OJMO

Open Journal of Mathematical Optimization

Axel Flinth, Frédéric de Gournay & Pierre Weiss

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

Volume 6 (2025), article no. 3 (27 pages)

https://doi.org/10.5802/ojmo.39

Article submitted on October 2, 2023, revised on August 20, 2024,
accepted on January 27, 2025.

CENTRE MERSENNE

Open Journal of Mathematical Optimization is a member of the
Centre Mersenne for Open Scientific Publishing
http://www.centre-mersenne.org/
e-ISSN: 2777-5860

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

Axel Flinth
Department of Mathematics and Mathematical Statistics, Umeå University, Sweden

Frédéric de Gournay
Institut de Mathématiques de Toulouse (IMT), Université de Toulouse, CNRS, INSA, France

Pierre Weiss
Institut de Recherche en Informatique de Toulouse (IRIT), Université de Toulouse, CNRS, Centre de Biologie Intégrative (CBI), Laboratoire de biologie Moléculaire, Cellulaire et Développement (MCD), France

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) 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]

Digital Object Identifier 10.5802/ojmo.39

Keywords Total variation, measure spaces, Frank–Wolfe.

Acknowledgments 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. AF was supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

1 Introduction

This article is concerned with the following class of problems:

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

Here, $\mathcal{M}(\Omega)$ is the set of Radon measures on a compact domain $\Omega \subseteq \mathbb{R}^D$, equipped with the weak-$*$ topology, and $A : \mathcal{M}(\Omega) \to \mathbb{R}^M$ a linear continuous operator (equivalently, A is an operator of the form $(A\mu)_m = \int_\Omega a_m d\mu$ for some continuous functions $a_m : \Omega \to \mathbb{R}$). $\|\cdot\|_{\mathcal{M}(\Omega)}$ is the total variation and $f : \mathbb{R}^M \to \mathbb{R}\cup\{+\infty\}$ is a convex, proper function.

1.1 Applications

This problem and its variants appear in various fields. In *inverse problems*, it is used heavily for sparse source localization and super-resolution [5, 6, 10, 15, 32]. 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 $L : \mathcal{B}(\Omega) \to \mathcal{M}(\Omega)$, where $\mathcal{B}(\Omega)$ is a suitably defined Banach space, consider the following problem:

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

[1] The 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

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

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

1.2 Literature review

The variable over which we optimize in $(\mathcal{P}(\Omega))$ is of infinite-dimensional nature. Therefore, it is non-trivial to numerically resolve it. However, there is a finite dimensional problem "hiding" within it. More concretely, the problems admit *sparse* solutions (see e.g. [4]) of the form

$$\mu^\star = \sum_{s=1}^{S} \alpha_s^\star \delta_{x_s^\star}, \tag{3}$$

where $\alpha_s^\star \in \mathbb{R}$ are weights, $x_s^\star \in \Omega$ are locations in which the Dirac deltas $\delta_{x_s} \in \mathcal{M}(\Omega)$ are concentrated and where the number of sources S satisfies $S \leq M$. Hence, although the problem inherently is infinite-dimensional, its solution can be parametrized by $O(S)$ parameters. This makes it plausible that efficient algorithms for solving $(\mathcal{P}(\Omega))$ should be possible to design. Let us give a brief overview of a few approaches here.

Lasserre hierarchies

If the functions a_m are of certain forms, e.g. (trigonometric) polynomials, it is possible is to use so-called Lasserre hierarchies, which are designed to solve near generic measure optimization problems [11, 24], to solve $(\mathcal{P}(\Omega))$. This approach, which was advocated for in e.g. [6], consists of recasting the problem as a finite-dimensional SDP, which then can be solved using various methods. This reformulation is elegant, but scales poorly for large M – the variable for the reformulated problem is generically of size $\geq C \cdot M^2$ for some constant C, which is much higher than the $O(S)$-dimensional structure (3). Recently, methods for mitigating this effect have been proposed [37], but their convergence is typically sublinear. We will not consider these methods further in this work.

Non-convex reparametrization

The structure (3) immediately suggests a way to numerically solve $(\mathcal{P}(\Omega))$: Given a number $N \geq M$ of particles, a set of locations $X = (x_1, \ldots, x_N)$ in Ω^N and a weight vector $\alpha \in \mathbb{R}^N$, we can define the mapping

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

and instead solve the following finite-dimensional problem

$$\inf_{\substack{X\in\Omega^N\\ \alpha\in\mathbb{R}^N}} \|\alpha\|_1 + f(A\mu(X, \alpha)). \tag{4}$$

Because of the known structure (3), the global minimum of (4) corresponds to a global minimum of $(\mathcal{P}(\Omega))$. However, (4) is – in contrast to $(\mathcal{P}(\Omega))$ non-convex – and its global optimization is hence challenging. Still, proximal gradient descent can be used successfully under additional assumptions.

One approach is to use overparametrization. That is, one uses $N \gg M$ particles initialized on a fine grid in Ω. If N is large enough, the method will converge globally, and locally with a linear speed [7, 8]. However, the condition $N \gg M$ again means that the number of variables used is unnecessarily large.

A different route is opened when an approximate solution to $\mathcal{P}(\Omega)$ is known – we can then use (4) to fine-tune it. By now, it possesses a rich convergence theory [17, 33]. In essence, if we have a good-enough approximation of each $(\alpha_s, x_s^\star)$ in (3), we obtain a linear convergence using only $S = O(M)$ variables. A variant of this approach that deserves to be mentioned is the sliding Frank–Wolfe algorithm in [13], where a gradient descent on (4) is launched after each step of a so-called Frank–Wolfe algorithm, which we will introduce next. Needless to say, since this method only works given a good initialization is known, there are no global convergence guarantees.

To conclude, the reparametrized methods cannot be proven to converge globally at a (locally) linear speed while keeping the number of variables under control. We will not consider them further in this work.

Exchange algorithms

A final popular approach to solve $(\mathcal{P}(\Omega))$, which is the main focus of this article, are *exchange algorithms*. Let us explain them a bit more generally.

Consider the dual problem $(\mathcal{D}(\Omega))$ to $(\mathcal{P}(\Omega))$:

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

where $\|A^*q\|_{L^\infty(\Omega)} \overset{\text{def.}}{=} \sup_{x\in\Omega} |A^*q|(x)$ and f^* is the Fenchel dual of f: $f^*(q) \overset{\text{def.}}{=} \sup_{x\in\mathbb{R}^M} \langle q, x\rangle - f(x)$. Its variable $q \in \mathbb{R}^M$ is finite-dimensional, 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 [20, 21, 29]. 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

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

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

Note that although the notation may suggest otherwise, these problems are standard, finite-dimensional convex problems: The space of measures $\mathcal{M}(\mathcal{V}_k)$ on $\mathcal{V}_k$ is of dimension $|\mathcal{V}_k| < \infty$, and the constraint $\|A^*q\|_{L^\infty(\mathcal{V}_k)} \leq 1$ only amounts to $|\mathcal{V}_k|$ linear constraints $\{|A^*q(x)| \leq 1, \forall\, x \in \mathcal{V}_k\}$. Hence, both problems can be solved with off-the-shelf solvers.

The $\mathcal{V}_k$ are still to be chosen. The simplest approach is to define $\mathcal{V}_k$ as a Euclidean grid with edge-length 2^{-k} [12, 31]. It can be proven that with this choice of $\mathcal{V}_k$, the solutions of $\mathcal{P}(\mathcal{V}_k)$ converge towards a solution of $\mathcal{P}(\Omega)$ as $k \to \infty$; in fact, Theorem 14 in this paper encompasses this. However, this choice of course entails an exponential explosion of the number of variables as $k \to \infty$.

