (d,e,f,g): sliding step iterations (with splitting), (h): final function.

4. Conclusion

In this part, we proposed an iterative algorithm for solving ($\mathcal{P}_\lambda(y)$), which is based on the conditional gradient (or Frank-Wolfe) algorithm. We studied the convergence of its iterates, and presented a natural analog of the "sliding step" introduced in [Denoyelle et al., 2019]. Unlike in the latter work, we left the theoretical analysis of this step open, and in particular could not prove that a so-called critical point of the objective can be reached at every iteration.

We then discussed the practical implementation of this theoretical algorithm, and presented two oracles to carry out its main operations. The first allows to approximate generalized Cheeger sets, and builds up on the method proposed in [Carlier et al., 2009]. It consists in optimizing the vertices of a polygon with a shape gradient-like algorithm, initialized with the output of the latter method. To support our approach, we also proved the existence of generalized Cheeger sets in the class of polygons with a given number of sides, and investigated their proximity with their continuous counterpart in a radial setting. The second oracle we proposed allows to implement the sliding step mentioned above. We discussed the topological changes that can appear during the resulting evolution, and argued that handling them in a consistent way is natural but challenging.

Finally, we assessed the performance of the propose numerical method on a few recovery examples, providing comparisons with fixed-grid methods. Our approach is particularly suited to reconstruct simple images (i.e. which are the superposition of a few simple shapes), whereas on more complex natural images, existing methods perform better.

Part 4

Conclusion

We end this manuscript by presenting natural avenues for future research. We first discuss those related to theoretical recovery guarantees, before turning to numerical methods.

1. Recovery guarantees

Let us start by recalling our contributions to the theoretical analysis of total variation regulatization. In Part 2, we described the structure of the exposed faces of the total variation unit ball, and introduced the class of M-simple functions, which are the sparse objects naturally associated to this regularizer. We also considered solutions of the prescribed curvature problem, and studied their stability under variations of the associated curvature functional. Finally, we introduced a condition under which the jump set of an M-simple function can be exactly recovered in a low noise regime.

Sufficient identifiability conditions. In our opinion, an important question which remains mostly open is that of identifiability. Given a function u_0 and a measurement operator Φ, being able to provide sufficient conditions ensuring u_0 is the unique solution of $(\mathcal{P}_0(y_0))$ is highly desirable. As already mentioned, the only related work we are aware of is [Bredies and Vicente, 2019]. A major difference with the sparse spikes setting is that even the identifiability of a single atom, i.e. of $u_0 = \mathbf{1}_E$ with E simple, seems difficult to study. Having a clear understanding of this problem could, subsequently, allow to answer the following question: if two atoms E, F are separately identifiable and $a, b \in \mathbb{R}^*$, how does the relative position of E and F impact the identifiability of $a\,\mathbf{1}_E + b\,\mathbf{1}_F$?

Pre-certificates. Considering the literature on sparse spikes recovery, the construction of pre-certificates (i.e. good candidates for being a dual certificate, possibly of minimial norm) is highly important for studying both identifiability and noise robustness. We have discussed in Section 3.4.2 possible ways to construct such objects in a simplified setting. This investigation should however be furthered. In particular, the pre-certificate we described only uses the first order optimality condition "in average", and other constraints could hence be considered. To

ensure a notion of pre-certificate is relevant, a natural way would be to find out if, as proved in [Duval and Peyré, 2015] in the sparse spikes setting, the fact it is a true certificate is necessary for (noiseless) exact support recovery.

Error bounds. The noise robustness analysis we presented in Part 2 is not quantitative. It does not provide bounds on the distance between solutions of $(\mathcal{P}_0(y_0))$ and $(\mathcal{P}_\lambda(y_0 + w))$ in terms of the noise level $\|w\|_{\mathcal{H}}$. In the sparse spikes setting, such bounds were derived by analyzing the behaviour of dual certificates in the so-called far and near regions (see for instance [Azaïs et al., 2015]). In our case, it seems that the near region analysis is closely linked to the coercivity constant of the second order shape derivative at optimal sets. What seems more challenging is to quantify how the relevant geometric functional behaves in the far region, i.e. for sets that are not smooth normal deformations of a minimizer. We argue that this could be related to the derivation of a quantitative inequality for our functional (we briefly discussed the case of the isoperimetric inequality in Section 2.1.1). To prove such a result, one would need to adapt the selection principle of [Cicalese and Leonardi, 2012], and then use the analysis provided in Section 2.2 to deal with smooth deformations of a minimizer.

Stability for the total variation flow. We argue that the stability analysis of Sections 2.2 and 3.1 could possibly be used to study the L^2 gradient flow of the total variation (see for instance [Bellettini et al., 2002]). A relevant question concerning this evolution could be: if the flow is initialized with an M-simple function, is the solution also M-simple for sufficiently small times? If so, is the length of the chains preserved? In the particular case where the flow is initialized with the indicator function of a calibrable set, it has been shown in the above reference that the solution remains proportional to the initialization. Answering the more general question formulated above is, to the best of our knowledge, an open problem. Given an initial data $u_0 \in \mathrm{L}^2(\mathbb{R}^2)$, an idea to apply the results of Sections 2.2 and 3.1 would be to consider the minimizing movement scheme

$$\min_{u \in \mathrm{L}^2(\mathbb{R}^2)} \mathrm{TV}(u) + \frac{1}{2\tau}\|u - u_0\|_{\mathrm{L}^2}\,, \tag{58}$$

and to prove a stability result in the spirit of Theorem 2.27 when $\tau \to 0^+$. The main difference with our setting is that, in (58), no advantage can be taken from the regularity of the measurement operator, and whether dual solutions converge smoothly to the minimal norm limit certificate is unclear. Finally, let us mention that, although it is even less clear whether our results could be applied in this setting, one could also consider the same question for the Wasserstein gradient flow of the total variation [Carlier and Poon, 2019].

Model bias. In Part 2, we argued that the class of sparse objects associated to the total variation are M-simple functions. Regularizing an inverse problem with the total variation will hence often produce M-simple reconstructions. In practical applications, the sought-after image might be approximately, but not exactly piecewise constant. An interesting question would hence be to investigate what class of functions can be well approximated with s-sparse M-simple functions, and how this approximation behaves when s grows. This is related to the notion of model bias, i.e. to the error one makes by assuming the unknown signal has a specific structure. Such an analysis has been conducted for sparse vectors (see for instance [Foucart and Rauhut, 2013, Section 2.1]). A

related topic is also the study of nonlinear wavelet approximation (see [Mallat, 2008, Section 9.2]), and more specifically of how the approximation error decreases when the number of coefficients grows.

2. Numerical resolution

In Part 3, we proposed an algorithm for solving ($\mathcal{P}_\lambda(y)$), which is based on the conditional gradient (or Frank Wolfe) algorithm. It iteratively constructs a sequence of iterates that provably converges to a solution of ($\mathcal{P}_\lambda(y)$). Each iterate is a linear combination of the atoms promoted by the total variation (i.e. indicator functions of simple sets). Then, we discussed the implementation of this theoretical algorithm, and proposed to numerically represent the above-mentioned atoms by indicator functions of simple polygons. We finally assessed the performance of our approach on a few recovery examples, providing comparisons with traditional grid-based methods.

Convergence in finite time. In [Denoyelle et al., 2019, Theorem 3], a finite time convergence result is proved for the so-called sliding Frank-Wolfe algorithm. It holds under the assumption that the dual certificate associated to the noisy problem is non-degenerate. A crucial ingredient of the proof is that a critical point of the objective is reached at the end of each sliding step. As explained above, we cannot guarantee this in our setting. Still, assuming this is true, it would be interesting to investigate whether the proof of [Denoyelle et al., 2019] can be adapted to Algorithm 3.

Topology changes. In Section 1.4, we presented the so-called sliding step (in the spirit of [Denoyelle et al., 2019]). We formally described the associated gradient flow, but its study (existence, uniqueness) seems challenging. In Section 3.1 we introduced a possible way to mimic this evolution numerically. We explained that this could lead to the apparition of singularities, and that we could not handle them consistently with our approach. Even if one can circumvent this issue by stopping the evolution before the appearance of a singularity (which does not break any convergence guarantee), it seems preferable to allow them, in order to yield iterates that are as sparse as possible. This topic is related to the more global issue of handling topology changes with purely Lagrangian methods. In the shape optimization community, Eulerian methods (such as the level set method) are preferred to achieve this, as they can treat complicated deformations of a shape in a very robust way [Allaire et al., 2021]. In a recent work [Lévy, 2022], Bruno Lévy proposed a new Lagrangian free-surface mesh representation for fluid simulation. He also mentioned these ideas could possibly be used for shape optimization. It would be interesting to investigate whether these tools can allow for a robust implementation of the sliding step.

Approximation of maximal elements of exposed faces. We mentioned above that the construction of pre-certificates is highly important to derive identifiability and noise robustness guarantees. Given a pre-certificate, one can numerically ensure it is a true certificate by solving the associated Cheeger problem, and showing its value is (approximately) one. If this is the case, it is desirable to find which functions are associated to this certificate, i.e. to obtain a complete description of the face it exposes. One can always find an element of this face by solving the associated Cheeger problem, but how to approximate the complete set of maximizers seems challenging. We argue that a possible way to achieve this would be to adapt the ideas

of [Buttazzo et al., 2007, Carlier et al., 2009]. These works propose to introduce a small strictly convex penalization of the objective, in order to select some "maximal" solution of the Cheeger problem. Their setting however slightly differs from ours (the weight functions they deal with are non-negative), and finding out whether their techniques can be adapted to our case would be interesting.

Practical applications. To conclude this part, let us discuss how the work presented here can be relevant for practical applications. The experiments conducted in Section 3.2 suggest that traditional grid-based methods perform better on complex natural images, while our approach is particularly well-suited for reconstructing simple piecewise constant ones. It is hence natural to wonder in which situations one would be interested in the recovery of such signals, and in obtaining a continuous domain representation of them. We mentioned in Section 1.1.1 that this could potentially be the case in cell imaging (see Figure 2). However, it seems natural that the choice of appropriate measurement operators and the relevance of this approach should extensively be discussed with practitioners.

References

[Acerbi et al., 2013] Acerbi, E., Fusco, N., and Morini, M. (2013). Minimality via Second Variation for a Nonlocal Isoperimetric Problem. *Communications in Mathematical Physics*, 322(2):515–557.

[Alberti et al., 2019] Alberti, G. S., Ammari, H., Romero, F., and Wintz, T. (2019). Dynamic Spike Superresolution and Applications to Ultrafast Ultrasound Imaging. *SIAM Journal on Imaging Sciences*, 12(3):1501–1527.

