Towards Off-the-grid Algorithms for Total Variation Regularized Inverse Problems

Yohann de Castro, Vincent Duval, Romain Petit

Towards off-the-grid algorithms for total variation regularized inverse problems

Yohann De Castro · Vincent Duval · Romain Petit

July 9, 2022

Abstract We introduce an algorithm to solve linear inverse problems regularized with the total (gradient) variation in a gridless manner. Contrary to most existing methods, that produce an approximate solution which is piecewise constant on a fixed mesh, our approach exploits the structure of the solutions and consists in iteratively constructing a linear combination of indicator functions of simple polygons.

Keywords Off-the-grid imaging · Inverse problems · Total variation

1 Introduction

By promoting solutions with a certain specific structure, the regularization of a variational inverse problem is a way to encode some prior knowledge on the signals to recover. Theoretically, it is now well understood which regularizers tend to promote signals or images which are sparse, low rank or piecewise constant. Yet, paradoxically enough, most numerical solvers are not designed with that goal in mind, and the targeted structural property (sparsity, low rank or piecewise constancy) only appears "in the limit", when the algorithm converges.

Several recent works have focused on incorporating structural properties in optimization algorithms. In the context of ℓ^1-based sparse spikes recovery, it was proposed to switch from, e.g. standard proximal methods (which require the introduction of an approximation grid) to algorithms which operate directly in a continuous domain: interior point methods solving a reformulation of the problem [Candès and Fernandez-Granda, 2014, Castro et al., 2017] or a Frank-Wolfe / conditional gradient algorithm [Bredies and Pikkarainen, 2013] approximating a solution in a greedy way. More generally, the conditional gradient algorithm has drawn a lot of interest from the data science community, for it provides iterates which are a sum of a small number of atoms which are promoted by the regularizer (see the review paper [Jaggi, 2013]).

In the present work, we explore the extension of these fruitful approaches to the total (gradient) variation regularized inverse problem

$$\min_{u\in \mathrm{L}^2(\mathbb{R}^2)} T_\lambda(u) \stackrel{\text{def.}}{=} \frac{1}{2}\left\|\Phi u - y\right\|^2 + \lambda \left|\mathrm{D}u\right|(\mathbb{R}^2)\,, \qquad (\mathcal{P}_\lambda)$$

where $|\mathrm{D}u|(\mathbb{R}^2)$ denotes the total variation of (the gradient of) u and $\Phi\colon \mathrm{L}^2(\mathbb{R}^2) \to \mathbb{R}^m$ is a continuous linear map such that

$$\forall u \in \mathrm{L}^2(\mathbb{R}^2),\ \Phi u = \int_{\mathbb{R}^2} u(x)\,\varphi(x)\,\mathrm{d}x\,, \qquad (1)$$

with $\varphi \in \left[\mathrm{L}^2(\mathbb{R}^2)\right]^m \cap C^0(\mathbb{R}^2, \mathbb{R}^m)$. Such variational problems have been widely used in imaging for the last

This work was supported by a grant from Région Ile-De-France and by the ANR CIPRESSI project, grant ANR-19-CE48-0017-01 of the French Agence Nationale de la Recherche.

Y. De Castro
Institut Camille Jordan, CNRS UMR 5208, École Centrale de Lyon, F-69134 Écully, France
E-mail: yohann.de-castro@ec-lyon.fr

V. Duval
CEREMADE, CNRS, UMR 7534, Université Paris-Dauphine, PSL University, 75016 Paris, France
INRIA-Paris, MOKAPLAN, 75012 Paris, France
E-mail: vincent.duval@inria.fr

R. Petit
CEREMADE, CNRS, UMR 7534, Université Paris-Dauphine, PSL University, 75016 Paris, France
INRIA-Paris, MOKAPLAN, 75012 Paris, France
E-mail: romain.petit@inria.fr

decades, following the pioneering work of Rudin, Osher and Fatemi [Rudin et al., 1992]. A typical application is the reconstruction of an unknown image u_0 from a set of noisy linear measurements $y = \Phi u_0 + w$, where $w \in \mathbb{R}^m$ is some additive noise.

The total variation term in $(\mathcal{P}_\lambda)$ is known to promote piecewise constant solutions. It has been shown that some solutions are sums of at most m indicator functions of simple sets (see [Bredies and Carioni, 2019, Boyer et al., 2019]). However, when u_0 is a simple piecewise constant image, there are evidences that solutions are usually made of a much smaller number of shapes. In such situations, it is highly desirable to design numerical solvers preserving this structure, and able to accurately estimate the jump set of solutions. This could be particularly relevant for specific applications, like astronomical and cell imaging.

1.1 Previous works

Many algorithms have been proposed to solve $(\mathcal{P}_\lambda)$. The vast majority of them rely on the introduction of a fixed spatial discretization, and of a discrete version of the total variation (see [Chambolle and Pock, 2021] for a review). These approaches often yield reconstruction artifacts, such as anisotropy or blur (see the previous reference, [Tabti et al., 2018], and the experiments section below). Most importantly, existing algorithms often fail to preserve the structure exhibited by solutions of $(\mathcal{P}_\lambda)$, which is discussed above.

To circumvent these issues, mesh adaptation techniques were introduced in [Viola et al., 2012, Bartels et al., 2021]. The refinement rules they 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.

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.

1.2 Contributions

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 construct an approximate solution built from the above-mentioned atoms, namely indicator functions of simple sets. As shown in the experiments section, this approach is particularly suited for reconstructing simple piecewise constant images. On more complex natural images, traditional grid-based methods perform better. In Sections 3 and 4, we introduce a theoretical iterative algorithm, whose output provably converges to a solution of $(\mathcal{P}_\lambda)$. The exploratory nature of our work lies in the numerical methods we propose to carry out several steps of this algorithm. Although experiments suggest they perform well, several questions concerning their theoretical analysis remain.

2 Preliminaries

In the following, for any function $u : \mathbb{R}^2 \to \mathbb{R}$, we shall use the notation

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

2.1 Functions of bounded variation and sets of finite perimeter

Let $u \in \mathrm{L}^1_{loc}(\mathbb{R}^2)$. The total variation of u is given by

$$J(u) \overset{\text{def.}}{=} \sup_{z \in C_c^\infty(\mathbb{R}^2, \mathbb{R}^2)} -\int_{\mathbb{R}^2} u \operatorname{div} z \quad \text{s.t.} \quad \|z\|_\infty \leq 1\,.$$

If $J(u)$ is finite, then u is said to have bounded variation, and the distributional gradient of u, denoted $\mathrm{D}u$, is a finite vector-valued Radon measure. We moreover have $|\mathrm{D}u|(\mathbb{R}^2) = J(u) < +\infty$.

A measurable set $E \subset \mathbb{R}^2$ is said to be of finite perimeter if $P(E) \overset{\text{def.}}{=} J(\mathbf{1}_E) < +\infty$. 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 is moreover such that $\|\nu_E(x)\| = 1$.

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 the following, we always consider such a representative and consequently obtain $\operatorname{Supp}(\mathrm{D}\mathbf{1}_E) = \overline{\partial^* E} = \partial E$.

We now introduce the notion of indecomposable and simple sets, which are the measure-theoretic analogues of connected and simply connected sets (see [Ambrosio

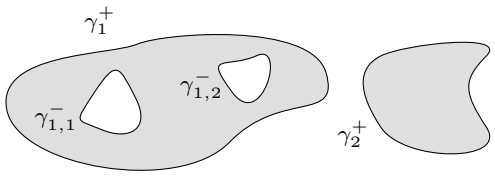

Fig. 1: Decomposition of a set (gray area) as in (2), (3)

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 with $P(E) = P(A) + P(B)$. We say that E is indecomposable if it is not decomposable. Any indecomposable set of finite measure whose complement is also indecomposable is called simple. If $E \subset \mathbb{R}^2$ has finite perimeter and finite measure it can be decomposed (up to Lebesgue negligible sets) into an at most countable union of pairwise disjoint indecomposable sets, i.e.

$$E = \bigcup_{i\in I} E_i,\ P(E) = \sum_{i\in I} P(E_i) \text{ and } \forall i,\ |E_i| > 0\,. \tag{2}$$

Each E_i can in turn be decomposed as

$$\begin{aligned} E_i &= \mathrm{int}(\gamma_i^+) \setminus \bigcup_{j\in J_i} \mathrm{int}(\gamma_{i,j}^-)\,, \\ \text{with } P(E_i) &= P(\mathrm{int}(\gamma_i^+)) + \sum_{j\in J_i} P(\mathrm{int}(\gamma_{i,j}^-))\,, \end{aligned} \tag{3}$$

where for all $i \in I$ and $j \in J_i$, γ_i^+ and $\gamma_{i,j}^-$ are rectifiable Jordan curves.

2.2 Subdifferential of the total variation

In the rest of this document, J is considered as a mapping from $\mathrm{L}^2(\mathbb{R}^2)$ to $\mathbb{R} \cup \{+\infty\}$. This mapping is convex, proper and lower semi-continuous. We have the following useful characterizations of $\partial J(0)$:

$$\begin{aligned} \partial J(0) &= \left\{ \eta \in \mathrm{L}^2(\mathbb{R}^2) \,\middle|\, \right. \\ &\qquad \left. \forall u \in \mathrm{L}^2(\mathbb{R}^2),\ \left| \int_{\mathbb{R}^2} \eta\, u \right| \le |\mathrm{D}u|(\mathbb{R}^2) \right\}, \\ \partial J(0) &= \left\{ \eta \in \mathrm{L}^2(\mathbb{R}^2) \,\middle|\, \forall E \subset \mathbb{R}^2,\ 0 < |E| < +\infty \right. \\ &\qquad \left. \text{and } P(E) < +\infty \implies \left| \int_{\mathbb{R}^2} \eta\, \frac{\mathbf{1}_E}{P(E)} \right| \le 1 \right\}. \end{aligned}$$

Moreover, the subdifferential of J at $u \in \mathrm{L}^2(\mathbb{R}^2)$ is given by:

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

We also have the following useful result:

Proposition 1 (see e.g. [Chambolle et al., 2016]) *Let $u \in \mathrm{L}^2(\mathbb{R}^2)$ be such that $J(u) < \infty$ and $\eta \in \mathrm{L}^2(\mathbb{R}^2)$. Then $\eta \in \partial J(u)$ if and only if $\eta \in \partial J(0)$ and the level sets of u satisfy*

$$\begin{cases} \forall t > 0,\ P(U^{(t)}) = \displaystyle\int_{U^{(t)}} \eta\,, \\ \forall t < 0,\ P(U^{(t)}) = -\displaystyle\int_{U^{(t)}} \eta\,. \end{cases} \tag{4}$$

2.3 Dual problem and dual certificates

The Fenchel-Rockafellar dual of ($\mathcal{P}_\lambda$) is the following finite dimensional problem

$$\max_{p\in\mathbb{R}^m} \langle p, y\rangle - \frac{\lambda}{2}\|p\|^2 \quad \text{s.t.} \quad \Phi^* p \in \partial J(0)\,, \tag{$\mathcal{D}_\lambda$}$$

which has a unique solution (it is in fact equivalent to the projection of $\frac{y}{\lambda}$ on the closed convex set of vectors p such that $\Phi^* p \in \partial J(0)$). Moreover, strong duality holds as stated by the following proposition

Proposition 2 *Problems ($\mathcal{P}_\lambda$) and ($\mathcal{D}_\lambda$) have the same value and any solution u_λ of ($\mathcal{P}_\lambda$) is linked with the unique solution p_λ of ($\mathcal{D}_\lambda$) by the extremality condition*

$$\begin{cases} \Phi^* p_\lambda \in \partial J(u_\lambda)\,, \\ p_\lambda = -\dfrac{1}{\lambda}\left(\Phi u_\lambda - y\right)\,. \end{cases} \tag{5}$$

Remark 1 Proposition 2 implies in particular that all solutions of ($\mathcal{P}_\lambda$) have the same total variation and the same image by Φ.

2.4 Distributional curvature

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

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

If $E \subset \mathbb{R}^2$ is a set of finite perimeter, then the distributional curvature vector of E is $\mathbf{H}_E : C_c^\infty(\mathbb{R}^2, \mathbb{R}^2) \to \mathbb{R}$ defined by

$$\forall T \in C_c^\infty(\mathbb{R}^2, \mathbb{R}^2),\ \langle \mathbf{H}_E, T\rangle = \int_{\partial^* E} \mathrm{div}_E\, T\, d\mathcal{H}^1\,,$$

where $\mathrm{div}_E\, T$ denotes the tangential divergence of T on E given by

$$\mathrm{div}_E\, T = \mathrm{div}\, T - (\mathrm{D}T\, \nu_E)\cdot \nu_E\,,$$

where DT denotes the differential of T. E is said to have locally integrable distributional curvature if there exists a function $H_E \in \mathrm{L}^1_{loc}(\partial^* E; \mathcal{H}^1)$ 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.

3 A modified Frank-Wolfe algorithm

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 $s_k \in \underset{s \in C}{\mathrm{Argmin}} f(x_k) + df(x_k)\,[s - x_k]$;
3 **if** $df(x_k)\,[s_k - x_k] = 0$ **then**
4 output $x^* \leftarrow x_k$, which is optimal;
5 **else**
6 $\gamma_k \leftarrow \frac{2}{k+2}$;
  // tentative update
7 $\tilde{x}_{k+1} \leftarrow x_k + \gamma_k (s_k - x_k)$;
  // final update
8 choose any x_{k+1} such that $f(x_{k+1}) \leq f(\tilde{x}_{k+1})$;
9 **end**
10 **end**

In the spirit of [Bredies and Pikkarainen, 2013, Boyd et al., 2017, Denoyelle et al., 2019] which introduced variants of the conditional gradient algorithm for sparse spikes recovery in a continuous domain, we derive a modified Frank-Wolfe algorithm allowing to iteratively solve ($\mathcal{P}_\lambda$) in a gridless manner.

3.1 Description

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

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 any $x_{k+1} \in C$ such that $f(x_{k+1}) \leq f(\tilde{x}_{k+1})$ instead.

Even though T_λ is not differentiable, it is possible to recast problem ($\mathcal{P}_\lambda$) into that framework by performing an epigraphical lift (see Appendix A). In this setting, the linear minimization step which is at the core of the algorithm amounts to solving the following problem

$$\min_{u \in \mathrm{L}^2(\mathbb{R}^2)} \int_{\mathbb{R}^2} \eta \, u \quad \text{s.t.} \quad |\mathrm{D}u|\,(\mathbb{R}^2) \leq 1 \, , \tag{6}$$

for an iteration-dependent function $\eta \in \mathrm{L}^2(\mathbb{R}^2)$. Denoting $u^{[k]}$ the k-th iterate, this function is given by

$$\eta^{[k]} \overset{\text{def.}}{=} -\frac{1}{\lambda} \Phi^* \left(\Phi u^{[k]} - y \right) .$$

As is usual when using the Frank-Wolfe algorithm, we notice that since the objective of (6) is linear and the total variation unit ball is convex and compact (in the weak $\mathrm{L}^2(\mathbb{R}^2)$ topology), at least one of its extreme points is optimal. A result due to Fleming [Fleming, 1957] (see also [Ambrosio et al., 2001]) states that those extreme points are exactly the functions of the form $\pm \mathbf{1}_E / P(E)$ where $E \subseteq \mathbb{R}^2$ is a simple set with $0 < |E| < +\infty$. This means the linear minimization step can be carried out by finding a simple set solving the following geometric variational problem:

$$\max_{E \subseteq \mathbb{R}^2} \frac{\left| \int_E \eta \right|}{P(E)} \quad \text{s.t.} \quad 0 < |E| < +\infty, \; P(E) < +\infty \, . \tag{7}$$

Since Problem (7) is reminiscent of the Cheeger problem [Parini, 2011], which, given a domain $\Omega \subseteq \mathbb{R}^2$, consists in finding the subsets E of Ω minimizing the ratio $P(E)/|E|$, we refer to it as the "Cheeger problem" in the rest of the paper, and to any of its solutions as a "Cheeger set".

In view of the above, we derive Algorithm 2, which produces a sequence of functions that are linear combinations of indicators of simple sets, and which is a valid application of Algorithm 1 to ($\mathcal{P}_\lambda$), in the sense that Proposition 3 holds.

Remark 2 We use here a so-called "fully corrective" variant of Frank-Wolfe, meaning that instead of choosing the next iterate $u^{[k+1]}$ as a convex combination of $\pm \mathbf{1}_{E_*} / P(E_*)$ and the previous iterate $u^{[k]}$ as in Line 7 of Algorithm 1, we optimize (Line 10 of Algorithm 2) the objective over $\mathrm{Vect}\left((\mathbf{1}_{E_i})_{i=1}^{N^{[k]}+1} \right)$, which decreases the objective more than the standard update, and hence does not break convergence guarantees.

