



HAL
open science

Grid is Good: Adaptive Refinement Algorithms for Off-the-Grid Total Variation Minimization

Axel Flinth, Frédéric de Gournay, Pierre Weiss

► **To cite this version:**

Axel Flinth, Frédéric de Gournay, Pierre Weiss. Grid is Good: Adaptive Refinement Algorithms for Off-the-Grid Total Variation Minimization: Grid is good. 2023. hal-03937286

HAL Id: hal-03937286

<https://inria.hal.science/hal-03937286>

Preprint submitted on 13 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Grid is Good: Adaptive Refinement Algorithms for Off-the-Grid Total Variation Minimization

Axel Flinth, Frédéric de Gournay and Pierre Weiss

January 13, 2023

Abstract

We propose an adaptive refinement algorithm to solve total variation regularized measure optimization problems. The method iteratively constructs dyadic partitions of the unit cube based on i) the resolution of discretized dual problems and ii) on the detection of cells containing points that violate the dual constraints. The detection is based on upper-bounds on the dual certificate, in the spirit of branch-and-bound methods. The interest of this approach is that it avoids the use of heuristic approaches to find the maximizers of dual certificates. We prove the convergence of this approach under mild hypotheses and a linear convergence rate under additional non-degeneracy assumptions. These results are confirmed by simple numerical experiments.

1

1 Introduction

We develop and analyze an algorithm to solve the following problem:

$$\inf_{\mu \in \mathcal{M}(\Omega)} J(\mu) \stackrel{\text{def.}}{=} \|\mu\|_{\mathcal{M}} + f(A\mu), \quad (\mathcal{P}(\Omega))$$

where $\Omega \subset \mathbb{R}^D$ is a compact domain, $\mathcal{M}(\Omega)$ is the set of Radon measures on Ω , $A : \mathcal{M}(\Omega) \rightarrow \mathbb{R}^M$ is a continuous operator, $\|\cdot\|_{\mathcal{M}}$ is the total variation and $f : \mathbb{R}^M \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex, proper function. A key property of this problem is that it admits *sparse* solutions (see e.g. [4]) of the form

$$\mu^* = \sum_{s=1}^S \alpha_s^* \delta_{x_s^*},$$

where $\alpha_s^* \in \mathbb{R}$ are weights, $x_s^* \in \Omega$ are locations and where the number of sources S satisfies $S \leq M$. This feature shows that a finite dimensional problem is somewhat hidden in $(\mathcal{P}(\Omega))$. It makes it possible to design specific and efficient numerical procedures, for the challenging measure optimization problem $\mathcal{P}(\Omega)$.

Applications This problem and its variants appear in various fields. In *inverse problems*, it is used heavily for sparse source localization and super-resolution [10, 5, 32, 6, 15]. It is also used in *optimal control* with sparse controls [9, 23]. In *approximation theory*, a “generalized” version of this problem was revisited recently in [35]. Given a surjective Fredholm operator $L : \mathcal{B}(\Omega) \rightarrow \mathcal{M}(\Omega)$, where $\mathcal{B}(\Omega)$ is a suitably defined Banach space, consider the following problem:

$$\inf_{u \in \mathcal{B}(\Omega)} \|Lu\|_{\mathcal{M}} + f(Au). \quad (1)$$

¹The title of this article is a reference to the seminal paper of J. Tropp on sparse approximation [34], which was itself a reference to the movie Wall Street of Oliver Stone.

The solutions of this problem are (generalized) splines with free knots [35]. Following [18] and letting L^+ denote a pseudo-inverse of L , this problem can be rephrased as

$$\inf_{\substack{\mu \in \mathcal{M}(\Omega) \\ u_K \in \ker(L)}} \|\mu\|_{\mathcal{M}} + f(A(L^+\mu + u_K)), \quad (2)$$

which is an instance of $(\mathcal{P}(\Omega))$.

Exchange algorithms In this work, we will introduce and study a variant of an *exchange algorithm*. To explain its principle, let us introduce the dual problem $(\mathcal{D}(\Omega))$ to $(\mathcal{P}(\Omega))$:

$$\sup_{\substack{q \in \mathbb{R}^M \\ \|A^*q\|_{L^\infty(\Omega)} \leq 1}} -f^*(q), \quad (\mathcal{D}(\Omega))$$

where $\|A^*q\|_{L^\infty(\Omega)} \stackrel{\text{def.}}{=} \sup_{x \in \Omega} |A^*q|(x)$. This dual formulation will play a pivotal role in our analysis. It involves the optimization of a finite dimensional variable $q \in \mathbb{R}^M$, subject to an infinite number of linear constraints $\{|A^*q|(x)| \leq 1, \forall x \in \Omega\}$. It is therefore called a *semi-infinite* program [21, 29, 20]. The first algorithm proposed to tackle it is usually attributed to Remez and his exchange algorithm [30]. It dates back to the 1930's and was adapted to a specific problem of the form $\mathcal{P}(\Omega)$. The general idea is to define a sequence of discretization sets $(\mathcal{V}_k)_{k \in \mathbb{N}}$, where \mathcal{V}_k is a finite set of points (vertices) in Ω . We can then define the discretized primal and dual problems as follows

$$\inf_{\mu \in \mathcal{M}(\mathcal{V}_k)} \|\mu\|_{\mathcal{M}} + f(A\mu) \quad (\mathcal{P}(\mathcal{V}_k))$$

$$\sup_{\substack{q \in \mathbb{R}^M \\ \|A^*q\|_{L^\infty(\mathcal{V}_k)} \leq 1}} -f^*(q). \quad (\mathcal{D}(\mathcal{V}_k))$$

Both problems are finite dimensional convex problem, and can be solved with off-the-shelf solvers. The simplest approach is to define \mathcal{V}_k as a Euclidean grid with edge-length 2^{-k} [31, 12]. Proving the convergence of this approach is rather straightforward. A problem however is the numerical complexity which explodes rapidly while k increases.

A lighter adaptive method consists in using the dual variable q_k to construct \mathcal{V}_{k+1} . It satisfies $|A^*q_k|(x) \leq 1$ for $x \in \mathcal{V}_k$ by construction. However, there may exist locations $x \in \Omega \setminus \mathcal{V}_k$ with $|A^*q_k|(x) > 1$. Such points are candidates to be added to \mathcal{V}_{k+1} . Perhaps the most popular approach in this class is the Frank-Wolfe [19] approach. It consists in adding only the global maximizer of $|A^*q_k|$ at each step. It is described precisely in Algorithm 1. It was revived in signal processing thanks to Bredies et al in [5]. Its connection with the exchange

Algorithm 1 The Frank-Wolfe Algorithm

1: **Input:**

- Initial discretization set \mathcal{V}_0 , set $k = 0$

2: **WHILE** a stopping criterion is not satisfied

- 3: 1) Determine a solution q_k of $\mathcal{D}(\mathcal{V}_k)$
- 4: 2) Determine $x_k^* \stackrel{\text{def.}}{=} \operatorname{argmax}_{x \in \Omega} |A^*q_k|(x)$.
- 5: 3) Set $\mathcal{V}_{k+1} = \mathcal{V}_k \cup \{x_k^*\}$.

6: **Output:**

- The dual solution q_k of $(\mathcal{D}(\mathcal{V}_k))$.
-

algorithms was recalled in [16]. A linear convergence theory was developed independently by Walter and Pieper in [27] and by the authors in [17]. Despite nice theoretical properties, this approach suffers from one major issue, which is the main motivation for the present paper:

How can we find the maximizer $x_k^* \stackrel{\text{def.}}{=} \operatorname{argmax}_{x \in \Omega} |A^* q_k|(x)$?

This problem has no reason to be simple: it is nonconvex and depends highly on the properties of the linear forms a_m . As far as we could judge, it is usually tackled with heuristics, which are often swept under the carpet. For instance, multiple gradient or Newton ascents are launched in parallel starting from a set of points covering Ω sufficiently finely. At the end of the process, the point with the largest value is then kept as an approximation of x_k^* . Our experience is that tuning the hyper-parameters in this quest for the global minimizer is time consuming and can represent a real headache for the optimizer. The main objective of this paper is to tackle this issue, by looking for approximate, easily detectable maximizers only.

Alternative solvers Finally, let us mention that alternative algorithms with a different flavor are available. A possibility is to use the Lasserre hierarchies, which are designed to solve near generic measure optimization problems [24, 11]. They however scale poorly for large M and will not be considered further in this work. Another possibility is to use continuous optimization procedures. The idea is to parameterize the measure μ by the locations and weights of its masses. Given a number $N \geq M$ of particles, a set of locations $X = (x_1, \dots, x_N)$ in Ω^N and a weight vector $\alpha \in \mathbb{R}^N$, we can define the mapping

$$\mu(X, \alpha) \stackrel{\text{def.}}{=} \sum_{n=1}^N \alpha_n \delta_{x_n}.$$

Plugging it in the primal problem $\mathcal{P}(\Omega)$, we obtain the following non-convex, finite dimensional problem

$$\inf_{\substack{X \in \Omega^N \\ \alpha \in \mathbb{R}^N}} \|\alpha\|_1 + f(A\mu(X, \alpha)). \quad (3)$$

It can be solved with continuous optimization routines such as proximal gradient descents. The difficulty with this type of approach is the initialization: how to set the number N of particles, their locations and weights? A possibility is to discretize the domain Ω finely. A nice global convergence theory has then been developed in [8, 7]. The price to pay is a large number of variables $N \gg M$. This is somewhat unsatisfactory since we know in advance that a solution supported on at most M points exists. A possibility to avoid this pitfall consists in using an exchange algorithm to find an approximate solution to $\mathcal{P}(\Omega)$, and then fine-tune it with a continuous optimization routine on (3). It now possesses a rich convergence theory [33, 17]. It can be used in conjunction with the algorithm proposed in this work. A variant of this approach is to launch the continuous optimization technique after every step in the Frank-Wolfe algorithm, and not only use it to fine-tune in the end. This approach was introduced and coined the sliding Frank-Wolfe algorithm in [13].

2 Preliminaries

We will work on the domain $\Omega = [0, 1]^D$ for simplicity. This is not a real restriction, since we can fit any compact domain of \mathbb{R}^D to $\Omega = [0, 1]^d$ using a non-degenerate affine transformation. Using this map, we may push forward measurement functions and measures, in a bijective fashion. We let $\mathcal{M}(\Omega)$ denote the space of Radon measures of bounded total variation on Ω and $\mathcal{C}_0(\Omega)$ the space of continuous functions on Ω vanishing on the boundary. Note that if we equip $\mathcal{M}(\Omega)$ with the total variation norm $\|\cdot\|_{\mathcal{M}}$ and $\mathcal{C}_0(\Omega)$ with the supremum norm, $\mathcal{M}(\Omega)$ can be identified with the dual of $\mathcal{C}_0(\Omega)$. For a subset ω of \mathbb{R}^D , we let $\operatorname{vol}(\omega)$ denote its volume (or Lebesgue measure). The notation $\llbracket 1, N \rrbracket$ indicates the set of integers from 1 to N . The relation $f \lesssim g$ indicates that f is dominated by g up to positive multiplicative constant. The relation $f \asymp g$ indicates that f and g are equivalent, i.e. that there exists two constants $0 < c_1 \leq c_2$ such that $c_1 g \leq f \leq c_2 g$.

2.1 Cells and cell partitions

The proposed algorithms rely on the use of 2^D -trees. We iteratively partition hypercubes in 2^D equal parts. For instance, we will use binary trees in 1D, quadtrees in 2D, and octrees in 3D. Let us define some objects.

Definition 1 (Cells, vertices, edge-length). We call a subset $\omega \subseteq [0, 1]^d$ a dyadic cell (or simply a cell) if it is of the form

$$\omega = x + 2^{-J} \cdot [0, 1]^d,$$

where J is a non-negative integer and $x \in (2^{-J} \cdot \{0, 1, \dots, 2^J - 1\})^d$.

For a cell ω , we let $\text{vert}(\omega)$ denote its vertices and $|\omega|$ denote its edge-length.

Definition 2 (Cell partition). A cell partition Ω_k is a collection of cells such that $\Omega = \cup_{\omega \in \Omega_k} \omega$ and $\text{vol}(\omega' \cap \omega) = 0$ for all $\omega, \omega' \in \Omega_k$ with $\omega \neq \omega'$. That is to say, two cells in Ω_k can only have faces in common.

2.2 Measurement operator

Throughout the paper, we will work under the following assumption.

Assumption 1 (Continuous operator). The operator $A : \mathcal{M}(\Omega) \rightarrow \mathbb{R}^M$ is weak-* continuous. Equivalently, the measurement functionals a_m^* defined by $\langle a_m^*, \mu \rangle = (A(\mu))_m$ are given for all $\mu \in \mathcal{M}(\Omega)$ by

$$\langle a_m^*, \mu \rangle = \int_{\Omega} a_m d\mu,$$

for functions $a_m \in \mathcal{C}_0(\Omega)$.

Given $q \in \mathbb{R}^M$, $A^*q = \sum_{m=1}^M q_m a_m$ is a continuous function. Assuming that $a_m \in \mathcal{C}^r(\Omega)$, we let $(A^*q)'$, $(A^*q)''$, $(A^*q)^{(r)}$ denote its derivative, its Hessian and its r -th tensor derivative. We define the following constants

$$\kappa_r \stackrel{\text{def.}}{=} \sup_{\|q\|_2 \leq 1} \sup_{x \in \Omega} \|(A^*q)^{(r)}(x)\|_{2 \rightarrow 2}, \quad (4)$$

where r indicates the derivative's order and where $\|\cdot\|_{2 \rightarrow 2}$ is the spectral norm.