[Allaire et al., 2021] Allaire, G., Dapogny, C., and Jouve, F. (2021). Chapter 1 - Shape and topology optimization. In Bonito, A. and Nochetto, R. H., editors, *Handbook of Numerical Analysis*, volume 22 of *Geometric Partial Differential Equations - Part II*, pages 1–132. Elsevier.

[Almgren, 1975] Almgren, F. J. (1975). Existence and regularity almost everywhere of solutions to elliptic variational problems with constraints. *Bulletin of the American Mathematical Society*, 81(1):151–154.

[Alter et al., 2005] Alter, F., Caselles, V., and Chambolle, A. (2005). A characterization of convex calibrable sets in Rn. *Mathematische Annalen*, 332(2):329–366.

[Ambrosio, 2010] Ambrosio, L. (2010). *Corso introduttivo alla teoria geometrica della misura e alle superfici minime*. Scuola Normale Superiore.

[Ambrosio et al., 2001] Ambrosio, L., Caselles, V., Masnou, S., and Morel, J.-M. (2001). Connected components of sets of finite perimeter and applications to image processing. *Journal of the European Mathematical Society*, 3(1):39–92.

[Ambrosio et al., 2000] Ambrosio, L., Fusco, N., and Pallara, D. (2000). *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Mathematical Monographs. Oxford University Press, Oxford, New York.

[Azaïs et al., 2015] Azaïs, J.-M., de Castro, Y., and Gamboa, F. (2015). Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 38(2):177–195.

[Bach, 2013] Bach, F. (2013). Learning with Submodular Functions: A Convex Optimization Perspective. *Foundations and Trends® in Machine Learning*, 6(2-3):145–373.

[Bartels et al., 2021] Bartels, S., Tovey, R., and Wassmer, F. (2021). Singular solutions, graded meshes, and adaptivity for total-variation regularized minimization problems.

[Bellettini et al., 2002] Bellettini, G., Caselles, V., and Novaga, M. (2002). The Total Variation Flow in RN. *Journal of Differential Equations*, 184(2):475–525.

[Benning and Burger, 2013] Benning, M. and Burger, M. (2013). Ground states and singular vectors of convex variational regularization methods. *Methods and Applications of Analysis*, 20(4):295–334.

[Betzig et al., 2006] Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J., and Hess, H. F. (2006). Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science*, 313(5793):1642–1645.

[Boulanger et al., 2014] Boulanger, J., Gueudry, C., Münch, D., Cinquin, B., Paul-Gilloteaux, P., Bardin, S., Guérin, C., Senger, F., Blanchoin, L., and Salamero, J. (2014). Fast high-resolution 3D total internal reflection fluorescence microscopy by incidence angle scanning and azimuthal averaging. *Proceedings of the National Academy of Sciences of the United States of America*, 111(48):17164–17169.

[Boyd et al., 2017] Boyd, N., Schiebinger, G., and Recht, B. (2017). The Alternating Descent Conditional Gradient Method for Sparse Inverse Problems. *SIAM Journal on Optimization*, 27(2):616–639.

[Boyer et al., 2019] Boyer, C., Chambolle, A., De Castro, Y., Duval, V., De Gournay, F., and Weiss, P. (2019). On Representer Theorems and Convex Regularization. *SIAM Journal on Optimization*, 29(2):1260–1281.

[Boyer et al., in preparation] Boyer, C., Chambolle, A., De Castro, Y., Duval, V., De Gournay, F., and Weiss, P. (in preparation).

[Brasco et al., 2014] Brasco, L., Lindgren, E., and Parini, E. (2014). The fractional Cheeger problem. *Interfaces and Free Boundaries*, 16(3):419–458.

[Bredies and Carioni, 2019] Bredies, K. and Carioni, M. (2019). Sparsity of solutions for variational inverse problems with finite-dimensional data. *Calculus of Variations and Partial Differential Equations*, 59(1):14.

[Bredies et al., 2022a] Bredies, K., Carioni, M., and Fanzon, S. (2022a). A superposition principle for the inhomogeneous continuity equation with Hellinger–Kantorovich-regular coefficients. *Communications in Partial Differential Equations*, 0(0):1–47.

[Bredies et al., 2021] Bredies, K., Carioni, M., Fanzon, S., and Romero, F. (2021). On the extremal points of the ball of the Benamou–Brenier energy. *Bulletin of the London Mathematical Society*, 53(5):1436–1452.

[Bredies et al., 2022b] Bredies, K., Carioni, M., Fanzon, S., and Romero, F. (2022b). A Generalized Conditional Gradient Method for Dynamic Inverse Problems with Optimal Transport Regularization. *Foundations of Computational Mathematics*.

[Bredies and Fanzon, 2020] Bredies, K. and Fanzon, S. (2020). An optimal transport approach for solving dynamic inverse problems in spaces of measures. *ESAIM: Mathematical Modelling and Numerical Analysis*, 54(6):2351–2382.

[Bredies and Pikkarainen, 2013] Bredies, K. and Pikkarainen, H. K. (2013). Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(1):190–218.

[Bredies and Vicente, 2019] Bredies, K. and Vicente, D. (2019). A perfect reconstruction property for PDE-constrained total-variation minimization with application in Quantitative Susceptibility Mapping. *ESAIM: Control, Optimisation and Calculus of Variations*, 25:83.

[Bucur and Fragalà, 2016] Bucur, D. and Fragalà, I. (2016). A Faber–Krahn Inequality for the Cheeger Constant of N-gons. *The Journal of Geometric Analysis*, 26(1):88–117.

[Burger and Osher, 2004] Burger, M. and Osher, S. (2004). Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411–1421.

[Buttazzo et al., 2007] Buttazzo, G., Carlier, G., and Comte, M. (2007). On the selection of maximal Cheeger sets. *Differential and Integral Equations*, 20(9):991–1004.

[Candes and Tao, 2005] Candes, E. and Tao, T. (2005). Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215.

[Candès and Fernandez-Granda, 2013] Candès, E. J. and Fernandez-Granda, C. (2013). Super-Resolution from Noisy Data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254.

[Candès and Fernandez-Granda, 2014] Candès, E. J. and Fernandez-Granda, C. (2014). Towards a Mathematical Theory of Super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956.

[Candes and Tao, 2006] Candes, E. J. and Tao, T. (2006). Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425.

[Carlier and Comte, 2007] Carlier, G. and Comte, M. (2007). On a weighted total variation minimization problem. *Journal of Functional Analysis*, 250(1):214–226.

[Carlier et al., 2009] Carlier, G., Comte, M., and Peyré, G. (2009). Approximation of maximal Cheeger sets by projection. *ESAIM: Mathematical Modelling and Numerical Analysis*, 43(1):139–150.

[Carlier and Poon, 2019] Carlier, G. and Poon, C. (2019). On the total variation Wasserstein gradient flow and the TV-JKO scheme. *ESAIM: Control, Optimisation and Calculus of Variations*, 25:42.

[Caselles et al., 2009] Caselles, V., Facciolo, G., and Meinhardt, E. (2009). Anisotropic Cheeger Sets and Applications. *SIAM Journal on Imaging Sciences*, 2(4):1211–1254.

[Catala, 2020] Catala, P. (2020). *Positive Semidefinite Relaxations for Imaging Science*. Theses, Ecole Normale Supérieure.

[Catala et al., 2019] Catala, P., Duval, V., and Peyré, G. (2019). A Low-Rank Approach to Off-the-Grid Sparse Superresolution. *SIAM Journal on Imaging Sciences*, 12(3):1464–1500.

[Chambolle et al., 2010a] Chambolle, A., Caselles, V., Cremers, D., Novaga, M., and Pock, T. (2010a). An Introduction to Total Variation for Image Analysis. In *An Introduction to Total Variation for Image Analysis*, pages 263–340. De Gruyter.

[Chambolle et al., 2016] Chambolle, A., Duval, V., Peyré, G., and Poon, C. (2016). Geometric properties of solutions to the total variation denoising problem. *Inverse Problems*, 33(1):015002.

[Chambolle et al., 2010b] Chambolle, A., Giacomini, A., and Lussardi, L. (2010b). Continuous limits of discrete perimeters. *ESAIM: Mathematical Modelling and Numerical Analysis*, 44(2):207–230.

[Chambolle and Pock, 2011] Chambolle, A. and Pock, T. (2011). A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145.

[Chambolle and Pock, 2021a] Chambolle, A. and Pock, T. (2021a). Chapter 6 - Approximating the total variation with finite differences or finite elements. In Bonito, A. and Nochetto, R. H., editors, *Handbook of Numerical Analysis*, volume 22 of *Geometric Partial Differential Equations - Part II*, pages 383–417. Elsevier.

[Chambolle and Pock, 2021b] Chambolle, A. and Pock, T. (2021b). Learning Consistent Discretizations of the Total Variation. *SIAM Journal on Imaging Sciences*, 14(2):778–813.

[Chandrasekaran et al., 2012] Chandrasekaran, V., Recht, B., Parrilo, P. A., and Willsky, A. S. (2012). The Convex Geometry of Linear Inverse Problems. *Foundations of Computational Mathematics*, 12(6):805–849.

[Chen et al., 1998] Chen, S. S., Donoho, D. L., and Saunders, M. A. (1998). Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61.

[Chizat, 2021] Chizat, L. (2021). Convergence Rates of Gradient Methods for Convex Optimization in the Space of Measures.

[Chizat, 2022] Chizat, L. (2022). Sparse optimization on measures with over-parameterized gradient descent. *Mathematical Programming*, 194(1):487–532.

[Cicalese and Leonardi, 2012] Cicalese, M. and Leonardi, G. P. (2012). A Selection Principle for the Sharp Quantitative Isoperimetric Inequality. *Archive for Rational Mechanics and Analysis*, 206(2):617–643.

[Claerbout and Muir, 1973] Claerbout, J. F. and Muir, F. (1973). Robust modeling with erratic data. *GEOPHYSICS*, 38(5):826–844.

[Cohen et al., 2009] Cohen, A., Dahmen, W., and DeVore, R. (2009). Compressed sensing and best k-term approximation. *Journal of the American Mathematical Society*, 22(1):211–231.

[Condat, 2016] Condat, L. (2016). Fast projection onto the simplex and the ℓ_1 ball. *Mathematical Programming*, 158(1):575–585.

[Condat, 2017] Condat, L. (2017). Discrete Total Variation: New Definition and Minimization. *SIAM Journal on Imaging Sciences*, 10(3):1258–1290.

[Dambrine and Lamboley, 2019] Dambrine, M. and Lamboley, J. (2019). Stability in shape optimization with second variation. *Journal of Differential Equations*, 267(5):3009–3045.

