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[1], Vincent Duval[2,3], and Romain Petit[2,3]

[1] Institut Camille Jordan, CNRS UMR 5208, École Centrale de Lyon
F-69134 Écully, France
[2] CEREMADE, CNRS, UMR 7534, Université Paris-Dauphine, PSL University
75016 Paris, France
[3] INRIA-Paris, MOKAPLAN, 75012 Paris, France

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", at the convergence of the algorithm.

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 sum-of-squares reformulation of the problem [7] or a Frank-Wolfe / conditional gradient algorithm [6] which approximates 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 [11]).

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

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

where $\Phi\colon \mathrm{L}^2(\mathbb{R}^2) \to \mathbb{R}^m$ is a continuous linear map such that

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

with $\varphi \in \left[L^2(\mathbb{R}^2)\right]^m \cap C^0(\mathbb{R}^2, \mathbb{R}^m)$ and $|\mathrm{D}u|(\mathbb{R}^2)$ denotes the total variation of (the gradient of) u. Such variational problems have been widely used in image processing for the last decades, following the pioneering works of Rudin, Osher and Fatemi [15].

Many algorithms have been proposed to solve (1). With the notable exception of [17], most of them rely on the introduction of a fixed discrete grid, which yields reconstruction artifacts such as anisotropy or blur (see the experiments in [16]). On the other hand, it is known that some solutions of (1) are sums of a finite number of indicator functions of simply connected sets [4, 5], yielding piecewise constant images. Our goal is to design an algorithm which does not suffer from some grid bias while providing solutions built from the above-mentioned atoms.

2 A modified Frank-Wolfe algorithm

In the spirit of [3, 6, 9] 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 (1) in a gridless manner.

2.1 Description

The Frank-Wolfe algorithm allows to minimize a convex differentiable function f over a weakly compact convex subset C of a Banach space. It is possible to recast problem (1) into that framework, e.g. by using an epigraphical lift (see [9, Sec. 4.1] for the case of the sparse spikes problem), or using a direct analysis (see [6]).

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. In our case, the minimization of the linearized function is of the form

$$\inf_{u \in \mathrm{L}^2(\mathbb{R}^2)} \int_{\mathbb{R}^2} \eta\, u \quad \text{s.t.} \quad |\mathrm{D}u|\left(\mathbb{R}^2\right) \leq 1 \tag{3}$$

for an iteration-dependent function $\eta \in \mathrm{L}^2(\mathbb{R}^2)$. Since the objective 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 [10] (see also [2]) 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 (the measure-theoretic analog of simply connected sets) with $0 < |E| < +\infty$ and $P(E) < \infty$, and $P(E)$ denotes the perimeter of E. This means the linear minimization step can be carried out by finding a simple set solving the following geometric variational problem:

$$J(E) \stackrel{\text{def}}{=} \sup_{E \subseteq \mathbb{R}^2} \frac{\left|\int_E \eta\right|}{P(E)} \quad \text{s.t.} \quad 0 < P(E) < \infty. \tag{4}$$

As a result, the iterates produced by the algorithm may be constructed as linear combinations of indicators of simple sets. Since Problem (4) is reminiscent of the well-known Cheeger problem [13], 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.

An important feature of the Frank-Wolfe algorithm is that, when building the next iterate, one can pick any admissible point that decreases the objective more than the standard update, without breaking convergence guarantees. Exploiting this, [3, 6, 9] have introduced "sliding steps" which help identify the sparse structure of the sought-after signal. Following the above-mentioned works, we derive an update step which provably decreases the objective more than the standard update, and derive Algorithm 1. As already mentioned, adding the sliding step to the Frank-Wolfe algorithm does not break its convergence properties. In particular the following property holds (see [11]):

Algorithm 1: Modified Frank-Wolfe algorithm

Data: measurement operator Φ, obs. y, max. num. of iter. n, reg. param. λ
Result: function u^*