2.3 Set distances

The “distance” between two sets X_1 and X_2 in \mathbb{R}^D is defined by

$$\text{dist}(X_1, X_2) \stackrel{\text{def.}}{=} \inf_{x_1 \in X_1, x_2 \in X_2} \|x_1 - x_2\|_2.$$

Notice that dist is not a proper distance. In particular, it does not satisfy the triangle inequality. For a point $x \in \mathbb{R}^D$ and a set $X \subset \mathbb{R}^D$, we will use the shorthand notation

$$\text{dist}(x, X) \stackrel{\text{def.}}{=} \text{dist}(\{x\}, X).$$

The *Hausdorff distance* between X_1 and X_2 is defined by

$$\text{dist}_{\mathcal{H}}(X_1|X_2) \stackrel{\text{def.}}{=} \sup_{x_2 \in X_2} \inf_{x_1 \in X_1} \|x_1 - x_2\|_2.$$

Notice that this distance is asymmetric: in general $\text{dist}_{\mathcal{H}}(X_1|X_2) \neq \text{dist}_{\mathcal{H}}(X_2|X_1)$. The following inequality will play an important role in the analysis.

Proposition 1 (Triangle inequality for set distances). For any triple of sets X_1, X_2, X_3 in \mathbb{R}^D we have

$$\text{dist}(X_1, X_2) \leq \text{dist}_{\mathcal{H}}(X_1|X_3) + \text{dist}(X_3, X_2). \quad (5)$$

2.4 Primal, dual and existence

Our results will be established under the following assumptions on f and A .

Assumption 2 (A convexity assumption). *The function $f : \mathbb{R}^M \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex lower semi-continuous function with $\text{int}(\text{dom}(f)) \neq \emptyset$.*

Assumption 3 (Coercivity). *The functional J is coercive, meaning that $J(\mu) \rightarrow \infty$ when $\|\mu\|_{\mathcal{M}(\Omega)} \rightarrow \infty$.*

Notice that Assumption 3 is granted if f is lower-bounded. The following result relates the primal and the dual.

Proposition 2 (Existence and strong duality). *Let $\mathcal{V} \subseteq \Omega$ denote a subset of Ω . Assume that there exists $\mu \in \mathcal{M}(\Omega)$ supported on \mathcal{V} with $A\mu \in \text{int}(\text{dom}(f))$. Then, under Assumptions 1, 2 and 3, the following statements hold true:*

- *The primal problem ($\mathcal{P}(\mathcal{V})$) has a nonempty set of solutions, bounded in total variation norm.*
- *The dual ($\mathcal{D}(\mathcal{V})$) has a nonempty set of solutions, which is also bounded.*
- *The following strong duality result holds*

$$\min_{\mu \in \mathcal{M}(\mathcal{V})} \|\mu\|_{\mathcal{M}(\mathcal{V})} + f(A\mu) = \max_{q \in \mathbb{R}^M, \|A^*q\|_{L^\infty(\mathcal{V})} \leq 1} -f^*(q).$$

- *Let (μ^*, q^*) denote a primal-dual pair. They are related by the following primal-dual relationships:*

$$A^*q^* \in \partial_{\|\cdot\|_{\mathcal{M}}}(\mu^*) \text{ and } -q^* \in \partial f(A\mu^*). \quad (6)$$

The left inclusion in (6) implies that the support of a solution μ^* satisfies: $\text{supp}(\mu^*) \subseteq \{x \in \Omega, |A^*q^*(x)| = 1\}$.

Remark 1. *Strong duality may hold under different assumptions. For instance, if f is polyhedral (allowing the hard constraint $A\mu = b$), then strong duality holds [3], but the dual solution set may be unbounded. Similarly, the coercivity of J is not absolutely needed. If J has a finite dimensional constancy space, the primal solution set still exists, but may be unbounded as well. In both cases, the unboundedness of either the primal or dual problem requires extra technicalities and assumptions in the proofs that we decided to discard.*

3 Main results

This section contains our main findings. All the proofs are post-poned to the appendix.

3.1 The algorithm

The algorithm we propose consists in designing a sequence of cell partitions $(\Omega_k)_{k \in \mathbb{N}}$ of Ω . At each step of the algorithm, Ω_{k+1} is constructed by dividing a few cells in Ω_k . Let \mathcal{V}_k denote the set of vertices of the partition Ω_k :

$$\mathcal{V}_k \stackrel{\text{def}}{=} \{\text{vert}(\omega) \mid \omega \in \Omega_k\}.$$

At step k , we solve $(\mathcal{D}(\mathcal{V}_k))$, to obtain a solution q_k of the discretized dual problem. It satisfies $\|A^*q_k\|_{L^\infty(\mathcal{V}_k)} \leq 1$ by construction. However, it might be infeasible for the problem $(\mathcal{D}(\Omega))$, i.e. $\|A^*q_k\|_{L^\infty(\Omega)} > 1$. This suggests that we should detect the cells $\omega \in \Omega_k$ for which $\|A^*q_k\|_{L^\infty(\omega)} > 1$ and subdivide them. To this end, we suppose that we have access to a set of *candidate cells* $\Omega_k^* \subset \Omega_k$ that are likely to satisfy $\|A^*q_k\|_{L^\infty(\Omega)} > 1$. To make it easier to control the growth of $|\mathcal{V}_k|$ with the iteration number k , we propose to not refine all candidates in Ω_k^* , but only the largest of them. The complete solver is described in Algorithm 2. One iteration of the algorithm is displayed in Figure 1.

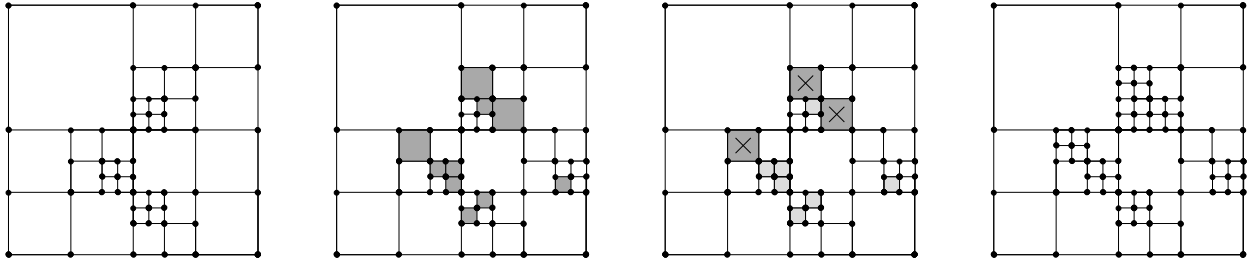


Figure 1: The refinement . From left to right: i) a cell partition, ii) the partition with the candidate cells grayed out, iii) only the candidate cells with largest diameter are selected for refinement and iv) the new resulting cell partition.

Algorithm 2 Adaptive Refinement Algorithm

1: **Input:**

- Operator A
- Initial partition Ω_0 (e.g. $\Omega_0 = \{\Omega\}$)
- Target precision J
- Solver for the discretized primal ($\mathcal{P}(\mathcal{V}_k)$) and dual ($\mathcal{D}(\mathcal{V}_k)$)
- Set $k = 1$

2: **WHILE** $\max_{\omega \in \Omega_{k-1}^*} |\omega| \geq 2^{-J}$

- 1) Determine a solution q_k of ($\mathcal{D}(\mathcal{V}_k)$)
- 2) Determine the candidate cells Ω_k^* .
- 3) Subdivide the cells in Ω_k^* with largest diameter
- 4) $k = k + 1$

7: **Output:**

- The dual solution q_k of ($\mathcal{D}(\mathcal{V}_k)$).
 - The primal solution μ_k of ($\mathcal{P}(\mathcal{V}_k)$).
-

3.2 Assumptions on the selection process

In this paragraph, we discuss how to construct the set of candidate cells Ω_k^* . Following [17], we let X_k denote the set of local maximizers of $|A^*q_k|$ exceeding 1, i.e.

$$X_k \stackrel{\text{def.}}{=} \{x \in \Omega \mid x \text{ is a local maximizer of } |A^*q_k|, |A^*q_k(x)| \geq 1\}.$$

To obtain convergence guarantees we will work under the following hypothesis.

Assumption 4 (Generic convergence assumption). *For each $x \in X_k$, at least one of the cells in Ω_k containing x is a candidate for refinement. In other terms, Ω_k^* satisfies*

$$X_k \subset \bigcup_{\omega \in \Omega_k^*} \omega.$$

The above condition is too weak to allow a control of the numerical complexity. For instance the choice $\Omega_k^* = \Omega_k$ – which corresponds to refine uniformly at each iteration – obeys Assumption 4, but leads to an exponential growth of $|\Omega_k|$. Our complexity analysis will therefore rely upon the following extra hypothesis.

Assumption 5 (Second order approximation). *There exists a constant $\kappa > 0$ independent of the iteration k and the cell ω such that for every $\omega \in \Omega_k^*$, we have*

$$\|A^*q_k\|_{L^\infty(\omega)} \geq 1 - \kappa|\omega|^2,$$

where $|\omega|$ is the edge-length of the cell ω .

3.3 Construction of selection processes

In this section, we discuss how to construct rules satisfying assumptions 4 and 5.

3.3.1 Ideal selection

An obvious refinement rule that obeys both assumptions (4) and (5) is to let

$$\Omega_{k,\text{ideal}}^* = \{\omega \in \Omega_k \mid \omega \cap X_k \neq \emptyset\}.$$

Indeed, Assumption 4 can be reformulated as $\{\omega \in \Omega_k \mid \omega \cap X_k \neq \emptyset\} \subseteq \Omega_k^*$, which certainly is true for the above choice. As for Assumption 5, note that for all ω with $X_k \cap \omega \neq \emptyset$, we have $\|A^*q_k\|_{L^\infty} \geq 1 \geq 1 - \kappa|\omega|^2$, since ω contains a cell point where $|A^*q_k|$ succeeds one.

In order to apply this rule, we however need to know which cells contain the elements of X_k . As discussed in the introduction, this is in general infeasible.

3.3.2 Upper-bounds selections

To obtain resolvable but still powerful enough selection processes, we will instead rely on the design of *simple to evaluate upper-bounds* $\overline{|A^*q_k|}(\omega) \in \mathbb{R}$ satisfying $\overline{|A^*q_k|}(\omega) \geq \|A^*q_k\|_{L^\infty(\omega)}$. Equipped with such an upper-bound, we can define the candidate cells as:

$$\Omega_k^* = \left\{ \omega \in \Omega_k \mid \overline{|A^*q_k|}(\omega) \geq 1 \right\}.$$

By construction, this selection process guarantees Assumption 4. Indeed, a cell ω with $\omega \cap X_k$ obviously obeys $\|A^*q_k\|_{L^\infty(\omega)} \geq 1$, and therefore also $\overline{|A^*q_k|}(\omega) \geq \|A^*q_k\|_{L^\infty(\omega)} \geq 1$.

Notice that this principle is similar to a branch-and-bound approach [25]. We base our decisions on upper-bounds, which secure that some regions of space can be safely neglected. An important difference lies in the fact that the objective function $|A^*q_k|$ varies at each iteration, meaning that one region which might have been discarded at one iteration can be refined some iterations later. Let us describe two such upper bounds.

Definition 3 (A first order selection process). *Assume that $a_m \in C^1(\Omega)$ for all m . Let us define*

$$\overline{|A^*q_k|}(\omega) = \inf_{v \in \text{vert}(\omega)} |A^*q_k(v)| + \kappa_1(q_k, \omega) \text{diam}(\omega), \quad (7)$$

with

$$\kappa_1(q_k, \omega) = \sum_{m=1}^M |q_k[m]| \|a'_m\|_{L^\infty(\omega)}. \quad (8)$$

The first order selection process is defined as

$$\Omega_{k,1}^* = \left\{ \omega \in \Omega_k \mid \overline{|A^*q_k|}(\omega) \geq 1 \right\}$$

Proposition 3. *The first order selection process $\Omega_{k,1}^*$ satisfies Assumption 4. However it may not respect Assumption 5.*

In order to satisfy both Assumption 4 and Assumption 5, we will use second order selection rules.

Definition 4 (A second order selection process). *Assume that $a_m \in C^2(\Omega)$ for all m . For any cell ω , and all $q_k \in \mathbb{R}^M$, define*

$$\overline{|A^*q_k|}(\omega) \stackrel{\text{def.}}{=} \inf_{v \in \text{vert}(\omega)} \sup_{x \in \omega} |A^*q_k(v) + \langle (A^*q_k)'(v), x - v \rangle| + \frac{\kappa_2(q_k, \omega)}{2} \|x - v\|^2, \quad (9)$$

with

$$\kappa_2(q_k, \omega) = \sum_{m=1}^M |q_k[m]| \sup_{x \in \omega} \|a_m''(x)\|_{2 \rightarrow 2}. \quad (10)$$

The second order selection process is defined as

$$\Omega_{k,2}^* = \left\{ \omega \in \Omega_k \mid \overline{|A^*q_k|}(\omega) \geq 1 \right\} \quad (11)$$

Proposition 4. *If the sequence $(q_k)_{k \in \mathbb{N}}$ is uniformly bounded, then the second order selection process $\Omega_{k,2}^*$ satisfies both Assumption 4 and 5.*

Importantly, notice that the problem

$$\sup_{x \in \omega} |A^*q_k(v) + \langle (A^*q_k)'(v), x - v \rangle| + \frac{\kappa_2(q_k, \omega)}{2} \|x - v\|^2,$$

consists in maximizing a convex function over a polyhedron. A solution is therefore attained on $\text{vert}(\omega)$, and can be evaluated in constant time per cell.