Remark 3 Line 10 of Algorithm 2 can always be reduced to the resolution of a LASSO-type problem (possibly changing $E^{[k+1]}$ and constraining the sign of the components of a). Indeed, given $N \in \mathbb{N}^*$ and $E_1, ..., E_N$ a collection of simple sets, assuming that we have

$$\forall a \in \mathbb{R}^N, \quad \left| \mathrm{D} \left(\sum_{i=1}^N a_i \, \mathbf{1}_{E_i} \right) \right| (\mathbb{R}^2) = \sum_{i=1}^N |a_i| \, P(E_i) \, , \tag{8}$$

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

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 $\quad E_* \leftarrow \underset{E \text{ simple}}{\text{Argmax}} \frac{\left|\int_E \eta^{[k]}\right|}{P(E)}$ s.t. $0 < |E| < +\infty$;
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 $\quad\quad a^{[k+1]} \leftarrow \underset{a \in \mathbb{R}^{N^{[k]}+1}}{\text{argmin}} T_\lambda\left(\sum_{i=1}^{N^{[k]}+1} a_i \mathbf{1}_{E_i^{[k+1]}}\right)$;
11 $\quad\quad$ remove atoms with zero amplitude;
12 $\quad\quad N^{[k+1]} \leftarrow$ number of atoms in $E^{[k+1]}$;
13 $\quad\quad u^{[k+1]} \leftarrow \sum_{i=1}^{N^{[k+1]}} a_i^{[k+1]} \mathbf{1}_{E_i^{[k+1]}}$;
14 $\quad$ **end**
15 **end**

then we get that

$$T_\lambda(u) = \frac{1}{2}||\Phi_E\, a - y||^2 + \lambda \sum_{i=1}^{N} P(E_i)\,|a_i|\,,$$

with

$$\Phi_E \overset{\text{def.}}{=} \left[\left(\int_{E_i} \varphi_j\right)_{\substack{1\leq i\leq N\\ 1\leq j\leq m}}\right]^T \in \mathbb{R}^{m\times N}\,.$$

Hence, finding the vector a minimizing $T_\lambda(u)$ with the sets $E_1, ..., E_N$ fixed amounts to solving a finite dimensional least squares problem with a weighted ℓ^1 norm penalization (the weights are here the perimeters of the sets $(E_i)_{i=1}^N$).

Identity (8) holds as soon as $\mathcal{H}^1(\partial^* E_i \cap \partial^* E_j) = 0$ for every $i \neq j$. Although this is generically satisfied, and that we never observe experimentally configurations where this fails, we describe in Appendix B how to change $E^{[k+1]}$ to reduce Line 10 to a LASSO-type problem at the price of constraining the sign of the components of a.

Remark 4 The stopping condition is here replaced by

$$\sup_E \frac{\left|\int_E \eta^{[k]}\right|}{P(E)} \leq 1, \text{ with } \eta^{[k]} = -\frac{1}{\lambda}\Phi^*\left(\Phi u^{[k]} - y\right),$$

which is equivalent to $\eta^{[k]} \in \partial J(0)$. Since the optimality of $a^{[k]}$ at Line 10 always ensures $\int_{\mathbb{R}^2} \eta^{[k]}\, u^{[k]} = J(u^{[k]})$, this yields $\eta^{[k]} \in \partial J(u^{[k]})$ and hence (5) holds, which means $u^{[k]}$ solves $(\mathcal{P}_\lambda)$.

3.2 Convergence results

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

Proposition 3 *Let $(u^{[k]})_{k\geq 0}$ be a sequence produced by Algorithm 2. Then there exists $C > 0$ such that for any solution u^* of Problem $(\mathcal{P}_\lambda)$,*

$$\forall k \in \mathbb{N}^*,\ T_\lambda(u^{[k]}) - T_\lambda(u^*) \leq \frac{C}{k}\,. \tag{9}$$

Remark 5 As discussed in [Jaggi, 2013], the linear minimization step (solving (6) or equivalently (7)) can be solved approximately. In fact if there exists $\delta > 0$ such that for every k the set computed at Line 5 is an ϵ_k-maximizer of (7) with $\epsilon_k = \frac{\gamma}{k+2}\delta$, then

$$\forall k \in \mathbb{N}^*,\ T_\lambda(u^{[k]}) - T_\lambda(u^*) \leq \frac{2\,\gamma}{k+2}(1+\delta)\,, \tag{10}$$

where γ is the curvature constant of the objective used in the reformulation of $(\mathcal{P}_\lambda)$. One can in fact show that this curvature constant is smaller than a quantity which is proportional to $\|\Phi\|^2\ \left(\|y\|^2/\lambda\right)^2$.

We first provide a general property of minimizing sequences (see e.g. [Iglesias et al., 2018] for a proof), which hence applies to the sequence of iterates produced by Algorithm 2.

Proposition 4 *Let $(u_n)_{n\geq 0}$ be a minimizing sequence for $(\mathcal{P}_\lambda)$. Then there exists a subsequence (not relabeled) which converges weakly in $\mathrm{L}^2(\mathbb{R}^2)$ and strongly in $\mathrm{L}^1_{loc}(\mathbb{R}^2)$ to a solution u_* of $(\mathcal{P}_\lambda)$. Moreover, we have $\mathrm{D}u_n \overset{*}{\rightharpoonup} \mathrm{D}u_*$ and $|\mathrm{D}u_n|(\mathbb{R}^2) \to |\mathrm{D}u|(\mathbb{R}^2)$.*

We now provide additional properties of sequences produced by Algorithm 2. We first begin by noticing that if $(u^{[k]})_{k\geq 0}$ is such a sequence, then the optimality condition at Line 10 ensures that

$$\forall k,\ \forall i \in \{1, ..., N^{[k]}\},\ P(E_i^{[k]}) = \left|\int_{E_i^{[k]}} \eta^{[k]}\right|.$$

But from Proposition 3 and Proposition 4 we have the existence of a (not relabeled) subsequence which converges strongly in $\mathrm{L}^1_{loc}(\mathbb{R}^2)$ and weakly in $\mathrm{L}^2(\mathbb{R}^2)$ towards a solution u^* of $(\mathcal{P}_\lambda)$. The weak convergence of $(u^{[k]})_{k\geq 0}$ in $\mathrm{L}^2(\mathbb{R}^2)$ implies that $\lim_{n\to+\infty} \Phi u^{[k]} = \Phi u^*$, which in turns yields the strong convergence in $\mathrm{L}^2(\mathbb{R}^2)$ of $(\eta^{[k]})_{k\geq 0}$ towards the solution η^* of $(\mathcal{D}_\lambda)$. We can then use the following lemma to show all the sets $E_i^{[k]}$ are included in some common ball.

Lemma 1 *Let $(\eta_k)_{k\geq 0}$ be a sequence of functions converging strongly to η_∞ in $\mathrm{L}^2(\mathbb{R}^2)$. For all $k \geq 0$, we denote*

$$\mathcal{F}_k \overset{\text{def.}}{=} \left\{ E \text{ simple} \,\middle|\, 0 < |E| < +\infty,\ P(E) = \left| \int_E \eta_k \right| \right\},$$

and $\mathcal{F} = \cup_{k\geq 0}\mathcal{F}_k$. Then there exist positive real numbers R and C such that

$$\forall E \in \mathcal{F},\ P(E) \leq C \text{ and } E \subset B(0,R)\,.$$

Proof. This proof is based on [Chambolle et al., 2016, Section 5].

Upper bound on the perimeter: the family of functions $\{\eta_k^2,\ k \in \mathbb{N}\} \cup \{\eta_\infty^2\}$ being equi-integrable, for all $\epsilon > 0$ there exists $R_1 > 0$ such that

$$\forall k,\ \int_{\mathbb{R}^2\setminus B(0,R_1)} \eta_k^2 \leq \epsilon^2\,.$$

Let $E \in \mathcal{F}$. Then there exists k s.t. $P(E) = \left|\int_E \eta_k\right|$ and we have:

$$\begin{aligned}\left|\int_E \eta_k\right| &\leq \left|\int_{E\cap B(0,R_1)} \eta_k\right| + \left|\int_{E\setminus B(0,R_1)} \eta_k\right| \\ &\leq \sqrt{|B(0,R_1)|}\ ||\eta_k||_{L^2} \\ &\quad + \sqrt{|E\setminus B(0,R_1)|}\sqrt{\int_{\mathbb{R}^2\setminus B(0,R_1)} \eta_k^2} \\ &\leq \sup_k ||\eta_k||_{L^2}\sqrt{|B(0,R_1)|} + \epsilon\sqrt{|E\setminus B(0,R_1)|}\,.\end{aligned}$$

Moreover,

$$\sqrt{|E\setminus B(0,R_1)|} \leq \frac{1}{\sqrt{c_2}}(P(E) + P(B(0,R_1)))\,,$$

where $c_2 \overset{\text{def.}}{=} 4\pi$ is the isoperimetric constant. Hence taking $\epsilon \overset{\text{def.}}{=} \frac{\sqrt{c_2}}{2}$ and defining

$$C \overset{\text{def.}}{=} 2\left(\sqrt{|B(0,R_1)|}\sup_k ||\eta_k||_{L^2} + \frac{1}{2}P(B(0,R_1))\right),$$

we have $P(E) \leq \frac{1}{2}P(E) + \frac{C}{2}$ and hence $P(E) \leq C$.

Inclusion in a ball: we still take $\epsilon = \frac{\sqrt{c_2}}{2}$ and fix a real $R_2 > 0$ such that $\int_{\mathbb{R}^2\setminus B(0,R_2)} \eta_k^2 \leq \epsilon^2$ for all k. Now let $E \in \mathcal{F}$ and k such that

$$P(E) = \left|\int_E \eta_k\right|.$$

Let us show that $E \cap B(0,R_2) \neq \emptyset$. By contradiction, if $E \cap B(0,R_2) = \emptyset$, we would have:

$$\begin{aligned}P(E) = \left|\int_{E\setminus B(0,R_2)} \eta_k\right| &\leq \sqrt{\int_{\mathbb{R}^2\setminus B(0,R_2)} \eta_k^2}\ \sqrt{|E|} \\ &\leq \frac{\epsilon}{\sqrt{c_2}}P(E)\,.\end{aligned}$$

Dividing by $P(E)$ (which is positive since $0 < |E| < \infty$) yields a contradiction. Since E is simple, the perimeter bound yields $\mathrm{diam}(E) \leq C$, which shows $E \subset B(0,R)$ with $R \overset{\text{def.}}{=} C + R_2$. □

We have now shown there exists $R > 0$ such that for all k we have $\mathrm{Supp}(u^{[k]}) \subset B(0,R)$, which means the strong L^1_{loc} convergence of $(u^{[k]})_{k\geq 0}$ towards u^* is in fact a strong L^1 convergence. This slightly improved convergence result is summarized in the following proposition:

Proposition 5 *Let $(u_n)_{n\geq 0}$ be a sequence produced by Algorithm 2. Then there exists a (not relabeled) subsequence and $R > 0$ such that $Supp(u_n) \subset B(0,R)$ for all n. Moreover, this subsequence converges strongly in $\mathrm{L}^1(\mathbb{R}^2)$ to a solution u_* of $(\mathcal{P}_\lambda)$ (and by Proposition 4 weakly in $\mathrm{L}^2(\mathbb{R}^2)$, with moreover $\mathrm{D}u_n \overset{*}{\rightharpoonup} \mathrm{D}u_*$ and $|\mathrm{D}u_n|(\mathbb{R}^2) \to |\mathrm{D}u|(\mathbb{R}^2)$).*

Corollary 1 *Let $(u_n)_{n\geq 0}$ be a subsequence such as in Proposition 5. Up to another extraction, for almost every $t \in \mathbb{R}$, we have*

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

where[1]

$$\liminf_{n\to+\infty} \partial U_n^{(t)} \overset{\text{def.}}{=} \left\{x \in \mathbb{R}^2 \,\middle|\, \limsup_{n\to+\infty} \mathrm{dist}(x, \partial U_n^{(t)}) = 0\right\}.$$

Proof. The strong convergence of $(u_n)_{n\geq 0}$ towards a solution u_* in $\mathrm{L}^1(\mathbb{R}^2)$ and Fubini's theorem give

$$0 = \lim_{n\to+\infty} \int_{\mathbb{R}^2} |u_n - u_*| = \lim_{n\to+\infty} \int_{\mathbb{R}} \left|U_n^{(t)} \triangle U_*^{(t)}\right| dt\,.$$

Hence, up to the extraction of a further subsequence, that we do not relabel, we get that

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

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

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

[1] For more details on this type of set convergence, see e.g. [Rockafellar and Wets, 1998, Chapter 4].

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

$$\forall n \in \mathbb{N},\ B(x,r) \cap \partial U^{(t)}_{\varphi(n)} = \emptyset\,.$$

Hence for all n, we either have

$$B(x,r) \subset U^{(t)}_{\varphi(n)} \quad \text{or} \quad B(x,r) \subset \mathbb{R}^2 \setminus U^{(t)}_{\varphi(n)}\,.$$

If $B(x,r) \subset U^{(t)}_{\varphi(n)}$ for a given n then

$$\begin{aligned}\left|U^{(t)}_{\varphi(n)} \triangle U^{(t)}_*\right| &\geq \left|U^{(t)}_{\varphi(n)} \setminus U^{(t)}_*\right| \geq \left|B(x,r) \setminus U^{(t)}_*\right| \\ &\geq C\,|B(x,r)|\,.\end{aligned}$$

The last inequality, which is a weak regularity property of $U^{(t)}_*$, holds for all r smaller than some $r_0 > 0$, for some constant C that is independent of r and x (see [Chambolle et al., 2016, Prop. 7]). We can in the same way show

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

if $B(x,r) \subset \mathbb{R}^2 \setminus U^{(t)}_{\varphi(n)}$ and hence get the inequality for all n. Using that $\lim_{n\to+\infty} |U^{(t)}_n \triangle U^{(t)}_*| = 0$, we get a contradiction. □

4 Sliding step

Several works [Bredies and Pikkarainen, 2013, Rao et al., 2015, 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 very end of an iteration, the gradient flow of the mapping

$$(a,E) \mapsto T_\lambda\left(\sum_{i=1}^{N^{[k+1]}} a_i \mathbf{1}_{E_i}\right) \tag{11}$$

initialized with $(a^{[k+1]}, E^{[k+1]})$, so as to find a set of parameters at which the objective is smaller. Formally, this would correspond[2] to finding a curve

$$t \mapsto (a_i(t), E_i(t))_{i=1}^{N^{[k+1]}}$$

such that for all t

$$\begin{cases} a_i'(t) = -\lambda\left(\operatorname{sign}(a_i(t))\,P(E_i(t)) - \displaystyle\int_{E_i(t)} \eta(t)\right), \\ V_i(t) = -\lambda\,|a_i(t)|\left(H_{E_i(t)} - \operatorname{sign}(a_i(t))\,\eta(t)\right), \end{cases} \tag{12}$$

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

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

The study of this gradient flow (existence, uniqueness) is out of the scope of this paper.

For our purpose, it is enough to introduce a sliding step which improves the objective by performing a local descent on

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

initialized with $(a^{[k+1]}, E^{[k+1]})$, that is to find a set of parameters $(a_i, E_i)_{i=1}^{N^{[k+1]}}$ such that E_i is simple for all i with

$$T_\lambda\left(\sum_{i=1}^{N^{[k+1]}} a_i\,\mathbf{1}_{E_i}\right) \leq T_\lambda\left(\sum_{i=1}^{N^{[k+1]}} a_i^{[k+1]}\,\mathbf{1}_{E_i^{[k+1]}}\right). \tag{13}$$

The resulting algorithm, which is Algorithm 3, is a valid application of Algorithm 1 to $(\mathcal{P}_\lambda)$. Moreover, Line 14 ensures that all convergence guarantees derived for Algorithm 2 remain valid.

The sliding step (Line 13 of Algorithm 3) was first introduced in [Bredies and Pikkarainen, 2013]. It allows in practice to considerably improve the convergence speed of the algorithm, and also produces sparser solutions: if the solution is expected to be a linear combination of a few indicator functions, removing the sliding step will typically produce iterates made of a much larger number of indicator functions, the majority of them correcting the crude approximations of the support of the solution made during the first iterations.

In [Denoyelle et al., 2019], the introduction of this 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 every iteration, a "critical point" of the objective can be reached at the end of the sliding step. In our case, the above mentioned existence issues make the adaptation of these results difficult. However, if the existence of a curve (formally) satisfying (12) could be guaranteed for all times, then one would expect it to converge when t goes to infinity to a critical point of the mapping defined in (11), in the sense of the following definition.

[2] The formulas given in (12) can be formally obtained by using the notion of shape derivative, see [Henrot and Pierre, 2018, Chapter 5].