$u^{[0]} \leftarrow 0$;
for k *from* 0 *to* $n-1$ **do**
  $\eta^{[k]} \leftarrow \frac{1}{\lambda}\Phi^*\left(y - \Phi u^{[k]}\right)$;
  $E_*^{[k]} \leftarrow \operatorname*{argmax}_E \frac{1}{P(E)}\left|\int_E \eta^{[k]}\right|$;
  if $\left|\int_{E_*^{[k]}} \eta^{[k]}\right| \leq P(E_*^{[k]})$ **then**
    output $u^* \leftarrow u^{[k]}$, which is optimal;
  else
    $E^{[k+1]} \leftarrow (E_1^{[k]}, ..., E_{N^{[k]}}^{[k]}, E_*^{[k]})$;
    $a^{[k+1]} \leftarrow \operatorname*{argmin}_{a \in \mathbb{R}^{N^{[k]}+1}} T_\lambda\left(\sum_{i=1}^{N^{[k]}+1} a_i \mathbf{1}_{E_i^{[k+1]}}\right)$;
    remove atoms with zero amplitude, $N^{[k+1]} \leftarrow$ num. of atoms in $E^{[k+1]}$;
    update $(E^{[k+1]}, a^{[k+1]})$ by performing a local descent on
      $(E, a) \mapsto T_\lambda\left(\sum_{i=1}^{N^{[k+1]}} a_i \mathbf{1}_{E_i}\right)$ initialized with $(E^{[k+1]}, a^{[k+1]})$;
    $u^{[k+1]} \leftarrow \sum_{i=1}^{N^{[k+1]}} a_i^{[k+1]} \mathbf{1}_{E_i^{[k+1]}}$;
  end
end

Proposition 1. *Let $(u^{[k]})_{k\geq 0}$ be a sequence produced by Algorithm 1. Then there exists $C > 0$ such that for any solution u_* of Problem* (1)*,*

$$\forall k \in \mathbb{N}^*,\ T_\lambda(u^{[k]}) - T_\lambda(u_*) \leq \frac{C}{k}. \tag{5}$$

Moreover, $(u^{[k]})_{k\geq 0}$ *has an accumulation point in the weak* $\mathrm{L}^2(\mathbb{R}^2)$ *topology and that point is a solution of Problem* (1).

The introduction of the sliding step (Line 11 of Algorithm 1), first proposed in [6], allows to considerably improve the convergence speed in practice (the variant proposed in [9], that we use here, even provides finite time convergence in the setting of the sparse spikes problem). It 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 over the first iterations. The "local descent" mentioned at Line 11 is discussed in the next section.

2.2 Implementation

The implementation of Algorithm 1[4] requires two oracles to carry out the operations described on Lines 4 and 11: a first one that, given a weight function η, returns a solution of (4), and a second one that, given a set of initial amplitudes and atoms, returns another such set which is obtained by performing a local descent on the objective. Our approach for designing these oracles relies on polygonal approximations: we look for a maximizer of J among polygons, and perform the sliding step by finding a sequence of amplitudes and polygons so as to iteratively decrease the objective. This choice is mainly motivated by our goal to solve (1) "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.

Given an integer $n \geq 3$, we denote by $\mathcal{X}_n$ the set of all $x \in (\mathbb{R}^2)^n$ such that the polygon E_x defined by the list of vertices x is simple and counterclockwise oriented. We also define J_n by setting, for all $x \in \mathcal{X}_n$, $J_n(x) \overset{\text{def.}}{=} J(E_x)$. The area and the perimeter functionals, and hence J_n, are continuously differentiable on the open set $\mathcal{X}_n$.

Polygonal approximation of Cheeger sets. Our method for carrying out Line 4 in Algorithm 1 consists of several steps. First, we solve a discrete version of (3), where the minimization is performed over the set of piecewise constant functions on a given mesh. Then, we extract a level set of the solution, and obtain a simple polygon whose edges are located on the edges of the mesh. Finally, we use a first order method initialized with the previously obtained polygon to locally maximize J_n. Given a simple polygon $E_{x^t} = (x_1^t, ..., x_n^t)$ and a step size α^t, the next iterate is defined as the polygon $E_{x^{t+1}} = (x_1^{t+1}, ..., x_n^{t+1})$ such that[5]:

$$x_j^{t+1} \overset{\text{def.}}{=} x_j^t - \frac{\alpha^t}{P(E_{x^t})^2}\Big(P(E_{x^t})\big(w_j^{t-}\nu_{j-1}^t + w_j^{t+}\nu_j^t\big) + \left[\int_{E_{x^t}} \eta\right](\tau_j^t - \tau_{j-1}^t)\Big) \quad (6)$$

[4] A repository containing a complete implementation of Algorithm 1 can be found online at `https://github.com/rpetit/tvsfw`.