[De Castro et al., 2021] De Castro, Y., Duval, V., and Petit, R. (2021). Towards Off-the-grid Algorithms for Total Variation Regularized Inverse Problems. In *Scale Space and Variational Methods in Computer Vision*, Lecture Notes in Computer Science, pages 553–564. Springer.

[De Castro et al., 2022] De Castro, Y., Duval, V., and Petit, R. (2022). Towards Off-the-grid Algorithms for Total Variation Regularized Inverse Problems.

[De Castro et al., in preparation] De Castro, Y., Duval, V., and Petit, R. (in preparation). Exact recovery of the support of piecewise constant images via total variation regularization.

[De Castro and Gamboa, 2012] De Castro, Y. and Gamboa, F. (2012). Exact reconstruction using Beurling minimal extrapolation. *Journal of Mathematical Analysis and Applications*, 395(1):336–354.

[De Castro et al., 2017] De Castro, Y., Gamboa, F., Henrion, D., and Lasserre, J. (2017). Exact Solutions to Super Resolution on Semi-Algebraic Domains in Higher Dimensions. *IEEE Transactions on Information Theory*, 63(1):621–630.

[Delfour and Zolesio, 2011] Delfour, M. C. and Zolesio, J.-P. (2011). *Shapes and Geometries: Metrics, Analysis, Differential Calculus, and Optimization, Second Edition*. SIAM.

[Denoyelle, 2018] Denoyelle, Q. (2018). *Theoretical and Numerical Analysis of Super-Resolution Without Grid*. PhD thesis, Université Paris Dauphine - PSL.

[Denoyelle et al., 2019] Denoyelle, Q., Duval, V., Peyre, G., and Soubies, E. (2019). The Sliding Frank-Wolfe Algorithm and its Application to Super-Resolution Microscopy. *Inverse Problems*.

[Dolzmann and Müller, 1995] Dolzmann, G. and Müller, S. (1995). Microstructures with finite surface energy: The two-well problem. *Archive for Rational Mechanics and Analysis*, 132(2):101–141.

[Donoho, 1992] Donoho, D. L. (1992). Superresolution via Sparsity Constraints. *SIAM Journal on Mathematical Analysis*, 23(5):1309–1331.

[Dossal and Mallat, 2005] Dossal, C. and Mallat, S. (2005). Sparse spike deconvolution with minimum scale. In *Proceedings of SPARS*, pages 123–126.

[Duval, 2022] Duval, V. (2022). *Faces and Extreme Points of Convex Sets for the Resolution of Inverse Problems*. Habilitation à diriger des recherches, Université Paris Dauphine - PSL.

[Duval and Peyré, 2015] Duval, V. and Peyré, G. (2015). Exact Support Recovery for Sparse Spikes Deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355.

[Duval and Peyré, 2017] Duval, V. and Peyré, G. (2017). Sparse regularization on thin grids I: The Lasso. *Inverse Problems*, 33(5):055008.

[Duval and Tovey, 2021] Duval, V. and Tovey, R. (2021). Dynamical Programming for off-the-grid dynamic Inverse Problems. *arXiv preprint arXiv:2112.11378*.

[Elad et al., 2007] Elad, M., Milanfar, P., and Rubinstein, R. (2007). Analysis versus synthesis in signal priors. *Inverse Problems*, 23(3):947–968.

[Evans and Gariepy, 2015] Evans, L. C. and Gariepy, R. F. (2015). *Measure Theory and Fine Properties of Functions, Revised Edition*. CRC Press.

[Fadili et al., 2019] Fadili, J., Garrigos, G., Malick, J., and Peyré, G. (2019). Model Consistency for Learning with Mirror-Stratifiable Regularizers. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, pages 1236–1244. PMLR.

[Fadili et al., 2018] Fadili, J., Malick, J., and Peyré, G. (2018). Sensitivity Analysis for Mirror-Stratifiable Convex Functions. *SIAM Journal on Optimization*, 28(4):2975–3000.

[Fernandez-Granda, 2013] Fernandez-Granda, C. (2013). Support detection in super-resolution. In *Proceedings of the 10th International Conference on Sampling Theory and Applications*.

[Figalli et al., 2010] Figalli, A., Maggi, F., and Pratelli, A. (2010). A mass transportation approach to quantitative isoperimetric inequalities. *Inventiones mathematicae*, 182(1):167–211.

[Fleming, 1957] Fleming, W. H. (1957). Functions with generalized gradient and generalized surfaces. *Annali di Matematica Pura ed Applicata*, 44(1):93–103.

[Flinth et al., 2021] Flinth, A., de Gournay, F., and Weiss, P. (2021). On the linear convergence rates of exchange and continuous methods for total variation minimization. *Mathematical Programming*, 190(1):221–257.

[Foucart and Rauhut, 2013] Foucart, S. and Rauhut, H. (2013). *A Mathematical Introduction to Compressive Sensing*. Springer Science & Business Media.

[Fuchs, 2004] Fuchs, J.-J. (2004). On sparse representations in arbitrary redundant bases. *IEEE Transactions on Information Theory*, 50(6):1341–1344.

[Fuglede, 1989] Fuglede, B. (1989). Stability in the isoperimetric problem for convex or nearly spherical domains in Rn. *Transactions of the American Mathematical Society*, 314(2):619–638.

[Fujishige, 2005] Fujishige, S. (2005). *Submodular Functions and Optimization*. Elsevier.

[Fusco, 2015] Fusco, N. (2015). The quantitative isoperimetric inequality and related topics. *Bulletin of Mathematical Sciences*, 5(3):517–607.

[Fusco et al., 2008] Fusco, N., Maggi, F., and Pratelli, A. (2008). The sharp quantitative isoperimetric inequality. *Annals of mathematics*, 168(3):941–980.

[Getreuer, 2012a] Getreuer, P. (2012a). Total Variation Deconvolution using Split Bregman. *Image Processing On Line*, 2:158–174.

[Getreuer, 2012b] Getreuer, P. (2012b). Total Variation Inpainting using Split Bregman. *Image Processing On Line*, 2:147–157.

[Giusti, 1984] Giusti (1984). *Minimal Surfaces and Functions of Bounded Variation*. Monographs in Mathematics. Birkhäuser Basel.

[Gribonval and Nielsen, 2003] Gribonval, R. and Nielsen, M. (2003). Sparse representations in unions of bases. *IEEE Transactions on Information Theory*, 49(12):3320–3325.

[Harchaoui et al., 2015] Harchaoui, Z., Juditsky, A., and Nemirovski, A. (2015). Conditional gradient algorithms for norm-regularized smooth convex optimization. *Mathematical Programming*, 152(1):75–112.

[Henrot, 2006] Henrot, A. (2006). *Extremum Problems for Eigenvalues of Elliptic Operators*. Birkhäuser Verlag, Basel - Boston - Berlin.

[Henrot and Pierre, 2018] Henrot, A. and Pierre, M. (2018). *Shape Variation and Optimization : A Geometrical Analysis*. Number 28 in Tracts in Mathematics. European Mathematical Society.

[Hintermüller et al., 2014] Hintermüller, M., Rautenberg, C. N., and Hahn, J. (2014). Functional-analytic and numerical issues in splitting methods for total variation-based image reconstruction. *Inverse Problems*, 30(5):055014.

[Hofmann et al., 2007] Hofmann, B., Kaltenbacher, B., Pöschl, C., and Scherzer, O. (2007). A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Problems*, 23(3):987–1010.

[Hormann and Agathos, 2001] Hormann, K. and Agathos, A. (2001). The point in polygon problem for arbitrary polygons. *Computational Geometry*, 20(3):131–144.

[Huang et al., 2008] Huang, B., Wang, W., Bates, M., and Zhuang, X. (2008). Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science (New York, N.Y.)*, 319(5864):810–813.

[Iglesias et al., 2018] Iglesias, J. A., Mercier, G., and Scherzer, O. (2018). A note on convergence of solutions of total variation regularized linear inverse problems. *Inverse Problems*, 34(5):055011.

[Ionescu and Lachand-Robert, 2005] Ionescu, I. R. and Lachand-Robert, T. (2005). Generalized Cheeger sets related to landslides. *Calculus of Variations and Partial Differential Equations*, 23(2):227–249.

[Ionescu and Lupaşcu-Stamate, 2019] Ionescu, I. R. and Lupaşcu-Stamate, O. (2019). Boundary variation method for the generalized Cheeger problem. *Applied Numerical Mathematics*, 140:199–214.

[Jaggi, 2013] Jaggi, M. (2013). Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization. In *International Conference on Machine Learning*, pages 427–435. PMLR.

[Jarret et al., 2022] Jarret, A., Fageot, J., and Simeoni, M. (2022). A Fast and Scalable Polyatomic Frank-Wolfe Algorithm for the LASSO. *IEEE Signal Processing Letters*, 29:637–641.

[Kuttler and Sigillito, 1984] Kuttler, J. R. and Sigillito, V. G. (1984). Eigenvalues of the Laplacian in Two Dimensions. *SIAM Review*, 26(2):163–193.

[Laville et al., 2021] Laville, B., Blanc-Féraud, L., and Aubert, G. (2021). Off-The-Grid Variational Sparse Spike Recovery: Methods and Algorithms. *Journal of Imaging*, 7(12):266.

[Laville et al., 2022] Laville, B., Blanc-Féraud, L., and Aubert, G. (2022). Off-the-grid curve reconstruction through divergence regularisation: An extreme point result.

[Leonardi, 2015] Leonardi, G. P. (2015). An Overview on the Cheeger Problem. In Pratelli, A. and Leugering, G., editors, *New Trends in Shape Optimization*, International Series of Numerical Mathematics, pages 117–139. Springer International Publishing, Cham.

[Lévy, 2022] Lévy, B. (2022). Partial optimal transport for a constant-volume Lagrangian mesh with free boundaries. *Journal of Computational Physics*, 451:110838.

[Maggi, 2012] Maggi, F. (2012). *Sets of Finite Perimeter and Geometric Variational Problems: An Introduction to Geometric Measure Theory*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge.

[Mallat, 2008] Mallat, S. (2008). *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press.

[März et al., 2022] März, M., Boyer, C., Kahn, J., and Weiss, P. (2022). Sampling Rates for $$\ell Î$$-Synthesis. *Foundations of Computational Mathematics*.

[Ongie and Jacob, 2016] Ongie, G. and Jacob, M. (2016). Off-the-Grid Recovery of Piecewise Constant Images from Few Fourier Samples. *SIAM Journal on Imaging Sciences*, 9(3):1004–1041.