A lighter adaptive method consists of using the dual variable q_k to construct $\mathcal{V}_{k+1}$. The dual variable 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[2]. It consists of adding only the global maximizer of $|A^*q_k|$ at each step. It is described precisely in Algorithm 1.

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 $x_k^\star \overset{\text{def.}}{=} \operatorname{argmax}_{x\in\Omega} |A^*q_k|(x)$.
5: 3) Set $\mathcal{V}_{k+1} = \mathcal{V}_k \cup \{x_k^\star\}$.
6: **Output:**
   - The dual solution q_k of $(\mathcal{D}(\mathcal{V}_k))$.

It was revived in signal processing thanks to Bredies et al. in [5]. Its connection with the exchange algorithms was recalled in [16]. A linear convergence theory was developed independently by Walter and Pieper in [27] and by the authors in [17]. In brief, it states that the approach generically converges globally, and under certain regularity conditions converges locally linearly, in the sense that no more than $O(\log(\epsilon^{-1}))$ iterations are needed to obtain a solution within an error of ϵ. Furthermore, the grids grow at a speed of $O(S)$ – in particular, an error of ϵ can be achieved while all grids fulfill $|\mathcal{V}_k| \leq S\log(\epsilon^{-1})$.

[2] The version of the Frank–Wolfe algorithm we present here is not the "standard" one. In its original form, the updates of the optimization variables are not made by fully resolving a discretized dual, but rather by making a line search between then current primal iterate μ_k and the new candidate $\delta_{x_k^\star}$. It is therefore more correct to say that we consider the *fully corrected* Frank–Wolfe algorithm.

Despite its 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 $x_k^\star \overset{\text{def.}}{=} \operatorname{argmax}_{x\in\Omega} |A^* q_k|(x)$?

Although this question is most often glossed over in treatises of Frank–Wolfe-type algorithms, this problem has no reason to be simple. It is nonconvex and depends highly on the properties of the functions a_m. Hence, from a fundamental point of view, the Frank–Wolfe algorithm in general is *totally infeasible*.

The problem can be, and is usually, of course tackled with heuristics; 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 $x_k^\star$. Our experience is however that tuning the hyper-parameters in this quest for the global minimizer is time consuming and can represent a real headache for the optimizer. Hence, circumventing the calculation of $x_k^\star$ is not only a theoretical, but also a practical, issue.

1.3 Contribution

In this article, we will show that the correct answer to our question of how to find $x_k^\star$ in fact is "*We do not need to*". Indeed, we will develop a modified exchange algorithm where the grid-update is made without determining $x_k^\star$, while keeping the convergence properties intact. We subdivide Ω into collection Ω_k of hypercubes ω, and let $\mathcal{V}_k$ denote the vertices of the ω. To construct the next $\mathcal{V}_{k+1}$, or equivalently Ω_{k+1}, we determine a set $\Omega_k^\star$ of candidate cells ω where $\sup_{x\in\omega} |A^* q_k(x)|$ is likely to exceed 1. We then subdivide the largest cells in $\Omega_k^\star$ to construct Ω_{k+1}. See Figure 1.

The description of the process of selecting the $\Omega_k^\star$ is intentionally kept vague in the above description, in order to showcase the potential flexibility of our approach. It allows for the design of selection rules tailor-made to special problems and known properties of the solution $\mu^\star$. We will in the paper describe conditions on the process that will imply convergence towards a solution $\mu^\star$ of $(\mathcal{P}(\Omega))$ (Theorem 14), and conditions that will secure a linear convergence rate similar to the results cited above (Theorem 16). These conditions will be surprisingly simple to satisfy: Indeed, choosing Ω_k^* as the cells in which a simple (trivial to evaluate) linear upper bound of $|A^* q_k|$ is larger than one, will be enough to secure the linear rate. Hence, the determination of the $x_k^\star$ can be completely circumvented while still obtaining the same convergence rates as the best results known previously for algorithms of the Frank–Wolfe type. This is the main contribution of this article.

An additional minor contribution is that we have slightly weakened some of the technical assumptions about f and A compared to [17]. In particular, our arguments can be used to sharpen the pendant of Theorem 14 there.[3]

Many of our arguments will follow similar lines as [17] – in particular, we will utilize many technical results derived there. Still, the proof of the convergence of the modified algorithm comes with significant new technical hurdles and subtleties.

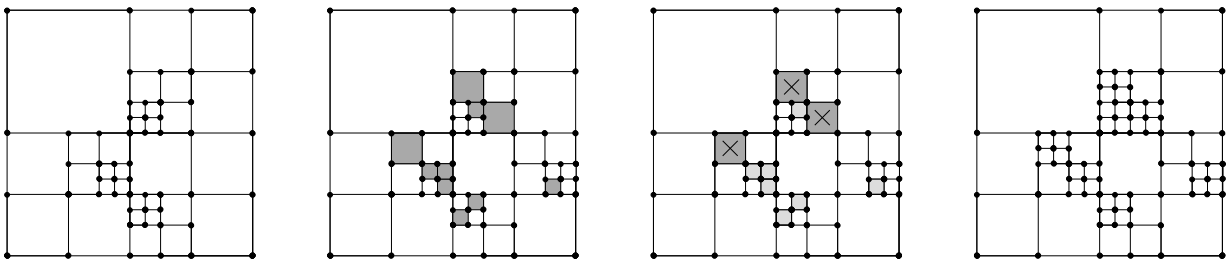


Figure 1 The refinement. From left to right: (i) a cell partition, (ii) the partition with the candidate cells grayed out, (iii) only the candidate cells with largest diameter are selected for refinement and (iv) the new resulting cell partition.

[3] For convenience for the reader, we collect all the assumptions to prove our main results, and in particular minor differences in detail between the assumptions of [17] and this paper, in Appendix B.

Outline

In Section 2, we will introduce the notation and some basic results used in the rest of the paper. In Section 3, we introduce the algorithm formally, discuss a few methods for determining the candidate cells $\Omega_k^\star$, and prove our main results (which gives conditions on f and A as well as the procedure for determining $\Omega_k^\star$ under which the algorithm converges linearly). In Section 4, we present some numerical experiments to illustrate our theory. Most proofs will be postponed to Appendix A.

2 Preliminaries

We will work on the domain $\Omega = [0,1]^D$ for simplicity. This is indeed is not a real restriction, since there always exists a non-degenerate affine transform $\varphi : \mathbb{R}^D \to \mathbb{R}^D$ with $\varphi(\Omega) \subseteq [0,1]^D$. Using this transform, we may push measurement functions and measures forward, in a bijective fashion. We let $\mathcal{M}(\Omega)$ denote the space of Radon measures of bounded total variation on Ω and $\mathcal{C}(\Omega)$ the space of continuous functions on Ω. Note that if we equip $\mathcal{M}(\Omega)$ with the total variation norm $\|\cdot\|_{\mathcal{M}(\Omega)}$ and $\mathcal{C}(\Omega)$ with the supremum norm, $\mathcal{M}(\Omega)$ can be identified with the dual of $\mathcal{C}(\Omega)$. For a subset ω of $\mathbb{R}^D$, we let $\mathrm{vol}(\omega)$ denote its volume (Lebesgue measure). The notation $[\![1, N]\!]$ 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 $\omega \subseteq [0,1]$ 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 $x \in (2^{-J} \cdot \{0, 1, \ldots, 2^J - 1\})^d$.