Remark 2. *The values $\kappa_1(q_k, \omega)$ and $\kappa_2(q_k, \omega)$ can be replaced by any upper-bound on the Lipschitz constant of A^*q_k and the Lipschitz constant of $(A^*q_k)'$ respectively. In particular, it is possible to use the global bounds $\kappa_1(q_k, \omega) = \kappa_1 \|q_k\|_2$ and $\kappa_2(q_k, \omega) = \kappa_2 \|q_k\|_2$, where κ_1 and κ_2 are defined in equation (4).*

3.3.3 Combining upper-bounds and lower-bounds on the gradient norm

The larger Ω_k^* , the higher the chances of selecting unwanted cells for refinement. To reduce the cardinality of the candidate cells, it makes sense to only refine the cells where the function $|A^*q_k|$ might surpass 1 and where the gradient's norm $\|(A^*q_k)'(x)\|_2$ might cancel. Indeed, Assumption 4 only requires the *local maximizers* of $|A^*q_k|$ that surpass 1 to be selected. In cells containing maximizers, the gradient of $|A^*q_k|$ vanishes. Consequently we can design a selection process based on lower bounds of the gradient.

Definition 5 (Second order selection process with first order gradient). *Assume that $a_m \in C^2(\Omega)$ for all m . Define $\overline{|A^*q_k|}(\omega)$ and $\kappa_2(q_k, \omega)$ as in (9) and (10) respectively. For any cell ω , and all $q_k \in \mathbb{R}^M$, define*

$$\underline{\|\nabla A^*q_k\|_2}(\omega) = \sup_{v \in \text{vert}(\omega)} \|(A^*q_k)'(v)\|_2 - \kappa_2(q_k, \omega) \text{diam}(\omega) \quad (12)$$

The second order selection process with first order gradient is defined as

$$\Omega_{k,2,1}^* = \left\{ \omega \in \Omega_k \mid \overline{|A^*q_k|}(\omega) \geq 1 \text{ and } \underline{\|\nabla A^*q_k\|_2}(\omega) \leq 0 \right\}.$$

Proposition 5. *If the sequence $(q_k)_{k \in \mathbb{N}}$ is uniformly bounded, then the selection process $\Omega_{k,2,1}^*$ satisfies both Assumption 4 and 5. Incorporating gradient lower bounds reduces the cardinality of Ω_k^* in the sense that the cardinality of Ω_k^* in Definition 5 is non greater than the one of Ω_k^* in Definition 4.*

Remark 3. *Following Remark 2, the definition of $\kappa_2(q_k, \omega)$ given in (10) can be replaced by any upper bound of the Lipschitz constant of the gradient of $|A^*q_k|$. Moreover, we use a first order Taylor expansion for $\underline{\|\nabla A^*q_k\|_2}$. It is possible to replace or even combine this lower bound with bounds stemming from higher order Taylor expansions.*

3.4 Generic convergence guarantees

To obtain a generic convergence result, we first need to prove that the algorithm is well defined. To this end, we introduce the following assumption.

Assumption 6 (Well-posedness of the algorithm). *The initial set of vertices \mathcal{V}_0 is admissible in the sense that there exists $\mu \in \mathcal{M}(\mathcal{V}_0)$ with $A\mu \in \text{int}(\text{dom}(f))$.*

Theorem 6. *Under Assumptions 1, 2, 3, 4 and 6, the sequences $(\mu_k)_{k \in \mathbb{N}}$ and $(q_k)_{k \in \mathbb{N}}$ defined in Algorithm (2) are well-defined. They contain subsequences that converge weakly to solutions μ^* and q^* of $(\mathcal{P}(\Omega))$ and $(\mathcal{D}(\Omega))$, as well as in optimal function value. If either the primal or dual solution is unique, the whole corresponding sequence converges.*

Proof. The proof of this theorem is postponed to Section A.1. □

Remark 4. *In [17], additional assumptions were made (either smoothness of f or surjectivity of A restricted to $\mathcal{M}(\mathcal{V}_0)$) to obtain generic convergence. The reason we can remove that assumption is based on a refined analysis.*

Remark 5. *The proof of this result relies on the fact that the sequence $(\mathcal{V}_k)_{k \in \mathbb{N}}$ is nested. In exchange algorithms, it is possible to not only add, but also discard points from \mathcal{V}_k to construct \mathcal{V}_{k+1} . The obvious interest is to reduce the numerical complexity. We do not know if it possible to adapt the algorithm and the proof to allow for points suppression as well.*

3.5 Linear convergence rates

Having established the generic convergence result, we move on to providing an eventual linear convergence rate under additional regularity conditions. We first need a couple of additional regularity conditions on f , A and the primal-dual solution pair, which are similar to those in [17] and [27].

Assumption 7 (Linear convergence conditions).

- The functionals a_m are twice differentiable: $a_m \in C_0^2(\Omega)$ for all $1 \leq m \leq M$.
- The function f is convex, differentiable with an L -Lipschitz gradient.

Following [15], we also require the following condition.

Assumption 8 (Non-degenerate source condition). *We say that the non-degenerate source condition [15] holds if we have the following:*

- The solution μ^* of $(\mathcal{P}(\Omega))$ is unique and supported on $S \in N$ points

$$\mu^* = \sum_{s=1}^S \alpha_s^* \delta_{x_s^*}$$

for some $\alpha_s^* \in \mathbb{R}$ and $x_s^* \in \Omega$. In what follows, we let $X^* \stackrel{\text{def.}}{=} \{x_1^*, \dots, x_S^*\}$.

- The dual certificate $|A^*q^*|$ is only equal to 1 in the points x_1^*, \dots, x_S^* and is strictly concave around those points. This ensures the existence of a parameter $\gamma > 0$ and a radius $R > 0$ with

$$B(x_{s_1}^*, R) \cap B(x_{s_2}^*, R) = \emptyset, \quad \forall s_1 \neq s_2 \tag{13}$$

$$\text{sign}(A^*q^*(x))(A^*q^*)'' \preceq -\gamma \text{Id} \quad \text{for } x \text{ with } \text{dist}(x, X^*) \leq R,$$

$$|A^*q^*(x)| \leq 1 - \frac{\gamma R^2}{2} \quad \text{for } x \text{ with } \text{dist}(x, X^*) \geq R. \tag{14}$$

This last assumption is generic, given that the solution is unique. It is a condition that has appeared in the literature as a mean to prove recovery of sparse measures using problems of the form $(\mathcal{P}(\Omega))$ – see e.g. [28, 6]. We can now formulate our main result.

Theorem 7. *Under Assumptions 3, 4, 5, 7 and 8, Algorithm 2 eventually converges linearly. That is, there exists constants $k_0 \in \mathbb{N}$, $c > 0$ (depending on A , f and μ^*) such that the algorithm terminates in no more than $k = k_0 + cSJ$ iterations. For sufficiently large J , we further have*

$$\begin{aligned} |\mathcal{V}_k| &= O(J), && \text{controlled complexity} \\ \text{dist}_{\mathcal{H}}(X^* | \text{vert}(\Omega_k^*)) &\lesssim 2^{-J}, && \text{controlled localization} \\ \|q_k - q^*\|_2 &\lesssim 2^{-J}, && \text{certificate on the dual} \\ J(\mu_k) - J(\mu^*) &\lesssim 2^{-2J}, && \text{certificate on the primal.} \end{aligned}$$

Proof. The proof of this theorem is quite technical, and is therefore postponed to Section A.2. It relies on a few technical inequalities from the companion paper [17], but differs significantly to account for the discretization procedure.

Informally, it is built using the following arguments. First, appealing to the generic convergence result and the fact that only finitely many cells have an edge length larger than any fixed value $\delta > 0$, we argue that after warming period of at most k_0 iterations, q_k is close to q^* , and no cells with an edge-length larger than a critical value can be active. Once that happens, the algorithm will only be able to refine cells close to the maximizers X^* . This results in a multiscale refinement of local regions around the sought-for locations, see e.g. Figure 3. \square

Remark 6. *We did not keep track of the constants in the above inequalities to simplify the reading. While some of them are explicit, others, like the time k_0 to reach a linear convergence rate are not.*

Remark 7. *We can replace μ_k , the solution of $(\mathcal{P}(\mathcal{V}_k))$, by the solution $\tilde{\mu}_k$ of $\mathcal{P}(\text{vert}(\Omega_k^*))$ defined as*

$$\tilde{\mu}_k \stackrel{\text{def.}}{=} \inf_{\mu \in \mathcal{M}(\text{vert}(\Omega_k^*))} \|\mu\|_{\mathcal{M}} + f(A\mu). \quad (15)$$

and still have $J(\tilde{\mu}_k) - J(\mu^) \lesssim 2^{-2J}$. The interest of this alternative problem is that the cardinality of $\text{vert}(\Omega_k^*)$ is significantly smaller than that of \mathcal{V}_k , helping to reduce the numerical complexity.*

Remark 8. *Our algorithm relies on dyadic subdivision of cells. We therefore cannot expect the algorithm to converge faster than linearly. In that regard, Theorem 7 is optimal.*

4 Numerical experiments

In this section, we aim at illustrating our main findings through some simple numerical experiments. We consider problems of sparse source recovery problem with filtered measurements. That is, given a ground truth $\bar{\mu}$, we set $y = A\bar{\mu}$ and let $f(q) = \frac{1}{2}\|q - y\|_2^2$. This yields

$$f^*(q') \stackrel{\text{def.}}{=} \sup_{q \in \mathbb{R}^M} \langle q, q' \rangle - \frac{1}{2}\|q - y\|_2^2 = \frac{1}{2}\|q'\|_2^2 + \langle q', y \rangle.$$

We consider Gaussian measurements functions of the form

$$a_m(x) = \frac{1}{2\pi\sigma} \exp\left(\frac{-\|x - z_m\|_2^2}{2\sigma^2}\right)$$

for some value $\sigma > 0$. To properly define our selection procedures in (10), we need an upper-bound on the second order derivatives.

Proposition 8. *Define*

$$\kappa_{2,m}(\omega) = \frac{a_m(\text{dist}(z_m, \omega))}{\sigma^4} \max\left(\sigma^2, (\text{dist}(z_m, \omega) + \sqrt{D}|\omega|)^2\right)$$

then $\kappa_{2,m}(\omega) \geq \sup_{x \in \omega} \|a_m''(x)\|_{2 \rightarrow 2}$. We can choose $\kappa_2(q_k, \omega) = \sum_{m=1}^M |q_k|[m]\kappa_{2,m}(\omega)$, in Proposition 10 to define the second order candidates Ω_k^ .*

4.1 Implementation details

We implement our algorithm in Python using the numpy package. To solve the discretized dual ($\mathcal{D}(\mathcal{V}_k)$), we rely on the SCS solver of the CVXPY package [14, 1]. The selection procedures are defined and implemented as described in the main text. To assess the convergence rates, we compute the exact solutions of the primal problem by running a fixed step gradient descent in the parameter space, see equation (3), initialized in the 'ground truth measures' we specify. This is sound, since the true solution lies close to them (see e.g. [28]).

4.2 1D-experiments

The problem We set $y = A\bar{\mu}$ with $\bar{\mu} \stackrel{\text{def.}}{=} 8\delta_{1/3} - 9\delta_{2/3}$. We choose the Dirac mass locations to lie at $1/3$ and $2/3$, since these points are the hardest to reach with dyadic partitions. The sampling locations and σ -parameter are set to $z_m \stackrel{\text{def.}}{=} m/M$ and $\sigma = 2/M$, with $M = 20$.

Second-order upper bound The behavior of the second order selection process algorithm of Definition 4 is displayed in Table 1a and Figure 2. As can be seen in the figure, the algorithm starts by a burn-in period of 4 iterations. This transient behavior explains why the linear convergence rate only occurs after a finite number of iterations. Then, only cells in a neighborhood of $\{1/3, 2/3\}$ are refined. The Table 1 clearly indicates that the distance $\text{dist}_{\mathcal{H}}(\mathcal{V}_k, X^*)$ decays exponentially fast, illustrating Theorem 7 and the linear convergence rate. Observe that less than 300 vertices are enough to obtain a precision 10^{-6} , while a uniform refinement would require 10^6 vertices. This illustrates the huge computational/memory advantage of this adaptive method.

Second-order upper bound with first order gradient We turn our attention to the second order selection process with first order gradient, see Definition 5. The results are displayed in Table 1b and Figure 3. In comparison to the previous test, a lower bound on the gradient is used to reduce the cardinality of Ω_k^* . Two measures of complexity can be used to compare the approaches: i) the cardinality $|\mathcal{V}_k|$ needed to reach a given accuracy $\text{dist}_{\mathcal{H}}(\mathcal{V}_k, X^*)$, or ii) the number of iterations to reach the same accuracy. Reducing the cardinality of Ω_k^* can be detrimental to the second notion of complexity. For example, compare Figure 3f and Figure 2f. The cells that are not flagged for refinement in Figure 3f are flagged in Figure 3g and refined at Iteration 7. Iteration 7 can be seen as a failed *zwischenzug* iteration that loses a tempo. However, for the first notion of complexity, the conclusion is different. We see that for this particular example, adding a gradient lower bounds allows reaching the target precision in Table 1b with less than half the number of vertices for the vanilla second order bound. A full complexity analysis would require a fine analysis of the quadratic programming solver, which is out of the scope of this paper.