[5] Details on the first variations of perimeter and area can be found in [12, sec. 17.3].

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

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

We compute the weights w_j^{t+}, w_j^{t-} and the integral of η over E_{x^t} using numerical integration methods. We found that integrating the weight function η on each triangle of a sufficiently fine triangulation of E_{x^t} yields good numerical results (this triangulation must be updated at each iteration, and sometimes re-computed from scratch to avoid the presence of ill-shaped triangles).

Comments. Two potential concerns regarding the above procedure are whether iterates remain simple polygons and whether it converges to a global maximizer of J_n. As the mesh used to solve the discrete version of (3) gets finer, the level sets of any of the solutions can be guaranteed to be arbitrarily close (in terms of the Lebesgue measure of the symmetric difference) to a solution of (4) (see [8]).

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 (4), we do not expect nor observe any change of topology during the optimization.

Moreover, even if the non-concavity of J_n makes its maximization difficult, the above initialization allows us to start our local descent with a polygon that hopefully lies in the basin of attraction of a global maximizer. However, let us stress that at Line 4 of Algorithm 1, we only need to find a set with near optimal energy J in (4) (see [11, Algorithm 2]).

An interesting problem is to quantify the distance (e.g. in the Haussdorff sense) of a maximizer of J_n to a maximizer of J. We discuss in Section 4 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

$$w_j^+ = w_j^- = \frac{\int_{E_x} \eta}{P(E_x)} \tan\left(\frac{\theta_j}{2}\right) \tag{7}$$

for all j, where θ_j is the angle between $x_j - x_{j-1}$ and $x_{j+1} - x_j$ (the j-th exterior angle of the polygon). This can be seen as a discrete version of the following first order optimality condition for solutions of (4):

$$\eta = \frac{\int_E \eta}{P(E)} \, H_E \text{ on } \partial^* E \tag{8}$$

where H_E denotes the distributional curvature of E. Note that (8) 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 [1] or [13, Prop. 2.4]).

Sliding step. The implementation of the local descent (Line 11 in Algorithm 1) is similar to what is described above for refining crude approximations of Cheeger sets. We use a first order method to minimize the function that maps a list of amplitudes and a list of polygons (each seen as a list of vertices) to the objective value of the linear combination of indicator functions they define. Given a step size α^t, N polygons $E_{x_1^t}, ..., E_{x_N^t}$, and a vector of amplitudes $a^t \in \mathbb{R}^N$, 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:

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

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

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

Comments. Unlike the local descent we perform to approximate Cheeger sets, the sliding step may tend to induce topology changes (see Section 3.2 for an example). 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) [3, 6, 9]. 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 $d > k$.

Let us stress that these dimension changes have a clear interpretation in terms of the topology of the sets which evolve through the descent: they typically correspond to a splitting. Our current implementation does not allow to handle them in a consistent way, and finding a way to deal with them "off-the-grid" is the subject of an ongoing work.

It is important to note that not allowing topological changes during the descent is not an issue, since all convergence guarantees of Algorithm 1 are preserved as soon as the output of the sliding step decreases the energy more than the standard Frank-Wolfe update. One can 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 handle topological changes.