For a cell ω, we let $\mathrm{vert}(\omega)$ 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 $\mathrm{vol}(\omega' \cap \omega) = 0$ 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 linear operator $A : \mathcal{M}(\Omega) \to \mathbb{R}^M$ is 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 $a_m \in \mathcal{C}(\Omega)$.

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 it r-th tensor derivative. We define the following constants

$$\kappa_r \stackrel{\text{def.}}{=} \sup_{\|q\|_2 \leq 1} \sup_{x \in \Omega} \|(A^*q)^{(r)}(x)\|, \tag{5}$$

where r indicates the derivative's order and where $\|\cdot\|$ is the canonical norm for r-forms, i.e.

$$\|T\| = \sup_{\|u_1\|, \ldots, \|u_r\| \leq 1} |T(u_1, \ldots, u_r)|$$

2.3 Set distances

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

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

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

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

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

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

Notice that this distance is asymmetric: in general $\mathrm{dist}_{\mathcal{H}}(X_1|X_2) \neq \mathrm{dist}_{\mathcal{H}}(X_2|X_1)$. The following inequality will play an important role in the analysis. Its proof can be found in Appendix A.3

▶ **Proposition 3** (Triangle inequality for set distances). *For any triple of sets X_1, X_2, X_3 in $\mathbb{R}^D$ we have*

$$\mathrm{dist}(X_1, X_2) \leq \mathrm{dist}_{\mathcal{H}}(X_1|X_3) + \mathrm{dist}(X_3, X_2). \tag{6}$$

2.4 Primal, dual and existence of solutions

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

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

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

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

▶ **Proposition 4** (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 $A\mu \in \mathrm{int}(\mathrm{dom}(f))$. 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 $(\mu^\star, q^\star)$ denote a primal-dual pair. They are related by the following primal-dual relationships:*

$$A^* q^\star \in \partial_{\|\cdot\|_{\mathcal{M}(\Omega)}}(\mu^\star) \text{ and } -q^\star \in \partial f(A\mu^\star). \tag{7}$$

The left inclusion in (7) implies that the support of a solution $\mu^\star$ satisfies: $\mathrm{supp}(\mu^\star) \subseteq \{x \in \Omega, |A^* q^\star(x)| = 1\}$.

▶ Remark 5. Strong duality may hold under assumptions different than Assumptions 2 and 3. 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 solution set requires extra technicalities and assumptions in the proofs, which we decided to discard.

3 Main results

This section contains our main findings. All the proofs are postponed to the appendix.

3.1 The algorithm

The algorithm we propose consists of 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:

$$\mathcal{V}_k \overset{\text{def.}}{=} \{\text{vert}(\omega) \,|\, \omega \in \Omega_k\}.$$

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 is most probably 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* $\Omega_k^\star \subset \Omega_k$ 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 $\Omega_k^\star$, but only the largest of them. Note that we formulate the algorithm for an abstract candidate cell selection process – we will later give conditions on it to guarantee (linear) convergence, and give examples of simple processes that fulfil them. The complete solver is described in Algorithm 2. One iteration of the algorithm is displayed in Figure 1 in the previous section.

3.1.1 Solution of $(\mathcal{D}(\mathcal{V}_k))$

Our algorithm is oblivious of the solvers of the discretized primals and duals, and our analysis is completely independent of the choice of them. Consequently, our statements are about the number of "meta-iterations" k needed to obtain an approximate solution up to a certain tolerance.

Our rationale behind this approach is that any other (fully corrective) Frank–Wolfe approach, which is the most natural method to compare with, would also solve $(\mathcal{D}(\mathcal{V}_k))$ at each iterations. Hence, counting the number of meta-iterations is the fairest comparison to the results known in the literature. Indeed, the interesting aspect of our result is that our fully discretized version of the Frank–Wolfe algorithm both has the same asymptotical cost as the standard one in terms of number of $(\mathcal{D}(\mathcal{V}_k))$-problems needed to be solved.

With that said, it should be remarked that the complexity of solving $(\mathcal{D}(\mathcal{V}_k))$ grows with $|\mathcal{V}_k|$. For example, if f is a quadratic form, the problem $(\mathcal{D}(\mathcal{V}_k))$ can be solved exactly with $O(L^2\,|\mathcal{V}_k|^4)$ arithmetic operations, where L is the number of bits needed to describe the input[36]. Note however that this involves resolving the problem from scratch each time, which is wasteful given that q_k should be a good approximation of a solution to the next dual problem $\mathcal{D}(\mathcal{V}_{k+1})$. Exploring ways to efficiently solve $(\mathcal{D}(\mathcal{V}_k))$ is an interesting line of research, which we however postpone to future work.

Algorithm 2 Adaptive Refinement Algorithm

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

WHILE $\max_{\omega\in\Omega_{k-1}^*} |\omega| \geq 2^{-K_{\text{end}}}$
- (1) Determine a solution q_k of $(\mathcal{D}(\mathcal{V}_k))$
- (2) Determine the candidate cells $\Omega_k^\star$.
- (3) Subdivide the cells in $\Omega_k^\star$ with largest diameter
- (4) $k = k + 1$

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 $\Omega_k^\star$. Following [17], we let X_k denote the set of local maximizers of $|A^* q_k|$ exceeding 1, i.e.

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

We will in the next section show that the following property of the selection process is sufficient to obtain convergence of the algorithm.

▶ **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, $\Omega_k^\star$ satisfies*

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

The above condition is too weak to allow a control of the numerical complexity. For instance the choice $\Omega_k^\star = \Omega_k$ – which corresponds to refine uniformly at each iteration – obeys Assumption 4, but leads to an exponential growth of $|\Omega_k|$. We will show that the following extra condition suffices to guarantee that the $\mathcal{V}_k$ grow at a linear rate, proportional to S.

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

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

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

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

Indeed, Assumption 4 can be reformulated as $\{\omega \in \Omega_k \,|\, \omega \cap X_k \neq \emptyset\} \subseteq \Omega_k^*$, 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 $\|A^* q_k\|_{L^\infty(\omega)} \geq 1 \geq 1 - \kappa |\omega|^2$, since ω contains a cell point where $|A^* q_k|$ exceeds 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:

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

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 secures 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 6** (A first order selection process). *Assume that $a_m \in C^1(\Omega)$ for all m. Let us define*

$$\overline{|A^*q_k|}_1(\omega) = \inf_{v \in \mathrm{vert}(\omega)} |A^*q_k(v)| + \kappa_1(q_k, \omega)\,\mathrm{diam}(\omega), \tag{8}$$

with

$$\kappa_1(q_k, \omega) = \sum_{m=1}^{M} |q_k[m]| \|a_m'\|_{L^\infty(\omega)}. \tag{9}$$

The first order selection process is defined as

$$\Omega_{k,1}^\star = \left\{ \omega \in \Omega_k \;\middle|\; \overline{|A^*q_k|}_1(\omega) \geq 1 \right\}$$

▶ **Proposition 7.** *The first order selection process $\Omega_{k,1}^\star$ 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 8** (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*

$$\overline{|A^*q_k|}_2(\omega) \overset{\text{def.}}{=} \inf_{v \in \mathrm{vert}(\omega)} \sup_{x \in \omega} |A^*q_k(v) + \langle (A^*q_k)'(v), x - v\rangle| + \frac{\kappa_2(q_k, \omega)}{2}\|x - v\|^2, \tag{10}$$

with

$$\kappa_2(q_k, \omega) = \sum_{m=1}^{M} |q_k[m]| \sup_{x \in \omega} \|a_m''(x)\|_{2\to 2}. \tag{11}$$

The second order selection process is defined as

$$\Omega_{k,2}^\star = \left\{ \omega \in \Omega_k \;\middle|\; \overline{|A^*q_k|}_2(\omega) \geq 1 \right\}. \tag{12}$$

▶ **Proposition 9.** *If the sequence $(q_k)_{k\in\mathbb{N}}$ is uniformly bounded, then the second order selection process $\Omega_{k,2}^\star$ 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 of maximizing a convex function over a polyhedron. A solution is therefore attained on $\mathrm{vert}(\omega)$, and can be evaluated in constant time per cell.