Algorithm 3: modified Frank-Wolfe algorithm applied to ($\mathcal{P}_\lambda$) (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 $\quad E_* \leftarrow \underset{E \text{ simple}}{\text{Argmax}} \frac{\left|\int_E \eta^{[k]}\right|}{P(E)}$ s.t. $0 < |E| < +\infty$;
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 $\quad\quad a^{[k+1]} \leftarrow \underset{a \in \mathbb{R}^{N^{[k]}+1}}{\text{argmin}} T_\lambda \left(\sum_{i=1}^{N^{[k]}+1} a_i \mathbf{1}_{E_i^{[k+1]}} \right)$;
11 $\quad\quad$ remove atoms with zero amplitude;
12 $\quad\quad N^{[k+1]} \leftarrow$ number of atoms in $E^{[k+1]}$;
13 $\quad\quad$ perform a local descent on $(a, E) \mapsto T_\lambda \left(\sum_{i=1}^{N^{[k+1]}} a_i \mathbf{1}_{E_i} \right)$ initialized with $(a^{[k+1]}, E^{[k+1]})$;
14 $\quad\quad$ repeat the operations of Lines 10-12;
15 $\quad\quad u^{[k+1]} \leftarrow \sum_{i=1}^{N^{[k+1]}} a_i^{[k+1]} \mathbf{1}_{E_i^{[k+1]}}$;
16 $\quad$ **end**
17 **end**

Definition 1 Let $N \in \mathbb{N}^*$, $a \in \mathbb{R}^N$ and $E_1, ..., E_N$ be subsets of $\mathbb{R}^2$ such that $|E_i| < +\infty$, $P(E_i) < +\infty$ for all $i \in \{1, ..., N\}$ and (8) holds. We say that $(a_i, E_i)_{i=1}^N$ is a critical point of the mapping

$$(a, E) \mapsto T_\lambda \left(\sum_{i=1}^{N} a_i \mathbf{1}_{E_i} \right)$$

if for all $i \in \{1, ..., N\}$ we either have $a_i \neq 0$ and

$$\begin{cases} P(E_i) = \text{sign}(a_i) \int_{E_i} \eta \,, \\ H_{E_i} = \text{sign}(a_i)\, \eta \,, \end{cases} \tag{14}$$

or $a_i = 0$ and $\left|\int_{E_i} \eta\right| \leq P(E_i)$, where

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

In Remark 6, we discuss how assuming a critical point is indeed reached at the end of the sliding step for every iteration could be used to derive additional properties of sequences produced by Algorithm 3. We stress that if, for a given iteration, a critical point is reached at the end of the sliding step, then Line 14 can be skipped, since the first equality in (14) and the inequality given above in the case of a zero amplitude ensure $a^{[k+1]}$ is already optimal for the problem to be solved.

Remark 6 As mentioned above, the introduction of the sliding step is supposed to allow the derivation of improved convergence properties. If its output is a critical point in the sense of Definition 1, a first remark we can make is that for all $i \in \{1, ..., N^{[k]}\}$ the set $E_i^{[k]}$ has distributional curvature $\text{sign}(a_i^{[k]})\, \eta^{[k]}$. This can be exploited to obtain "uniform" density estimates for the level sets of $u^{[k]}$ in the spirit of [Maggi, 2012, Corollary 17.18]. One could then wonder whether this weak regularity of the level sets could be used to prove

$$\limsup_{n \to +\infty} \partial U_n^{(t)} \subseteq \partial U_*^{(t)} \,, \tag{15}$$

where

$$\limsup_{n \to +\infty} \partial U_n^{(t)} \overset{\text{def.}}{=} \left\{ x \in \mathbb{R}^2 \,\middle|\, \liminf_{n \to +\infty} \text{dist}(x, \partial U_n^{(t)}) = 0 \right\} .$$

This, combined with the result of Corollary 1 and the fact $(\partial U_n^{(t)})_{n \geq 0}$ is uniformly bounded would mean that

$$\lim_{n \to +\infty} \partial U_n(t) = \partial U_*^{(t)}$$

in the Hausdorff sense (see [Rockafellar and Wets, 1998] for more details).

A major obstacle towards this result is that, although Lemma 1 provides a uniform upper bound on the perimeter of the atoms involved in the definition of the iterates, to our knowledge, it does not seem possible to derive such a bound for the perimeter of their level sets, which prevents one from using the potential weak-* convergence of $\mathrm{D}\mathbf{1}_{U_n^{(t)}}$ towards $\mathrm{D}\mathbf{1}_{U_*^{(t)}}$.

5 Implementation

The implementation[3] of Algorithm 3 requires two oracles to carry out the operations described on Lines 5 and 13 (recall Line 10 can always be reduced to a LASSO-type problem which can efficiently be solved by existing solvers): a first one that, given a weight function η, returns a solution of (7), and a second one that, given a collection of real numbers and simple sets, returns another such collection with a lower objective value. Our approach for designing these oracles relies on polygonal approximations: we fix an integer $n \geq 3$ (that might be

[3] A complete implementation of Algorithm 3 can be found online at `https://github.com/rpetit/tvsfw` (see also `https://github.com/rpetit/PyCheeger`).

iteration-dependent), look for a maximizer of $\mathcal{J}$ defined by

$$\mathcal{J}(E) \overset{\text{def.}}{=} \frac{1}{P(E)} \left| \int_E \eta \right|$$

among simple polygons with at most n sides, and perform the sliding step by finding a collection of real numbers and simple polygons satisfying (13). This choice is mainly motivated by our goal to solve ($\mathcal{P}_\lambda$) "off-the-grid", which naturally leads us to consider purely Lagrangian methods which do not rely on the introduction of a pre-defined discrete grid.

5.1 Polygonal approximation of Cheeger sets

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[4] $\cup_{i=1}^n [x_i, x_{i+1}]$ is a Jordan curve. It hence divides the plane in two regions, one of which is bounded. We denote this region E_x (it is hence a simple polygon). When x spans $\mathcal{X}_n$, E_x spans $\mathcal{P}_n$, the set of simple polygons with at most n sides. The sets we wish to approximate in this section (in order to carry out Line 5 in Algorithm 3) are the maximizers of $\mathcal{J}$ over $\mathcal{P}_n$. We prove their existence in Appendix C.

The approximation method presented thereafter consists of several steps. First, we solve a discrete version of (6), where the minimization is performed over the set of piecewise constant functions on a fixed grid. Then, we extract a level set of the solution, and obtain a simple polygon whose edges are located on the edges of the grid. Finally, we use a first order method initialized with the previously obtained polygon to locally maximize $\mathcal{J}$.

5.1.1 Fixed grid step

Every solution of (6) has its support included in some ball (indeed if u solves (6), then there exists α such that $\alpha\,\eta \in \partial J(u)$, and the result follows from Proposition 1 and Lemma 1). We can hence solve (6) in $[-R, R]^2$ (with Dirichlet boundary conditions) for a sufficiently large $R > 0$. We now proceed as in [Carlier et al., 2009]. 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

$$\partial_x^h u_{i,j} \overset{\text{def.}}{=} u_{i+1,j} - u_{i,j} \qquad \partial_y^h u_{i,j} \overset{\text{def.}}{=} u_{i,j+1} - u_{i,j} \tag{16}$$

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 now define

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

and set

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

We then solve the following discretized version of (6) for increasingly small values of h

$$\min_{u\in E^h} h^2 \left\langle \overline{\eta}^h, u \right\rangle \quad \text{s.t.} \quad J^h(u) \leq 1, \tag{17}$$

where $\overline{\eta}^h = \left(\frac{1}{h^2} \int_{C_{i,j}^h} \eta\right)_{(i,j)\in[1,N]^2}$ and $(C_{i,j}^h)_{(i,j)\in[1,N]^2}$ is a partition of $[-R, R]^2$ composed of squares of equal size, i.e.

$$C_{i,j}^h \overset{\text{def.}}{=} [-R+(i-1)h, -R+ih]\times[-R+(j-1)h, -R+jh].$$

For convenience reasons, we will also use the above expression to define $C_{i,j}^h$ if i or j belongs to $\{0, N + 1\}$.

In practice we solve (17) using the primal-dual algorithm introduced in [Chambolle and Pock, 2011]: we take (τ, σ) such that $\tau\,\sigma\,\|D\|^2 < 1$ holds with $D \overset{\text{def.}}{=} h\nabla^h$ and define

$$\begin{cases} \phi^{n+1} = \text{prox}_{\sigma\|\cdot\|_{2,\infty}}(\phi^n + \sigma\, D\bar{u}^n), \\ u^{n+1} = (u^n - \tau\, D^*\phi^{n+1}) - \tau\, h^2\, \bar{\eta}^h, \\ \bar{u}^{n+1} = 2\, u^{n+1} - u^n, \end{cases} \tag{18}$$

where $\text{prox}_{\sigma\|\cdot\|_{2,\infty}}$ is given by:

$$\text{prox}_{\sigma\|\cdot\|_{2,\infty}}(\phi) = \phi - \sigma\, \text{proj}_{\{\|\cdot\|_{2,1}\leq 1\}}\left(\frac{\phi}{\sigma}\right).$$

See [Condat, 2016] for the computation of the projection onto the $(2, 1)$-unit ball.

The following proposition shows that, when the grid becomes finer, solutions of (17) converge to a solution of (6). 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 in Appendix D for the sake of completeness.

Proposition 6 *Let u^h be the piecewise constant function on $(C_{i,j}^h)_{(i,j)\in[1,N]^2}$, extended to 0 outside $[-R, R]^2$ associated to a solution of (17). 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 (6) when $h \to 0$, with moreover $\mathrm{D}u^h \overset{*}{\rightharpoonup} \mathrm{D}u$.*

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

Since we are interested in finding a simple set E that approximately solves (7), and now have a good way of approximating solutions of (6), we make use of the following result:

Proposition 7 *Let u be a solution of (6). Then the level sets of u are such that for all $t \in \mathbb{R}^*$ with $|U^{(t)}| > 0$, the set $U^{(t)}$ solves (7).*

Proof. This is a direct consequence of Proposition 1. □

If we have v^h converging strongly in $\mathrm{L}^1(\mathbb{R}^2)$ to a solution v^* of (6), then up to the extraction of a (not relabeled) subsequence, 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 (7). However, this convergence only implies that

$$\limsup_{k\to\infty} \mathcal{J}(E_k) \leq \mathcal{J}(E_*) \,,$$

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 Remark 5, we know we have to design a numerical method that allows to find a set at which the value of $\mathcal{J}$ is arbitrarily close to $\mathcal{J}(E_*)$. This hence motivates the introduction of the refinement step described in the next subsection.

As a final remark, we note that, even for k large enough, E_k could be non-simple. However, using the notations of Section 2.1, since for every set of finite perimeter E, $\mathcal{J}(E)$ is a convex combination of the

$$\left(\mathcal{J}(\mathrm{int}(\gamma_i^+))\right)_{i\in I} \,, \quad \left(\mathcal{J}(\mathrm{int}(\gamma_{i,j}^-))\right)_{i\in I, j\in J_i} \,,$$

there is a simple set F in the decomposition of E which is such that $\mathcal{J}(F) \geq \mathcal{J}(E)$. 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.

5.1.2 Refinement step

We use a shape gradient algorithm (see [Allaire et al., 2021]) to refine the output of the fixed grid step. It consists in iteratively constructing a sequence of simple polygons by finding at each step a displacement of steepest ascent for $\mathcal{J}$, along which the vertices of the previous polygon are moved. Given $x^t \in \mathcal{X}_n$ and a step size α^t, we define the next iterate by:

$$\begin{aligned}
x_j^{t+1} &\overset{\text{def.}}{=} x_j^t + \alpha^t \, \theta_j^t \,, \\
\theta_j^t &\overset{\text{def.}}{=} \frac{1}{P(E_{x^t})} \left(\theta_{\mathrm{area},j}^t - \frac{\int_{E_{x^t}} \eta}{P(E_{x^t})} \, \theta_{\mathrm{per},j}^t \right) \,, \\
\theta_{\mathrm{area},j}^t &\overset{\text{def.}}{=} w_j^{t-} \nu_{j-1}^t + w_j^{t+} \nu_j^t \,, \\
\theta_{\mathrm{per},j}^t &\overset{\text{def.}}{=} -(\tau_j^t - \tau_{j-1}^t) \,,
\end{aligned} \tag{19}$$

where, for all j, τ_j^t and ν_j^t are respectively the unit tangent and outer normal vectors on $[x_j^t, x_{j+1}^t]$ and

$$\begin{aligned}
w_j^{t+} &\overset{\text{def.}}{=} \int_{[x_j^t, x_{j+1}^t]} \eta(x) \, \frac{||x - x_{j+1}^t||}{||x_j^t - x_{j+1}^t||} \, d\mathcal{H}^1(x) \,, \\
w_j^{t-} &\overset{\text{def.}}{=} \int_{[x_j^t, x_{j-1}^t]} \eta(x) \, \frac{||x - x_{j-1}^t||}{||x_j^t - x_{j-1}^t||} \, d\mathcal{H}^1(x) \,.
\end{aligned}$$

One can actually show that the displacement θ^t we apply to the vertices of E_{x^t} is such that

$$\theta^t = \underset{||\theta||\leq 1}{\mathrm{Argmax}} \lim_{\alpha\to 0^+} \frac{\mathcal{J}(E_{x^t+\alpha\theta}) - \mathcal{J}(E_{x^t})}{\alpha} \,, \tag{20}$$

i.e. that it is the displacement of steepest ascent for $\mathcal{J}$ at E_{x^t}. We provide a proof of this result in Appendix E.

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 $\mathcal{J}$ 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 (7) (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 $\mathcal{J}$ could have non-optimal critical points[5], the above initialization allows us to start our local descent with a polygon

[5] Here critical point is to be understood in the sense that the limit appearing in (20) is equal to zero for every θ.

that hopefully lies in the basin of attraction of a global maximizer. Additionally, we stress again that to carry out Line 5 of Algorithm 3, thanks to Remark 5, we only need to find a set with near optimal value in (7).

An interesting problem is to quantify the distance (e.g. in the Hausdorff sense) of a maximizer of $\mathcal{J}$ over $\mathcal{P}_n$ to a maximizer of $\mathcal{J}$. We discuss in Section 7 the simpler case of radial measurements. In the general case, if the sequence of polygons defined above converges to a simple polygon E_x, then E_x is such that

$$w_j^+ = w_j^- = \frac{\int_{E_x} \eta}{P(E_x)} \tan\left(\frac{\theta_j}{2}\right) \tag{21}$$

for all j, where θ_j is the j-th exterior angle of the polygon (the angle between $x_j - x_{j-1}$ and $x_{j+1} - x_j$). This can be seen as a discrete version of the following first order optimality condition for solutions of (7):

$$\eta = \frac{\int_E \eta}{P(E)} \, H_E \text{ on } \partial^* E \,. \tag{22}$$

Note that (22) 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 [Alter et al., 2005] or [Parini, 2011, Prop. 2.4]).

5.2 Sliding step

The implementation of the sliding step (Line 13 in Algorithm 3) is similar to what is described above for refining crude approximations of Cheeger sets. We use a first order optimization method on the mapping

$$(a, x) \mapsto T_\lambda \left(\sum_{i=1}^N a_i \, \mathbf{1}_{E_{x_i}} \right). \tag{23}$$

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 set $u^t \overset{\text{def.}}{=} \sum_{i=1}^N a_i^t \, \mathbf{1}_{E_{x_i^t}}$ and perform the following update:

$$\begin{aligned}
a_i^{t+1} &\overset{\text{def.}}{=} a_i^t - \alpha^t \, h_i^t \,, \\
h_i^t &\overset{\text{def.}}{=} \left\langle \Phi \mathbf{1}_{E_{x_i^t}}, \Phi u^t - y \right\rangle + \lambda \, P\left(E_{x_i^t}\right) \, \operatorname{sign}\left(a_i^t\right), \\
x_{i,j}^{t+1} &\overset{\text{def.}}{=} x_{i,j}^t - \alpha^t \, \theta_{i,j}^t \,, \\
\theta_{i,j}^t &\overset{\text{def.}}{=} a_i^t \left[\theta_{\text{data},i,j}^t - \lambda \operatorname{sign}(a_i^t) \left(\tau_{i,j}^t - \tau_{i,j-1}^t \right) \right], \\
\theta_{\text{data},i,j}^t &\overset{\text{def.}}{=} \langle \Phi u^t - y, \, w_{i,j}^{t-} \rangle \, \nu_{i,j-1}^t + \langle \Phi u^t - y, \, w_{i,j}^{t+} \rangle \, \nu_{i,j}^t \,,
\end{aligned}$$

where $\tau_{i,j}^t$, $\nu_{i,j}^t$ are respectively the unit tangent and outer normal vectors on the edge $[x_{i,j}^t, x_{i,j+1}^t]$ and

$$\begin{aligned}
w_{i,j}^{t+} &\overset{\text{def.}}{=} \int_{[x_{i,j}^t, x_{i,j+1}^t]} \varphi(x) \, \frac{||x - x_{i,j+1}^t||}{||x_{i,j}^t - x_{i,j+1}^t||} \, d\mathcal{H}^1(x) \,, \\
w_{i,j}^{t-} &\overset{\text{def.}}{=} \int_{[x_{i,j}^t, x_{i,j-1}^t]} \varphi(x) \, \frac{||x - x_{i,j-1}^t||}{||x_{i,j}^t - x_{i,j-1}^t||} \, d\mathcal{H}^1(x) \,.
\end{aligned}$$

Using the notations of Section 5.1.2, the complexity of each iteration is of order $\mathcal{O}\left(N \, m \left(|\mathcal{T}| \, |\mathcal{S}_T| + n \, |\mathcal{S}_L|\right)\right)$.