3 Numerical experiments

3.1 Recovery examples

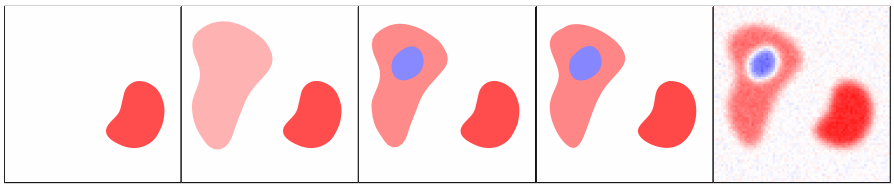
Here, we investigate the practical performance of Algorithm 1. We focus on the case where Φ is a sampled Gaussian convolution operator i.e.:

$$\forall x \in \mathbb{R}^2,\ \varphi(x) = \left(\exp\left(-\frac{||x - x_i||^2}{2\sigma^2} \right) \right)_{i \in \{1,\dots,m\}}$$

for a given $\sigma > 0$ and a sampling grid $(x_i)_{i\in\{1,\dots,m\}}$. The noise is drawn from a multivariate Gaussian with zero mean and isotropic covariance matrix $\tau^2 I_m$. We take the regularization parameter λ of the order of $\sqrt{2\log(m)\,\tau^2}$.

Numerically certifying that a given function is an approximate solution of (1) 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 (1) with $y = \Phi u_0 + w$ are close to u_0, modulo the regularization effects of the total variation.

Our first experiment consists in recovering a function u_0 that is a linear combination of three indicator functions (see Figure 1). 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 (Line 11) 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.



(a) $u^{[1]}$ (b) $u^{[2]}$ (c) $u^{[3]}$ (d) u_0 (e) observations

Fig. 1: First experiment ($u^{[k]}$ is the k-th iterate produced by Algorithm 1)

The second experiment (see Figure 2) 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. 3) 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. 4 the results of an experiment on a more challenging task, which consists in reconstructing a natural grayscale image.

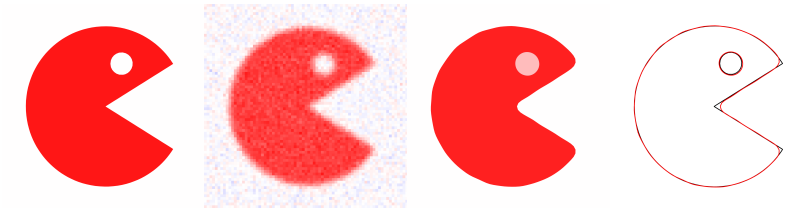


Fig. 2: From left to right: unknown function, observations, final estimate, gradients support (red: estimate, black: unknown)

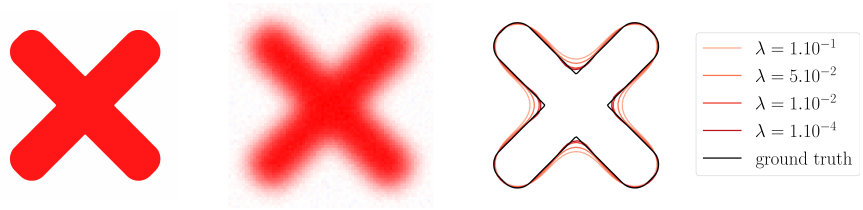


Fig. 3: Left: unknown function, middle: observations, right: output of the algorithm for different values of λ

Fig. 4: From left to right: original image, observations, and iterates $u^{[k]}$ ($k \in \{2, 4, 11\}$) produced by the algorithm

3.2 Topology changes during the local descent

Here, we illustrate the changes of topology that may occur during the sliding step (Line 11 of Algorithm 1). All relevant plots are given in Figure 5. 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 4 of the first iteration covers the two disks (see (c)).

In this setting, our implementation of the local descent (Line 11) converges to a function similar to (f)[6], and we obtain a valid update that decreases the objective more than the standard Frank-Wolfe update. The next iteration of the

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

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).

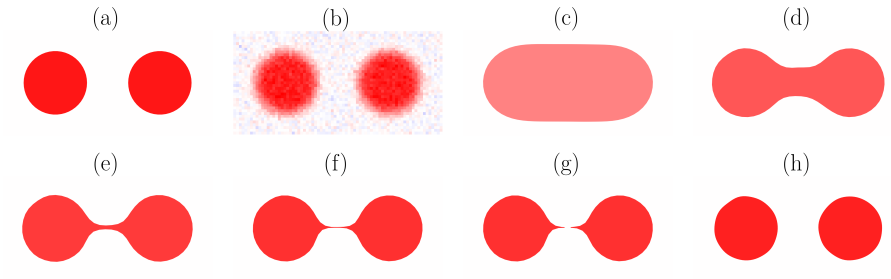


Fig. 5: Topology change experiment. (a): unknown signal, (b): observations, (c): weighted Cheeger set, (d,e,f,g): sliding step iterations (with splitting), (h): final function.