Also note that the assumption that $(q_k)_{k\in\mathbb{N}}$ is bounded is weak – and particular always is true in the case of the sequence q_k converging.

▶ Remark 10. 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 Lipshitz 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 (5).

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

The larger $\Omega_k^\star$, 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 $|A^\star q_k|$ that surpass 1 to be selected. In cells containing maximizers, the gradient of $|A^\star q_k|$ vanishes. Consequently we can design a selection process based on lower bounds of the gradient.

▶ **Definition 11** (Second order selection process with first order gradient). *Assume that $a_m \in C^2(\Omega)$ for all m. Define $\overline{|A^*q_k|}_2(\omega)$ and $\kappa_2(q_k, \omega)$ as in (10) and (11) respectively. For any cell ω, and all $q_k \in \mathbb{R}^M$, define*

$$\underline{\|\nabla A^*q_k\|_2}(\omega) = \sup_{v \in \mathrm{vert}(\omega)} \|(A^*q_k)'(v)\|_2 - \kappa_2(q_k, \omega)\,\mathrm{diam}(\omega). \tag{13}$$

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

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

▶ **Proposition 12.** *If the sequence $(q_k)_{k\in\mathbb{N}}$ is uniformly bounded, then the selection process $\Omega^\star_{k,2,1}$ satisfies both Assumption 4 and 5. Incorporating gradient lower bounds reduces the cardinality of $\Omega^\star_k$ in the sense that the cardinality of $\Omega^\star_k$ in Definition 11 is not greater than the one of $\Omega^\star_k$ in Definition 8.*

▶ Remark 13. Following Remark 10, the definition of $\kappa_2(q_k, \omega)$ given in (11) 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 – i.e. that the discretized problems $(\mathcal{P}(\mathcal{V}_k))$ and $(\mathcal{D}(\mathcal{V}_k))$ all have solutions. 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 $A\mu \in \mathrm{int}(\mathrm{dom}(f))$.*

▶ **Theorem 14.** *Under Assumptions 1, 2, 3, 4 and 6, the sequences $(\mu_k)_{k\in\mathbb{N}}$ and $(q_k)_{k\in\mathbb{N}}$ in Algorithm (2) are well-defined. They contain subsequences that converge weakly to solutions $\mu^\star$ and $q^\star$ 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 15. 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: $a_m \in C^2_0(\Omega)$ for all $1 \le m \le M$.*
- *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 $\mu^\star$ of $(\mathcal{P}(\Omega))$ is unique and supported on $S \in \mathbb{N}$ points*

$$\mu^\star = \sum_{s=1}^{S} \alpha^\star_s \delta_{x^\star_s}$$

  for some $\alpha^\star_s \in \mathbb{R}$ and $x^\star_s \in \Omega$. In what follows, we let $X^\star \overset{\text{def.}}{=} \{x^\star_1, \dots, x^\star_S\}$.
- *The* dual certificate *$|A^*q^\star|$ is only equal to 1 in the points $x^\star_1, \dots, x^\star_S$ and is strictly concave around those points. This ensures the existence of a parameter $\gamma > 0$ and a radius $R > 0$ with*

$$B(x^\star_{s_1}, R) \cap B(x^\star_{s_2}, R) = \emptyset, \quad \forall\, s_1 \neq s_2 \tag{14}$$

$$\mathrm{sign}(A^*q^\star(x))(A^*q^\star)'' \preccurlyeq -\gamma\,\mathrm{Id} \quad \textit{for } x \textit{ with } \mathrm{dist}(x, X^\star) \le R,$$

$$|A^*q^\star(x)| \le 1 - \frac{\gamma R^2}{2} \quad \textit{for } x \textit{ with } \mathrm{dist}(x, X^\star) \ge R. \tag{15}$$

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. [6, 28]. We can now formulate our main result.

▶ **Theorem 16.** *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 $\mu^\star$) such that the algorithm terminates in no more than $k = k_0 + cK_{\text{end}}$ iterations. We further have*

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

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 [17], and utilizes similar ideas, but still 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 $q^\star$, 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 $X^\star$. This results in a multiscale refinement of local regions around the sought-for locations, see e.g. Figure 3. ◀

▶ Remark 17. We did not keep track of the constants in the above inequalities. Some of them could theoretically be extracted from the proof, but others, like the time k_0 to reach a linear convergence rate, cannot.

▶ Remark 18. We can replace the final output μ_k, the solution of $(\mathcal{P}(\mathcal{V}_k))$, by the solution $\widetilde{\mu}_k$ of $\mathcal{P}(\operatorname{vert}(\Omega_k^\star))$ defined as

$$\widetilde{\mu}_k \overset{\text{def.}}{=} \inf_{\mu \in \mathcal{M}(\operatorname{vert}(\Omega_k^\star))} \|\mu\|_{\mathcal{M}} + f(A\mu), \tag{16}$$

and still have $J(\widetilde{\mu}_k) - J(\mu^\star) \lesssim 2^{-2K_{\text{end}}}$. The interest of this alternative problem is that the cardinality of $\operatorname{vert}(\Omega_k^\star)$ is significantly smaller than that of $\mathcal{V}_k$, helping to reduce the numerical complexity of the resolution of the final $(\mathcal{P}(\mathcal{V}_k))$. Note however that the same trick cannot be used for solution of the discretized duals $(\mathcal{D}(\mathcal{V}_k))$ – here, the entire $\mathcal{V}_k$ needs to be used each time.

▶ Remark 19. Our algorithm relies on dyadic subdivision of cells. We therefore cannot expect the algorithm to converge faster than linearly. In that regard, Theorem 16 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

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

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 (11), we need an upper-bound on the second order derivatives.

▶ **Proposition 20.** *Define*

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

Then, $\kappa_{2,m}(\omega) \geq \sup_{x \in \omega} \|a_m''(x)\|_{2\to 2}$. We can choose $\kappa_2(q_k, \omega) = \sum_{m=1}^M |q_k|[m]\kappa_{2,m}(\omega)$, in Proposition 11 to define the second order candidates $\Omega_k^\star$.

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 [1, 14]. 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 (4), initialized in the "ground truth measures" we specify. This is sound, since the true solution lies close to them (see e.g. [28]).

The code used in the experiments is released at `https://github.com/usinedepain/griddedTV`.

4.2 1D-experiments

4.2.1 The problem

We set $y = A\bar{\mu}$ with $\bar{\mu} \overset{\text{def.}}{=} 8\delta_{1/3} - 9\delta_{2/3}$. 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 $z_m \overset{\text{def.}}{=} m/M$ and $\sigma = 2/M$, with $M = 20$.