[Pan et al., 2017] Pan, H., Blu, T., and Vetterli, M. (2017). Towards Generalized FRI Sampling With an Application to Source Resolution in Radioastronomy. *IEEE Transactions on Signal Processing*, 65(4):821–835.

[Parini, 2011] Parini, E. (2011). An introduction to the Cheeger problem. *Surveys in Mathematics and its Applications*, 6:9–21.

[Pavani et al., 2009] Pavani, S. R. P., Thompson, M. A., Biteen, J. S., Lord, S. J., Liu, N., Twieg, R. J., Piestun, R., and Moerner, W. E. (2009). Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proceedings of the National Academy of Sciences*, 106(9):2995–2999.

[Peyré, 2021] Peyré, G. (2021). Mathematical Foundations of Data Sciences.

[Phelps, 1993] Phelps, R. R. (1993). *Convex Functions, Monotone Operators and Differentiability*, volume 1364 of *Lecture Notes in Mathematics*. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Polya and Szegö, 1951] Polya, G. and Szegö, G. (1951). *Isoperimetric Inequalities in Mathematical Physics*, volume 27. Princeton University Press.

[Poon, 2019] Poon, C. (2019). An introduction to sparse spikes recovery via the BLASSO.

[Rockafellar, 1970] Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press.

[Rockafellar and Wets, 1998] Rockafellar, R. T. and Wets, R. J.-B. (1998). *Variational Analysis*. Grundlehren Der Mathematischen Wissenschaften. Springer-Verlag, Berlin Heidelberg.

[Rudin et al., 1992] Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268.

[Rudin, 1986] Rudin, W. (1986). *Real and Complex Analysis*. McGraw Hill, New York, 3rd edition edition.

[Rust et al., 2006] Rust, M. J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, 3(10):793–796.

[Santos et al., 2014] Santos, M. C. D., Déturche, R., Vézy, C., and Jaffiol, R. (2014). Axial nanoscale localization by normalized total internal reflection fluorescence microscopy. *Optics Letters*, 39(4):869–872.

[Scherzer et al., 2008] Scherzer, O., Grasmair, M., Grossauer, H., Haltmeier, M., and Lenzen, F. (2008). *Variational Methods in Imaging*, volume 167 of *Applied Mathematical Sciences*. Springer.

[Schiebinger et al., 2018] Schiebinger, G., Robeva, E., and Recht, B. (2018). Superresolution without separation. *Information and Inference: A Journal of the IMA*, 7(1):1–30.

[Shroff et al., 2008] Shroff, H., Galbraith, C. G., Galbraith, J. A., and Betzig, E. (2008). Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics. *Nature Methods*, 5(5):417–423.

[Soubies et al., 2014] Soubies, E., Blanc-Féraud, L., Schaub, S., and Aubert, G. (2014). A 3D model with shape prior information for biological structures reconstruction using multiple-angle total internal reflection fluorescence microscopy. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pages 608–611.

[Stein, 1970] Stein, E. M. (1970). *Singular Integrals and Differentiability Properties of Functions (PMS-30)*. Princeton University Press.

[Stein, 1993] Stein, E. M. (1993). *Harmonic Analysis (PMS-43): Real-Variable Methods, Orthogonality, and Oscillatory Integrals. (PMS-43)*. Princeton University Press.

[Tabti et al., 2018] Tabti, S., Rabin, J., and Elmoata, A. (2018). Symmetric Upwind Scheme for Discrete Weighted Total Variation. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1827–1831.

[Tang et al., 2013] Tang, G., Bhaskar, B. N., Shah, P., and Recht, B. (2013). Compressed Sensing Off the Grid. *IEEE Transactions on Information Theory*, 59(11):7465–7490.

[Tibshirani, 1996] Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

[Tropp, 2006] Tropp, J. (2006). Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, 52(3):1030–1051.

[Vaiter et al., 2015] Vaiter, S., Golbabaee, M., Fadili, J., and Peyré, G. (2015). Model selection with low complexity priors. *Information and Inference: A Journal of the IMA*, 4(3):230–287.

[Vaiter et al., 2018] Vaiter, S., Peyré, G., and Fadili, J. (2018). Model Consistency of Partly Smooth Regularizers. *IEEE Transactions on Information Theory*, 64(3):1725–1737.

[Viola et al., 2012] Viola, F., Fitzgibbon, A., and Cipolla, R. (2012). A unifying resolution-independent formulation for early vision. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 494–501.

Appendix

A. Sets of finite perimeter

In this section, we collect a few useful definitions and properties related to sets of finite perimeter. We refer the reader to e.g. [Maggi, 2012] for further details, and to [Ambrosio et al., 2001] for more information concerning M-connected components and related notions.

Reduced boundary. The reduced boundary $\partial^* E$ of a set of finite perimeter E is defined as the set of points $x \in \operatorname{Supp}(|\mathrm{D}\mathbf{1}_E|)$ at which

$$\nu_E(x) \overset{\text{def.}}{=} \lim_{r \to 0^+} -\frac{\mathrm{D}\mathbf{1}_E(B(x,r))}{|\mathrm{D}\mathbf{1}_E|(B(x,r))}$$

exists and satisfies $|\nu_E(x)| = 1$.

Choice of representative. From [Giusti, 1984, Proposition 3.1], we know that if E has finite perimeter, there exists a Lebesgue representative of E with the property that

$$\forall x \in \partial E,\ 0 < |E \cap B(x,r)| < |B(x,r)|\,.$$

In particular, it does not have isolated points. In the following, we always consider such a representative and consequently obtain

$$\operatorname{Supp}(\mathrm{D}\mathbf{1}_E) = \overline{\partial^* E} = \partial E\,.$$

Distributional curvature. If $E \subset \mathbb{R}^2$ is a set of finite perimeter, then the distributional curvature vector of E is $\mathbf{H}_E : \mathrm{C}_c^\infty(\mathbb{R}^2, \mathbb{R}^2) \to \mathbb{R}$ defined by

$$\forall T \in \mathrm{C}_c^\infty(\mathbb{R}^2, \mathbb{R}^2),\ \langle \mathbf{H}_E, T\rangle = \int_{\partial^* E} \operatorname{div}_E T \, d\mathcal{H}^1\,,$$

where $\operatorname{div}_E T$ denotes the tangential divergence of T on E given by

$$\operatorname{div}_E T = \operatorname{div} T - (\mathrm{D}T\,\nu_E)\cdot \nu_E\,,$$

and $\mathrm{D}T$ denotes the differential of T. A set E is said to have locally integrable distributional curvature if there exists a function $H_E \in \mathrm{L}^1_{\mathrm{loc}}(\partial^* E; \mathcal{H}^1)$ such that

$$\mathbf{H}_E = H_E\, \nu_E\, \mathcal{H}^1 \llcorner \partial^* E\,.$$

For instance, if E is an open set with C^2 boundary, it has a locally summable distributional curvature which is given by the (classical) scalar mean curvature.

M-connected components, holes, exterior. If $E \subset \mathbb{R}^2$ has finite perimeter it can be decomposed (up to Lebesgue negligible sets) into an at most countable union of pairwise disjoint indecomposable sets of positive measure, i.e.

$$E = \bigcup_{i\in I} E_i,\ P(E) = \sum_{i\in I} P(E_i) \text{ and } \forall i,\ |E_i| > 0\,. \tag{59}$$

The collection $\{E_i\}_{i\in I}$ is unique (up to Lebesgue negligible sets), denoted $\mathcal{CC}^M(E)$, and its elements are called the *M-connected components* of E. Any M-connected component of $\mathbb{R}^2 \setminus E$ is called a hole of E. The exterior of E, denoted by $\operatorname{ext}(E)$, is defined as the unique (modulo Lebesgue negligble sets) M-connected component of $\mathbb{R}^2 \setminus E$ with infinite measure.

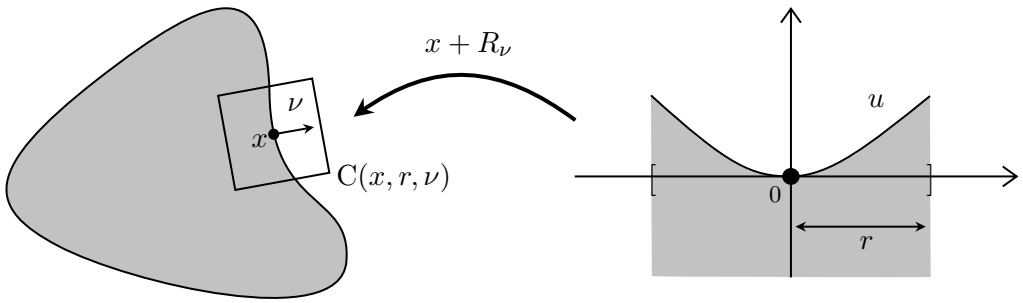


Figure 22 – Illustration of Definition B.1.

B. Smooth sets: definition and convergence

In this section, we give some definitions and facts about smooth sets. Although these notions are standard (most of them are for example discussed in [Delfour and Zolesio, 2011]), we recall them here to keep the exposition (mostly) self-contained.

Smooth sets. We say that a set is smooth if it is locally the hypograph of a smooth function. If $r > 0$, $x \in \mathbb{R}^2$ and $\nu \in \mathbb{S}^1$, we denote by $\mathrm{C}(x_0, r, \nu)$ the square with axis ν and side r centered at x_0, defined as follows:

$$\mathrm{C}(x_0, r, \nu) \overset{\text{def.}}{=} x_0 + \left\{ x \in \mathbb{R}^2 \,\middle|\, |\langle x, \nu \rangle| < r,\ |x - \langle x, \nu \rangle \nu| < r \right\}. \tag{60}$$

With this notation, we have $\mathrm{C}(0, r, e_2) = (-r, r)^2$. We also denote by R_ν the rotation which maps e_2 to ν (and hence $(-r, r)^2$ to $\mathrm{C}(0, r, \nu)$). Given a mapping $u : (-r, r) \to \mathbb{R}$ we define

$$\begin{aligned} \mathrm{graph}(u) &\overset{\text{def.}}{=} \{(z, u(z)),\ z \in (-r, r)\}, \\ \mathrm{hypograph}(u) &\overset{\text{def.}}{=} \{(z, t) \,|\, z \in (-r, r),\ -r < t < u(z)\}. \end{aligned}$$

Definition B.1

Let $k \geq 1$. A set $E \subset \mathbb{R}^2$ is said to be of class C^k if for every $x \in \partial E$ there exists $r > 0$, $\nu \in \mathbb{S}^1$ and $u \in \mathrm{C}^k([-r, r])$ such that

$$\begin{cases} R_\nu^{-1}(\partial E - x) \cap (-r, r)^2 = \mathrm{graph}(u), \\ R_\nu^{-1}(\mathrm{int}\, E - x) \cap (-r, r)^2 = \mathrm{hypograph}(u). \end{cases}$$

As stated in [Delfour and Zolesio, 2011, Theorem 5.2], if E is bounded, the compactness of ∂E allows to take r in Definition B.1 independent of x, and the family $(u_x)_{x \in \partial E}$ uniformly bounded in C^k with equicontinuous k-th derivative.