4 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. To be more precise we assume that the measurement operator is given by (2) with $\varphi : \mathbb{R}^2 \to \mathbb{R}_+^*$ a positive radial function with $\varphi(x) = \tilde{\varphi}(||x||)$. We may invoke the following assumption:

Assumption 1. The function $f : r \mapsto r\,\tilde{\varphi}(r)$ is continuously differentiable on $\mathbb{R}_+^*$, there exists $\rho_0 > 0$ s.t. $f'(r) > 0$ on $]0, \rho_0[$, $f'(r) < 0$ on $]\rho_0, +\infty[$, and $rf(r) \to 0$ when $r \to +\infty$.[7]

We first explain what each step of Algorithm 1 should theoretically return, without worrying about approximations made for implementation matters. Then, we compare those with the output at each step of the practical algorithm. The proofs of all results, which are omitted here, will be given in a longer version of this work in preparation.

4.1 Theoretical behavior of the algorithm

The first step of Algorithm 1 consists in solving the Cheeger problem (4) associated to $\eta \overset{\text{def}}{=} \frac{1}{\lambda}\Phi^* y = \frac{y}{\lambda}\varphi$ (or equivalently to φ). Using the Steiner symmetrization

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

principle, and denoting by $B(0,R)$ the disk of radius R centered at the origin, we may state:

Proposition 2. *All the solutions of the Cheeger problem* (4) *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* $R \mapsto \left[\int_0^R r\,\tilde{\varphi}(r)\,dr\right]/R$.

The second step (Line 9) of the algorithm then consists in solving

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

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

$$a^* = \frac{\operatorname{sign}(y)}{\int_{E^*}\varphi}\left(|y| - \lambda\,\frac{P(E^*)}{\int_{E^*}\varphi}\right)^+ \tag{9}$$

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

The next step should be the sliding one (Line 11). However, in this specific setting, one can show that the constructed function is already optimal, as stated by the following proposition:

Proposition 3. *Under Assumption 1, Problem* (1) *has a unique solution, which is of the form* $a^*\,\mathbf{1}_{E^*}$ *with* $E^* = B(0,R^*)$ *the solution of the Cheeger problem given by Prop. 2, and* a^* *given by* (9).

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.

4.2 Study of implementation approximations

In practice, instead of solving (4), we look for an n-gon minimizing J, for some given integer n. Investigating the proximity of this n-gon with $B(0,R^*)$ is hence natural. Solving classical geometric variational problems restricted to the set of n-gons is more involved, as the Steiner symmetrization procedure does not preserve n-gons in general [14, Sec. 7.4]. However, using a trick from Pólya and Szegö, one may prove:

Proposition 4. *Let* $n \in \{3,4\}$. *Then all the maximizers of* J_n *are regular* n*-gons inscribed in a circle centered at the origin.*

Our proof does not extend to $n \geq 5$, but the following conjecture is natural:

Conjecture 1. The result of Prop. 4 holds for all $n \geq 3$.

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

Proposition 5. *Under Assumption 1, we have that* $||\mathcal{J}_n - \mathcal{J}||_\infty = O\left(\frac{1}{n^2}\right)$. *Moreover, if* f *is of class* C^2 *and* $f''(\rho_0) < 0$ *, then for* n *large enough, there exists a unique maximizer* R_n^* *of* $\mathcal{J}_n$*, 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 2.2, is a polygon that is obtained by locally maximizing J_n using a first order method. Even if we carefully initialize this first order method, the possible existence of non-optimal critical points makes its analysis challenging. However, in our setting (a radial weight function), the critical points of J_n coincide with the global maximizers (at least for small n).