4.3 2D experiments

The problem In this section, we assume that the sampling points z_m lie on a Euclidean grid. More precisely, we suppose that $\sqrt{M} \in \mathbb{N}$ and that each index $m \in \llbracket 1, M \rrbracket$ can be decomposed as $m = (m_1, m_2) \in \llbracket 0, \sqrt{M} - 1 \rrbracket^2$ and $z_m = \frac{1}{\sqrt{M}}(m_1, m_2)$ with $M = 15$. $\bar{\mu}$ is chosen as

$$\bar{\mu} = -9\delta_{(1/3, 1/3)} + 8\delta_{(1/3, 2/3)} + 5\delta_{(2/3, 2/3)}.$$

Results We begin by showcasing the behaviour of the algorithm when the second-order upper bound is used. Table 2a and Figure 4 summarize the algorithm's behavior. The conclusions are similar to the previous section and consistent with Theorem 7: after a burn-in period, the grid is refined in a multi-scale fashion, only around the support X^* of the solution μ^* . To control the complexity of our algorithm we refine only the cells with largest diameter. The effect of this strategy is striking in $2D$, where the algorithm spends some iterations to refine larger cells only. See Figure 4, iterations 6, 8, 9, 10. At these iterations, it is not the cells containing the maximizers X_k which are refined, but only the largest ones which were not refined in the previous iterations. Yet, the table indicates a clear advantage of this adaptive method: about 3000 vertices are sufficient to reach a precision 10^{-4} ,

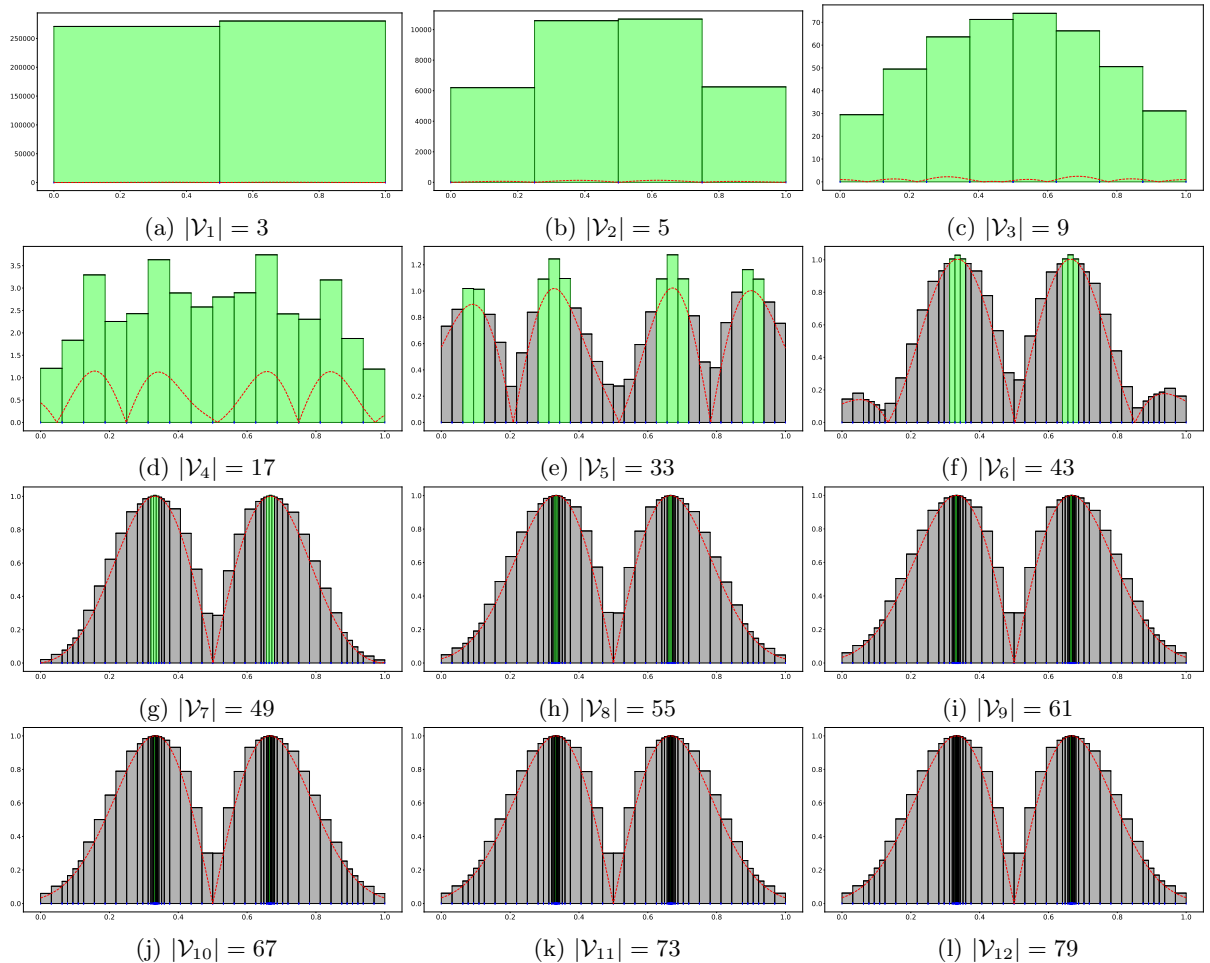


Figure 2: The behavior of the adaptive refinement algorithm with a second order selection process, on a 1D sparse recovery problem. The set Ω_k^* is displayed in green, the function $|A^*q_k|$ is displayed in dashed red, the upper-bound $\overline{|A^*q_k|}$ is the piecewise-constant function. Observe that it always dominates $|A^*q_k|$. The algorithm starts with a burn-in period of 4 iterations. There, it refines all cells uniformly since the upper-bound is highly inaccurate. After a while, only the cells around the locations X^* get refined in a multiscale fashion.

while the same guarantee would be obtained only with 10^8 vertices for a uniform refinement. The results when a lower bound of the gradient is added are displayed in Table 2b and Figure 5. For this example, there is no increase in the number of iterations, and only a slight decrease of the number of vertices is observed. Again, a more detailed analysis of the effects of gradient-including rules is beyond the scope of this paper.

5 Perspectives

In this work, we proposed an alternative to the Frank-Wolfe algorithm for infinite dimensional total variation regularization. This adaptive refinement approach has a significant advantage: it does not require to search for the maximizers of a non-convex function at each iteration. Instead, it progressively discards regions of the space, in a certified manner, resembling a branch-and-bound approach. The only prerequisite to implement it is the computation of upper bounds on the largest eigenvalues of the measurement functions Hessians. We proved that the method has great adaptivity properties. It converges generically under weak assumptions, and its rate of

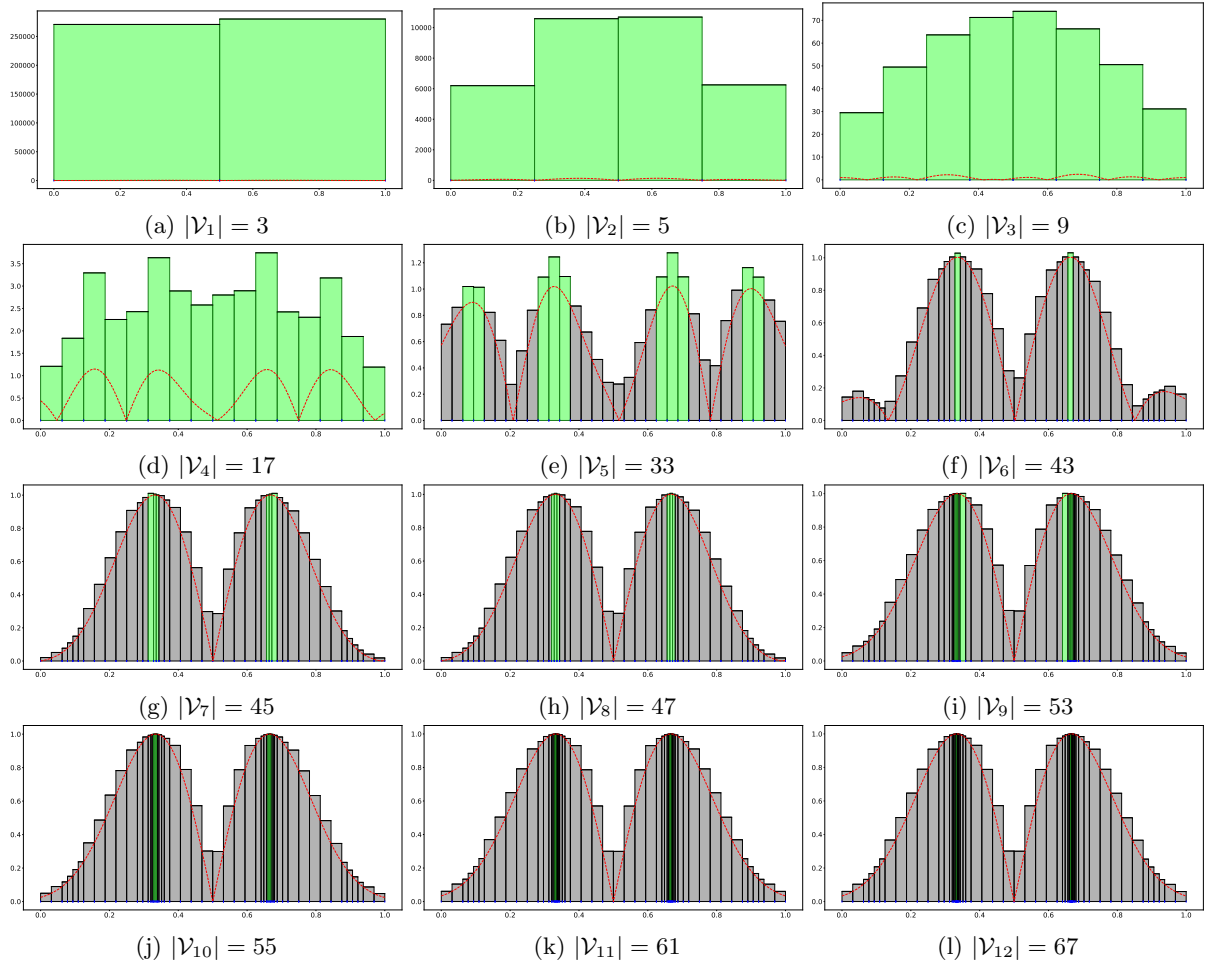


Figure 3: The adaptive refinement algorithm with a second order selection process and first order gradient. The setting is the same as the one of Figure 2. The cardinality of the set of candidates Ω_k^* is smaller, as can be seen when comparing Figure 3f with Figure 2f.

convergence is linear under stronger regularity assumptions. To the best of our knowledge, this is as good as the best existing results for the Frank-Wolfe algorithm, and we cannot expect more from this dichotomic approach.

Despite these assets, some parts of the algorithm still require some analysis. In particular, the solution of a finite dimensional convex problem needs to be computed at each iteration. In this work, we assumed that this could be achieved with an arbitrary accuracy. A complete theory should account for approximation errors and for the complexity of the sub-problems.

On a more positive side, the scope of this approach is possibly significantly wider than total variation regularization. Up to some adjustments, we believe that the method could be extended to more general sparse measure optimization problems. In particular, we think of other regularizers that promote sparse solutions, such as problems defined over the cone of nonnegative measures, or over the set probability measures.

A Proofs

Here, we include the proofs omitted in the main text. We begin by proving the main results, i.e. the generic (Theorem 6) and linear (Theorem 7) convergence results, and save the proofs of smaller, technical propositions

Iteration	$ \mathcal{V}_k $	primal	$\text{dist}_{\mathcal{H}}(\mathcal{V}_k, X^*)$
0	2	3.80563e+03	3.3e-01
1	3	3.79912e+03	1.7e-01
2	5	9.39226e+02	8.3e-02
3	9	3.01878e+01	4.2e-02
4	17	1.84675e+01	2.1e-02
5	33	1.72061e+01	1.0e-02
6	43	1.70209e+01	5.1e-03
7	49	1.69895e+01	2.7e-03
8	55	1.69826e+01	1.2e-03
9	61	1.69810e+01	7.2e-04
10	67	1.69873e+01	2.6e-04
11	73	1.69828e+01	2.3e-04
12	79	1.69806e+01	1.9e-05
13	89	1.69811e+01	1.9e-05
14	105	1.69806e+01	1.9e-05
15	132	1.69805e+01	1.2e-05
16	162	1.69805e+01	4.3e-06
17	208	1.69805e+01	3.5e-06
18	272	1.69805e+01	4.6e-07

(a) Refinement rule with second-order bounds.

Iteration	$ \mathcal{V}_k $	primal	$\text{dist}_{\mathcal{H}}(\mathcal{V}_k, X^*)$
0	2	3.80563e+03	3.3e-01
1	3	3.79912e+03	1.7e-01
2	5	9.39226e+02	8.3e-02
3	9	3.01878e+01	4.2e-02
4	17	1.84675e+01	2.1e-02
5	33	1.72061e+01	1.0e-02
6	43	1.70209e+01	5.1e-03
7	45	1.69895e+01	2.7e-03
8	47	1.69895e+01	2.7e-03
9	53	1.69826e+01	1.2e-03
10	55	1.69826e+01	1.2e-03
11	61	1.69810e+01	7.2e-04
12	67	1.69873e+01	2.6e-04
13	73	1.69828e+01	2.3e-04
14	79	1.69806e+01	1.9e-05
15	83	1.69831e+01	1.9e-05
16	87	1.69816e+01	1.9e-05
17	92	1.69806e+01	1.2e-05
18	96	1.69805e+01	4.3e-06
19	98	1.69805e+01	4.3e-06
20	100	1.69805e+01	4.3e-06
21	104	1.69805e+01	3.5e-06
...
31	125	1.69805e+01	3.5e-06
32	128	1.69805e+01	4.6e-07

(b) Refinement rule with second-order upper bounds and gradient lower bound.