Convergence of smooth sets. Now, we turn to the definition of C^k convergence for smooth sets.

Definition B.2

Let E be a bounded set of class C^k with $k \geq 1$. We say that $(E_n)_{n\geq 0}$ converges to E in C^k if there exists $r > 0$ and $n_0 \in \mathbb{N}$ such that

- *for every $n \geq n_0$ we have $\partial E_n \subset \bigcup_{x\in\partial E} \mathrm{C}(x, r, \nu_E(x))$*
- *for every $n \geq n_0$ and $x \in \partial E$ there exists $u_{n,x} \in \mathrm{C}^k([-r, r])$ such that:*

$$\begin{cases} R^{-1}_{\nu_E(x)}\left(\partial E_n - x\right) \cap (-r, r)^2 = \mathrm{graph}(u_{n,x}) \\ R^{-1}_{\nu_E(x)}\left(\mathrm{int}\, E_n - x\right) \cap (-r, r)^2 = \mathrm{hypograph}(u_{n,x}) \end{cases}$$

- *denoting $(u_x)_{x\in\partial E}$ functions satisfying*

$$\begin{cases} R^{-1}_{\nu_E(x)}\left(\partial E - x\right) \cap (-r, r)^2 = \mathrm{graph}(u_x) \\ R^{-1}_{\nu_E(x)}\left(\mathrm{int}\, E - x\right) \cap (-r, r)^2 = \mathrm{hypograph}(u_x) \end{cases}$$

we have $\lim_{n\to+\infty} \sup_{x\in\partial E} \|u_{n,x} - u_x\|_{\mathrm{C}^k([-r,r])} = 0$

In Section 2.2, we make use of the following result, which states that, if a sequence of sets converges in C^k, the boundary of its terms can eventually be written as a normal deformation of the boundary of the limit set. A related statement can be found for instance in [Acerbi et al., 2013, Theorem 4.2]. The authors of this last work refer to [Almgren, 1975], which, according to them, essentially contains the result. We provide a detailed proof below for the sake of completeness.

Proposition B.3

If $(E_n)_{n\geq 0}$ converges to a bounded set E in C^k with $k \geq 2$, then for n large enough there exists a mapping $\varphi_n \in \mathrm{C}^{k-1}(\partial E)$ such that $\partial E_n = (Id + \varphi_n\, \nu_E)(\partial E)$, and $\|\varphi_n\|_{\mathrm{C}^{k-1}(\partial E)} \to 0$.

Proof : In all the following, we denote by ν (instead of ν_E) the outward unit normal to E. If $u \in \mathrm{C}^k(\overline{\Omega})$ and $\delta > 0$, we also use the notation

$$B_{\mathrm{C}^k}(u, \delta) \overset{\text{def.}}{=} \left\{ v \in \mathrm{C}^k(\overline{\Omega}) \,\middle|\, \|u - v\|_{\mathrm{C}^k(\overline{\Omega})} < \delta \right\}.$$

We divide the proof in two main steps.

Step 1: we first prove that, for every $x \in \partial E$, there exists $n_0 \in \mathbb{N}$ and two open neighborhoods U, V of x such that, for every $n \geq n_0$, there exists $\varphi_n : \partial E \cap U \to \mathbb{R}$ of class C^{k-1} with $\partial E_n \cap V = (Id + \varphi_n\, \nu)(\partial E \cap U)$ and $\|\varphi_n\|_{\mathrm{C}^{k-1}(\partial E\cap U)} \to 0$.

By the C^k convergence of $(E_n)_{n\geq 0}$ towards E, we can fix $x \in \partial E$, $r > 0$ and $n_0 \in \mathbb{N}$ such that, in $\mathrm{C}(x, r, \nu_E(x))$, the set E is the hypograph of $u \in \mathrm{C}^k([-r, r])$ and E_n is the hypograph of $u_n \in \mathrm{C}^k([-r, r])$ for every $n \geq n_0$, with $\|u_n - u\|_{\mathrm{C}^k([-r,r])} \to 0$. By an abuse of notation, the normal ν is given in local coordinates by

$$\forall z \in (-r, r),\ \nu(z, u(z)) \overset{\text{def.}}{=} \frac{1}{\sqrt{1 + u'(z)^2}}(-u'(z), 1)\,.$$

We notice that if $(z,t) \in (-r,r)\times\mathbb{R}$ and $v : (-r,r) \to \mathbb{R}$, then $(z,u(z))+t\,\nu(z,u(z)) \in \mathrm{graph}(v)$ if and only if

$$\left| z - \frac{t}{\sqrt{1+u'(z)^2}} u'(z) \right| < r \tag{61}$$

and $f(v,z,t) = 0$ with

$$f(v,z,t) \stackrel{\text{def.}}{=} u(z) + \frac{t}{\sqrt{1+u'(z)^2}} - v\left(z - \frac{t}{\sqrt{1+u'(z)^2}} u'(z) \right).$$

If $(z,t) \in (-r',r')^2$ with $r' \stackrel{\text{def.}}{=} r/2$, then (61) holds and $f(v,z,t)$ is hence well defined.

We claim that f is of class C^{k-1} on $\mathrm{C}^{k-1}([-r,r]) \times (-r',r')^2$. To see this, one should first notice that the two mappings

$$(z,t) \mapsto u(z) + \frac{t}{\sqrt{1+u'(z)^2}}\,, \qquad (z,t) \mapsto z - \frac{t}{\sqrt{1+u'(z)^2}} u'(z)\,,$$

are of class C^{k-1} on $(-r',r')^2$. It is then sufficient to observe that, if $g : (-r',r')^2 \to (-r,r)$ is of class C^{k-1}, then $(v,y) \mapsto v(g(y))$ is of class C^{k-1} on $\mathrm{C}^{k-1}([-r,r]) \times (-r',r')^2$.

Now, we know f is of class C^{k-1}. Moreover, we have $f(u,z,0) = 0$ for every $z \in (-r',r')$, and in particular $f(u,0,0) = 0$. Using that $u'(0) = 0$, we also get $\partial_t f(u,0,0) = 1 \neq 0$. From the implicit function theorem, we hence obtain the existence of $\delta_1 > 0$, of $\delta_2 \in (0,r')$, of an open neighborhood $W \subset (-r',r')$ of 0, and of a mapping $\psi : B_{\mathrm{C}^{k-1}}(u,\delta_1) \times (-\delta_2,\delta_2) \to W$ of class C^{k-1} such that, for every (v,z,t) in $\mathrm{C}^{k-1}([-r,r]) \times (-r',r')^2$:

$$\begin{aligned} &\big[(v,z,t) \in B_{\mathrm{C}^{k-1}(u,\delta_1)} \times (-\delta_2,\delta_2) \times W \text{ and } f(v,z,t) = 0\big] \\ \iff &\big[(v,z) \in B_{\mathrm{C}^{k-1}(u,\delta_1)} \times (-\delta_2,\delta_2) \text{ and } t = \psi(v,z)\big]\,. \end{aligned}$$

Since $f(u,z,0) = 0$, we obtain $\psi(u,z) = 0$ for every $z \in (-\delta_2,\delta_2)$. Now, using the fact ψ is of class C^{k-1}, we obtain that all its derivatives up to order $k-1$ are uniformly continuous on every compact subset of $B_{\mathrm{C}^{k-1}}(u,\delta_1) \times (-\delta_2,\delta_2)$. By the Ascoli-Arzelá theorem $\mathrm{C}^k([-r,r])$ is compactly embedded into $\mathrm{C}^{k-1}([-r,r])$, and we hence have that, for every $\delta_2' < \delta_2$, the family of functions $\psi_v : z \mapsto \psi(v,z)$ converges in $\mathrm{C}^{k-1}([-\delta_2',\delta_2'])$ towards 0 when $\|v-u\|_{\mathrm{C}^k([-r,r])} \to 0$.

Now, we define

$$\begin{aligned} &F : (-r,r)^2 \to \mathbb{R}^2 \\ &\qquad (z,t) \mapsto (z,u(z)) + t\,\nu(z,u(z)) = \left(z - \frac{t}{\sqrt{1+u'(z)^2}} u'(z),\ u(z) + \frac{t}{\sqrt{1+u'(z)^2}} \right). \end{aligned}$$

We have that F is of class C^{k-1} and $DF(0,0) = Id$. There hence exists two open neighborhoods U_1, U_2 of 0 such that F is a C^{k-1} diffeomorphism from $U_1 \times U_2$ to $F(U_1 \times U_2)$.

We take $\delta_2' < \delta_2$ small enough to have $(-\delta_2',\delta_2') \subset U_1$, and $\delta_1' < \delta_1$ small enough to have $\|\psi_v\|_\infty \leq \gamma$ for every $v \in \mathrm{B}_{\mathrm{C}^k}(u,\delta_1')$, with $(-\gamma,\gamma) \subset U_2 \cap W$. We claim that, for every $v \in \mathrm{B}_{\mathrm{C}^k}(u,\delta_1')$, we have

$$\{F(z,\psi_v(z)),\ z \in (-\delta_2',\delta_2')\} = \mathrm{graph}(v) \cap V' \quad \text{with} \quad V' \stackrel{\text{def.}}{=} F((-\delta_2',\delta_2') \times (-\gamma,\gamma))\,. \tag{62}$$

By construction, the set on the left-hand side is included in the one on the right-hand side. Now, if $y \in \mathrm{graph}(v) \cap V'$, then denoting $(z,t) \stackrel{\text{def.}}{=} F^{-1}(y)$, we have that $z \in (-\delta_2',\delta_2')$. Moreover, we also have $(v,z,t) \in B_{\mathrm{C}^{k-1}}(u,\delta_1) \times (-\delta_2,\delta_2) \times W$ and $f(v,z,t)$, which shows $t = \psi_v(z)$, and hence (62).