Comments. We first stress that the above update is similar to the evolution formally described in (12). Now, unlike the local optimization we perform to approximate Cheeger sets, the sliding step may tend to induce topology changes (see Section 6.2 for an example). This is of course linked to the possible appearance of singularities mentioned in Section 4. 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 $d > k$.

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.

6 Numerical experiments

6.1 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 $(\mathcal{P}_\lambda)$ 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 $(\mathcal{P}_\lambda)$ with $y = \Phi u_0 + w$ are all close to u_0, modulo the regularization effects of the total variation.

We also assess the performance of our algorithm by comparing its output to that of a primal dual algorithm minimizing a discretized version of $(\mathcal{P}_\lambda)$ on a pixel grid, where the total variation term is replaced by the discrete isotropic total variation or Condat's discrete total variation[6]. 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.

Our first experiment consists in recovering a function u_0 that is a linear combination of three indicator functions (see Figures 2 and 3). 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 significantly refined during the sliding step of the third iteration, when enough atoms have been introduced.

The second experiment (see Figure 4) 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 (gradient) variation regularization are noticeable: corners are slightly rounded, and there is a "loss of contrast" in the eye of the pacman.

The third experiment (Fig. 5) 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. 6 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 (17) 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 triangle integrals using the Hammer-Marlowe-Stroud scheme of order 5 (7 points).

[6] Condat's total variation is introduced in [Condat, 2017]. See also [Chambolle and Pock, 2021] for a review of discretizations of the total variation.

6.2 Topology changes during the sliding step

Here, we illustrate the changes of topology that may occur during the sliding step (Line 13 of Algorithm 3). All relevant plots are given in Figure 7. 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)[7], 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 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).

7 The case of a single radial measurement

In this section, we study a particular setting, where the number of observations m is equal to 1, and the unique sensing function is radial, i.e. the measurement operator is given by (1) with $\varphi : \mathbb{R}^2 \to \mathbb{R}$ a radial function[8]. We first state a proposition about the solutions of $(\mathcal{P}_\lambda)$ in this setting, before carrying on with results that will

[7] This only occurs when λ is small enough. For higher values of λ, the output is similar to (d) or (e).

[8] We say that $f : \mathbb{R}^2 \to \mathbb{R}$ is radial if there exists $g : [0, +\infty[\to \mathbb{R}$ such that $f(x) = g(\|x\|)$ for almost every $x \in \mathbb{R}^2$.

Fig. 2: 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

require more assumptions on φ. Unless otherwise specified, sets that differ by a Lebesgue negligible set and functions that are equal almost everywhere are identified.

For every $u \in \mathrm{L}^2(\mathbb{R}^2)$, we define the radialisation $\tilde{u}$ of u by

$$\tilde{u}(x) = \int_{\mathbb{S}^1} u(\|x\| e)\, d\mathcal{H}^1(e)\,.$$

We note that in our setting Φu only depends on u through $\tilde{u}$, that is:

$$\Phi u = \int_{\mathbb{R}^2} \varphi\, u = \int_{\mathbb{R}^2} \tilde{\varphi}\, \tilde{u}\,.$$

Using the fact $|\mathrm{D}\tilde{u}|(\mathbb{R}^2) \leq |\mathrm{D}u|(\mathbb{R}^2)$ for any $u \in \mathrm{L}^2(\mathbb{R}^2)$ such that $|\mathrm{D}u|(\mathbb{R}^2) < +\infty$ with equality if and only if u is radial (see Appendix F.1 for a proof of this statement), we may state the following result:

Proposition 8 *Every solution of* $(\mathcal{P}_\lambda)$ *is radial, and there exists a solution that is proportional to the indicator of a disk centered at the origin.*

Proof. The first part of the result is a direct consequence of the above statements. Then, using [Boyer et al., 2019, Corollary 2 and Theorem 2], we have that there exists a solution of $(\mathcal{P}_\lambda)$ which is proportional to the indicator function of a simple set. The result follows from the fact that every simple set whose indicator function is radial is a disk centered at the origin. □

We will now assume φ is positive, continuous and decreasing[9] along rays. For any $r \in \mathbb{R}_+$, we will denote by an abuse of notation $\tilde{\varphi}(r)$ the value of $\tilde{\varphi}$ at any point $x \in \mathbb{R}^2$ such that $\|x\| = r$. We may also invoke the following assumption:

Assumption 1. The function $f : r \mapsto r\,\tilde{\varphi}(r)$ is continuously differentiable on $\mathbb{R}_+^*$, $rf(r) \to 0$ when $r \to +\infty$, and there exists $\rho_0 > 0$ such that $f'(r) > 0$ on $]0, \rho_0[$ and $f'(r) < 0$ on $]\rho_0, +\infty[$.[10]

[9] In all the following, by decreasing we mean *strictly* decreasing.

[10] Assumption 1 is for example satisfied by $\varphi : x \mapsto \exp\left(-||x||^2/(2\sigma^2)\right)$ for any $\sigma > 0$.

In the rest of this section we first explain what each step of Algorithm 2 should theoretically return in this particular setting, without worrying about approximations made for implementation matters. Then, we compare those with the output of each step of the practical algorithm.

7.1 Theoretical behavior of the algorithm

The first step of Algorithm 3 consists in solving the Cheeger problem (7) associated to $\eta \overset{\text{def}}{=} \frac{1}{\lambda}\Phi^* y = \frac{y}{\lambda}\varphi$ (or equivalently to φ). To describe the solutions of this problem, we rely on Steiner symmetrization. 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

$$E^s_\nu \overset{\text{def.}}{=} \{x \in \mathbb{R}^2 \,|\, |\langle x, \nu^\perp\rangle| \leq \mathcal{L}^1(E_{\nu,\langle x,\nu\rangle})/2\}\,,$$

where $\mathcal{L}^1$ denotes the Lebesgue measure on $\mathbb{R}$. 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). Using this, and denoting by $B(0,R)$ the disk of radius R centered at the origin, we may state[11]:

Proposition 9 *All the solutions of the Cheeger problem (7) associated to* $\eta \overset{def}{=} \varphi$ *are disks centered at the origin. Under Assumption 1 the unique solution is the disk* $B(0,R^*)$ *with* R^* *the unique maximizer of*

$$R \mapsto \left[\int_0^R r\,\tilde{\varphi}(r)\,dr\right]/R\,.$$

[11] This result can be proved using the radialisation operation previously introduced. We here however rely on classical arguments used in the analysis of geometric variational problems, which we will moreover also use later in this section.

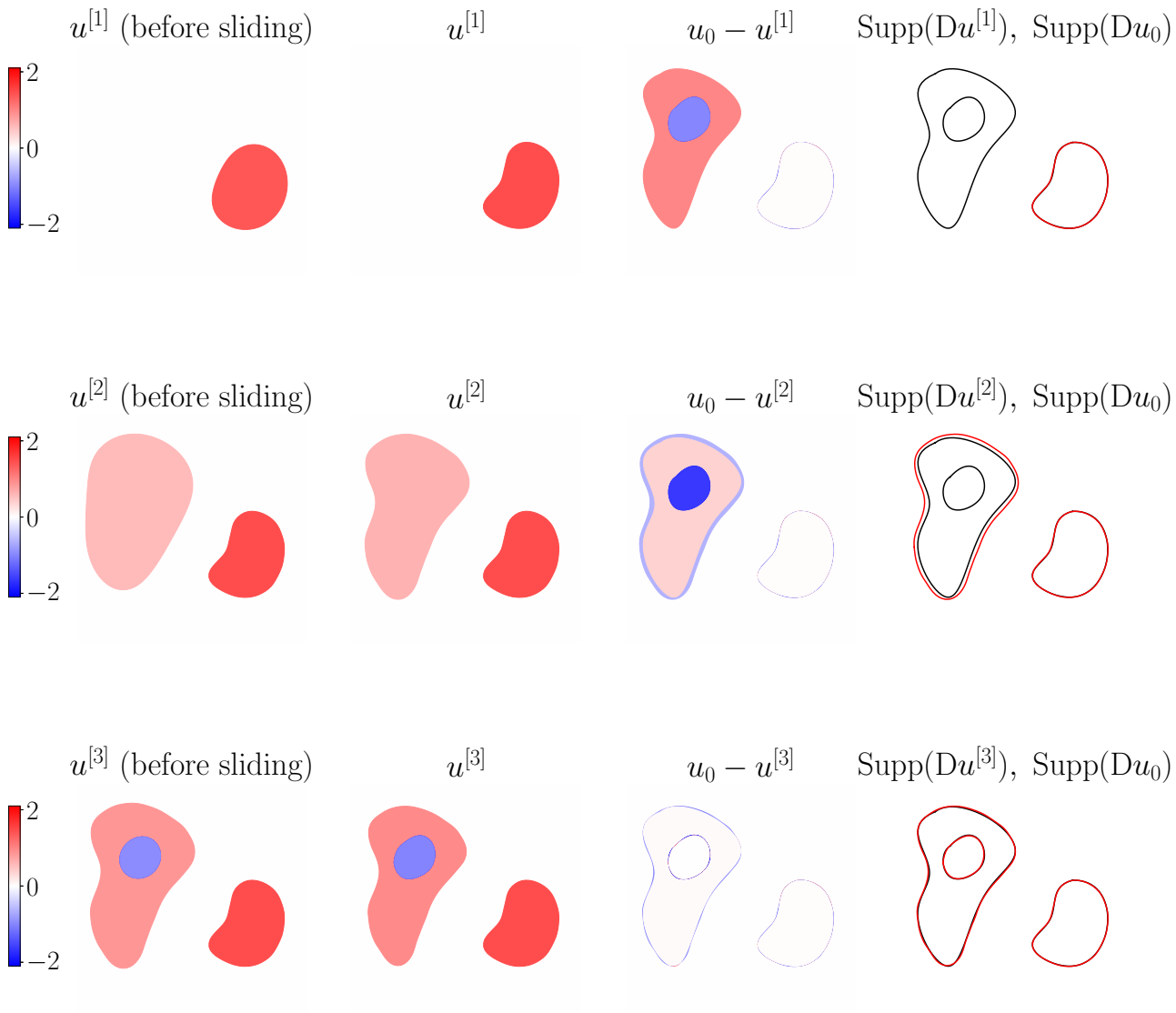

Fig. 3: Unfolding of Algorithm 3 for the first experiment ($u^{[k]}$ denotes the k-th iterate)

Fig. 4: 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)

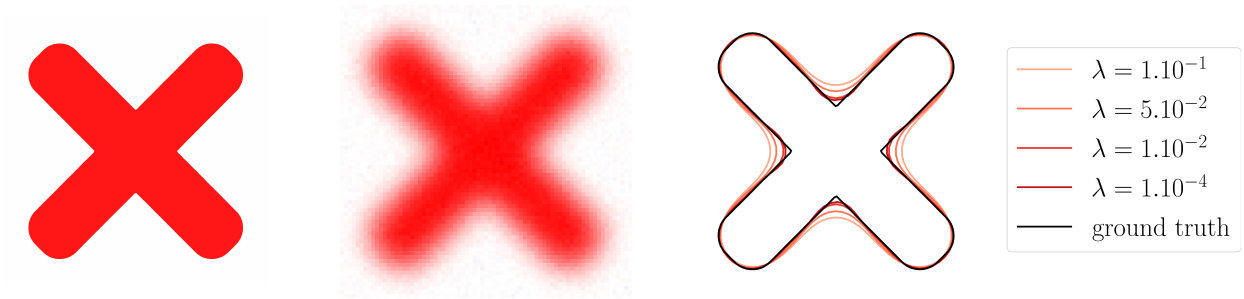

Fig. 5: Left: unknown function, middle: observations, right: output of Algorithm 3 for different values of λ

Fig. 6: 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

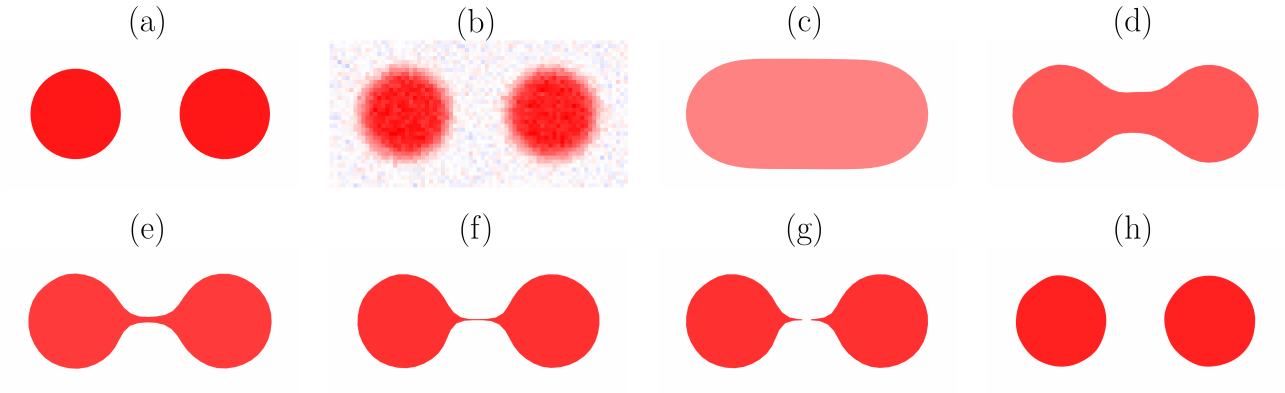

Fig. 7: Topology change experiment. (a): unknown signal, (b): observations, (c): weighted Cheeger set, (d,e,f,g): sliding step iterations (with splitting), (h): final function.

Proof. We first stress that existence of solutions was already briefly discussed in Section 3.1 (it can either be obtained by purely geometric arguments, or by showing the existence of solutions of (6) by the direct method of calculus of variations and then using Krein-Milman theorem).

Now if $E \subset \mathbb{R}^2$ is such that $0 < P(E) < +\infty$ and $\nu \in \mathbb{S}^1$ we have (see Lemma 9):

$$\frac{\int_E \eta}{P(E)} \leq \frac{\int_{E^s_\nu} \eta}{P(E^s_\nu)},$$

with equality if and only if $|E \triangle E^s_\nu| = 0$. Hence if E^* solves (7), 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 ball centered at the origin.

Now for any $R > 0$ we have

$$\mathcal{G}(R) \stackrel{\text{def.}}{=} \frac{\int_{B(0,R)} \varphi}{P(B(0,R))} = \frac{1}{R} \int_0^R r\, \tilde{\varphi}(r) dr\,,$$

and the last part of the result follows from a simple analysis of the variations of $\mathcal{G}$ under Assumption 1, which is given in Appendix F.

□

The second step (Line 10) of the algorithm then consists in solving

$$\inf_{a \in \mathbb{R}} \frac{1}{2} \left(a \int_{E^*} \varphi - y \right)^2 + \lambda P(E^*) |a|\,, \tag{24}$$

where $E^* = B(0, R^*)$. The solution a^* has a closed form which writes:

$$a^* = \frac{\mathrm{sign}(y)}{\int_{E^*} \varphi} \left(|y| - \lambda \frac{P(E^*)}{\int_{E^*} \varphi} \right)^+ , \tag{25}$$

where $x^+ = \max(x, 0)$.

The next step should be the sliding one (Line 13). However, in this specific setting, one can show that the constructed function is already optimal, as stated by the following proposition:

Proposition 10 *Under Assumption 1, Problem* $(\mathcal{P}_\lambda)$ *has a unique solution* $a^* \mathbf{1}_{E^*}$ *with* $E^* = B(0, R^*)$ *the solution of the Cheeger problem given by Prop. 9, and* a^* *given by* (25).

Proof. If $u^* \in \mathrm{L}^2(\mathbb{R}^2)$ solves $(\mathcal{P}_\lambda)$ then

$$\Phi^* p^* = p^* \varphi \in \partial J(u^*) ,$$

with $p^* = -\frac{1}{\lambda}(\Phi u^* - y)$. Now from Proposition 1 we know $p^* \varphi \in \partial J(u^*)$ implies that $p^* \varphi \in \partial J(0)$ and that the level sets of u^* satisfy

$$P(U_*^{(t)}) = \left| \int_{U_*^{(t)}} p^* \varphi \right| .$$

This means that the non trivial level sets of u^* are all solutions of the Cheeger problem associated to $p^* \varphi$ (or equivalently to φ), and are hence equal to $B(0, R^*)$. This shows there exists $a \in \mathbb{R}$ such that $u^* = a \mathbf{1}_{B(0,R^*)}$, and the result easily follows. □

To summarize, with a single observation and a radial sensing function, a solution is found in a single iteration, and its support is directly identified by solving the Cheeger problem.