Table 1: Algorithm’s behavior for the 1D super-resolution problem. Here, we set $\sigma = 2/M$, $M = 20$.

to the end.

A.1 Proof of Theorem (6)

The proof of Theorem 6 builds on similar arguments as the ones in the companion paper [17, Thm 3.1], with some modifications related to the discretization and the fact that the assumptions have been slightly weakened.

Proof of Theorem 6. i) *Well-posedness* Under Assumptions 2, 1, 3 and 6, we can apply Proposition 2 to ensure the existence of the primal-dual pair (μ_0, q_0) . For the next iterates $k \geq 1$, the measure μ in Assumption 6 still satisfies $\mu \in \mathcal{M}(\mathcal{V}_k)$ by nestedness of the sequence $(\mathcal{V}_k)_{k \in \mathbb{N}}$. Hence we can apply Proposition 2 again.

ii) *Existence of the limit of the primal solutions.* First remark that the sequence $(J(\mu_k))_{k \in \mathbb{N}}$ is non-increasing since the sets \mathcal{V}_k are nested. Since J is coercive, the sequence $(\|\mu_k\|_{\mathcal{M}(\Omega)})_{k \in \mathbb{N}}$ is bounded. Hence there exists a subsequence $(\mu_k)_{k \in \mathbb{N}}$, which we do not relabel, that weak-* converges towards a measure μ_∞ .

iii) *Existence of the limit of the dual solutions.*

Let $C_k \stackrel{\text{def.}}{=} \{q \in \mathbb{R}^M, \|A^*q\|_{L^\infty(\mathcal{V}_k)} \leq 1\}$. The fact that $\mathcal{V}_{k+1} \supseteq \mathcal{V}_k$ implies that $C_{k+1} \subseteq C_k$. Therefore any solution q_k of $(\mathcal{D}(\mathcal{V}_k))$ belongs to C_0 . By Assumption 6, f^* is coercive on C_0 (see the proof 2). Moreover, the sequence $(f^*(q_k))_{k \in \mathbb{N}}$ is nondecreasing, bounded above by $f^*(q^*)$. Therefore all the vectors q_k belong to

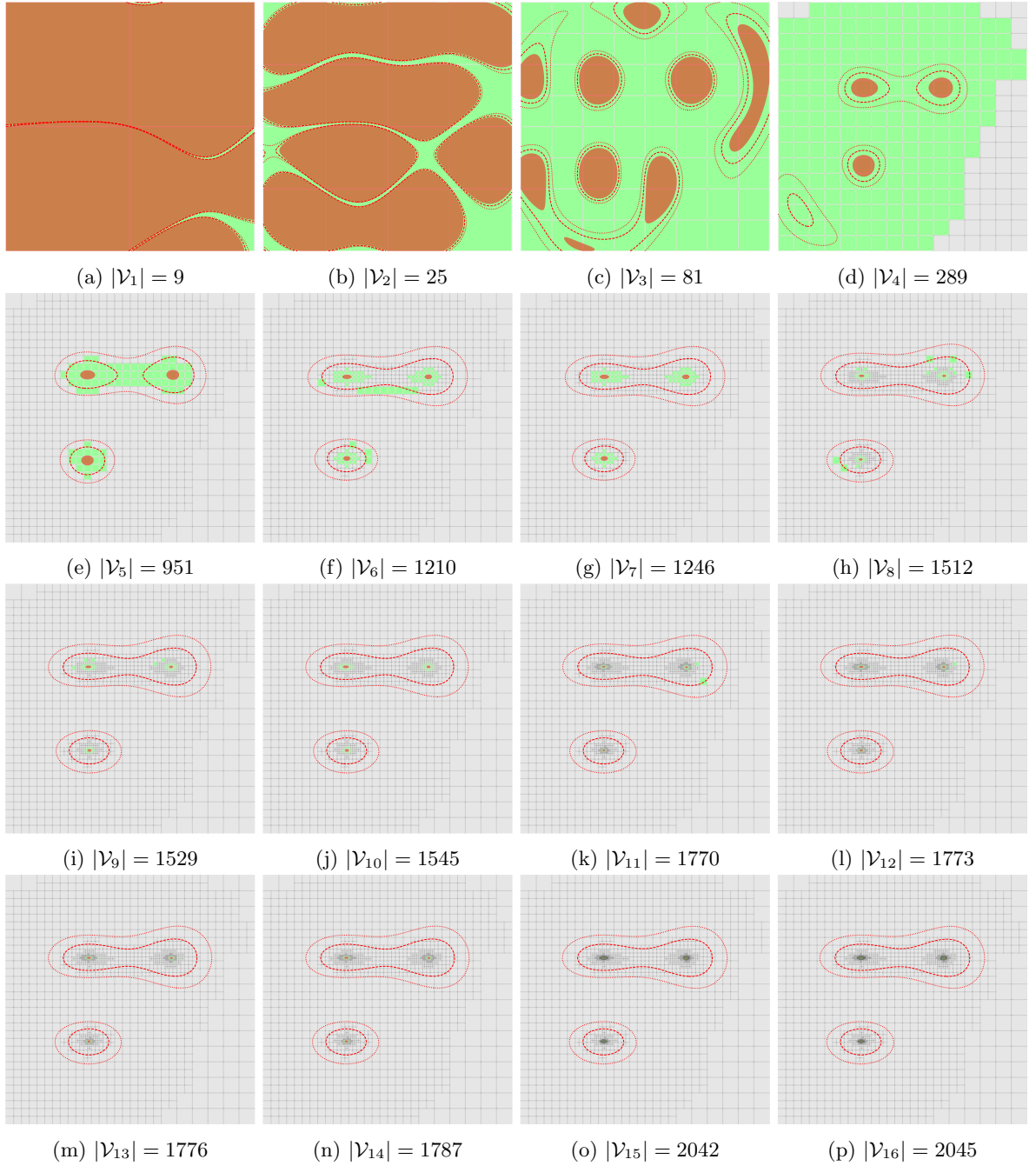


Figure 4: Our algorithm's behavior on a 2D example for second-order selection rules. The set Ω_k^* is displayed in green, the superlevelset 1 of $|A^*q_k|$ is filled with red, the level-set 0.9 is represented with a dashed red line and the levelset 0.75 with a dotted red line. The algorithm starts with a burn-in period of 3 iterations. There, it refines all cells uniformly since the upper-bound is highly inaccurate. Then, only the cells around the locations of X^* get refined in a multiscale fashion. Remember that only the cells in Ω_k^* with largest diameter are refined. This explains the behavior of the algorithm between, e.g. Figures 4f and 4g.

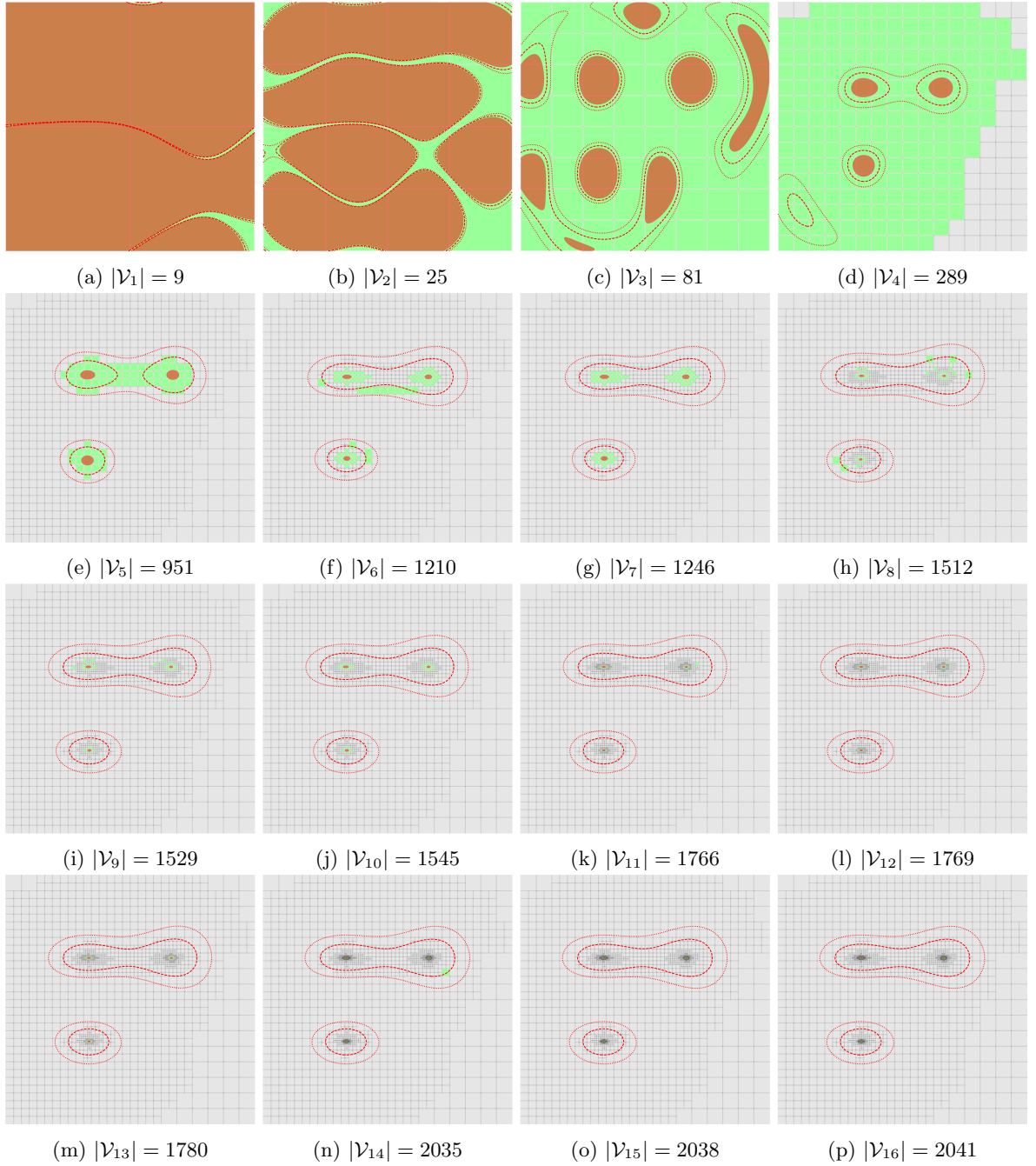


Figure 5: The adaptive refinement algorithm's behavior on a 2D example for second-order selection rules with first order gradient. The behavior quadtree are very similar to those in Figure 4.

the level set $\{q \in \mathbb{R}^M, f^*(q) \leq f^*(q^*)\}$, which is bounded by coercivity of f^* . Up to a subsequence, $(q_k)_{k \in \mathbb{N}}$ converges to a limit point q_∞ .

- iv) *Equicontinuity of $(A^* q_k)_{k \in \mathbb{N}}$.* As another technical lemma, we prove that the set $(A^* q_k)_{k \in \mathbb{N}}$ is equicontinuous. Let $\epsilon > 0$ be arbitrary. Since the functions $a_m \in \mathcal{C}_0(\Omega)$ all are uniformly continuous, there exists a $\delta > 0$

Iteration	$ \mathcal{V}_k $	primal	$\text{dist}_{\mathcal{H}}(\mathcal{V}_k, X^*)$
0	4	1.35942e+03	4.7e-01
1	9	9.42990e+02	2.4e-01
2	25	1.53313e+02	1.2e-01
3	81	3.01429e+01	6.0e-02
4	289	2.31285e+01	3.0e-02
5	951	2.21082e+01	1.6e-02
6	1210	2.19244e+01	7.7e-03
7	1246	2.19244e+01	7.7e-03
8	1512	2.18916e+01	4.6e-03
9	1529	2.18955e+01	4.6e-03
10	1545	2.18956e+01	4.6e-03
11	1770	2.18836e+01	2.2e-03
12	1773	2.18870e+01	2.2e-03
13	1776	2.18870e+01	2.2e-03
14	1787	2.18870e+01	2.2e-03
15	2042	2.18795e+01	6.7e-04
16	2045	2.18795e+01	6.7e-04
17	2315	2.18778e+01	4.4e-04
18	2647	2.18770e+01	2.7e-04
19	3126	2.18766e+01	1.2e-04

(a) Refinement rule with second-order bounds.

Iteration	$ \mathcal{V}_k $	primal	$\text{dist}_{\mathcal{H}}(\mathcal{V}_k, X^*)$
0	4	1.35942e+03	4.7e-01
1	9	9.42990e+02	2.4e-01
2	25	1.53313e+02	1.2e-01
3	81	3.01429e+01	6.0e-02
4	289	2.31285e+01	3.0e-02
5	951	2.21082e+01	1.6e-02
6	1210	2.19244e+01	7.7e-03
7	1246	2.19244e+01	7.7e-03
8	1512	2.18916e+01	4.6e-03
9	1529	2.18955e+01	4.6e-03
10	1545	2.18956e+01	4.6e-03
11	1766	2.18870e+01	2.2e-03
12	1769	2.18870e+01	2.2e-03
13	1780	2.18870e+01	2.2e-03
14	2035	2.18795e+01	6.7e-04
15	2038	2.18795e+01	6.7e-04
16	2041	2.18795e+01	6.7e-04
17	2318	2.18778e+01	4.4e-04
18	2623	2.18770e+01	2.7e-04
19	3007	2.18766e+01	1.2e-04

(b) Refinement rule with second-order upper bounds and gradient lower bound.