4.2.2 Second-order upper bound

The behavior of the second order selection process algorithm of Definition 8 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 $\text{dist}_{\mathcal{H}}(\mathcal{V}_k, X^\star)$ decays exponentially fast, illustrating Theorem 16 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.

4.2.3 Second-order upper bound with first order gradient

We turn our attention to the second order selection process with first order gradient, see Definition 11. 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 $\Omega_k^\star$. Two measures of complexity can be used to compare the approaches: i) the cardinality $|\mathcal{V}_k|$ needed to reach a given accuracy $\text{dist}_{\mathcal{H}}(\mathcal{V}_k, X^\star)$, or ii) the number of iterations to reach the same accuracy. Reducing the cardinality of $\Omega_k^\star$ 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

4.3.1 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 $m \in [\![1, M]\!]$ can be decomposed as $m = (m_1, m_2) \in [\![0, \sqrt{M} - 1]\!]^2$ and $z_m = \frac{1}{\sqrt{M}}(m_1, m_2)$ with $M = 16$. $\bar{\mu}$ is chosen as

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

4.3.2 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 16: after a burn-in period, the grid is refined in a multi-scale fashion, only around the support $X^\star$ of the solution $\mu^\star$. 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

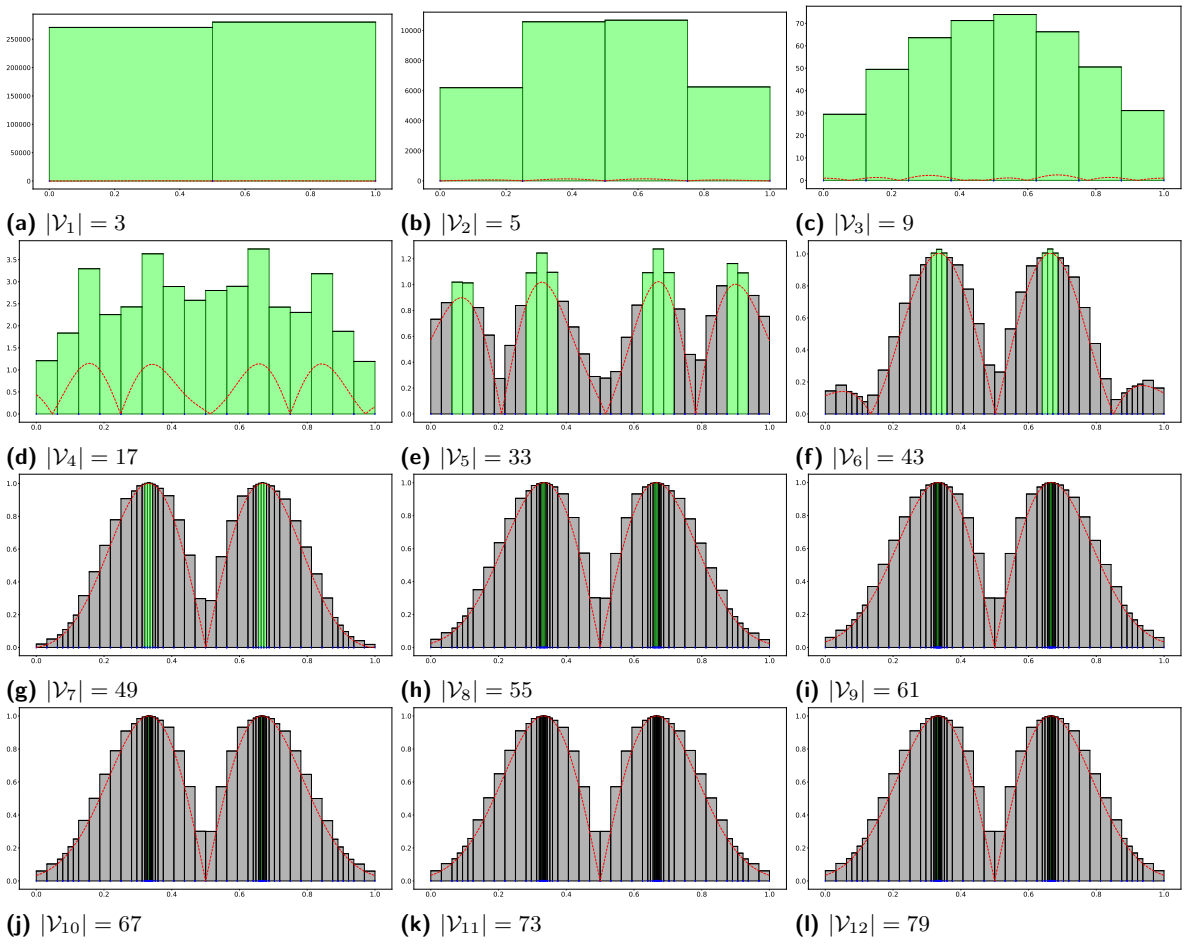


Figure 2 The behavior of the adaptive refinement algorithm with a second order selection process, on a 1D sparse recovery problem. The set $\Omega_k^\star$ is displayed in green, the function $|A^*q_k|$ is displayed in dashed red, the upper-bound $\overline{|A^*q_k|}$ is the piecewise-constant function. 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 $X^\star$ get refined in a multiscale fashion.

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}, 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 a modified 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 is to implement 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 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