Finally, defining

$$U' \stackrel{\text{def.}}{=} (-\delta_2', \delta_2') \times (-r, r),\ \ U \stackrel{\text{def.}}{=} x + R_{\nu_E(x)} U',\ \ V \stackrel{\text{def.}}{=} x + R_{\nu_E(x)} V',$$
$$\varphi_v : y \mapsto \psi_v((R^{-1}_{\nu_E(x)}(y - x))_1),$$

and taking n_0 large enough to have $u_n \in \mathrm{B}_{\mathrm{C}^k}(u, \delta_1')$ for every $n \geq n_0$, we obtain

$$\partial E_n \cap V = (Id + \varphi_n\, \nu)(\partial E \cap U)$$

and $\|\varphi_n\|_{\mathrm{C}^{k-1}(\partial E \cap U)} \to 0$ with $\varphi_n \stackrel{\text{def.}}{=} \varphi_{u_n}$.

Step 2: from now on, we make the dependencies on x explicit. We know from step 1 that for every $x \in \partial E$ there exists $n_0(x) \in \mathbb{N}$, a mapping $\varphi_{n,x} : \partial E \cap U_x \to \mathbb{R}$, and two open neighborhoods U_x, V_x of x such that, for every $n \geq n_0(x)$

$$\partial E_n \cap V_x = (Id + \varphi_{n,x}\, \nu)(\partial E \cap U_x).$$

Moreover, $\|\varphi_{n,x}\|_{\mathrm{C}^{k-1}(\partial E \cap U_x)} \to 0$.

Let us first show that if $y \in \partial E \cap U_{x_1} \cap U_{x_2}$ and $n \geq \max(n_0(x_1), n_0(x_2))$ then

$$\varphi_{n,x_1}(y) = \varphi_{n,x_2}(y).$$

We assume without loss of generality that $|\varphi_{n,x_1}(y)| \leq |\varphi_{n,x_2}(y)|$. From step 1, we know that $\varphi_{n,x_2}(y)$ is the unique $t \in (-\gamma_{x_2}, \gamma_{x_2})$ such that $y + t\,\nu(y) \in \partial E_n \cap \mathrm{C}(x_2, \delta_2, \nu(x_2))$. But by definition $y + \varphi_{n,x_1}(y)\,\nu(y) \in \partial E_n$, and since $|\varphi_{n,x_1}(y)| \leq |\varphi_{n,x_2}(y)|$ we also have that $\varphi_{n,x_1}(y) \in (-\gamma_{x_2}, \gamma_{x_2})$ and $y + \varphi_{n,x_1}(y)\,\nu(y) \in \mathrm{C}(x_2, \delta_2, \nu(x_2))$. We therefore obtain that $\varphi_{n,x_1}(y) = \varphi_{n,x_2}(y)$.

Now, we cover ∂E by $\bigcup_{x \in \partial E} U_x \cap V_x$ and obtain the existence of a finite set I such that $E \subset \bigcup_{i \in I} U_{x_i} \cap V_{x_i}$. For every $n \geq \max_{i \in I} n_0(x_i)$ we define φ_n globally using the results above, and obtain

$$\|\varphi_n\|_{\mathrm{C}^{k-1}(\partial E)} \leq \max_{i \in I} \|\varphi_{n,x_i}\|_{\mathrm{C}^{k-1}(\partial E \cap U_{x_i})} \to 0.$$

By construction, we also have $(Id + \varphi_n\, \nu)(\partial E) = \partial E_n$, which concludes the proof. ∎

Normal deformations of smooth sets. Considering the result of Proposition B.3, one could wonder if, given a sufficiently smooth $\varphi : \partial E \to \mathbb{R}$ in a neighborhood of 0, there is a unique smooth set F satisfying

$$\partial F = (Id + \varphi\, \nu_E)(\partial E).$$

This is indeed the case, as shown below.

Lemma B.4

If E is a bounded set of class C^k ($k \geq 2$), then there exists $C > 0$ such that, for every φ in $\mathrm{C}^{k-1}(\partial E)$, the mapping $\varphi\, \nu_E$ can be extended to $\xi_\varphi \in \mathrm{C}^{k-1}(\mathbb{R}^2, \mathbb{R}^2)$ with

$$\|\xi_\varphi\|_{\mathrm{C}^{k-1}(\mathbb{R}^2, \mathbb{R}^2)} \leq C\, \|\varphi\|_{\mathrm{C}^{k-1}(\partial E)}.$$

Proof : We use the arguments of [Dambrine and Lamboley, 2019, Section 4]. Since E is bounded and of class C^k, there exists an open bounded set $\Omega \supset \partial E$ of class C^1, such that the projection $\pi_{\partial E}$ on ∂E belongs to $\mathrm{C}^{k-1}(\overline{\Omega})$. For every $\varphi \in \mathrm{C}^{k-1}(\partial E)$, one can then define $\xi_\varphi = (\varphi\, \nu_E) \circ \pi_{\partial E}$ to obtain an extension of $\varphi\, \nu_E$ which belongs to $\mathrm{C}^{k-1}(\overline{\Omega})$. Using Faà di Bruno's formula, one obtains the existence of $C > 0$ such that

$$\forall \varphi \in \mathrm{C}^{k-1}(\partial E),\ \|\xi_\varphi\|_{\mathrm{C}^{k-1}(\Omega,\mathbb{R}^2)} \leq C\, \|\varphi\|_{\mathrm{C}^{k-1}(\partial E)}\,.$$

One can then use an extension theorem (see for instance [Stein, 1970, Chapter 6, Theorem 5]) to obtain the result with $\Omega = \mathbb{R}^2$, up to a modification of C. ■

Proposition B.5

Let E be a bounded open set of class C^k (with $k \geq 2$). There exists $c > 0$ such that, for every $\varphi \in \mathrm{C}^{k-1}(\partial E)$ with $\|\varphi\|_{\mathrm{C}^{k-1}(\partial E)} \leq c$, there is a unique bounded open set of class C^{k-1}, denoted E_φ, satisfying

$$\partial E_\varphi = (Id + \varphi\, \nu_E)(\partial E)\,. \tag{63}$$

Moreover, there exists an extension ξ_φ of $\varphi\, \nu_E$ such that $E_\varphi = (Id + \xi_\varphi)(E)$ and

$$\|\xi_\varphi\|_{\mathrm{C}^{k-1}(\mathbb{R}^2,\mathbb{R}^2)} < 1\,.$$

In particular, E_φ is C^{k-1}-diffeomorphic to E.

Proof : Let us first prove the existence of E_φ. Let C be as in Lemma B.4. If $c < 1/C$, then for every φ such that $\|\varphi\|_{\mathrm{C}^{k-1}(\partial E)} \leq c$, we have $\|\xi_\varphi\|_{\mathrm{C}^{k-1}}(\mathbb{R}^2, \mathbb{R}^2) < 1$. As a result, we obtain that $f_\varphi \overset{\text{def.}}{=} Id + \xi_\varphi$ is a C^{k-1} diffeomorphism. We hence have that $E_\varphi \overset{\text{def.}}{=} f_\varphi(E)$ is a bounded open set of class C^{k-1} and satisfies $\partial E_\varphi = f_\varphi(\partial E) = (Id + \varphi\, \nu_E)(\partial E)$.

Let us now turn to uniqueness. If F and G are two bounded open sets of class C^{k-1} satisfying (63), in particular we have that $\partial F = \partial G$. Since F and G are sufficiently smooth and bounded, they have finite perimeter and $\partial^M F = \partial F = \partial G = \partial^M G$. Using Proposition C.6 below (whose proof is the object of Appendix C), we obtain that $F = G$ up to a Lebesgue negligible set. Using that F and G are sufficiently smooth open sets, this finally yields $F = G$. ■

If E and c are as in Proposition B.5, for every $\epsilon \leq c$, we say that a set F is a C^{k-1}-normal deformation of E of size at most ϵ if there exists φ such that $\|\varphi\|_{\mathrm{C}^{k-1}}(\partial E) \leq \epsilon$ and $F = E_\varphi$. Likewise, if E is such that E^c satisfies the assumptions of Proposition B.5, we use the same terminology for F if $F^c = (E^c)_\varphi$.

C. Characterization of sets from their boundary

The aim of this section is to prove the following result, which, loosely speaking, states that sets of finite perimeter and finite measure are characterized by their measure-theoretic boundary. In all the following $d \geq 2$ is a fixed integer and $\mathcal{H}^s$ denotes the s-dimensional Hausdorff measure on $\mathbb{R}^d$.

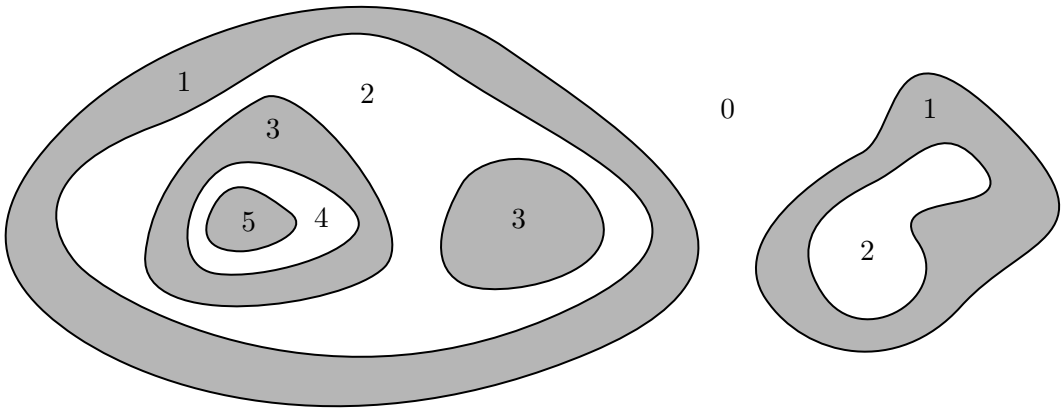


Figure 23 – The topographic function of a set E (gray area).

Proposition C.6

Let $E, F \subset \mathbb{R}^d$ be sets of finite perimeter with finite measure. If $\partial^M E = \partial^M F \pmod{\mathcal{H}^{d-1}}$ then we have $E = F \pmod{\mathcal{H}^d}$.

Although this result is certainly well-known, we were not able to find it clearly stated or used, and hence decided to include a proof for the sake of completeness. To prove it, we rely on a tool introduced in [Ambrosio et al., 2001, Section 7]: the topographic function (see Figure 23 for an illustration). To be more precise, if E and F are as in Proposition C.6 and u and v denote their topographic function, we prove that $u = v$ almost everywhere, which, since $\mathbf{1}_E = u \bmod 2$ and $\mathbf{1}_F = v \bmod 2$, yields the result. This is done by induction, i.e. by proving that

$$\{u = k\} = \{v = k\} \pmod{\mathcal{H}^d}$$

for every $k \in \mathbb{N}$. Before diving into the proof, we state a few useful lemmas.