Table 2: The adaptive refinement Algorithm's behavior for the 2D super-resolution problem.

with the property

$$\|x - y\|_2 < \delta \Rightarrow |a_m(x) - a_m(y)| < \frac{\epsilon}{\sup_k \|q_k\|_1} \text{ for all } m.$$

Consequently,

$$\begin{aligned} \|x - y\|_2 < \delta \Rightarrow |(A^*q_k)(x) - (A^*q_k)(y)| &= \left| \sum_{m=1}^M (a_m(x) - a_m(y))q_k(m) \right| \leq \sum_{m=1}^M |a_m(x) - a_m(y)| |q_k(m)| \\ &< \frac{\epsilon}{\sup_k \|q_k\|_1} \sum_{m=1}^M |q_k(m)| \leq \epsilon. \end{aligned} \quad (16)$$

- v) *Feasibility of q_∞ .* Due to the convergence of $(q_k)_{k \in \mathbb{N}}$, the sequence $(A^*q_k)_{k \in \mathbb{N}}$ is converging strongly to A^*q_∞ . We will now prove that $\|A^*q_\infty\|_{L^\infty(\Omega)} \leq 1$. Towards a contradiction, assume that $\|A^*q_\infty\|_{L^\infty(\Omega)} = 1 + 2\epsilon$ for an $\epsilon > 0$. By convergence of (A^*q_k) , we can conclude that there exists $k_0 \in \mathbb{N}$ such that for $k \geq k_0$, $\|A^*q_k\|_{L^\infty(\Omega)} \geq 1 + \epsilon$. Set δ as in (16). The set X_k is not empty and there exists a cell ω that contains a point of X_k and satisfies $\|A^*q_k\|_{L^\infty(\omega)} \geq 1 + \epsilon$. By Assumption 4 this cell belongs to Ω_k^* . It must further satisfy $\text{diam}(\omega) \geq 2\delta$. If not, all points in x have a distance to \mathcal{V}_k smaller than δ . Since we have $|A^*q_k|(x) \leq 1$ for $x \in \text{vert}(\omega)$, the equicontinuity of the A^*q_k imply that $|A^*q_k(x)| \leq 1 + \epsilon$ for all $x \in \omega$, which is a contradiction. Hence, for all $k \geq k_0$, there exists $\omega \in \Omega_k^*$ such that $\text{diam}(\omega) \geq 2\delta$. Let $(\omega_k)_{k \in \mathbb{N}}$ denote a sequence of refined cells in Ω_k . Since we pick the active cells of largest diameter, we must have $\text{diam}(\omega_k) \geq 2\delta$ for all $k \geq k_0$. Since all the ω_k 's belong to a compact set Ω , there is a finite number of cells with diameter larger than 2δ . Hence, we can extract a subsequence of (ω_k) which is constant. This is a contradiction, because the cells (ω_k) are refined and cannot appear twice.

vi) *Convergence to a solution.* Overall, we proved that the primal-dual pair (μ_∞, q_∞) is feasible. It remains to prove that it is actually a solution. Here, we reproduce the argument of [17] for completeness. Let us first remark that $\|\mu_\infty\|_{\mathcal{M}} + f(A\mu_\infty) \geq -f^*(q_\infty)$ by weak duality. To prove the second inequality, first notice that the weak-* continuity of A implies that $A\mu_k \rightarrow A\mu_\infty$. Assumption 2 furthermore implies that f is lower semi-continuous. As a supremum of linear functions, so is f^* . Since also $q_k \rightarrow q_\infty$, we conclude

$$f^*(q_\infty) + f(A\mu_\infty) \leq \liminf_{k \rightarrow \infty} f^*(q_k) + f(A\mu_k).$$

Assumptions 2, 1 together with Proposition 2 imply exact duality of the discretized problems. This means $f^*(q_k) + f(A\mu_k) = -\|\mu_k\|_{\mathcal{M}}$. Since the norm is weak-* l.s.c. , we thus obtain

$$\liminf_{k \rightarrow \infty} f^*(q_k) + f(A\mu_k) = \liminf_{k \rightarrow \infty} -\|\mu_k\|_{\mathcal{M}} \leq -\liminf_{k \rightarrow \infty} \|\mu_k\|_{\mathcal{M}} \leq -\|\mu_\infty\|_{\mathcal{M}}.$$

Reshuffling these inequalities yields $\|\mu_\infty\|_{\mathcal{M}} + f(A\mu_\infty) \leq -f^*(q_\infty)$, i.e., the reverse inequality. Thus, μ_∞ and q_∞ fulfill the duality conditions, and are solutions. The final claim follows from a standard subsequence argument. \square

A.2 Proof of Theorem (7)

In this section, we prove the main theoretical result of the paper, which is Theorem (7). To proceed to the final result, we begin by proving a set of intermediary results. The first one is a list of useful inequalities. Many of them are direct adaptations from [17].

Proposition 9. *The following inequalities hold under Assumption 7:*

$$\|q_k - q^*\|_2 \lesssim \text{dist}_{\mathcal{H}}(\mathcal{V}_k | X_k) \tag{17}$$

$$\|q_k - q^*\|_2^2 \lesssim \max(\text{dist}_{\mathcal{H}}(X_k | X^*), \text{dist}_{\mathcal{H}}(\mathcal{V}_k | X^*)) \cdot \text{dist}_{\mathcal{H}}(\mathcal{V}_k | X^*) \tag{18}$$

Proof. All those inequalities come directly from [17]. The first inequality (17) comes from Lemma 3.5, the second (18) from Lemma 3.6. \square

Proposition 10 (A list of useful inequalities). *Under Assumptions 7 and 8, there exists a $k_0 \in \mathbb{N}$ with the property that for $k \geq k_0$, the following inequalities are true:*

$$\text{dist}_{\mathcal{H}}(X^* | X_k) \lesssim \|q_k - q^*\|_2 \tag{19}$$

$$\text{dist}_{\mathcal{H}}(X_k | X^*) = \text{dist}_{\mathcal{H}}(X^* | X_k) \tag{20}$$

$$\text{dist}_{\mathcal{H}}(X_k | X^*) \lesssim \text{dist}_{\mathcal{H}}(\mathcal{V}_k | X_k) \tag{21}$$

$$\text{dist}_{\mathcal{H}}(\mathcal{V}_k | X_k) \asymp \text{dist}_{\mathcal{H}}(\mathcal{V}_k | X^*) \tag{22}$$

$$\|q_k - q^*\|_2 \lesssim \text{dist}_{\mathcal{H}}(\mathcal{V}_k | X^*) \tag{23}$$

$$f(\mu_k) - f(\mu^*) \lesssim \text{dist}_{\mathcal{H}}(\mathcal{V}_k | X^*)^2. \tag{24}$$

Proof. The first inequalities (20) and (19) are simple consequences of Proposition 3.7 in the companion paper [17], together with the fact that, by the generic convergence result, q_k converges to q^* .

Inequality (21) is a combination of (17), (20) and (19).

To prove inequality (22), let us start by proving that $\text{dist}_{\mathcal{H}}(\mathcal{V}_k | X_k) \gtrsim \text{dist}_{\mathcal{H}}(\mathcal{V}_k | X^*)$. We have by the triangular inequality

$$\text{dist}_{\mathcal{H}}(\mathcal{V}_k | X^*) \leq \text{dist}_{\mathcal{H}}(\mathcal{V}_k | X_k) + \text{dist}_{\mathcal{H}}(X_k | X^*) \stackrel{(21)}{\lesssim} \text{dist}_{\mathcal{H}}(\mathcal{V}_k | X_k).$$

Let us prove the converse inequality $\text{dist}_{\mathcal{H}}(\mathcal{V}_k | X_k) \lesssim \text{dist}_{\mathcal{H}}(\mathcal{V}_k | X^*)$. To this end, first combine (18), (20) and (19) to get

$$\text{dist}_{\mathcal{H}}(X^* | X_k)^2 \lesssim \max(\text{dist}_{\mathcal{H}}(X^* | X_k), \text{dist}_{\mathcal{H}}(\mathcal{V}_k | X^*)) \cdot \text{dist}_{\mathcal{H}}(\mathcal{V}_k | X^*). \tag{25}$$

Regardless which of the expressions $\text{dist}_{\mathcal{H}}(X^*|X_k)$ and $\text{dist}_{\mathcal{H}}(\mathcal{V}_k|X^*)$ is larger, this inequality yields $\text{dist}_{\mathcal{H}}(X^*|X_k) \lesssim \text{dist}_{\mathcal{H}}(\mathcal{V}_k|X^*)$. Combining this and the triangular inequality, we get

$$\text{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k) \leq \text{dist}_{\mathcal{H}}(\mathcal{V}_k|X^*) + \text{dist}_{\mathcal{H}}(X^*|X_k) \lesssim \text{dist}_{\mathcal{H}}(\mathcal{V}_k|X^*).$$

The inequality (22) together with (22) now implies (23). Since the inequality (24) is a direct consequence of Proposition 3.12 in [17], we can conclude the proof. \square

Let us introduce the following shorthand notation to design a neighborhood of X^* of width $r > 0$:

$$\mathcal{B}_r \stackrel{\text{def}}{=} \bigcup_{x \in X^*} B(x, r).$$

Proposition 11 (Approximate nondegeneracy of $|A^*q_k|$). *Under Assumptions 7 and 8, there exists $k_1 \in \mathbb{N}$, which we can assume to be larger than k_0 , such that for all $k \geq k_1$, A^*q_k satisfies the approximate nondegenerate source condition:*

i) *The balls $B(x_s^*, R)$ contain exactly one local maximizer $x_{k,s}$ of $|A^*q_k|$ for each $1 \leq s \leq S$.*

ii) *Within these balls, $|A^*q_k|$ is strongly concave:*

$$|A^*q_k|''(x) \preceq -\frac{\gamma}{2} \text{Id}, \quad \forall x \in \mathcal{B}_R. \quad (26)$$

iii) *Outside of these balls, we have:*

$$|A^*q_k|(x) \leq 1 - \frac{\gamma R^2}{4}, \quad \forall x \in \Omega \setminus \mathcal{B}_R. \quad (27)$$

iv) *Finally*

$$\sup_{x \in \Omega} |A^*q_k|(x) \leq 1 + c_2 \text{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k)^2. \quad (28)$$

Proof. The convergence of $(q_k)_{k \in \mathbb{N}}$ to q^* and the fact that the functions $a_m \in C^2(\Omega)$ imply

$$A^*q_k \rightarrow A^*q^*, \quad (A^*q_k)' \rightarrow (A^*q^*)' \quad \text{and} \quad (A^*q_k)'' \rightarrow (A^*q^*)'' \quad \text{uniformly.}$$

The conclusion of the three first points follows from the nondegeneracy of A^*q^* in Assumption 8. To obtain the last, let ω_s denote the cell containing the point $x_{k,s}$ in X_k closest to x_s^* . We have by definition $(A^*q_k)'(x_{k,s}) = 0$. Let v_s denote a vertex of ω_s closest to $x_{k,s}$. To conclude, we can use a second-order Taylor expansion with the mean-value form of the remainder. It reads:

$$|A^*q_k|(v_s) = |A^*q_k|(x_{k,s}) + \langle |A^*q_k|'(x_{k,s}), v_s - x_{k,s} \rangle + \frac{1}{2} \langle |A^*q_k|''(\xi)(v_s - x_{k,s}), (v_s - x_{k,s}) \rangle$$

for some point ξ in the segment $[v, x_{k,s}]$. By construction $|A^*q_k|(v_s) \leq 1$, $|A^*q_k|'(x_{k,s}) = 0$. Moreover $(|A^*q_k|''(\xi))_{k \in \mathbb{N}}$ is uniformly bounded. This yields for all s :

$$1 \geq |A^*q_k|(x_{k,s}) - c_2 \|v_s - x_{k,s}\|_2^2 \geq |A^*q_k|(x_{k,s}) - c_2 \text{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k)^2.$$

Taking the maximum over the different $1 \leq s \leq S$ gives the result (28). \square

The above proposition translates to the following result for the upper-bound.

Proposition 12 (Finite time behavior of the upper-bound). *Under Assumptions 7 and 8, there exists $k_2 \geq k_1$ and some positive constants c_1, c_2, c_3 such that for all $k \geq k_2$ and for all cell ω :*

$$\sup_{x \in \omega} |A^*q_k|(x) \leq \begin{cases} 1 - c_1 \text{dist}(\omega, X_k)^2 + c_2 \text{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k)^2 & \text{if } \text{dist}(\omega, X^*) \leq R, \\ 1 - c_3 R^2 & \text{if } \text{dist}(\omega, X^*) \geq R. \end{cases} \quad (29)$$

Proof. Take a cell ω with $\text{dist}(\omega, X^*) \geq R$. For $k \geq k_1$, the upper-bound (27) is valid. Hence, we obtain the second bound in inequality (29) for all $k \geq k_1$ and $c_3 = \frac{\gamma}{4}$.