7.2 Study of implementation approximations

In practice, instead of solving (7), we look for an element of $\mathcal{P}_n$ (a simple polygon with at most n sides) maximizing $\mathcal{J}$, for some given integer $n \geq 3$. It is hence natural to investigate the proximity of this optimal polygon with $B(0, R^*)$. 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 [Pólya and Szegö, 1951, Sec. 7.4]. However, using a trick from Pólya and Szegö, one may prove:

Proposition 11 *Let* $n \in \{3, 4\}$. *Then all the maximizers of* $\mathcal{J}$ *over* $\mathcal{P}_n$ *are regular and inscribed in a circle centered at the origin.*

Proof.

Triangles: let E^* be a maximizer of $\mathcal{J}$ among triangles. Then the Steiner symmetrization of E^* with respect to any of its heights through the origin (see Figure 8) is still a triangle, and Lemma 9 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: 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 9) is still a simple quadrilateral. We can then proceed exactly as for triangles to prove any maximizer E^* of $\mathcal{J}$ 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 11 holds for all $n \geq 3$.

For $n \in \{3, 4\}$ or if Conjecture 1 holds, it remains to compare the optimal polygons with $B(0, R^*)$. If we define $\mathcal{G}(R) \overset{\text{def.}}{=} \mathcal{J}(B(0, R))$ and $\mathcal{G}_n(R)$ the value of $\mathcal{J}$ at any regular n-gon inscribed in a circle of radius R centered at the origin, then we can state the following result (its proof is given in Appendix F.3):

Proposition 12 *Under Assumption 1, we have that*

$$||\mathcal{G}_n - \mathcal{G}||_\infty = O\left(\frac{1}{n^2}\right) .$$

Moreover, if f *is of class* C^2 *and* $f''(\rho_0) < 0$ *, then for* n *large enough* $\mathcal{G}_n$ *has a unique maximizer* R_n^* *and*

$$|R_n^* - R^*| = O\left(\frac{1}{n}\right) .$$

If φ *is the function defined by*

$$\varphi : x \mapsto exp\left(-||x||^2/(2\sigma^2)\right) ,$$

then this last result holds for all $n \geq 3$.

Now, the output of our method for approximating Cheeger sets, described in Section 5, is a polygon that is obtained by locally maximizing $\mathcal{J}$ using a first order method. Even if we carefully initialize this first order

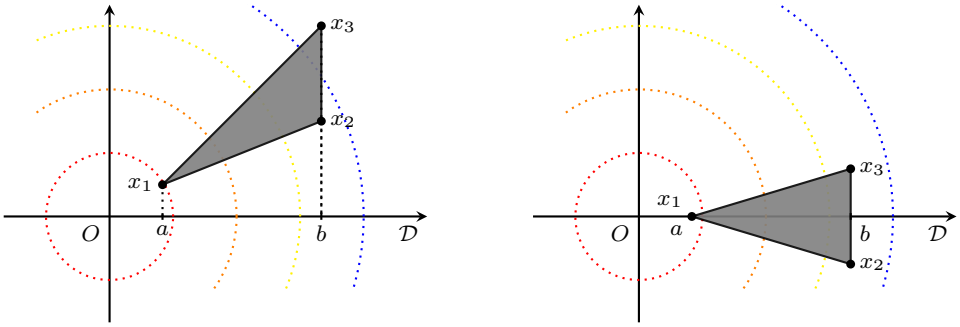

Fig. 8: Steiner symmetrization of triangles

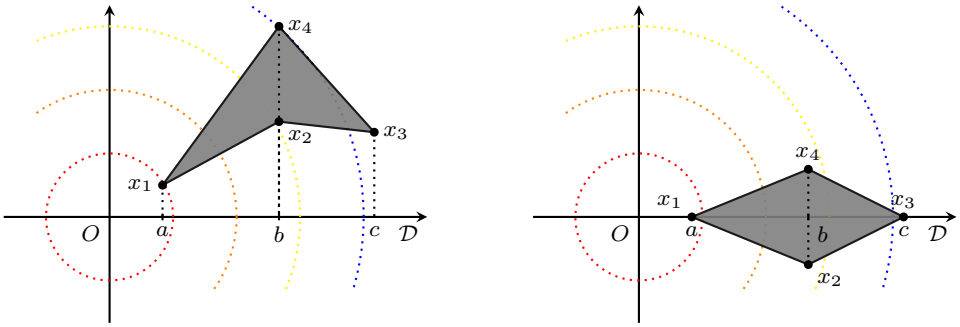

Fig. 9: Steiner symmetrization of quadrilaterals

method, the possible existence of non-optimal critical points makes its analysis challenging. However, in our setting (a radial weight function), the simple polygons that are critical points[12] of $\mathcal{J}$ coincide with its global maximizers over $\mathcal{P}_n$ (at least for small n). The proof of this result is given in Appendix F.4.

Proposition 13 *Let $n \in \{3, 4\}$. Under Assumption 1, if f is of class C^2 and $f''(\rho_0) < 0$, the elements of $\mathcal{P}_n$ that are critical points of $\mathcal{J}$ are the regular n-gons inscribed in the circle of radius R_n^* centered at the origin.*

We make the following conjecture:

Conjecture 2 The result stated in Proposition 13 holds for all $n \geq 3$.

If $n \in \{3, 4\}$, or if Conjecture 2 holds, we may therefore expect our polygonal approximation to be at Hausdorff distance of order $O\left(\frac{1}{n}\right)$ to $B(0, R^*)$.

8 Conclusion

As shown in the present exploratory work, solving total variation regularized inverse problems in a gridless manner is highly beneficial, as it allows to preserve structural properties of their solutions, which cannot be achieved by traditional numerical solvers. The price to pay for going "off-the-grid" is an increased complexity of the analysis and the implementation of the algorithms. Furthering their theoretical study and improving their practical efficiency and reliability is an interesting avenue for future research. Investigating extensions to higher dimensions (e.g. 3D) could also be promising. Although the computational cost of each step might be large, it seems that the proposed algorithm could be transposed to this new setting.

[12] We recall that critical point is here to be understood in the sense that the limit appearing in (20) is equal to zero for every θ.

Acknowledgements The authors thank Robert Tovey for carefuly reviewing the code used in the numerical experiments section, and for suggesting several modifications that significantly improved the results presented therein.

References

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.

Alter et al., 2005. Alter, F., Caselles, V., and Chambolle, A. (2005). Evolution of characteristic functions of convex sets in the plane by the minimizing total variation flow. *Interfaces and Free Boundaries*, 7(1):29–53.

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.

Bartels et al., 2021. Bartels, S., Tovey, R., and Wassmer, F. (2021). Singular solutions, graded meshes, and adaptivity for total-variation regularized minimization problems.

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.

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 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.

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.

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.

Castro et al., 2017. Castro, Y. D., 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.

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 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, 2021. Chambolle, A. and Pock, T. (2021). 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.

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.

Dautray and Lions, 2012. Dautray, R. and Lions, J.-L. (2012). *Mathematical Analysis and Numerical Methods for Science and Technology: Volume 1 Physical Origins and Classical Methods*. Springer Science & Business Media.

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*.

Duval, 2022. Duval, V. (2022). *Faces and extreme points of convex sets for the resolution of inverse problems*. Habilitation à diriger des recherches. In preparation.

Fleming, 1957. Fleming, W. H. (1957). Functions with generalized gradient and generalized surfaces. *Annali di Matematica Pura ed Applicata*, 44(1):93–103.

Giusti, 1984. Giusti (1984). *Minimal Surfaces and Functions of Bounded Variation*. Monographs in Mathematics. Birkhäuser Basel.

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.

Hormann and Agathos, 2001. Hormann, K. and Agathos, A. (2001). The point in polygon problem for arbitrary polygons. *Computational Geometry*, 20(3):131–144.

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.

Jaggi, 2013. Jaggi, M. (2013). Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization. In *International Conference on Machine Learning*, pages 427–435. PMLR.

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.

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.

Parini, 2011. Parini, E. (2011). An introduction to the Cheeger problem. *Surveys in Mathematics and its Applications*, 6:9–21.

Pólya and Szegö, 1951. Pólya, G. and Szegö, G. (1951). *Isoperimetric Inequalities in Mathematical Physics. (AM-27)*. Princeton University Press.

Rao et al., 2015. Rao, N., Shah, P., and Wright, S. (2015). Forward–Backward Greedy Algorithms for Atomic Norm Regularization. *IEEE Transactions on Signal Processing*, 63(21):5798–5811.

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.

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.

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.

A Derivation of Algorithm 2

See [Denoyelle et al., 2019, Sec. 4.1] for the case of the sparse spikes problem.

Lemma 2 *Problem* $(\mathcal{P}_\lambda)$ *is equivalent to*

$$\min_{(u,t)\in C} \quad \tilde{T}_\lambda(u,t) \overset{def}{=} \frac{1}{2}||\Phi u - y||^2 + \lambda t \qquad (\tilde{\mathcal{P}}_\lambda)$$

with

$$C \overset{\text{def.}}{=} \left\{(u,t) \in \mathrm{L}^2(\mathbb{R}^2)\times\mathbb{R} \,\middle|\, |\mathrm{D}u|(\mathbb{R}^2) \le t \le M\right\}.$$

and $M \overset{\text{def.}}{=} ||y||^2/(2\lambda)$*, i.e. if* u *is a solution of* $(\mathcal{P}_\lambda)$ *then we have that* $(u, |\mathrm{D}u|(\mathbb{R}^2))$ *is a solution of* $(\tilde{\mathcal{P}}_\lambda)$*, and conversely any solution of* $(\tilde{\mathcal{P}}_\lambda)$ *is of the form* $(u, |\mathrm{D}u|(\mathbb{R}^2))$ *with* u *a solution of* $(\mathcal{P}_\lambda)$*.*

Proof. If u^* is a solution of $(\mathcal{P}_\lambda)$, then

$$T_\lambda(u^*) \le T_\lambda(0) = ||y||^2/2\,.$$

Hence we have that $|\mathrm{D}u^*|(\mathbb{R}^2) \le M$, which shows the feasible set of $(\mathcal{P}_\lambda)$ can be restricted to functions u which are such that $|\mathrm{D}u|(\mathbb{R}^2) \le M$. It is then straightforward to show that the resulting program is equivalent to $(\tilde{\mathcal{P}}_\lambda)$, in the sense defined above. □

The objective $\tilde{T}_\lambda$ of $(\tilde{\mathcal{P}}_\lambda)$ is now convex, differentiable and we have for all $(u,t) \in \mathrm{L}^2(\mathbb{R}^2)\times\mathbb{R}$

$$d\tilde{T}_\lambda(u,t)\colon \mathrm{L}^2(\mathbb{R}^2)\times\mathbb{R} \to \mathbb{R}$$
$$(v,s) \mapsto \left[\int_{\mathbb{R}^2} \Phi^*(\Phi u - y)\, v\right] + \lambda s\,.$$

Moreover, the feasible set C is weakly compact. We can therefore apply Frank-Wolfe algorithm to $(\tilde{\mathcal{P}}_\lambda)$. The following result shows how the linear minimization step (Line 2 of Algorithm 1) one has to perform at step k amounts to solving the Cheeger problem (7) associated to $\eta \overset{\text{def.}}{=} -\frac{1}{\lambda}\Phi^*(\Phi u^{[k]} - y)$.

Proposition 14 *Let* $(u,t) \in C$ *and* $\eta \overset{\text{def.}}{=} -\frac{1}{\lambda}\Phi^*(\Phi u - y)$*. We also denote*

$$\alpha \overset{\text{def.}}{=} \sup_{E\subset\mathbb{R}^2} \frac{\left|\int_E \eta\right|}{P(E)} \quad s.t. \quad 0<|E|<+\infty,\ P(E)<+\infty\,. \qquad (26)$$

Then, if $\alpha \le 1$*,* $(0,0)$ *is a minimizer of* $d\tilde{T}_\lambda(u,t)$ *over* C*. Otherwise, there exists a simple set* E *achieving the supremum in* (26) *such that, denoting* $\epsilon = sign\left(\int_E \eta\right)$*,* $\left(\frac{\epsilon M}{P(E)}\mathbf{1}_E, M\right)$ *is a minimizer of* $d\tilde{T}_\lambda(u,t)$ *on* C*.*

Proof. The extreme points of C are $(0,0)$ and the elements of

$$\left\{\left(\pm\frac{M}{P(E)}\mathbf{1}_E, M\right) \;\middle|\; E \text{ is simple},\ 0<|E|<+\infty\right\}.$$

Since $d\tilde{T}_\lambda(u,t)$ is linear, it reaches its minimum on C at least at one of these extreme points. We hence have that

$$(0,0) \in \operatorname*{Argmin}_{(v,s)\in C} d\tilde{T}_\lambda(u,t)(v,s)$$

or that a minimizer can be found by finding an element of

$$\operatorname*{Argmin}_{\substack{E \text{ simple}\\ \epsilon\in\{-1,1\}}} \left\langle \Phi u - y, \frac{\epsilon M}{P(E)}\Phi\mathbf{1}_E\right\rangle + \lambda M$$
$$= \operatorname*{Argmin}_{\substack{E \text{ simple}\\ \epsilon\in\{-1,1\}}} \left\langle \Phi u - y, \frac{\epsilon}{\lambda P(E)}\Phi\mathbf{1}_E\right\rangle$$
$$= \operatorname*{Argmin}_{\substack{E \text{ simple}\\ \epsilon\in\{-1,1\}}} \frac{\epsilon}{P(E)}\int_E \frac{1}{\lambda}\Phi^*(\Phi u - y)\,.$$

This last problem is equivalent to finding an element of

$$\operatorname*{Argmax}_{E \text{ simple}} \frac{1}{P(E)}\left|\int_E \eta\right|,$$

in the sense that E^* is optimal for the latter if and only if the couple $\left(E^*, \operatorname{sign}\left(\int_{E^*}\eta\right)\right)$ is optimal for the former. We can moreover show that $(0,0)$ is optimal if and only if for all $E\subset\mathbb{R}^2$ such that $0<|E|<+\infty$ and $P(E)<+\infty$ we have:

$$\frac{1}{P(E)}\left|\int_E \eta\right| \le 1\,.$$

□

B Discussion on Line 10 of Algorithms 2 and 3

Consdering Appendix A and Algorithm 1, the standard Frank-Wolfe update at iteration k would be to take $u^{[k+1]}$ equal to $\tilde{u}^{[k+1]}$ with:

$$\tilde{u}^{[k+1]} \overset{\text{def.}}{=} (1-\gamma_k)\,u^{[k]} + \gamma_k \frac{M\,\epsilon_*}{P(E_*)}\mathbf{1}_{E_*}\,,$$

where $\gamma_k = \frac{2}{k+2}$, E_* is the set obtained at Line 5 and ϵ_* is the sign of $\int_{E_*}\eta^{[k]}$. Now, one can write $\tilde{u}^{[k+1]}$ as a linear combination of indicator functions of its level sets, and then apply the decomposition mentionned in Section 2.1 to each level set. This allows to find a family $(E_i)_{i=1}^N$ of simple sets of positive measure and $(a_i)_{i=1}^N \in \mathbb{R}^N$ such that $\tilde{u}^{[k+1]} = \sum_{i=1}^N a_i\,\mathbf{1}_{E_i}$ and

$$\left|\mathrm{D}\left(\sum_{i=1}^N a_i\,\mathbf{1}_{E_i}\right)\right|(\mathbb{R}^2) = \sum_{i=1}^N |a_i|\,P(E_i)\,.$$

Moreover, it is possible to prove (see [Duval, 2022]) that for every $i \neq j$

1. Either $E_i \subset E_j$, $E_j \subset E_i$ or $E_i \cap E_j = \emptyset$.
2. If $\operatorname{sign}(a_i) = \operatorname{sign}(a_j)$ and $E_i \cap E_j = \emptyset$ then it holds that $\mathcal{H}^1(\partial^* E_i \cap \partial^* E_j) = 0$.
3. If $\operatorname{sign}(a_i) = -\operatorname{sign}(a_j)$ and $E_i \subset E_j$ then it holds again that $\mathcal{H}^1(\partial^* E_i \cap \partial^* E_j) = 0$.

We hence deduce that for every $b \in \mathbb{R}^N$ such that

$$\forall i \in \{1,\dots,N\},\ \operatorname{sign}(a_i) = \operatorname{sign}(b_i)\,,$$

we have:

$$\left|\mathrm{D}\left(\sum_{i=1}^N b_i\,\mathbf{1}_{E_i}\right)\right|(\mathbb{R}^2) = \sum_{i=1}^N |b_i|\,P(E_i)\,. \qquad (27)$$

This shows that if $E^{[k+1]} = (E_1,\dots,E_N)$ and

$$a^{[k+1]} \in \operatorname*{Argmin}_{b\in\mathbb{R}^N} \frac{1}{2}||\Phi_E\, b - y||^2 + \lambda\sum_{i=1}^N P(E_i)\,|b_i|$$
$$\text{s.t.} \qquad \forall i \in \{1,\dots,N\},\ \operatorname{sign}(b_i) = \operatorname{sign}(a_i)\,, \qquad (28)$$

then, defining $u^{[k+1]} = \sum_{i=1}^N a_i^{[k+1]}\,\mathbf{1}_{E_i^{[k+1]}}$, we finally obtain $T_\lambda(u^{[k+1]}) \le T_\lambda(\tilde{u}^{[k+1]})$, which ensures the validity of this update.