Lemma C.7

Let $E, F \subseteq \mathbb{R}^d$ be of finite perimeter. If E is indecomposable and $\mathcal{H}^{d-1}(\partial^M F \cap \mathring{E}^M) = 0$ then we have $E \subseteq F \pmod{\mathcal{H}^d}$ or $E \subseteq F^c \pmod{\mathcal{H}^d}$.

Proof : Since $\operatorname{supp}(\mathrm{D}\mathbf{1}_F) = \partial^M F \pmod{\mathcal{H}^{d-1}}$ we have $|\mathrm{D}\mathbf{1}_F|(\mathring{E}^M) = 0$. The indecomposability of E hence yields, using a result of [Dolzmann and Müller, 1995], that $\mathbf{1}_F$ is a.e. equal to 0 on E, or a.e. equal to 1 on E, hence the result. ∎

In order to initialize our proof by induction, we need to prove the equality of the zero level set of the topographic functions. From the construction of [Ambrosio et al., 2001, Theorem 6] it is clear that, if u is the topographic function of a set of finite perimeter E with finite measure, then $\operatorname{ext}(E) = \{u = 0\}$. We can hence conclude using the following lemma.

Lemma C.8

Assume E and F are two sets of finite measure and finite perimeter in $\mathbb{R}^d$. Then, if

$$\partial^M E = \partial^M F \ (\mathrm{mod}\ \mathcal{H}^{d-1}),$$

we have $\mathrm{ext}(E) = \mathrm{ext}(F)\ (\mathrm{mod}\ \mathcal{H}^d)$.

Proof : We have:

$$\mathcal{H}^{d-1}\left(\partial^M F \cap \mathring{\mathrm{ext}}(E)^M\right) = \mathcal{H}^{d-1}\left(\partial^M E \cap \mathring{\mathrm{ext}}(E)^M\right) = 0\,.$$

Since $\mathrm{ext}(E)$ is indecomposable from Lemma C.7 we obtain $\mathrm{ext}(E) \subseteq F\ (\mathrm{mod}\ \mathcal{H}^d)$ (which is impossible since $\mathrm{ext}(E)$ has infinite measure and F has finite measure) or $\mathrm{ext}(E) \subseteq F^c\ (\mathrm{mod}\ \mathcal{H}^d)$. Since $\mathrm{ext}(E)$ is indecomposable, it is included (modulo H^d) in an M-connected component of F^c, which is necessarily $\mathrm{ext}(F)$ since $\mathrm{ext}(E)$ has infinite measure. We hence obtain

$$\mathrm{ext}(E) \subseteq \mathrm{ext}(F)\ (\mathrm{mod}\ \mathcal{H}^d)\,.$$

Exchanging E and F and applying the same argument yields the result. ∎

Before proving the main result, we need a last lemma, which is Lemma C.9. In the following, if u has bounded variation, we denote by J_u its approximate jump set and S_u its approximate discontinuity set (see [Ambrosio et al., 2000, Sections 3.6 and 3.7]).

Lemma C.9

Let $E \subseteq \mathbb{R}^d$ be a set of finite perimeter, u its topographic function and $k \in u(\mathbb{R}^2)$. Then for $\mathcal{H}^{d-1}$-almost every $x \in \partial^M\{u \leq k\}$ we have $x \in J_u$ and, denoting $\nu(x)$ the outward unit normal to $\{u \leq k\}$ at x, the following holds:

$$⨍_{B^+(x,r,\nu(x))} |u - (k+1)| \xrightarrow{r\to 0^+} 0, \qquad ⨍_{B^-(x,r,\nu(x))} |u - k| \xrightarrow{r\to 0^+} 0\,.$$

Proof : First, let us show that

$$\partial^M\{u \leq k\} \subseteq J_u\ (\mathrm{mod}\ \mathcal{H}^{d-1})\,. \tag{64}$$

Since $u \in \mathrm{BV}_{\mathrm{loc}}(\mathbb{R}^d, \mathbb{N})$ we have $\mathcal{H}^{d-1}(S_u \setminus J_u)$. Let $x \in S_u^c$ and assume the approximate limit $\tilde{u}(x)$ of u at x satisfies $\tilde{u}(x) \leq k$. We hence have:

$$⨍_{B(x,r)} \mathbf{1}_{\{u\geq k+1\}} \leq ⨍_{B(x,r)} (u-k)\,\mathbf{1}_{\{u\geq k+1\}} \leq ⨍_{B(x,r)} (u-\tilde{u}(x))\,\mathbf{1}_{\{u\geq k+1\}} \leq ⨍_{B(x,r)} |u-\tilde{u}(x)|\,.$$

Since the last term converges to 0 as $r \to 0^+$, we obtain that $x \in \{u \overset{\circ}{\leq} k\}^M$. Likewise if $\tilde{u}(x) > k$ we obtain that $x \in (\{u \overset{\circ}{\leq} k\}^c)^M$. This shows $\partial^M\{u \leq k\} \subseteq S_u$, and we finally obtain (64).

Now, let $x \in \partial^M\{u \leq k\} \cap J_u$ (which, from our assumptions, is non-empty). We denote by ν the outward unit normal to $\{u \leq k\}$, and have the existence of $a \neq b$ such that

$$⨍_{B^-(x,r,\nu(x))} |u - a| \xrightarrow{r\to 0^+} 0, \quad ⨍_{B^+(x,r,\nu(x))} |u - b| \xrightarrow{r\to 0^+} 0\,,$$

with $B^{\pm}(x,r,\nu(x)) = \{y \in B(x,r) \,|\, \pm\langle y-x,\nu\rangle > 0\}$. Since $|\mathrm{D}u| = \mathcal{H}^{d-1} \llcorner \partial^M E$ we also have $|a-b| = 1$. Let us show that $a \leq k$ and $b \geq k+1$. We have:

$$\begin{aligned} a &= \fint_{B^-(x,r,\nu(x))} (a-u) + \fint_{B^-(x,r,\nu(x))} u \\ &\leq \fint_{B^-(x,r,\nu(x))} |u-a| + 2\left[\fint_{B(x,r)} u\,\mathbf{1}_{\{u\leq k\}} + \fint_{B(x,r)} u\,(\mathbf{1}_{B^-(x,r,\nu(x))} - \mathbf{1}_{\{u\leq k\}})\right] \\ &\leq \fint_{B^-(x,r,\nu(x))} |u-a| + 2\left[k\,\frac{|\{u\leq k\}\cap B(x,r)|}{|B(x,r)|} + \fint_{B(x,r)} u\,(\mathbf{1}_{B^-(x,r,\nu(x))} - \mathbf{1}_{\{u\leq k\}})\right]. \end{aligned}$$

The first term converges to 0. The second converges to k assuming $x \in \{u \leq k\}^{(1/2)}$ (which is not restrictive since $\mathcal{H}^{d-1}(\partial^M\{u\leq k\} \setminus \{u\leq k\}^{(1/2)}) = 0$). To show that the last term converges to 0, we write

$$\begin{aligned} \fint_{B(x,r)} u\,(\mathbf{1}_{B^-(x,r,\nu(x))} - \mathbf{1}_{\{u\leq k\}}) &\leq \frac{|B^-(x,r,\nu(x))\setminus\{u\leq k\}|^{\frac{1}{d}}}{|B(x,r)|}\left[\int_{B(x,r)} |u|^{d/(d-1)}\right]^{\frac{d-1}{d}} \\ &= \left[\frac{|B^-(x,r,\nu(x))\setminus\{u\leq k\}|}{|B(x,r)|}\right]^{\frac{1}{d}}\left[\fint_{B(x,r)} |u|^{d/(d-1)}\right]^{\frac{d-1}{d}} \end{aligned}$$

The first term converges to 0 assuming $x \in \partial^*\{u\leq k\}$. Again, this is not restrictive since

$$\mathcal{H}^{d-1}(\partial^M\{u\leq k\} \setminus \partial^*\{u\leq k\}) = 0\,.$$

By [Ambrosio et al., 2000, Lemma 3.75] the second term is bounded for $\mathcal{H}^{d-1}$-a.e. x. We hence obtain $a \leq k$. By using the same arguments, one can show $b \geq k+1$ and we finally get $a = k$ and $b = k+1$. ■

Proposition C.10

Let $E, F \subset \mathbb{R}^d$ be sets of finite perimeter with finite measure and u, v their topographic functions. If $\partial^M E = \partial^M F \pmod{\mathcal{H}^{d-1}}$ then $u = v$ almost everywhere.

Proof : We already know that $\{u=0\} = \{v=0\}$. Let us prove by induction that $\{u=k\} = \{v=k\}$ for every $k \in \mathbb{N}$. We fix $k \geq 1$ and assume that $\{u=l\} = \{v=l\}$ for every $l \leq k$.

Let $A \in \mathcal{CC}^M(\{u = k+1\})$. Since A is indecomposable, u is constant on A and

$$|\mathrm{D}u| = \mathcal{H}^{d-1} \llcorner \partial^M E = \mathcal{H}^{d-1} \llcorner \partial^M F = |\mathrm{D}v|\,,$$

we obtain that v is a.e. constant on A. The construction of [Ambrosio et al., 2001, Theorem 6] also yields

$$\mathcal{H}^{d-1}(\partial^M\{v\leq k\} \cap \partial^M A) = \mathcal{H}^{d-1}(\partial^M\{u\leq k\} \cap \partial^M A) > 0\,.$$

Using Lemma C.9, we obtain that $v = k+1$ a.e. on A, and hence $A \subseteq \{v = k+1\} \pmod{\mathcal{H}^d}$. Since this holds for every $A \in \mathcal{CC}^M(\{u=k+1)\})$ we get $\{u=k+1\} \subseteq \{v=k+1\} \pmod{\mathcal{H}^d}$. Exchanging E and F yields $\{v=k+1\} \subseteq \{u=k+1\} \pmod{\mathcal{H}^d}$, and we can finally conclude. ■

D. Convergence of the level sets

The aim of this short section is to prove the following elementary result.

Proposition D.1

If $(\Omega, \Sigma, |.|)$ is a measure space and $\|u_n - u\|_{\mathrm{L}^1(\Omega)} \to 0$ then for almost every $t \in \mathbb{R}$ we have:

$$\lim_{n\to+\infty} \left|U_n^{(t)} \triangle U^{(t)}\right| \to 0\,. \tag{65}$$

Before giving its proof, let us stress that this property is certainly well-known, but we could not find it clearly stated or proved. In fact, using Fubini's theorem, one could directly obtain

$$\lim_{n\to+\infty} \int_0^{+\infty} \left|U_n^{(t)} \triangle U^{(t)}\right| dt = \lim_{n\to+\infty} \int_{\mathbb{R}^2} |u_n - u| = 0\,,$$

and hence prove the L^1 convergence of $f : t \mapsto \left|U_n^{(t)} \triangle U^{(t)}\right|$. This implies the pointwise convergence of f almost everywhere, but only up to the extraction of a subsequence. Lifting this requirement is in fact possible, as shown below.