To obtain the first inequality, consider a cell ω with $\text{dist}(\omega, X^*) \leq R$. Let $s \in \llbracket 1, S \rrbracket$ denote any index such that $B(x_s^*, R) \cap \omega \neq \emptyset$. Point i) in Proposition 11 implies the existence of a unique point $x_{k,s}$ in $X_k \cap B(x_s^*, R)$. Proposition 11, point iv) implies that $|A^*q_k(x_{k,s})| \leq 1 + c_2 \text{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k)^2$. Moreover, point ii) in Proposition 11 states that $|A^*q_k|$ is strongly concave in the balls $B(x_s^*, R)$. Therefore:

$$|A^*q_k|(x) \leq 1 + c_2 \text{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k)^2 - c_1 \text{dist}(x, X_k)^2 \text{ with } c_1 = \frac{\gamma}{4}, \forall x \in B(x_s^*, R) \cap \omega.$$

Using the above inequality and point iii) in Proposition 11 gives:

$$\begin{aligned} \sup_{x \in \omega \cap \mathcal{B}_R} |A^*q_k|(x) &\leq 1 + c_2 \text{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k)^2 - c_1 \text{dist}(\omega, X_k)^2, \\ \sup_{x \in \omega \cap \mathcal{B}_R^c} |A^*q_k|(x) &\leq 1 - \frac{\gamma R^2}{4} \end{aligned} \quad (30)$$

We now need to show that (30) actually implies

$$|A^*q_k|(x) \leq 1 + c_2 \text{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k)^2 - c'_1 \text{dist}(\omega, X_k)^2.$$

for $x \in \omega \cap \mathcal{B}_R^c$. To this end, first notice that point i) in Proposition 11 shows that $\text{dist}_{\mathcal{H}}(X_k|X^*) \leq R$. Therefore, if $\text{dist}(\omega, X^*) \leq R$, we get

$$\text{dist}(\omega, X_k) \stackrel{(5)}{\leq} \text{dist}(\omega, X^*) + \text{dist}_{\mathcal{H}}(X_k|X^*) \leq 2R.$$

Hence, we get $\text{dist}(\omega, X_k)^2 \lesssim R^2$, and by (30) we get

$$\sup_{x \in \omega \cap \mathcal{B}_R^c} |A^*q_k|(x) \leq 1 - \frac{\gamma R^2}{4} \leq 1 - c'_1 \text{dist}(\omega, X_k)^2$$

for some other constant c'_1 . In particular, we get

$$\sup_{x \in \omega} |A^*q_k|(x) \leq 1 + c_2 \text{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k)^2 - c'_1 \text{dist}(\omega, X_k)^2.$$

□

Proposition 13 (Structural properties of the partitions Ω_k). *For $k \geq k_2$, let Ω_k denote a partition generated by Algorithm 2. There exists a radius $r > 0$ such that any cell $\omega \in \Omega_k$ satisfies:*

i) $\text{dist}(\omega, X^*) \geq R \Rightarrow |\omega| \geq r.$

ii) $\text{dist}(\omega, X^*) \leq R \Rightarrow |\omega| \gtrsim \text{dist}(\omega, X^*).$

iii) For $\ell < 2^{-k_2} \cdot \min_{\omega \in \Omega_0} |\omega|$, we have $\min_{\omega \in \Omega_k} |\omega| < \ell \Rightarrow \text{dist}_{\mathcal{H}}(\mathcal{V}_k|X^*) \lesssim \ell.$

Proof of Proposition 13. Let k_2 be the number of iterations referenced in Proposition 12. It is clear that after k_2 iterations, all the cells have a diameter larger than $2^{-k_2} \cdot \min_{\omega \in \Omega_0} |\omega|$.

Let us establish point i) first. Let ω denote a cell in Ω_k with $\text{dist}(\omega, X^*) \geq R$. To be refined by the algorithm, this cell needs to verify the second order approximation Assumption 5:

$$\|A^*q_k\|_{L^\infty(\omega)} \geq 1 - \kappa|\omega|^2$$

On the other hand, Proposition 12 states that

$$\|A^*q_k\|_{L^\infty(\omega)} \leq 1 - c_3R^2$$

A necessary condition for refinement by Algorithm 2 is then $|\omega| \geq cR$ for some c . Taking $r = \min\left(\frac{c}{2}R, 2^{-k_2} \cdot \min_{\omega \in \Omega_0} |\omega|\right)$ proves i).

Point ii) is more technical. It follows from the following arguments.

1. Suppose that the partition Ω_k contains a cell ω with an edge-length $|\omega| \stackrel{\text{def.}}{=} \ell$ with $\ell < 2^{-k_2} \cdot \min_{\omega \in \Omega_0} |\omega|$.
2. The parent cell ω_p of ω must have been refined in some iteration k_ℓ before k but after k_2 . For this k_ℓ , the size of the parent cell is $|\omega_p| = 2\ell$.
3. Since we refine the largest cells in $\Omega_{k_\ell}^*$, it means that every cell in $\Omega_{k_\ell}^*$ has an edge-length smaller or equal than 2ℓ . By construction, every cell that contains an element of X_k is in $\Omega_{k_\ell}^*$, hence

$$\text{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k) \lesssim \ell.$$

4. By Proposition 10, equation (22), we have $\text{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k) \asymp \text{dist}_{\mathcal{H}}(\mathcal{V}_k|X^*)$. We get $\text{dist}_{\mathcal{H}}(\mathcal{V}_k|X^*) \lesssim \ell$ for $k = k_\ell$. By monotonicity of the sequence $(\text{dist}_{\mathcal{H}}(\mathcal{V}_k|X^*))_{k \in \mathbb{N}}$, we also get the inequality for all $k \geq k_\ell$.
5. Noticing that $\text{dist}(\omega_p, X^*) \leq \text{dist}(\omega, X^*) \leq R$, we apply Proposition 12 to ω_p to get

$$\|A^*q_k\|_{L^\infty(\omega_p)} \leq 1 - c_1 \text{dist}(\omega_p, X_k)^2 + c_2 \ell^2.$$

Since ω_p is refined at iteration k_ℓ , it belongs to $\Omega_{k_\ell}^*$ and needs to verify the second order approximation Assumption 5:

$$\|A^*q_k\|_{L^\infty(\omega_p)} \geq 1 - \kappa|\omega_p|^2 \geq 1 - 4\kappa\ell^2.$$

From the two previous inequalities, we get $\text{dist}(\omega_p, X_k) \lesssim \ell$.

6. To conclude, remark that

$$\begin{aligned} \ell &\gtrsim \text{dist}(\omega_p, X_k) \\ &\stackrel{(5)}{\geq} \text{dist}(\omega, X_k) - \text{dist}_{\mathcal{H}}(\omega|\omega_p) \geq \text{dist}(\omega, X_k) - \sqrt{D}\ell \\ &\stackrel{(5)}{\geq} \text{dist}(\omega, X^*) - \text{dist}_{\mathcal{H}}(X_k|X^*) - \sqrt{D}\ell \geq \text{dist}(\omega, X^*) - c_5\ell, \end{aligned}$$

for some $c_5 > 0$. This proves ii).

For point iii), repeat the first four arguments of point ii) to the cell that achieves $\min_{\omega \in \Omega_k} |\omega|$. □

Now that we established the geometrical structure of Ω_k , the remaining task is to count the number of cells in Ω_k .

Proposition 14 (Counting cells). *Let Ω_k denote a cell partition generated by Algorithm 2. Assume that*

$$\min_{\omega \in \Omega_k} |\omega| = 2^{-J}$$

for some $J \in \mathbb{N}$ with $J \geq k_2$ in Proposition 13. Then the number of cells in Ω_k satisfies:

$$|\Omega_k| \leq c_0 + c_1 S J$$

for some constants $c_0, c_1 > 0$ independent of J and S .

Proof. We decompose Ω_k as $\bigcup_{s=0}^S \Omega_{k,s}$ with

$$\begin{aligned} \Omega_{k,0} &\stackrel{\text{def.}}{=} \{\omega \in \Omega_k, \text{dist}(\omega, X^*) > R\}, \\ \Omega_{k,s} &\stackrel{\text{def.}}{=} \{\omega \in \Omega_k, \text{dist}(\omega, x_s^*) = \text{dist}(\omega, X^*), \text{dist}(\omega, X^*) \leq R\} \quad s \in \llbracket 1, S \rrbracket. \end{aligned}$$

In words, $\Omega_{k,s}$ is the set of cells in Ω_k closest to x_s^* , and at a distance smaller than R from X^* . We have

$$|\Omega_k| \leq \sum_{s=0}^S |\Omega_{k,s}|. \quad (31)$$

We first use point i) in Proposition 13 to control $|\Omega_{k,0}|$. It states that all cells in $\Omega_{k,0}$ have an edge-length larger than r . The volume of a cell of edge-length r is r^D . Since all the cells are disjoint and contained in Ω , we get

$$|\Omega_{k,0}| \leq \text{vol}(\Omega)/r^D = r^{-D}.$$

Now let us derive a bound for $|\Omega_{k,s}|$. Let $\omega(x)$ denote the cell in Ω_k containing x and $\ell \stackrel{\text{def.}}{=} 2^{-J}$. By assumption $|\omega(x)| \geq \ell$ for all $x \in \Omega$. Moreover for all $x \in \Omega$ such that $\omega(x) \in \Omega_{k,s}$ we have by Proposition 13, point ii) $|\omega(x)| \geq c \text{dist}(\omega(x), x_s^*)$. Therefore

$$|\omega(x)| \gtrsim \text{dist}(\omega(x), x_s^*) \stackrel{(5)}{\geq} \text{dist}(x, x_s^*) - \text{dist}_{\mathcal{H}}(x|\omega(x)) \geq \|x - x_s^*\|_2 - \sqrt{D}|\omega(x)|.$$

This gives $|\omega(x)| \gtrsim \|x - x_s^*\|_2$ for any $x \in \bigcup_{\omega \in \Omega_{k,s}} \omega$. Combining the two inequalities yields

$$|\omega(x)| \geq \max(\ell, c\|x - x_s^*\|_2), \quad \forall x \in \bigcup_{\omega \in \Omega_{k,s}} \omega.$$

for some $c \geq 0$. For each cell ω , we have $|\omega|^D = \int_{\omega} dx$. We continue as follows

$$\begin{aligned} |\Omega_{k,s}| &= \sum_{\omega \in \Omega_{k,s}} 1 = \sum_{\omega \in \Omega_{k,s}} \int_{\omega} |\omega(x)|^{-d} dx = \int_{\bigcup_{\omega \in \Omega_{k,s}} \omega} |\omega(x)|^{-D} dx \\ &\leq \int_{\bigcup_{\omega \in \Omega_{k,s}} \omega} \max(\ell, c\|x - x_s^*\|_2)^{-D} dx \\ &\leq \int_{\Omega} \max(\ell, c\|x - x_s^*\|_2)^{-D} dx \\ &\leq \int_{B(x_s^*, \sqrt{D})} \max(\ell, c\|x - x_s^*\|_2)^{-D} dx \\ &= \int_{B(x_s^*, \ell/c)} \ell^{-d} dx + \int_{B(x_s^*, \sqrt{D}) \setminus B(x_s^*, \ell/c)} (c\|x - x_s^*\|_2)^{-D} dx \\ &\lesssim 1 + \int_{\rho=\ell/c}^{\sqrt{D}} \rho^{-D} \rho^{D-1} d\rho \lesssim 1 + |\log_2(\ell)| \lesssim 1 + J. \end{aligned}$$

Summing up everything, we obtain $|\Omega_k| \lesssim c_0 + c_1 S J$ for some constants $c_0, c_1 \geq 0$. \square

We now gathered all the necessary ingredient to prove the complexity result.

Proof of Theorem 7. Take $J \geq k_2$. The Algorithm terminates whenever a cell of size $2^{-(J+1)}$ has to be refined. When it stops, all the cells therefore have a size larger than $2^{-(J+1)}$ by construction and $\min_{\omega \in \Omega_k} |\omega| = 2^{-(J+1)}$.

Proposition 14 therefore indicates that $|\Omega_k| \leq c_0 + c_1 S J$. Since at least one cell is refined per iteration, we reached the termination criterion for a number of iterations $k \leq c_0 + c_1 S J$. Point iii) in Proposition 13 allows us to conclude that $\text{dist}_{\mathcal{H}}(\mathcal{V}_k|X^*) \lesssim 2^{-J}$. The list of inequalities in Proposition 10 yield the conclusion. \square

A.3 Further proofs

Here, we collect proofs of the remaining, smaller and more technical propositions.

A.3.1 Proof of Proposition 1

Proof. For any $x_1 \in X_1$, $x_2 \in X_2$, $x_3 \in X_3$, we have $\|x_1 - x_2\|_2 \leq \|x_1 - x_3\|_2 + \|x_3 - x_2\|_2$. Taking the infimum over $x_1 \in X_1$ and the infimum over $x_2 \in X_2$ yields

$$\begin{aligned} \text{dist}(X_1, X_2) &\leq \inf_{x_1 \in X_1} \|x_1 - x_3\|_2 + \inf_{x_2 \in X_2} \|x_3 - x_2\|_2 \leq \sup_{x_3 \in X_3} \inf_{x_1 \in X_1} \|x_1 - x_3\|_2 + \inf_{x_2 \in X_2} \|x_3 - x_2\|_2 \\ &= \text{dist}_{\mathcal{H}}(X_1|X_3) + \inf_{x_2 \in X_2} \|x_3 - x_2\|_2. \end{aligned}$$

Taking the infimum over $x_3 \in X_3$, we obtain the claimed result. \square

A.3.2 Proof of Proposition 2

Proof. Under Assumptions 1 and 2, the function J is lower semi-continuous for the weak-* topology. The existence of a measure μ supported on \mathcal{V} with $J(\mu) < +\infty$ and the coercivity of J therefore ensures the existence of a primal solution. We then invoke Theorem [2, 9.8.1] to conclude on the existence of a dual solution, the extremality relationships and on the fact that there is no duality gap.

For the boundedness of the primal solution set in total variation norm, it suffices to use the fact that J is coercive, ensuring boundedness of its sub-level sets.