As a final note, let us mention that applying the decomposition mentionned in Section 2.1 to the level sets of $\tilde{u}^{[k+1]}$ is a computationally challenging task. However, we stress again that, generically, $\mathcal{H}^1(\partial^* E_i \cap \partial^* E_j) = 0$ for every $i \neq j$, so that the above procedure is never required in practice.

C Existence of maximizers of the Cheeger ratio among simple polygons with at most n sides

Let $\eta \in \mathrm{L}^2(\mathbb{R}^2) \cap C^0(\mathbb{R}^2)$ and $n \geq 3$. We want to prove the existence of maximizers of the Cheeger ratio $\mathcal{J}$ associated to η among simple polygons with at most n sides. We will in fact prove a slightly stronger result, namely the existence of maximizers of a relaxed energy which coincides with $\mathcal{J}$ on simple polygons, and the existence of a simple polygon maximizing this relaxed energy.

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:

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

where $\chi_x(y)$ denotes the index (or winding number) of any parametrization of the polygonal curve $[x_1, x_2], ..., [x_m, x_1]$ around $y \in \mathbb{R}^2$. 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

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

and hence, as soon as $\mathcal{P}(x) > 0$:

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

This naturally leads us to define

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

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

The function χ_x is constant on each connected component of $\mathbb{R}^2 \setminus \Gamma_x$ with $\Gamma_x \stackrel{\text{def.}}{=} \cup_{i=1}^m [x_i, x_{i+1}]$. It takes values in $\{-m, ..., m\}$ and is equal to zero on the only unbounded connected component. We also have $\partial \operatorname{supp}(\chi_x) \subset \Gamma_x$. 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

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

Now we define

$$\alpha \stackrel{\text{def.}}{=} \sup_{x \in \mathcal{Y}_n} \mathcal{J}(x) \,.$$

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, Lebesgue differentiation theorem ensures that

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

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

Lemma 3 *Let $C > 0$. There exists $R > 0$ and $c > 0$ such that*

$$\forall x \in \mathcal{Y}_n,\ \mathcal{J}(x) \geq C \implies \mathcal{P}(x) \geq c \text{ and } \|x_i\| \leq R \text{ for all } i.$$

Proof. The proof is similar to that of Lemma 1.

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

$$\int_{\mathbb{R}^2 \setminus B(0,R_1)} \eta^2 \leq \epsilon^2 \,. \tag{29}$$

Let $\epsilon > 0$ and $R_1 > 0$ such that (29) holds. We have

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

Now, taking

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

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

Inclusion in a ball: we take $\epsilon = \frac{\sqrt{c_2}}{4n}$ 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:

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

Dividing by $\mathcal{P}(x) > 0$ yields a contradiction. Now since

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

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

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

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

$$\begin{aligned}
\mathcal{P}(x) &\leq \frac{1}{C} |\mathcal{A}(x)| = \frac{1}{C} \left| \int_{\operatorname{supp}(\chi_x)} \eta \right| \\
&\leq \frac{1}{C} \sqrt{\int_{\operatorname{supp}(\chi_x)} \eta^2}\ \sqrt{|\operatorname{supp}(\chi_x)|} \\
&\leq \frac{\epsilon}{C\sqrt{c_2}}\, P(\operatorname{supp}(\chi_x)) \\
&\leq \frac{\epsilon}{C\sqrt{c_2}}\, \mathcal{P}(x) \,,
\end{aligned}$$

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

Applying Lemma 3 with e.g. $C = \alpha/2$, and defining

$$\mathcal{Y}'_n \overset{\text{def.}}{=} \left\{ x \in \mathbb{R}^{n\times 2} \,\middle|\, \mathcal{P}(x) \geq c \text{ and } \|x_i\| \leq R \text{ for all } i \right\},$$

we see that any maximizer of $\mathcal{J}$ over $\mathcal{Y}'_n$ (if it exists) is also a maximizer of $\mathcal{J}$ over $\mathcal{Y}_n$, and conversely.

Lemma 4 *Let $x \in \mathbb{R}^{n\times 2}$. Then for every $a \in \mathbb{R}^2$ we have*

$$\mathcal{A}(x) = \sum_{i=1}^{n} \operatorname{sign}(\det(x_i - a\ x_{i+1} - a)) \int_{a x_i x_{i+1}} \eta$$
$$= \sum_{i=1}^{n} \det(x_i - a\ x_{i+1} - a) \int_{T_1} \eta((x_i - a\ x_{i+1} - a)\, y)\, dy\,,$$

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

$$\chi_x = \sum_{i=1}^{n} \operatorname{sign}(\det(x_i - a\ x_{i+1} - a))\, \mathbf{1}_{a x_i x_{i+1}} \tag{30}$$

almost everywhere. First, 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

$$\operatorname{sign}(\det(x_i - a\ x_{i+1} - a)) = \operatorname{sign}\left((y-a)\cdot(x_{i+1}-x_i)^\perp\right).$$

The above hence shows that, if $y \in \mathbb{R}^2 \setminus \cup_{i=1}^n [x_i, x_{i+1}]$ does not belong to any of the segments $[a, x_i]$, evaluating the right hand side of (30) 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 (30) holds almost everywhere, and the result follows. □

From Lemma 4, we deduce that $\mathcal{A}$ is continuous on $\mathbb{R}^{n\times 2}$. This is also the case of $\mathcal{P}$. Now $\mathcal{Y}'_n$ is compact and included in $\mathcal{Y}_n$, hence the existence of maximizers of $\mathcal{J}$ over $\mathcal{Y}'_n$, which in turn implies the existence of maximizers of $\mathcal{J}$ over $\mathcal{Y}_n$.

Let us now show there exists a maximizer which belongs to $\mathcal{X}_n$. To do so, we rely on the following lemma

Lemma 5 *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*

$$\mathcal{J}(x) \leq \mathcal{J}(y)\,.$$

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

$$y = (x_1, ..., x_i, x_{j+1}, ..., x_m) \in \mathbb{R}^{m-(j-i)}\,,$$
$$z = (x_i, x_{i+1}, ..., x_{j-1}) \in \mathbb{R}^{j-i}\,.$$

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

$$y = (x_1, ..., x_i, x_{j+1}, ..., x_m) \in \mathbb{R}^{m-(j-i)}\,,$$
$$z = (x_i, x_{i+1}, ..., x_j) \in \mathbb{R}^{j-i+1}\,.$$

We again have $2 \leq m-(j-i) < m$, and since $(i,j) \neq (1,m)$, we have $j-i < m-1$ which shows $2 \leq j-i+1 < m$.

If there exists $i < j$ with $x_j \in]x_i, x_{i+1}[$: we necessarily have $j > i+1$. We define

$$y = (x_1, ..., x_i, x_j, ..., x_m) \in \mathbb{R}^{m-(j-i)+1}\,,$$
$$z = (x_{i+1}, ..., x_j) \in \mathbb{R}^{j-i}\,.$$

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

$$y = (x_1, ..., x_i, x', x_{j+1}, ..., x_m) \in \mathbb{R}^{m-(j-i)+1}\,,$$
$$z = (x', x_{i+1}, ..., x_j) \in \mathbb{R}^{j-i+1}\,.$$

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 $\mathcal{P}(x) = \mathcal{P}(y) + \mathcal{P}(z)$ and $\chi_x = \chi_y + \chi_z$ almost everywhere, which in turn gives that $\mathcal{A}(x) = \mathcal{A}(y) + \mathcal{A}(z)$. We hence get that $\mathcal{P}(y) = 0$ or $\mathcal{P}(z) = 0$, and in this case $\mathcal{J}(x) = \mathcal{J}(y)$ or $\mathcal{J}(x) = \mathcal{J}(z)$, or that $\mathcal{P}(y) > 0$ and $\mathcal{P}(z) > 0$, which yields

$$\frac{|\mathcal{A}(x)|}{\mathcal{P}(x)} \leq \frac{|\mathcal{A}(y)| + |\mathcal{A}(z)|}{\mathcal{P}(y) + \mathcal{P}(z)}$$
$$= \frac{\mathcal{P}(y)}{\mathcal{P}(y)+\mathcal{P}(z)} \frac{|\mathcal{A}(y)|}{\mathcal{P}(y)} + \frac{\mathcal{P}(z)}{\mathcal{P}(y)+\mathcal{P}(z)} \frac{|\mathcal{A}(z)|}{\mathcal{P}(z)}\,.$$

Hence $\mathcal{J}(x)$ is smaller than a convex combination of $\mathcal{J}(y)$ and $\mathcal{J}(z)$, which gives that it is smaller than $\mathcal{J}(y)$ or $\mathcal{J}(z)$. 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

$$\forall x \in \mathcal{Y}_n,\ \mathcal{J}(x_*) \geq \mathcal{J}(x)\,.$$

Indeed, repeatedly applying the above lemma starting with a maximizer x_* of $\mathcal{J}$ 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 $\mathcal{J}(x_*) = \mathcal{J}(x'_*)$, or that there exists $y \in \mathcal{Y}_2$ such that $\mathcal{J}(x_*) \leq \mathcal{J}(y)$, which is impossible since in that case $\mathcal{J}(y) = 0$ and $\mathcal{J}(x_*) = \alpha > 0$. We hence have $x'_* \in \mathcal{X}_m$ such that

$$\forall x \in \mathcal{Y}_n,\ \mathcal{J}(x'_*) = \mathcal{J}(x_*) \geq \mathcal{J}(x)\,.$$

We can finally build $x''_* \in \mathcal{X}_n$ such that $\mathcal{J}(x''_*) = \mathcal{J}(x'_*)$ by adding dummy vertices to x'_*, which finally allows to conclude.

D Proof of Proposition 6

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 $J(v) = h\,\|\nabla^h v\|_{1,1}$ where by abuse of notation $\nabla^h v$ is given by (16) with $v_{i,j}$ the value of v in $C_{i,j}$. Hence $J^h(u^h) \le 1$ for all h implies that $J(u^h)$ (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 $||\phi||_\infty \le 1$. The weak-* convergence of the gradients give us that

$$\begin{aligned}&\int_{\mathbb{R}^2} \phi\cdot d\mathrm{D}u = \lim_{h\to 0}\int_{\mathbb{R}^2}\phi\cdot d\mathrm{D}u^h\\ &= \lim_{h\to 0}\sum_{i=0}^N\sum_{j=0}^N \begin{pmatrix}\int_{C^h_{i,j}\cap C^h_{i+1,j}}\phi^{(1)}\,d\mathcal{H}^1\\ \int_{C^h_{i,j}\cap C^h_{i,j+1}}\phi^{(2)}\,d\mathcal{H}^1\end{pmatrix}\cdot\nabla^h u^h_{i,j}\,.\end{aligned}$$

One can moreover show there exists $C>0$ such that for h small enough and all (i,j) we have:

$$\left|\left[\int_{C^h_{i,j}\cap C^h_{i+1,j}}\phi^{(1)}\,d\mathcal{H}^1\right] - h\,\phi^{(1)}(x^h_{i+1,j+1})\right| \le Ch^2\,,$$
$$\left|\left[\int_{C^h_{i,j}\cap C^h_{i,j+1}}\phi^{(2)}\,d\mathcal{H}^1\right] - h\,\phi^{(2)}(x^h_{i+1,j+1})\right| \le Ch^2\,,$$

with $x_{i,j} \overset{\text{def.}}{=} (-R+i\,h,-R+j\,h)$. We use the above inequalities and the fact $\|\phi(x)\|\le 1$ for all x to obtain the existence of $C'>0$ such that for h small enough and for all (i,j) we have:

$$\left\|\begin{pmatrix}\int_{C^h_{i,j}\cap C^h_{i+1,j}}\phi^{(1)}\,d\mathcal{H}^1\\ \int_{C^h_{i,j}\cap C^h_{i,j+1}}\phi^{(2)}\,d\mathcal{H}^1\end{pmatrix}\right\|_2 \le h\sqrt{1+C'h}\,.$$

This finally yields

$$\begin{aligned}&\sum_{i=0}^N\sum_{j=0}^N \begin{pmatrix}\int_{C^h_{i,j}\cap C^h_{i+1,j}}\phi^{(1)}\,d\mathcal{H}^1\\ \int_{C^h_{i,j}\cap C^h_{i,j+1}}\phi^{(2)}\,d\mathcal{H}^1\end{pmatrix}\cdot\nabla^h u^h_{i,j}\\ &\le \sum_{i=0}^N\sum_{j=0}^N h\sqrt{1+C'h}\,\|\nabla^h u^h_{i,j}\| = \sqrt{1+C'h}\;J^h(u^h)\,,\end{aligned}$$

which gives

$$\int_{\mathbb{R}^2}\phi\cdot d\mathrm{D}u \le \limsup_{h\to 0}\sqrt{1+C'h}\;J^h(u^h)\le 1\,.$$

We now have to show that

$$\forall v\in\mathrm{L}^2(\mathbb{R}^2),\;J(v)\le 1 \implies \int_{\mathbb{R}^2}\eta\,u \le \int_{\mathbb{R}^2}\eta\,v\,.$$

Let $v\in C^\infty([-R,R]^2)$ be such that $J(v)\le 1$. We define

$$v^h \overset{\text{def.}}{=} \left(v\left(\left(i+\frac12\right)h,\left(j+\frac12\right)h\right)\right)_{(i,j)\in[1,N]^2}\,.$$

One can then show that

$$\lim_{h\to 0} J^h(v^h) = J(v) = 1\,,$$

so that for every $\delta>0$ we have $J^h\left(\frac{v^h}{1+\delta}\right)\le 1$ for h small enough. Now this yields

$$\begin{aligned}\int_{[-R,R]^2}\eta\,u &= \lim_{h\to 0}\int_{[-R,R]^2}\eta\,u^h\\ &\le \lim_{h\to 0}\int_{[-R,R]^2}\eta\,\frac{v^h}{1+\delta}\\ &= \int_{[-R,R]^2}\eta\,\frac{v}{1+\delta}\,.\end{aligned}$$

Since this holds for all $\delta>0$ we get that

$$\int_{[-R,R]^2}\eta\,u \le \int_{[-R,R]^2}\eta\,v\,. \tag{31}$$

Finally, if $v\in\mathrm{L}^2(\mathbb{R}^2)$ is such that $v=0$ outside $[-R,R]^2$ and $J(v)\le 1$, by standard approximation results (see [Ambrosio et al., 2000, remark 3.22]) we also have that (31) holds, and hence u solves (6). Finally, since u solves (6), its support is included in $[-R,R]^2$, which shows the strong $\mathrm{L}^1_{loc}(\mathbb{R}^2)$ convergence of (u^h) towards u^* in fact implies its strong $\mathrm{L}^1(\mathbb{R}^2)$ convergence.

E First variation of the perimeter and weighted area functionals for simple polygons

We stress that since $\mathcal{X}_n$ is open, for every $x\in\mathcal{X}_n$ the functions $h\mapsto P(E_{x+h})$ and $h\mapsto\int_{E_{x+h}}\eta$ are well-defined in a neighborhood of zero (for any locally integrable function η). We now compute the first variation of these two quantities.

Proposition 15 *Let $x\in\mathcal{X}_n$. Then we have*

$$P(E_{x+h}) = P(E_x) - \sum_{i=1}^n \langle h_i,\tau_i-\tau_{i-1}\rangle + o(\|h\|)\,, \tag{32}$$

where $\tau_i \overset{\text{def.}}{=} \frac{x_{i+1}-x_i}{\|x_{i+1}-x_i\|}$ is the unit tangent vector to $[x_i,x_{i+1}]$.