Proof : Let $t \geq 0$. We first prove that $|\{u_n \geq t\} \setminus \{u \geq t\}| \to 0$. For every $h > 0$ we have:

$$\begin{aligned}
\|u_n - u\|_{\mathrm{L}^1} &\geq \int_{t-h}^{t} |\{u_n \geq s\} \triangle \{u \geq s\}|\, ds \\
&\geq \int_{t-h}^{t} |\{u_n \geq s\} \setminus \{u \geq s\}|\, ds \\
&\geq \int_{t-h}^{t} |\{u_n \geq t\} \setminus \{u \geq t-h\}|\, ds \\
&= h\, |\{u_n \geq t\} \setminus \{u \geq t-h\}|\,.
\end{aligned}$$

Using that $|A \setminus B| \leq |A \setminus C| + |C \setminus B|$ for every triple of sets (A, B, C) we get:

$$|\{u_n \geq t\} \setminus \{u \geq t\}| \leq |\{u_n \geq t\} \setminus \{u \geq t-h\}| + |\{u \geq t-h\} \setminus \{u \geq t\}|\,.$$

We hence obtain:

$$|\{u_n \geq t\} \setminus \{u \geq t\}| \leq \frac{1}{h}\|u_n - u\|_{\mathrm{L}^1} + |\{u \geq t-h\} \setminus \{u \geq t\}|\,,$$

which yields

$$\limsup_{n\to+\infty} |\{u_n \geq t\} \setminus \{u \geq t\}| \leq |\{u \geq t-h\} \setminus \{u \geq t\}|\,. \tag{66}$$

Since $\{u \geq t-h\} \setminus \{u \geq t\}$ decreases as h decreases and $\bigcap_{h\geq 0} \{u \geq t-h\} \setminus \{u \geq t\} = \emptyset$, we obtain that the right-hand side in (66) converges to 0 when h goes to 0, and therefore that:

$$\lim_{n\to+\infty} |\{u_n \geq t\} \setminus \{u \geq t\}| = 0\,.$$

Let us now prove that $|\{u \geq t\} \setminus \{u_n \geq t\}| \to 0$. For every $h > 0$ it holds that:

$$\begin{aligned}
\|u_n - u\|_{\mathrm{L}^1} &\geq \int_{t}^{t+h} |\{u_n \geq s\} \triangle \{u \geq s\}|\, ds \\
&\geq \int_{t}^{t+h} |\{u \geq s\} \setminus \{u_n \geq s\}|\, ds \\
&\geq \int_{t}^{t+h} |\{u \geq t+h\} \setminus \{u_n \geq t\}|\, ds \\
&= h\, |\{u \geq t+h\} \setminus \{u_n \geq t\}|\,.
\end{aligned}$$

Using that $|B \setminus A| \leq |C \setminus A| + |B \setminus C|$ for every triple of sets (A, B, C) we get:

$$\begin{aligned}|\{u \geq t\} \setminus \{u_n \geq t\}| &\leq |\{u \geq t+h\} \setminus \{u_n \geq t\}| + |\{u \geq t\} \setminus \{u \geq t+h\}| \\ &\leq \frac{1}{h}\|u_n - u\|_{\mathrm{L}^1} + |\{u \geq t\} \setminus \{u \geq t+h\}|\,,\end{aligned}$$

and hence

$$\limsup_{n \to +\infty} |\{u \geq t\} \setminus \{u_n \geq t\}| \leq |\{u \geq t\} \setminus \{u \geq t+h\}|\,, \tag{67}$$

Since $\{u \geq t\} \setminus \{u \geq t+h\}$ decreases as h decreases and $\bigcap_{h \geq 0} \{u \geq t\} \setminus \{u \geq t+h\} = \{u = t\}$, we obtain that, for almost every $t \geq 0$, the right-hand side in (67) converges to 0 and therefore that:

$$\lim_{n \to +\infty} |\{u \geq t\} \setminus \{u_n \geq t\}| = 0\,.$$

We hence finally have $|\{u_n \geq t\} \triangle \{u \geq t\}| \to 0$ for almost every $t \geq 0$. The fact

$$|\{u_n \leq t\} \triangle \{u \leq t\}| \to 0$$

for almost every $t \leq 0$ follows from the same arguments. ∎

E. Proof of Lemma 3.2

We now give the proof of Lemma 3.2, whose content is recalled below.

Lemma 3.2

If $t^* > 0$, the extreme points of $C \overset{\text{def.}}{=} \{u \in \mathrm{L}^2(\mathbb{R}^2) \,|\, \mathrm{TV}(u) \leq t \leq t^*\}$ are:

- $(0, 0)\,,$
- $(\epsilon t^* \mathbf{1}_E / P(E), t^*)$ *with* $\epsilon \in \{-1, 1\}$, $E \subset \mathbb{R}^2$ *simple and* $0 < |E| < +\infty\,.$

Proof : Let $(u, t) \in \mathrm{ext}(C)$.

1. If $t = 0$ then $u = 0$.
2. Otherwise we have

$$(u, t) = (1 - t/t^*)(0, 0) + (t/t^*)(t^* u/t, t^*)\,,$$

which yields $t = t^*$. Considering that, for any v such that $\mathrm{TV}(v) \leq t^*$, we have:

$$(0, t^*) = \frac{1}{2}(-v, t^*) + \frac{1}{2}(v, t^*)\,,$$

we also obtain $\mathrm{TV}(u) > 0$. Now, since

$$(u, t^*) = (1 - \mathrm{TV}(u)/t^*)(0, t^*) + (\mathrm{TV}(u)/t^*)(t^* u/\mathrm{TV}(u), t^*)\,,$$

we get $\mathrm{TV}(u) = t^*$. Finally, let $\alpha \in (0, 1)$ and $u_1, u_2 \in \mathrm{L}^2(\mathbb{R}^2)$ with $\mathrm{TV}(u_1) \leq t^*$ and $\mathrm{TV}(u_2) \leq t^*$. If $u = (1 - \alpha)u_1 + \alpha u_2$ then writing $(u, t^*) = (1 - \alpha)(u_1, t^*) + \alpha(u_2, t^*)$ we obtain $u = u_1 = u_2$, which shows $u \in \mathrm{ext}(\{u \in \mathrm{L}^2(\mathbb{R}^2) \,|\, \mathrm{TV}(u) \leq t^*\})$. This shows that $u = \epsilon t^* \mathbf{1}_E / P(E)$ for some simple set E such that $0 < |E| < +\infty$ and $\epsilon \in \{-1, 1\}$.

Conversely, let (u,t), (u_1,t_1) and (u_2,t_2) be in C and $\alpha \in (0,1)$ with

$$(u,t) = (1-\alpha)(u_1,t_1) + \alpha(u_2,t_2)\,.$$

1. If $(u,t) = (0,0)$, since t_1, t_2 are nonnegative we obtain $t_1 = t_2 = 0$ and therefore $u_1 = u_2 = 0$. Hence $(0,0) \in \mathrm{ext}(C)$.
2. If $(u,t) = (\epsilon t^* \mathbf{1}_E/P(E), t^*)$ with $\epsilon \in \{-1,1\}$ and E a simple set such that $0 < |E| < +\infty$, then since $t_1 \leq t^*$ and $t_2 \leq t^*$ we get $t_1 = t_2 = t^*$. Moreover, using that

$$\epsilon t^* \mathbf{1}_E/P(E) \in \mathrm{ext}(\{u \in \mathrm{L}^2(\mathbb{R}^2) \,|\, \mathrm{TV}(u) \leq t^*\})\,,$$

we also obtain $u_1 = u_2 = u$. We therefore have $(\epsilon t^* \mathbf{1}_E/P(E), t^*) \in \mathrm{ext}(C)$. ∎

Résumé

On s'intéresse dans cette thèse à une famille de problèmes inverses, qui consistent à reconstruire une image à partir de mesures linéaires possiblement bruitées. On cherche à analyser les méthodes de reconstruction variationnelles utilisant un régulariseur spécifique, la variation totale (du gradient). Cette fonctionnelle est utilisée en imagerie depuis les travaux de Rudin Osher et Fatemi, menés en 1992. Alors qu'il est bien connu que sa minimisation produit des images constantes par morceaux, présentant une forme de parcimonie (elles sont composées d'un petit nombre de formes simples), ce point de vue n'a à notre connaissance pas été privilégié pour analyser les performances de ce régulariseur. Dans cette thèse, on se propose de mener cette étude.
Dans un premier temps, on considère les reconstructions obtenues par minimisation de la variation totale dans un régime de faible bruit, et on étudie leur proximité avec l'image inconnue. Puisque cette dernière est supposée parcimonieuse, on s'intéresse particulièrement à la structure de la reconstruction: est-elle elle-même parcimonieuse, est-elle composée du même nombre de formes, et ces formes sont-elles proches de celles présentes dans l'image inconnue ? Dans une seconde partie, on propose une méthode numérique pour résoudre les problèmes variationnels associés à ce régulariseur. On introduit un algorithme ne reposant pas sur l'introduction d'une discrétisation spatiale fixe. Ceci a l'avantage, contrairement aux techniques existantes, de n'introduire ni flou ni anisotropie dans les images reconstruites, et d'en produire une représentation parcimonieuse.

Mots clés

Problèmes inverses, variation totale, parcimonie.

Abstract

This thesis is devoted to the recovery of piecewise constant images from noisy linear measurements. We aim at analyzing variational reconstruction methods based on total (gradient) variation regularization. The total variation functional has been extensively used in imaging since the 90's. Its minimization is known to produce piecewise constant images, which hence have some kind of sparsity (they can be decomposed as the superposition of a few simple shapes). However, the performance of this regularizer has to our knowledge not extensively been studied from a sparse recovery viewpoint. This thesis aims at bridging this gap.
We first focus on noise robustness results. We assume that sought-after image is sparse, and study the structure of reconstructions in a low noise regime: are they sparse, made of the same number of shapes, and are these shapes close to those appearing in the unknown image? We then turn to numerical methods for total variation regularization. Existing techniques rely on the introduction of a fixed spatial discretization, which often yield reconstruction artifacts such as anisotropy or blur. We propose an algorithm which does not suffer from this grid bias, and produces a sparse representation of the reconstructed image.

Keywords

Inverse problems, total variation, sparsity.