Now let us prove the boundedness of the dual solution set. To this end, notice that by convexity, f is continuous at any point in $\text{int}(\text{dom}(f))$. In particular, f is continuous at $A\mu$. Using Proposition 1.3.9 in [22], we conclude that

$$g^*(q) \stackrel{\text{def.}}{=} f^*(q) - \langle A\mu, q \rangle \quad (32)$$

is coercive. We have $f^*(q) \geq g^*(q) - \|\mu\|_{\mathcal{M}(\mathcal{V})} \|A^*q\|_{L^\infty(\mathcal{V})}$. Hence f^* is coercive on the admissible set for which $f^*(q) \geq g^*(q) - \|\mu\|_{\mathcal{M}(\mathcal{V})}$. This ensures the boundedness of the dual solution set. \square

A.3.3 Proof of Proposition 3

Proof. Any choice of $\kappa_1(q_k, \omega)$ in (7) satisfying the Lipschitz inequality

$$\sup_{x_1, x_2 \in \omega} \frac{|A^*q_k(x_1)| - |A^*q_k(x_2)|}{\|x_1 - x_2\|_2} \leq \kappa_1(q_k, \omega)$$

also satisfies $\overline{|A^*q_k|}(\omega) \geq \|A^*q_k\|_{L^\infty(\omega)}$, i.e. Assumption 4. We have

$$\sup_{x_1, x_2 \in \omega} \frac{|A^*q_k(x_1)| - |A^*q_k(x_2)|}{\|x_1 - x_2\|_2} = \sup_{x \in \omega} |A^*q_k|'(x),$$

where we consider that $|A^*q_k|'(x) = 0$ on the points of non differentiability of $|A^*q_k|$. To obtain the expression (8), we use a Hölder inequality:

$$\sup_{x \in \omega} |A^*q_k|'(x) = \sup_{x \in \omega} \left| \sum_{m=1}^M q_k[m] a'_m(x) \right| \leq \sum_{m=1}^M |q_k[m]| \sup_{x \in \omega} |a'_m(x)|. \quad (33)$$

Showing that Assumption 5 is not always valid stems from the fact that we use a 0-th order Taylor expansion, with a remainder that is therefore of first order only. \square

A.3.4 Proofs of Proposition 4 and Proposition 5

Our proofs rely on the following well-known (see e.g. [26, Lemma 1.2.3, 1.2.4]) statements about Taylor-expansions. If $f : C \rightarrow \mathbb{R}$ is a function on a convex domain $C \subseteq \mathbb{R}^n$ with a κ_2 -Lipschitz continuous gradient, we have for x, y arbitrary

$$|f(x) - f(y) - \langle f'(y), x - y \rangle| \leq \kappa_2 \frac{\|x - y\|_2^2}{2} \quad (34)$$

$$\|f'(x) - f'(y)\|_2 \leq \kappa_2 \|x - y\|_2. \quad (35)$$

The value $\kappa_2(q_k, \omega)$ is an upper-bound on the Lipschitz constant of $(A^*q_k)'$ restricted to ω . Indeed,

$$\sup_{x \in \omega} \|(A^*q_k)''(x)\|_{2 \rightarrow 2} = \sup_{x \in \omega} \left\| \sum_{m=1}^M q_k[m] a_m''(x) \right\|_{2 \rightarrow 2} \leq \sum_{m=1}^M |q_k[m]| \sup_{x \in \omega} \|a_m''\|_{2 \rightarrow 2}(x). \quad (36)$$

Proof of Proposition 4

Proof. Now, let us prove that

$$\overline{|A^*q_k|}(\omega) - \kappa_2(q_k, \omega) \text{diam}(\omega)^2 \leq \|A^*q_k\|_{L^\infty(\omega)} \leq \overline{|A^*q_k|}(\omega). \quad (37)$$

By equation (34), we get

$$|A^*q(x) - A^*q(v) - \langle (A^*q)'(v), x - v \rangle| \leq \kappa_2(q_k, \omega) \frac{\|x - v\|_2^2}{2}.$$

It follows

$$|A^*q(x)| \leq |A^*q(v) + \langle (A^*q)'(v), x - v \rangle| + \kappa_2(q_k, \omega) \frac{\|x - v\|_2^2}{2},$$

Taking first the supremum in x and then the infimum over the vertices v yields the right-hand side of (37). We also have

$$\begin{aligned} |A^*q(x)| &\geq |A^*q(v) + \langle (A^*q)'(v), x - v \rangle| - \kappa_2(q_k, \omega) \frac{\text{diam}(\omega)^2}{2} \\ &\geq |A^*q(v) + \langle (A^*q)'(v), x - v \rangle| + \kappa_2(q_k, \omega) \frac{\|x - v\|_2^2}{2} - \kappa_2(q_k, \omega) \text{diam}(\omega)^2. \end{aligned}$$

Again, taking the supremum in x and then the infimum over the vertices v we obtain the left-hand side of (37). The right hand-side of (37) proves that the second order selection process satisfies Assumption 4, whereas the left hand-side proves that it obeys Assumption 5. \square

Proof of Proposition 5

Proof. We have for all $v \in \text{vert } \omega$ and $x \in \omega$

$$\|(A^*q_k)'(x) - (A^*q_k)'(v)\|_2 \leq \kappa_2(q_k, \omega) \|x - v\|_2,$$

which implies

$$\|(A^*q_k)'(x)\|_2 \geq \|(A^*q_k)'(v)\|_2 - \kappa_2(q_k, \omega) \|x - v\|_2 \implies \inf_{x \in \omega} \|(A^*q_k)'(x)\|_2 \geq \|(A^*q_k)'(v)\|_2 - \kappa_2(q_k, \omega) \text{diam}(\omega).$$

Since this is true for every $v \in \text{vert}(\omega)$, we get

$$\|\nabla A^*q_k\|_2(\omega) \leq \inf_{x \in \omega} \|\nabla A^*q_k(x)\|_2,$$

In other words, $\|\nabla A^*q_k\|_2$ is a lower bound of $\|(A^*q_k)'\|$. Since κ_2 is an upper bound of the Lipschitz constant of $\|(A^*)'q_k\|$, $\|\nabla A^*q_k\|_2(\omega)$ is a lower bound of $\|(A^*q_k)'\|$. It follows that any cell ω that contains a point of X_k will verify both $\overline{|A^*q_k|}(\omega) \geq 1$ and $\|\nabla A^*q_k\|_2(\omega) \leq 0$. Hence Ω_k^* verifies Assumption 4. It is clear that the Ω_k^* of Definition 5 is included in the Ω_k^* of Definition 4. Because the latter verifies Assumption 5, so does the former. \square

A.3.5 Proof of Proposition 8

Proof. We have

$$\begin{aligned} a'_m(x) &= -a_m(x) \left[\frac{x - z_m}{\sigma^2} \right] \\ a''_m(x) &= a_m(x) \left[-\frac{\sigma^2}{\sigma^4} \text{Id} + \frac{(x - z_m)(x - z_m)^T}{\sigma^4} \right]. \end{aligned}$$

For any $u \in \mathbb{R}^D$ and any $x \in \Omega$, we have

$$\begin{aligned} |\langle a''_m(x)u, u \rangle| &= \frac{a_m(x)}{\sigma^4} |-\sigma^2\|u\|_2^2 + \langle u, x - z_m \rangle^2| \\ &\leq \frac{a_m(x)}{\sigma^4} \max(\sigma^2, \|x - z_m\|_2^2) \|u\|_2^2. \end{aligned}$$

To conclude, it suffices to notice that for $x \in \omega$

$$\begin{aligned} a_m(x) &\leq a_m(\text{dist}(z_m, \omega)) \\ \|x - z_m\|_2 &\leq [\text{dist}(z_m, \omega) + \text{diam}(\omega)]. \end{aligned}$$

□

Acknowledgement

AF acknowledges support from the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. FdG and PW were supported by the ANR Micro-Blind (grant ANR-21-CE48-0008), by the ANR LabEx CIMI (grant ANR-11-LABX-0040) and the support of AI Interdisciplinary Institute ANITI funding, through the French “Investing for the Future PIA3” program under the Grant Agreement ANR-19-PI3A-0004.

References

- [1] Akshay Agrawal, Robin Verschueren, Steven Diamond, and Stephen Boyd. A rewriting system for convex optimization problems. *Journal of Control and Decision*, 5(1):42–60, 2018.
- [2] Hedy Attouch, Giuseppe Buttazzo, and Gérard Michaille. *Variational analysis in Sobolev and BV spaces: applications to PDEs and optimization*. SIAM, 2014.
- [3] John M. Borwein and Adrian S. Lewis. Partially finite convex programming, Part I: Quasi relative interiors and duality theory. *Mathematical Programming*, 57(1):15–48, May 1992.
- [4] Claire Boyer, Antonin Chambolle, Yohann De Castro, Vincent Duval, Frédéric De Gournay, and Pierre Weiss. On Representer Theorems and Convex Regularization. *SIAM Journal of Optimization*, 29(2):1260–1281, 2019.
- [5] Kristian Bredies and Hanna Katriina Pikkarainen. Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(1):190–218, 2013.
- [6] Emmanuel J Candès and Carlos Fernandez-Granda. Towards a Mathematical Theory of Super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.
- [7] Lenaïc Chizat. Sparse optimization on measures with over-parameterized gradient descent. *Mathematical Programming*, 194(1):487–532, 2022.

- [8] Lenaïc Chizat and Francis Bach. On the global convergence of gradient descent for over-parameterized models using optimal transport. In *Advances in neural information processing systems*, pages 3036–3046, 2018.
- [9] Christian Clason and Karl Kunisch. A measure space approach to optimal source placement. *Computational optimization and applications*, 53(1):155–171, 2012.
- [10] Yann De Castro and Fabrice Gamboa. Exact reconstruction using Beurling minimal extrapolation. *Journal of Mathematical Analysis and applications*, 395(1):336–354, 2012.
- [11] Yann De Castro, Fabrice Gamboa, Didier Henrion, and J-B Lasserre. Exact solutions to Super Resolution on semi-algebraic domains in higher dimensions. *IEEE Transactions on Information Theory*, 63(1):621–630, 2017.
- [12] Thomas Debarre, Quentin Denoyelle, and Julien Fageot. TV-based Spline Reconstruction with Fourier Measurements: Uniqueness and Convergence of Grid-Based Methods. *arXiv preprint arXiv:2202.05059*, 2022.
- [13] Quentin Denoyelle, Vincent Duval, Gabriel Peyré, and Emmanuel Soubies. The sliding frank–wolfe algorithm and its application to super-resolution microscopy. *Inverse Problems*, 36(1):014001, 2019.
- [14] Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.
- [15] Vincent Duval and Gabriel Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015.
- [16] Armin Eftekhari and Andrew Thompson. Sparse inverse problems over measures: Equivalence of the conditional gradient and exchange methods. *SIAM Journal on Optimization*, 2019. *to appear*.
- [17] Axel Flinth, Frédéric De Gournay, and Pierre Weiss. On the linear convergence rates of exchange and continuous methods for total variation minimization. *Mathematical Programming*, 190(1):221–257, 2021.
- [18] Axel Flinth and Pierre Weiss. Exact solutions of infinite dimensional total-variation regularized problems. *Information and Inference: A Journal of the IMA*, 8(3):407–443, September 2018.
- [19] Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2):95–110, 1956.
- [20] Rainer Hettich. A review of numerical methods for semi-infinite optimization. In *Semi-infinite programming and applications*, pages 158–178. Springer, 1983.
- [21] Rainer Hettich and Kenneth O Kortanek. Semi-infinite programming: theory, methods, and applications. *SIAM review*, 35(3):380–429, 1993.
- [22] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex analysis and minimization algorithms II: Advanced Theory and Bundle Methods*. Springer Berlin Heidelberg, 2013.
- [23] Karl Kunisch, Philip Trautmann, and Boris Vexler. Optimal control of the undamped linear wave equation with measure valued controls. *SIAM Journal on Control and Optimization*, 54(3):1212–1244, 2016.
- [24] Jean Bernard Lasserre. *An introduction to polynomial and semi-algebraic optimization*, volume 52. Cambridge University Press, 2015.
- [25] Eugene L Lawler and David E Wood. Branch-and-bound methods: A survey. *Operations research*, 14(4):699–719, 1966.
- [26] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2003.

- [27] Konstantin Pieper and Daniel Walter. Linear convergence of accelerated conditional gradient algorithms in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 27:38, 2021.
- [28] Clarice Poon, Nicolas Keriven, and Gabriel Peyré. Support Localization and the Fisher Metric for off-the-grid Sparse Regularization. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 1341–1350. PMLR, 16–18 Apr 2019.
- [29] Rembert Reemtsen and Stephan Görner. Numerical methods for semi-infinite programming: A survey. pages 195–262, 1998.
- [30] Eugene Y Remez. Sur la détermination des polynômes d’approximation de degré donnée. *Comm. Soc. Math. Kharkov*, 10(196):41–63, 1934.
- [31] Gongguo Tang, Badri Narayan Bhaskar, and Benjamin Recht. Sparse recovery over continuous dictionaries—just discretize. In *Signals, Systems and Computers, 2013 Asilomar Conference on*, pages 1043–1047. IEEE, 2013.
- [32] Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht. Compressed sensing off the grid. *IEEE transactions on information theory*, 59(11):7465–7490, 2013.
- [33] Yann Traonmilin and Jean-François Aujol. The basins of attraction of the global minimizers of the non-convex sparse spike estimation problem. *Inverse Problems*, 36(4):045003, 2020.
- [34] Joel A Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information theory*, 50(10):2231–2242, 2004.
- [35] Michael Unser, Julien Fageot, and John Paul Ward. Splines Are Universal Solutions of Linear Inverse Problems with Generalized TV Regularization. *SIAM Review*, 59(4):769–793, 2017.