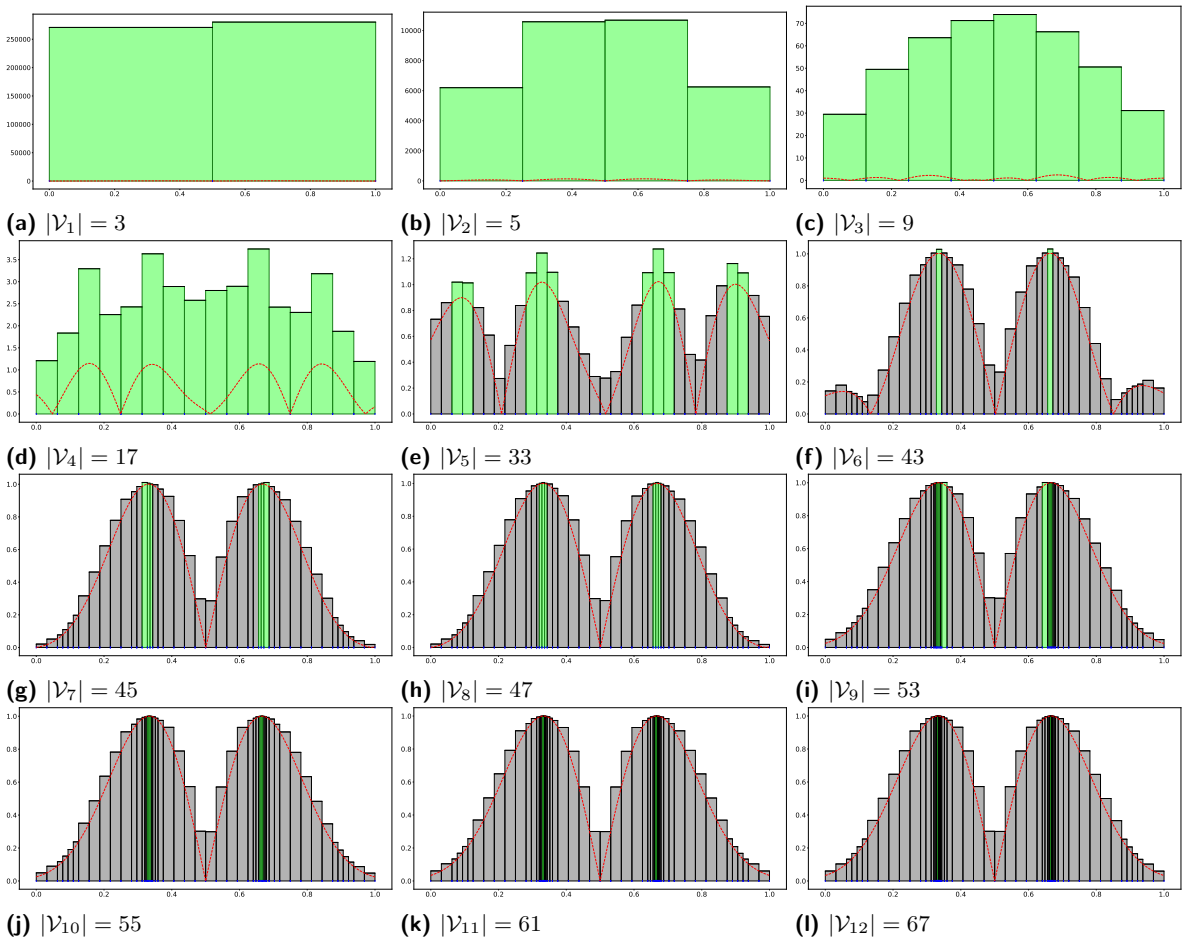


Figure 3 The adaptive refinement algorithm with a second order selection process and first order gradient. The setting is the same as the one of Figure 2. The cardinality of the set of candidates $\Omega_k^\star$ is smaller, as can be seen when comparing Figure 3f with Figure 2f.

could be achieved with an arbitrary accuracy. A complete theory should account for approximation errors and for the complexity of the sub-problems. We leave this for future work.

On a more positive side, the scope of this approach is possibly significantly wider than total variation regularization. Up to some adjustements, 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 14) and linear (Theorem 16) convergence results, and save the proofs of smaller, technical propositions to the end.

A.1 Proof of Theorem (14)

The proof of Theorem 14 will closely follow the argumentation of [17, Theorem 3.1]. Just as there, we will proceed in six steps: (i) Well-posedness, (ii) Existence of the limit of the primal solutions, (iii) Existence of the limit of the dual solutions, (iv) Equicontinuity of $(A^* q_k)_{k\in\mathbb{N}}$, (v) Feasibility of q_∞, (vi) Convergence to a solution. For completeness, we will present every step in its entirety, but let us still note that steps (i), (ii), (iv) and (vi) will be exactly the same as the proof in [17]. Step (iii) will differ slightly, due to the relaxed assumptions on the operator A and function f. (v) also differs, due to the modification of the algorithm, although the main idea is still the same.

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

(a) Refinement rule with second-order bounds.

Iteration	$\lvert\mathcal{V}_k\rvert$	primal	$\mathrm{dist}_{\mathcal{H}}(\mathcal{V}_k, X^\star)$
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

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

Iteration	$\lvert\mathcal{V}_k\rvert$	primal	$\mathrm{dist}_{\mathcal{H}}(\mathcal{V}_k, X^\star)$
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

Proof of Theorem 14. We will closely follow the proof idea of the corresponding Theorem 3.1 in

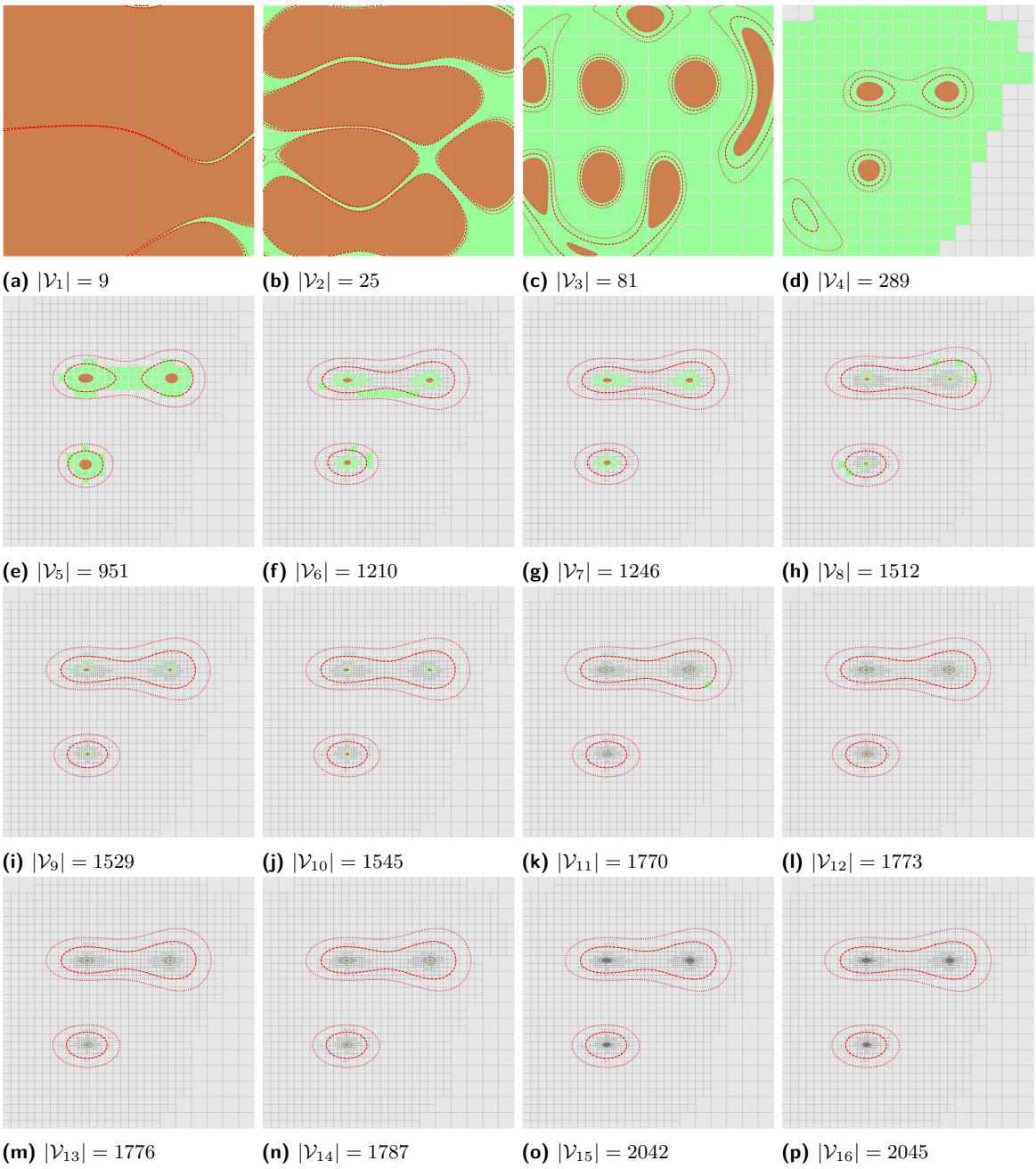
i. *Well-posedness.* Under Assumptions 1, 2, 3 and 6, we can apply Proposition 4 to ensure the existence of the primal-dual pair (μ_0, q_0). For the next iterates $k \geq 1$, the measure μ in Assumption 6 still satisfies $\mu \in \mathcal{M}(\mathcal{V}_k)$ by nestedness of the sequence $(\mathcal{V}_k)_{k\in\mathbb{N}}$. Hence we can apply Proposition 4 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 $C_k \overset{\text{def.}}{=} \left\{q \in \mathbb{R}^M, \|A^*q\|_{L^\infty(\mathcal{V}_k)} \leq 1\right\}$. 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 4). Moreover, the sequence $(f^*(q_k))_{k\in\mathbb{N}}$ is nondecreasing, bounded above by $f^*(q^\star)$. Therefore all the vectors q_k belong to the level set $\{q \in \mathbb{R}^M, f^*(q) \leq f^*(q^\star)\}$, 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 $a_m \in \mathcal{C}(\Omega)$ all are uniformly continuous, there exists a $\delta > 0$ with the property