Proof. If $\|h\|$ is small enough we have:

$$\begin{aligned}&P(E_{x+h}) = \sum_{i=1}^n \|x_{i+1}-x_i+h_{i+1}-h_i\|\\ &= \sum_{i=1}^n\sqrt{\|x_{i+1}-x_i+h_{i+1}-h_i\|^2}\\ &= \sum_{i=1}^n\|x_{i+1}-x_i\|\left(1+\frac{\langle x_{i+1}-x_i,h_{i+1}-h_i\rangle}{\|x_{i+1}-x_i\|^2}+o(\|h\|)\right)\\ &= P(E_x)+\sum_{i=1}^n\langle\tau_i,h_{i+1}-h_i\rangle+o(\|h\|)\,,\end{aligned}$$

and the result follows by re-arranging the terms in the sum. □

Proposition 16 *Let $x\in\mathcal{X}_n$ and $\eta\in C^0(\mathbb{R}^2)$. Then we have*

$$\int_{E_{x+h}}\eta = \int_{E_x}\eta + \sum_{i=1}^n\left\langle h_i, w_i^-\nu_{i-1}+w_i^+\nu_i\right\rangle + o(\|h\|)\,, \tag{33}$$

where ν_i is the outward unit normal to E_x on $]x_i,x_{i+1}[$ and

$$w_i^+ \overset{\text{def.}}{=} \int_{[x_i,x_{i+1}]}\eta(x)\,\frac{\|x-x_{i+1}\|}{\|x_i-x_{i+1}\|}\,d\mathcal{H}^1(x)\,,$$
$$w_i^- \overset{\text{def.}}{=} \int_{[x_{i-1},x_i]}\eta(x)\,\frac{\|x-x_{i-1}\|}{\|x_i-x_{i-1}\|}\,d\mathcal{H}^1(x)\,.$$

Proof. Our proof relies on the following identity (see Lemma 4 for a proof of a closely related formula):

$$\int_{E_x} \eta = \operatorname{sign}\left(\sum_{i=1}^n \det(x_i\ x_{i+1})\right) \sum_{i=1}^n \omega(x_i, x_{i+1})\,,$$

with

$$\omega(a_1,a_2) \overset{\text{def.}}{=} \det(a_1\ a_2) \int_{T_1} \eta((a_1\ a_2)\,y)\,dy\,,$$

where $T_1 \overset{\text{def.}}{=} \left\{(\alpha,\beta)\in(\mathbb{R}_+)^2 \;\middle|\; \alpha+\beta\leq 1\right\}$ is the unit triangle. Assuming $\eta\in C^1(\mathbb{R}^2)$ and denoting $\operatorname{adj}(A)$ the adjugate of a matrix A, we have:

$$\begin{aligned}
&\omega(a_1+h_1,a_2+h_2)\\
=\ &\omega(a_1,a_2)+\det(a_1\ a_2)\int_{T_1}\nabla\eta((a_1\ a_2)\,y)\cdot((h_1\ h_2)y)\,dy\\
+\ &\operatorname{tr}\left(\operatorname{adj}(a_1\ a_2)^T(h_1\ h_2)\right)\int_{T_1}\eta((a_1\ a_2)\,y)\,dy+o(\|h\|)\\
=\ &\omega(a_1,a_2)\\
+\ &\operatorname{sign}(\det(a_1\ a_2))\int_{Oa_1a_2}\nabla\eta(y)\cdot((h_1\ h_2)\,(a_1\ a_2)^{-1})\,y)\,dy\\
+\ &\frac{\operatorname{tr}\left(\operatorname{adj}(a_1\ a_2)^T(h_1\ h_2)\right)}{|\det(a_1\ a_2)|}\int_{Oa_1a_2}\eta(y)\,dy+o(\|h\|)\,.
\end{aligned}$$

Denoting $g(y)\overset{\text{def.}}{=}(h_1\ h_2)(a_1,a_2)^{-1}\,y$, we obtain:

$$\begin{aligned}
&\omega(a_1+h_1,a_2+h_2)\\
=\ &\omega(a_1,a_2)\\
+\ &\operatorname{sign}(\det(a_1\ a_2))\int_{Oa_1a_2}[\nabla\eta\cdot g+\eta\operatorname{div}g]+o(\|h\|)\\
=\ &\omega(a_1,a_2)\\
+\ &\operatorname{sign}(\det(a_1\ a_2))\int_{\partial(Oa_1a_2)}\eta\,(g\cdot\nu_{Oa_1a_2})\,d\mathcal{H}^1+o(\|h\|)\,,
\end{aligned}$$

where we used Gauss-Green theorem to obtain the last equality. Now if $\|h\|$ is small enough then

$$\sum_{i=1}^n \det(x_i+h_i\quad x_{i+1}+h_{i+1}) \text{ and } \sum_{i=1}^n \det(x_i\ x_{i+1})$$

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

$$\begin{aligned}
\epsilon &\overset{\text{def.}}{=} \operatorname{sign}\left(\sum_{i=1}^n\det(x_i\ x_{i+1})\right),\\
\omega_i &\overset{\text{def.}}{=} \int_{\partial^*(Ox_ix_{i+1})}\eta\,(g_i\cdot\nu_{Ox_ix_{i+1}})\,d\mathcal{H}^1\,.
\end{aligned}$$

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. □

F Results used in Section 7

F.1 Properties of the radialisation operator

The goal of this subsection, based on [Dautray and Lions, 2012, II.1.4], is to prove the following result:

Proposition 17 *Let $u\in\mathrm{L}^2(\mathbb{R}^2)$ be s.t. $|\mathrm{D}u|(\mathbb{R}^2)<+\infty$. Then $|\mathrm{D}\tilde{u}|(\mathbb{R}^2)\leq|\mathrm{D}u|(\mathbb{R}^2)$ with equality if and only if u is radial.*

First, one can show that for every $u\in\mathrm{L}^2(\mathbb{R}^2)$, the radialisation $\tilde{u}$ of u defined in Section 7 by

$$\tilde{u}(x)=\int_{\mathbb{S}^1}u(\|x\|\,e)\,d\mathcal{H}^1(e) \tag{34}$$

is well defined and belongs to $\mathrm{L}^2(\mathbb{R}^2)$. Then a change of variables in polar coordinates shows that, as stated in the following lemma, the radialisation operator is self-adjoint.

Lemma 6 *We have*

$$\forall u,v\in\mathrm{L}^2(\mathbb{R}^2),\quad\int_{\mathbb{R}^2}\tilde{u}(x)\,v(x)\,dx=\int_{\mathbb{R}^2}u(x)\,\tilde{v}(x)\,dx\,.$$

We now state a useful identity:

Lemma 7 *For every $\varphi\in C_c^\infty(\mathbb{R}^2\setminus\{0\},\mathbb{R}^2)$, we have:*

$$\langle\mathrm{D}\tilde{u},\varphi\rangle=\left\langle\mathrm{D}u,\widetilde{\varphi\cdot\frac{x}{\|x\|}}\right\rangle,$$

where $\varphi\cdot\frac{x}{\|x\|}$ denotes the mapping $x\mapsto\left(\varphi(x)\cdot\frac{x}{\|x\|}\right)$.

Proof. From Lemma 6 we get

$$\langle\mathrm{D}\tilde{u},\varphi\rangle=\int_{\mathbb{R}^2}\tilde{u}\operatorname{div}\varphi=\int_{\mathbb{R}^2}u\,\widetilde{\operatorname{div}\varphi}\,.$$

Using polar coordinates, defining

$$h(r,\theta)\overset{\text{def.}}{=}(r\cos(\theta),r\sin(\theta))\,,$$

we get

$$\begin{aligned}
(\operatorname{div}\varphi)\,(h(r,\theta))=\ &\frac{1}{r}\frac{\partial}{\partial r}(r(\varphi_r\circ h))(r,\theta)\\
&+\frac{1}{r}\frac{\partial}{\partial\theta}(\varphi_\theta\circ h)(r,\theta)\,,
\end{aligned}\tag{35}$$

where φ_r and φ_θ respectively denote the radial and orthoradial components of φ, i.e.

$$\varphi_r(x)=\varphi(x)\cdot\frac{x}{\|x\|}\text{ and }\varphi_\theta(x)=\varphi(x)\cdot\frac{x^\perp}{\|x\|}\,.$$

The second term in (35) has zero circular mean. Interchanging derivation and integration we get that the radialisation of the first term equals $\frac{1}{r}\frac{\partial}{\partial r}\left(r\,(\widetilde{\varphi_r}\circ h)\right)$, which yields

$$\left(\widetilde{\mathrm{div}\varphi}\right)(x) = \mathrm{div}\left(\varphi\cdot\widetilde{\frac{x}{\|x\|}}\right)(x)\,.$$

□

We now introduce the radial and orthoradial components of the gradient, which are Radon measures on $U \overset{\text{def.}}{=} \mathbb{R}^2\setminus\{0\}$ defined by

$$\forall\psi\in C_c^\infty(U),\ \ \langle \mathrm{D}_{\mathrm{rad}}u,\psi\rangle = \left\langle \mathrm{D}u, \psi\,\frac{x}{|x|}\right\rangle,$$
$$\langle \mathrm{D}_{\mathrm{orth}}u,\psi\rangle = \left\langle \mathrm{D}u, \psi\,\frac{x^\perp}{|x|}\right\rangle.$$

Proposition 18 *There exist two $|\mathrm{D}u|$-measurable mappings from U to $\mathbb{R}$, denoted g_{rad} and g_{orth}, such that*

$g_{\mathrm{rad}}^2 + g_{\mathrm{orth}}^2 \le 1$ *$|\mathrm{D}u|$-almost everywhere*

and

$$\begin{aligned}\forall\psi\in C_c^\infty(U),\ \ \langle \mathrm{D}_{\mathrm{rad}}u,\psi\rangle &= \int_U \psi(x)\,g_{\mathrm{rad}}(x)\,d|\mathrm{D}u|(x)\,,\\ \langle \mathrm{D}_{\mathrm{orth}}u,\psi\rangle &= \int_U \psi(x)\,g_{\mathrm{orth}}(x)\,d|\mathrm{D}u|(x)\,.\end{aligned}\tag{36}$$

Proof. The existence of the $|\mathrm{D}u|$-measurable mappings g_{rad} and g_{orth}, as well as (36), come from Lebesgue differentiation theorem and the fact $\mathrm{D}_{\mathrm{rad}}u$ and $\mathrm{D}_{\mathrm{orth}}u$ are absolutely continuous with respect to $\mathrm{D}u$. Now for every open set $A\subset U$ we have:

$$\begin{aligned}|\mathrm{D}u|(A) &= \sup\left\{\langle \mathrm{D}u,\varphi\rangle \;\middle|\; \varphi\in C_c^\infty(A,\mathbb{R}^2),\ \|\varphi\|_\infty\le 1\right\}\\ &= \sup\left\{\left\langle \mathrm{D}u,\varphi_1\frac{x}{\|x\|}\right\rangle + \left\langle \mathrm{D}u,\varphi_2\frac{x^\perp}{\|x\|}\right\rangle \text{ s.t. }\right.\\ &\qquad\left.\varphi_i\in C_c^\infty(A),\ \left\|\varphi_1^2+\varphi_2^2\right\|_\infty\le 1\right\}\\ &= \sup\left\{\langle \mathrm{D}_{\mathrm{rad}}u,\varphi_1\rangle + \langle \mathrm{D}_{\mathrm{orth}}u,\varphi_2\rangle \text{ s.t. }\right.\\ &\qquad\left.\varphi_i\in C_c^\infty(A),\ \left\|\varphi_1^2+\varphi_2^2\right\|_\infty\le 1\right\}.\end{aligned}$$

Hence for $\varphi_i\in C_c^\infty(A)$ such that $\|\varphi_1^2+\varphi_2^2\|_\infty\le 1$ we have:

$$\int_A 1\,d|\mathrm{D}u| \ge \int_A (g_{\mathrm{rad}}\,\varphi_1 + g_{\mathrm{orth}}\,\varphi_2)\,d|\mathrm{D}u|\,.$$

If we had $g_{\mathrm{rad}}^2+g_{\mathrm{orth}}^2>1$ on a set of non zero measure $|\mathrm{D}u|$, we would have a contradiction. □

We can now prove Proposition 17. Indeed, since $\{0\}$ is $\mathcal{H}^1$-negligible, we have that

$$|\mathrm{D}u|(\{0\}) = |\mathrm{D}\tilde{u}|(\{0\}) = 0\,,$$

and moreover

$$|\mathrm{D}\tilde{u}|(\mathbb{R}^2\setminus\{0\}) \le |\mathrm{D}_{\mathrm{rad}}u|(\mathbb{R}^2\setminus\{0\}) \le |\mathrm{D}u|(\mathbb{R}^2\setminus\{0\})\,.\tag{37}$$

The first equality comes from Lemma 7, while the second is easily obtained from the definition of $\mathrm{D}_{\mathrm{rad}}$. Now if we have $|\mathrm{D}_{\mathrm{rad}}u|(U) = |\mathrm{D}u|(U)$, then we get

$$\int_U g_{\mathrm{rad}}\,d|\mathrm{D}u| = \int_U \sqrt{g_{\mathrm{rad}}^2+g_{\mathrm{orth}}^2}\,d|\mathrm{D}u| = \int_U d|\mathrm{D}u|\,.$$

This yields $g_{\mathrm{orth}}=0$ (and $|g_{\mathrm{rad}}|=1$) $|\mathrm{D}u|$-almost everywhere. Hence $\mathrm{D}_{\mathrm{orth}}u=0$.

Let us now show this implies that u is radial. If we define $A\overset{\text{def.}}{=}]0,+\infty[\times]-\pi,\pi[$, then we have that the mapping given by $h:(r,\theta)\mapsto(r\cos\theta, r\sin\theta)$ is a C^∞-diffeomorphism from A to $\mathbb{R}^2\setminus(\mathbb{R}_-\times\{0\})$. Now if $\xi\in C_c^\infty(A)$ we have that $\xi\circ h^{-1}\in C_c^\infty(h(A))$ and

$$\begin{aligned}0 &= \left\langle \mathrm{D}_{\mathrm{orth}}u, \xi\circ h^{-1}\right\rangle\\ &= \int_{\mathbb{R}^2} u\,\mathrm{div}\left((\xi\circ h^{-1})\,\frac{x^\perp}{\|x\|}\right)\\ &= \int_0^{+\infty}\int_{-\pi}^{\pi}(u\circ h)(r,\theta)\left(\frac{1}{r}\frac{\partial}{\partial\theta}(\xi)(r,\theta)\right)r\,d\theta\,dr\,.\end{aligned}$$

This means that $\frac{\partial\theta}{\partial}(u\circ h)=0$ in the sense of distributions, and hence that there exists[13] a mapping $g:]0,+\infty[\to\mathbb{R}$ such that for almost every $(r,\theta)\in A$, $(u\circ h)(r,\theta)=g(r)$. We finally get $u(x)=g(\|x\|)$ for almost every $x\in h(A)$, which shows u is radial.

F.2 Lemmas used in the proof of Proposition 9

We take $\eta\overset{\text{def.}}{=}\varphi$ and keep the assumptions of Section 7.

Lemma 8 *Let $f:\mathbb{R}\to\mathbb{R}_+$ be square integrable, even and decreasing on $\mathbb{R}_+$. Then for every measurable set A such that $|A|<+\infty$ we have*

$$\int_A f \le \int_{A^s} f\,,$$

where $A^s\overset{\text{def.}}{=}[-\frac{|A|}{2},\frac{|A|}{2}]$. Moreover, equality holds if and only if $|A\triangle A^s|=0$.

Proof. We have

$$\int_A f = \int_0^{+\infty}|\{f\,\mathbf{1}_A\ge t\}|\,dt = \int_0^{+\infty}|\{f\ge t\}\cap A|\,dt\,.$$

For all $t>0$ there exists α such that $\{f\ge t\}=[-\alpha,\alpha]$, so that we have

$$\begin{aligned}|\{f\ge t\}\cap A| &= |[-\alpha,\alpha]\cap A| \le \min(2\alpha,|A|)\\ &= |[-\alpha,\alpha]\cap[-|A|/2,|A|/2]|\\ &= |\{f\ge t\}\cap A^s|\,.\end{aligned}$$

Hence

$$\int_A f \le \int_0^{+\infty}\left|\{f\ge t\}\cap A^S\right|dt = \int_{A^s} f\,.$$

[13] To see this, notice that if we convolve $u\circ h$ with an approximation of unity ρ_ϵ, then we have

$$\frac{\partial}{\partial\theta}\left((u\circ h)\star\rho_\epsilon\right) = (u\circ h)\star\frac{\partial}{\partial\theta}\rho_\epsilon = 0\,,$$

hence the smooth function $(u\circ h)\star\rho_\epsilon$ is equal to some function g_ϵ that depends only on r. Letting $\epsilon\to 0^+$, we see that for almost every (r,θ), $u\circ h$ only depends on r.

Now if $|A \triangle A^s| > 0$ then $|A \setminus A^s| = |A^s \setminus A| > 0$ and we have

$$\begin{aligned}\int_{A^s} f &= \int_{A\cap A^s} f + \int_{A^s\setminus A} f \\ &> \int_{A\cap A^s} f + f\left(\frac{|A|}{2}\right)|A^s \setminus A| \\ &\geq \int_{A\cap A^s} f + \int_{A\setminus A^s} f = \int_A f\,,\end{aligned}$$

which proves the second part of the result. □

Lemma 9 *Let $E \subset \mathbb{R}^2$ be s.t. $0 < |E| < \infty$ and $P(E) < \infty$. Then for any $\nu \in \mathbb{S}^1$, denoting E^s_ν the Steiner symmetrization of E with respect to the line through the origin directed by ν, we have*

$$\frac{\int_{E^s_\nu}\eta}{P(E^s_\nu)} \geq \frac{\int_E \eta}{P(E)}\,,$$

with equality if and only if $|E \triangle E^s_\nu| = 0$.