Proposition 6. *Let* $n \in \{3, 4\}$. *Under Assumption 1, if* f *is of class* C^2 *and* $f''(\rho_0) < 0$*, the critical points of* J_n *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 of Prop. 6 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 $B(0, R^*)$.

5 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.

Acknowledgments

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. RP warmly thanks the owners of Villa Margely for their hospitality during the writing of this paper.

References

1. Alter, F., Caselles, V., Chambolle, A.: Evolution of characteristic functions of convex sets in the plane by the minimizing total variation flow. Interfaces and Free Boundaries **7**(1), 29–53 (Mar 2005). https://doi.org/10.4171/IFB/112

2. Ambrosio, L., Caselles, V., Masnou, S., Morel, J.M.: Connected components of sets of finite perimeter and applications to image processing. Journal of the European Mathematical Society **3**(1), 39–92 (Feb 2001). https://doi.org/10.1007/PL00011302
3. Boyd, N., Schiebinger, G., Recht, B.: The Alternating Descent Conditional Gradient Method for Sparse Inverse Problems. SIAM Journal on Optimization **27**(2), 616–639 (Jan 2017). https://doi.org/10.1137/15M1035793
4. Boyer, C., Chambolle, A., De Castro, Y., Duval, V., de Gournay, F., Weiss, P.: On Representer Theorems and Convex Regularization. SIAM Journal on Optimization **29**(2), 1260–1281 (Jan 2019). https://doi.org/10.1137/18M1200750
5. Bredies, K., Carioni, M.: Sparsity of solutions for variational inverse problems with finite-dimensional data. Calculus of Variations and Partial Differential Equations **59**(1), 14 (Dec 2019). https://doi.org/10.1007/s00526-019-1658-1
6. Bredies, K., Pikkarainen, H.K.: Inverse problems in spaces of measures. ESAIM: Control, Optimisation and Calculus of Variations **19**(1), 190–218 (Jan 2013). https://doi.org/10.1051/cocv/2011205
7. Candès, E.J., Fernandez-Granda, C.: Towards a Mathematical Theory of Super-resolution. Communications on Pure and Applied Mathematics **67**(6), 906–956 (2014). https://doi.org/10.1002/cpa.21455
8. Carlier, G., Comte, M., Peyré, G.: Approximation of maximal Cheeger sets by projection. ESAIM: Mathematical Modelling and Numerical Analysis **43**(1), 139–150 (Jan 2009). https://doi.org/10.1051/m2an/2008040
9. Denoyelle, Q., Duval, V., Peyre, G., Soubies, E.: The Sliding Frank-Wolfe Algorithm and its Application to Super-Resolution Microscopy. Inverse Problems (2019). https://doi.org/10.1088/1361-6420/ab2a29
10. Fleming, W.H.: Functions with generalized gradient and generalized surfaces. Annali di Matematica Pura ed Applicata **44**(1), 93–103 (Dec 1957). https://doi.org/10.1007/BF02415193
11. Jaggi, M.: Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization. In: International Conference on Machine Learning. pp. 427–435. PMLR (Feb 2013)
12. Maggi, F.: Sets of Finite Perimeter and Geometric Variational Problems: An Introduction to Geometric Measure Theory. Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge (2012). https://doi.org/10.1017/CBO9781139108133
13. Parini, E.: An introduction to the Cheeger problem. Surveys in Mathematics and its Applications **6**, 9–21 (2011)
14. Pólya, G., Szegö, G.: Isoperimetric Inequalities in Mathematical Physics. No. 27 in Annals of Mathematics Studies, Princeton University Press (1951)
15. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena **60**(1), 259–268 (Nov 1992). https://doi.org/10.1016/0167-2789(92)90242-F
16. Tabti, S., Rabin, J., Elmoata, A.: Symmetric Upwind Scheme for Discrete Weighted Total Variation. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1827–1831 (Apr 2018). https://doi.org/10.1109/ICASSP.2018.8461736
17. Viola, F., Fitzgibbon, A., Cipolla, R.: A unifying resolution-independent formulation for early vision. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 494–501 (Jun 2012). https://doi.org/10.1109/CVPR.2012.6247713