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



(a) $|\mathcal{V}_1| = 9$ **(b)** $|\mathcal{V}_2| = 25$ **(c)** $|\mathcal{V}_3| = 81$ **(d)** $|\mathcal{V}_4| = 289$

(e) $|\mathcal{V}_5| = 951$ **(f)** $|\mathcal{V}_6| = 1210$ **(g)** $|\mathcal{V}_7| = 1246$ **(h)** $|\mathcal{V}_8| = 1512$

(i) $|\mathcal{V}_9| = 1529$ **(j)** $|\mathcal{V}_{10}| = 1545$ **(k)** $|\mathcal{V}_{11}| = 1770$ **(l)** $|\mathcal{V}_{12}| = 1773$

(m) $|\mathcal{V}_{13}| = 1776$ **(n)** $|\mathcal{V}_{14}| = 1787$ **(o)** $|\mathcal{V}_{15}| = 2042$ **(p)** $|\mathcal{V}_{16}| = 2045$

Figure 4 Our algorithm's behavior on a 2D example for second-order selection rules. The set $\Omega_k^\star$ is displayed in green, the superlevel set 1 of $|A^*q_k|$ is filled with brown, the level set 0.9 is represented with a dashed red line and the level set 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 $X^\star$ 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.

Consequently,

$$\|x-y\|_2 < \delta \;\Rightarrow\; |(A^*q_k)(x) - (A^*q_k)(y)| = \left|\sum_{m=1}^{M} (a_m(x) - a_m(y))q_k(m)\right| \leq \sum_{m=1}^{M} |a_m(x) - a_m(y)|\,|q_k(m)|$$
$$< \frac{\epsilon}{\sup_k \|q_k\|_1} \sum_{m=1}^{M} |q_k(m)| \leq \epsilon. \tag{17}$$

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 $\|A^*q_\infty\|_{L^\infty(\Omega)} = 1 + 2\epsilon$

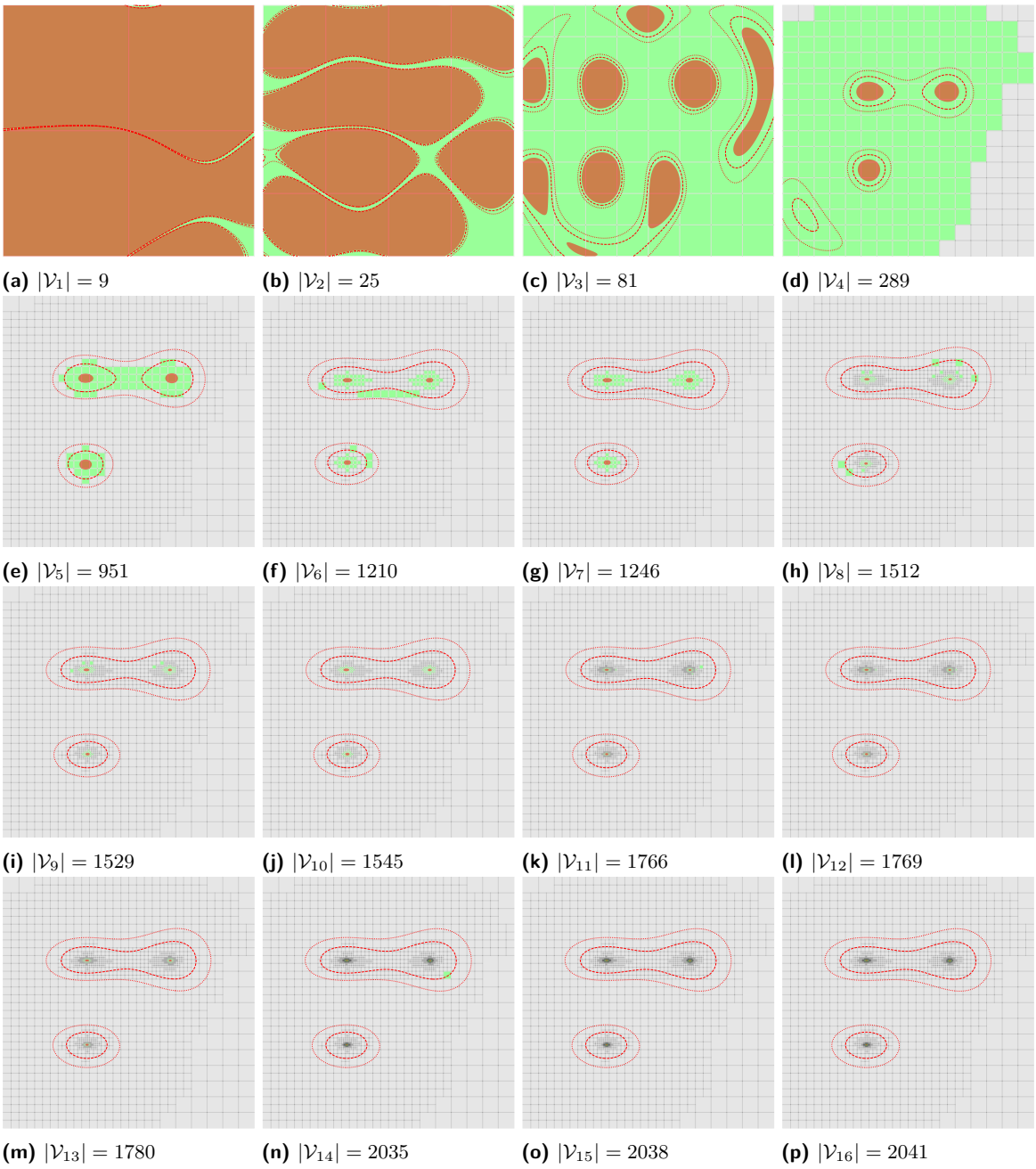


Figure 5 The adaptive refinement algorithm's behavior on a 2D example for second-order selection rules with first order gradient. The behavior quadtrees are very similar to those in Figure 4.

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 (17). 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 $\Omega_k^\star$. It must further satisfy $\operatorname{diam}(\omega) \geq 2\delta$. If not, all points in ω have a distance to $\mathcal{V}_k$ smaller than δ. Since we have $|A^*q_k|(x) \leq 1$ for $x \in \operatorname{vert}(\omega)$, the equicontinuity of the A^*q_k implies 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 $\omega \in \Omega_k^\star$ such that $\operatorname{diam}(\omega) \geq 2\delta$. 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 $\operatorname{diam}(\omega_k) \geq 2\delta$ 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) that 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 $\|\mu_\infty\|_{\mathcal{M}(\Omega)} + f(A\mu_\infty) \geq -f^*(q_\infty)$ by weak duality. To prove the second inequality, first notice

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

(a) Refinement rule with second-order bounds.