Proof. From [Maggi, 2012, theorem 14.4] we know that we have $P(E^s_\nu) \leq P(E)$. We now perform a change of coordinates in order to have $E^s_\nu = \left\{(x_1,x_2) \in \mathbb{R}^2 \,|\, |x_2| \leq \frac{\mathcal{L}^1(E_{x_1})}{2}\right\}$ 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, square integrable, even and decreasing on $\mathbb{R}_+$. We can hence apply Lemma 8 and get that

$$\int_E \eta \geq \int_{-\infty}^{+\infty}\left(\int_{(E_{x_1})^s} \eta(x_1,\cdot)\right)d_{x_1} = \int_{E^s_\nu}\eta\,. \tag{38}$$

Moreover, if $|E\triangle E^s_\nu| > 0$, then since

$$\begin{aligned}|E\triangle E^s_\nu| &= \int_0^{+\infty}\left(\int_0^{+\infty} |\mathbf{1}_E(x_1,x_2) - \mathbf{1}_{E^s_\nu}(x_1,x_2)|\,dx_2\right)dx_1 \\ &= \int_0^{+\infty}\left(\int_0^{+\infty} \left|\mathbf{1}_{E_{x_1}}(x_2) - \mathbf{1}_{(E_{x_1})^s}(x_2)\right|\,dx_2\right)dx_1 \\ &= \int_0^{+\infty} \left|E_{x_1}\triangle\left(E^s_{x_1}\right)\right|\,dx_1\,,\end{aligned}$$

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 (38) is strict. □

Lemma 10 *Under Assumption 1, the mapping*

$$\mathcal{G} : R \mapsto \frac{1}{R}\int_0^R r\,\tilde\varphi(r)\,dr$$

has a unique maximizer.

Proof. Since φ (and hence $\tilde\varphi$) is continuous, we have that $\mathcal{G}$ is C^1 on R^*_+ and

$$\mathcal{G}'(R) = \frac{R\,(R\,\tilde\varphi(R)) - \int_0^R r\,\tilde\varphi(r)\,dr}{R^2}\,.$$

Now an integration by part yields that for any continuously differentiable function $h :]0,+\infty[\to \mathbb{R}$ and for any $x > 0$ we have

$$H(x) \overset{\text{def.}}{=} x\,h(x) - \int_0^x h = \int_0^x t\,h'(t)\,dt\,,$$

which shows $H'(x) = x\,h'(x)$. This means the mappings

$$R \mapsto R\,(R\,\tilde\varphi(R)) - \int_0^R r\,\tilde\varphi(r)\,dr \text{ and } R \mapsto f(R) = R\,\tilde\varphi(R)$$

have the same variations. Under Assumption 1, it is then easy to show there exists $R_0 > 0$ such that $\mathcal{G}'(R_0) = 0$, $\mathcal{G}'$ is positive on $]0,R_0[$ and negative on $]R_0,+\infty[$, hence the result. □

F.3 Proof of Proposition 12

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 the boundary of a regular n-gon of radius R with a vertex at $(0,0)$ is given by $r(\theta) = R_n(\theta)$. Under Assumption 1 we have, for all $R > 0$:

$$\begin{aligned}&2\pi R\,\frac{\tan(\pi/n)}{\pi/n}\,|\mathcal{G}(R) - \mathcal{G}_n(R)| \\ &= \left|\frac{\tan(\pi/n)}{\pi/n}\int_0^{2\pi}\int_0^R r\tilde\varphi(r)\,dr\,d\theta \right. \\ &\qquad \left. - \int_0^{2\pi}\int_0^{R_n(\theta)} r\tilde\varphi(r)\,dr\,d\theta\right| \\ &= \left|\int_0^{2\pi}\int_{R_n(\theta)}^R r\tilde\varphi(r)\,dr\,d\theta \right. \\ &\qquad \left. - \left(1 - \frac{\tan(\pi/n)}{\pi/n}\right)\int_0^{2\pi}\int_0^R r\tilde\varphi(r)\,dr\,d\theta\right| \\ &\leq \left[2\pi \sup_{\theta\in[0,2\pi]} |R - R_n(\theta)|\,\|f\|_\infty \right. \\ &\qquad \left. + \left(1 - \frac{\tan(\pi/n)}{\pi/n}\right)2\pi R\,\|f\|_\infty\right] \\ &\leq \|f\|_\infty\left[(1-\cos(\pi/n)) + \left(1 - \frac{\tan(\pi/n)}{\pi/n}\right)\right].\end{aligned}$$

We hence obtain that $|\mathcal{G}(R) - \mathcal{G}_n(R)|_\infty = O\left(\frac{1}{n^2}\right)$.

Now assuming f is of class C^2 and $f''(\rho_0) < 0$ we want to prove that for n large enough, $\mathcal{G}_n$ has a unique maximizer R^*_n and $|R^*_n - R^*| = O\left(\frac{1}{n}\right)$. Denoting $\alpha_n(s) \overset{\text{def.}}{=} \frac{\cos(\pi/n)}{\cos(\pi s/n)}$, we

have:

$$\begin{aligned}\mathcal{G}_n(R) &= \frac{1}{2\pi R\,\frac{\tan(\pi/n)}{\pi/n}} \int_0^{2\pi}\int_0^{R_n(\theta)} r\tilde{\varphi}(r)\,dr\,d\theta \\ &= \frac{1}{2\pi R\,\frac{\tan(\pi/n)}{\pi/n}}\; n\int_0^{2\pi/n}\int_0^{R\frac{\cos(\pi/n)}{\cos(\theta-\pi/n)}} r\tilde{\varphi}(r)\,dr\,d\theta \\ &= \frac{\pi/n}{R\,\tan(\pi/n)}\;\int_0^1\int_0^{R\,\alpha_n(s)} r\tilde{\varphi}(r)\,dr\,ds \\ &= \frac{\pi/n}{\tan(\pi/n)}\,\frac{1}{R}\int_0^R r\left[\int_0^1 \alpha_n(s)^2\,\tilde{\varphi}(r\,\alpha_n(s))\,ds\right]dr\,.\end{aligned}$$

Considering Lemma 10 and defining

$$f_n : r \mapsto r\left[\int_0^1 \alpha_n(s)^2\,\tilde{\varphi}(r\,\alpha_n(s))\,ds\right],$$

we see that showing f_n' is positive on $]0,\rho_1[$ and negative on $]\rho_1,+\infty[$ for some ρ_1 is sufficient to prove $\mathcal{G}_n$ has a unique maximizer. Now we have

$$f_n'(r) = \int_0^1 \alpha_n(s)^2\left(\tilde{\varphi}(r\,\alpha_n(s)) + r\,\alpha_n(s)\,\tilde{\varphi}'(r\,\alpha_n(s))\right)ds\,.$$

The image of $[0,1]$ by $s\mapsto r\,\alpha_n(s)$ is $[r\cos(\pi/n)\,,r]$. Since the mapping $r\mapsto\tilde{\varphi}(r)+r\tilde{\varphi}'(r) = (r\tilde{\varphi})'(r)$ is positive on $]0,\rho_0[$ and negative on $]\rho_0,+\infty[$, we get that f_n' is positive on $]0,\rho_0[$ and negative on $]\rho_0/\cos(\pi/n)\,,+\infty[$ and it hence remains to investigate its sign on $[\rho_0,\rho_0/\cos(\pi/n)]$. But since f is of class C^2 and $f''(\rho_0)<0$ there exists $\epsilon>0$ s.t. $f''(r)<0$ on $]\rho_0-\epsilon,\rho_0+\epsilon[$. For n large enough, we hence have

$$[\rho_0\cos(\pi/n)\,,\rho_0/\cos(\pi/n)]\subset]\rho_0-\epsilon,\rho_0+\epsilon[\,,$$

which implies that

$$\forall r\in[\rho_0,\rho_0/\cos(\pi/n)],\;\; r\,\alpha_n(s)\in]\rho_0-\epsilon,\rho_0+\epsilon[\,,$$

and hence $f_n''(r)<0$. This finally shows there exists ρ_1 such that f_n' is positive on $]0,\rho_1[$ and negative on $]\rho_1,+\infty[$, and the result follows as in the proof of Lemma 10.

Now R^* and R_n^* and are respectively the unique solutions of $F(0,R)=0$ and $F(\pi/n,R)=0$ with

$$\begin{aligned}F(t,R) &\overset{\text{def.}}{=} \left[\int_0^R f_t\right] - R\,f_t(R)\,,\\ f_t(r) &\overset{\text{def.}}{=} r\int_0^1 \alpha(t,s)^2\,\tilde{\varphi}(r\,\alpha(t,s))\,ds\,,\\ \alpha(t,s) &\overset{\text{def.}}{=} \frac{\cos t}{\cos(ts)}\,.\end{aligned}$$

One can then show $\frac{\partial}{\partial R}F(0,R)=0$ if and only if $f_0'(R)=0$, i.e. if and only if $R=\rho_0$. But from the proof of Lemma 10 and the above, it is easy to see neither R^* nor R_n^* equals ρ_0. We can hence apply the implicit function theorem to finally get that $|R^*-R_n^*| = O\left(\frac{1}{n^2}\right)$.

F.4 Proof of Proposition 13

F.4.1 Triangles

Let T be a triangle. Up to a rotation of the axis, we can assume that there exist $a<b$ and two affine functions u,v such that $v\geq u$ and $u(a)=v(a)$ with

$$T=\{(x,y)\in\mathbb{R}^2\,|\,x\in[a,b],\;u(x)\leq y\leq v(x)\}\,.$$

The Steiner symmetrization T_s of T with respect to the line through the origin perpendicular to the side $\{b\}\times[u(b),v(b)]$ is hence obtained by replacing u and v in the definition of T by $(u-v)/2$ and $(v-u)/2$. For all $\theta\in[0,1]$, we define

$$\begin{aligned}u_\theta &\overset{\text{def.}}{=} (1-\theta)\,u+\theta\,(-v)\,,\\ v_\theta &\overset{\text{def.}}{=} \theta\,(-u)+(1-\theta)\,v\,,\end{aligned}$$

and

$$T_\theta \overset{\text{def.}}{=} \left\{(x,y)\in\mathbb{R}^2\,\middle|\,x\in[a,b],\;u_\theta(x)\leq y\leq v_\theta(x)\right\},$$

so that $T_{1/2}=T_s$. Let us now show that

$$\frac{d}{d\theta}\mathcal{J}\left(T_\theta\right)\Big|_{\theta=0}\leq 0\,,$$

with equality if and only if T is symmetric with respect to the symmetrization line.

Weighted area term: first, we have:

$$\int_{T_\theta}\eta = \int_a^b\left(\int_{u_\theta(x)}^{v_\theta(x)}\eta\left(\sqrt{x^2+y^2}\right)dy\right)dx\,.$$

Hence

$$\frac{d}{d\theta}\int_{T_\theta}\eta\Big|_{\theta=0} = -\int_a^b (u+v)(x)\left(g_x(|v|(x))-g_x(|u|(x))\right)dx\,,$$

with $g_x(p)\overset{\text{def.}}{=}\eta\left(\sqrt{x^2+p^2}\right)$. Our assumptions on η ensure that $p\mapsto g_x(p)$ is decreasing, so that $g_x(|v|(x))-g_x(|u|(x))$ and $|u|(x)-|v|(x)$ have the same sign. But since

$$-(u(x)+v(x)) \;\text{ and }\; -\left(|u|(x)-|v|(x)\right)$$

also have the same sign, we have that

$$-(u+v)(x)\left(g_x(|v|(x))-g_x(|u|(x))\right)<0\,,$$

unless $u(x)=v(x)$ or $u(x)=-v(x)$. Since u and v are affine and $u(a)=v(a)$, the first equality can not hold for any $x\in]a,b[$ (otherwise we would have $u=v$ on $[a,b]$ and T would be flat). Moreover, $u(x)=-v(x)$ almost everywhere on $[a,b]$ if and only if $T=T_s$. Hence $\frac{d}{d\theta}\int_{T_\theta}\eta\big|_{\theta=0}\leq 0$ with equality if and only if $T=T_s$.

Perimeter term: now, the perimeter of T_θ is given by

$$\begin{aligned}P(T_\theta) &= \int_a^b\sqrt{1+|\nabla u_\theta|^2}+\int_a^b\sqrt{1+|\nabla v_\theta|^2}+v_\theta(b)-u_\theta(b)\\ &= (b-a)\left(f(\nabla u_\theta)+f(\nabla v_\theta)\right)+\left(v(b)-u(b)\right),\end{aligned}$$

with $f(p)\overset{\text{def.}}{=}\sqrt{1+\|p\|^2}$, this last function being strictly convex. Now since

$$\begin{cases}\nabla u_\theta = \nabla u-\theta(\nabla u+\nabla v)\,,\\ \nabla v_\theta = \nabla v-\theta(\nabla u+\nabla v)\,,\end{cases}$$

we get

$$\begin{aligned}\frac{d}{d\theta}P(T_\theta)\Big|_{\theta=0} &= (b-a)\left[\nabla f(\nabla u)+\nabla f(\nabla v)\right]\cdot\left[-(\nabla u+\nabla v)\right]\\ &= -(b-a)\left[\nabla f(\nabla u)-\nabla f(-\nabla v)\right]\cdot\left[\nabla u-(-\nabla v)\right],\end{aligned}$$

and the strict convexity of f hence shows

$$\frac{d}{d\theta}P(T_\theta)\Big|_{\theta=0}\leq 0\,,$$

with equality if and only if $\nabla u=-\nabla v$, which means, up to a translation, that T is equal to T_s.

Applying the above arguments to all three sides finally yields the result.

F.4.2 Quadrilaterals

Let Q be a simple quadrilateral. Up to a rotation of the axis, we can assume that there exist $a < b < c$ and four affine functions u_1, v_1, u_2, v_2 such that

$$\begin{cases} v_1 \geq u_1,\ v_2 \geq u_2\,, \\ u_1(a) = v_1(a),\ u_2(c) = v_2(c)\,, \\ u_1(b) = u_2(b),\ v_1(b) = v_2(b)\,, \end{cases}$$

with $Q = T_1 \cup T_2$ and

$$T_i \overset{\text{def.}}{=} \{(x,y) \in \mathbb{R}^2 \,|\, x \in [a,b],\ u_i(x) \leq y \leq v_i(x)\}\,.$$

For all $\theta \in [0,1]$ and $i \in \{1,2\}$, we define

$$\begin{cases} u_{i,\theta} \overset{\text{def.}}{=} (1-\theta)\,u_i + \theta\,(-v_i)\,, \\ v_{i,\theta} \overset{\text{def.}}{=} \theta\,(-u_i) + (1-\theta)\,v_i\,, \end{cases}$$

and $Q_\theta = T_{1,\theta} \cup T_{2,\theta}$ with

$$T_{i,\theta} = \left\{(x,y) \in \mathbb{R}^2 \,|\, x \in [a,b],\ u_{i,\theta}(x) \leq y \leq v_{i,\theta}(x)\right\},$$

so that the Steiner symmetrization Q^s of Q with respect to the ligne through the origin perpendicular to the diagonal $\{b\} \times [u_1(b), v_1(b)]$ satisfies $Q_{1/2} = Q^s$.

Weighted area term: the fact $\frac{d}{d\theta}\int_{Q_\theta} \eta\big|_{\theta=0} \leq 0$ with equality if and only if $Q = Q^s$ can easily be deduced from the case of triangles using the fact that $\int_{Q_\theta} \eta = \int_{T_{1,\theta}} \eta + \int_{T_{2,\theta}} \eta$.

Perimeter term: now, the perimeter of Q_θ is given by:

$$\begin{aligned} P(Q_\theta) &= \int_a^b \sqrt{1+|\nabla u_{1,\theta}|^2} + \int_a^b \sqrt{1+|\nabla v_{1,\theta}|^2} \\ &\quad + \int_b^c \sqrt{1+|\nabla u_{2,\theta}|^2} + \int_b^c \sqrt{1+|\nabla v_{2,\theta}|^2} \\ &= (b-a)\,(f(\nabla u_{1,\theta}) + f(\nabla v_{1,\theta})) \\ &\quad + (c-b)\,(f(\nabla u_{2,\theta}) + f(\nabla v_{2,\theta})) \end{aligned}$$

with $f(p) \overset{\text{def.}}{=} \sqrt{1+\|p\|^2}$ as before. We then get

$$\begin{aligned} &\frac{d}{d\theta} P(Q_\theta)\Big|_{\theta=0} \\ &= -\,(b-a)\,[\nabla f(\nabla u_1) - \nabla f(-\nabla v_1)] \cdot [\nabla u_1 - (-\nabla v_1)] \\ &\quad -\,(c-b)\,[\nabla f(\nabla u_2) - \nabla f(-\nabla v_2)] \cdot [\nabla u_2 - (-\nabla v_2)]\,, \end{aligned}$$

and the strict convexity of f hence shows

$$\frac{d}{d\theta} P(Q_\theta)\Big|_{\theta=0} \leq 0\,,$$

with equality if and only if $\nabla u_1 = -\nabla v_1$ and $\nabla u_2 = -\nabla v_2$, which means, up to a translation, that Q is equal to Q^s.