Iteration	$\lvert\mathcal{V}_k\rvert$	primal	$\mathrm{dist}_{\mathcal{H}}(\mathcal{V}_k, X^\star)$
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

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

Iteration	$\lvert\mathcal{V}_k\rvert$	primal	$\mathrm{dist}_{\mathcal{H}}(\mathcal{V}_k, X^\star)$
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

that the weak-$*$-continuity of A implies that $A\mu_k \to A\mu_\infty$. Assumption 2 furthermore implies that f is lower semi-continuous. As a supremum of linear functions, so is f^*. Since also $q_k \to q_\infty$, we conclude

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

Assumptions 2, 1 together with Proposition 4 imply exact duality of the discretized problems. This means $f^*(q_k) + f(A\mu_k) = -\|\mu_k\|_{\mathcal{M}(\Omega)}$. Since the norm is weak-$*$-l.s.c., we thus obtain

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

Reshuffling these inequalities yields $\|\mu_\infty\|_{\mathcal{M}(\Omega)} + f(A\mu_\infty) \leq -f^*(q_\infty)$, 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. ◀

A.2 Proof of Theorem (16)

In this section, we prove the main theoretical result of the paper, which is Theorem 16. This is a counterpart of [17, Theorem 3.11].

A.2.1 A few important estimates

The modification of the algorithm will cause the argument to be quite different compared to the one given in [17]. However, it will still rest on a number of technical inequalities from said paper. We can carry these over without thought, since the assumptions on f and A under which we prove 16 are exactly the same as the ones under which [17, Theorem 3.11] was proven.

▶ **Proposition 21.** *The following inequalities hold under Assumption 7:*

$$\|q_k - q^\star\|_2 \lesssim \mathrm{dist}_{\mathcal{H}}(\mathcal{V}_k | X_k) \tag{18}$$

$$\|q_k - q^\star\|_2^2 \lesssim \max\left(\mathrm{dist}_{\mathcal{H}}(X_k | X^\star), \mathrm{dist}_{\mathcal{H}}(\mathcal{V}_k | X^\star)\right) \cdot \mathrm{dist}_{\mathcal{H}}(\mathcal{V}_k | X^\star). \tag{19}$$

Proof. These are direct reformulations of inequalities from [17]: The first inequality (18) comes from Lemma 3.5, the second (19) from Lemma 3.6. ◀

▶ **Proposition 22** (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:*

$$\operatorname{dist}_{\mathcal{H}}(X^\star|X_k) \lesssim \|q_k - q^\star\|_2 \tag{20}$$
$$\operatorname{dist}_{\mathcal{H}}(X_k|X^\star) = \operatorname{dist}_{\mathcal{H}}(X^\star|X_k) \tag{21}$$
$$\operatorname{dist}_{\mathcal{H}}(X_k|X^\star) \lesssim \operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k) \tag{22}$$
$$\operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k) \asymp \operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k|X^\star) \tag{23}$$
$$\|q_k - q^\star\|_2 \lesssim \operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k|X^\star) \tag{24}$$
$$f(\mu_k) - f(\mu^\star) \lesssim \operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k|X^\star)^2. \tag{25}$$

Proof. The first inequalities (21) and (20) are simple consequences of Proposition 3.7 in [17], together with the fact that, by the generic convergence result, q_k converges to $q^\star$.

Inequality (22) is a combination of (18), (21) and (20).

To prove inequality (23), let us start by proving that $\operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k) \gtrsim \operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k|X^\star)$. We have by the triangular inequality

$$\operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k|X^\star) \leq \operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k) + \operatorname{dist}_{\mathcal{H}}(X_k|X^\star) \overset{(22)}{\lesssim} \operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k).$$

Let us prove the converse inequality $\operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k) \lesssim \operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k|X^\star)$. To this end, first combine (19), (21) and (20) to get

$$\operatorname{dist}_{\mathcal{H}}(X^\star|X_k)^2 \lesssim \max\left(\operatorname{dist}_{\mathcal{H}}(X^\star|X_k), \operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k|X^\star)\right) \cdot \operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k|X^\star). \tag{26}$$

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

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

The inequality (23) together with (23) now implies (24). Since the inequality (25) is a direct consequence of Proposition 3.12 in [17], we can conclude the proof. ◀

The next proposition says that under our regularity conditions, the dual certificates A^*q_k will be well-behaved for late k. The argumentation is similar to [17, Proposition 3.7], but phrased in a language more suitable for the continued argumentation here. To phrase it conveniently, let us introduce the following shorthand notation to design a neighborhood of $X^\star$ of width $r > 0$:

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

where $B(x, r) = \{y \,|\, \|y - x\| < r\}$ is the ball of radius r around x.

▶ **Proposition 23** (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 $B(x_s^\star, R)$ 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:*

$$|A^*q_k|''(x) \preccurlyeq -\frac{\gamma}{2}\,\mathrm{Id}, \ \forall\, x \in \mathcal{B}_R. \tag{27}$$

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

$$|A^*q_k|(x) \leq 1 - \frac{\gamma R^2}{4}, \ \forall\, x \in \Omega \setminus \mathcal{B}_R. \tag{28}$$

iv. *Finally*

$$\sup_{x \in \Omega} |A^*q_k|(x) \leq 1 + c_2 \operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k)^2. \tag{29}$$

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

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

The conclusion of the three first points follows from the nondegeneracy of $A^*q^\star$ in Assumption 8. To obtain the last, let ω_s denote the cell containing the point $x_{k,s}$ in X_k closest to $x_s^\star$. 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 $|A^*q_k|(v_s) \le 1$, $|A^*q_k|'(x_{k,s}) = 0$. Moreover $(|A^*q_k|''(\xi))_{k\in\mathbb{N}}$ is uniformly bounded. This yields for all s:

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

Taking the maximum over the different $1 \le s \le S$ gives the result (29). ◀

A.2.2 Analyzing the modified algorithm

Having collected all the useful inequalities that we need from [17], we can start analyzing the behaviour of the algorithm we are interested in here. Let us begin by translating Proposition 23 into a bound of $|A^*q_k|$ on each cell ω for late iterations.

▶ **Proposition 24** (Finite time behavior of the upper-bound). *Under Assumptions 7 and 8, there exists $k_2 \ge k_1$ and some positive constants c_1, c_2, c_3 such that for all $k \ge k_2$ and for all cell ω:*

$$\sup_{x\in\omega}|A^*q_k|(x) \le \begin{cases} 1 - c_1\,\mathrm{dist}(\omega, X_k)^2 + c_2\,\mathrm{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k)^2 & \text{if } \mathrm{dist}(\omega, X^\star) \le R, \\ 1 - c_3R^2 & \text{if } \mathrm{dist}(\omega, X^\star) \ge R. \end{cases} \tag{30}$$

Proof. Take a cell ω with $\mathrm{dist}(\omega, X^\star) \ge R$. For $k \ge k_1$, the upper-bound (28) is valid. Hence, we obtain the second bound in inequality (30) for all $k \ge k_1$ and $c_3 = \frac{\gamma}{4}$.

To obtain the first inequality, consider a cell ω with $\mathrm{dist}(\omega, X^\star) \le R$. Let $s \in [\![1, S]\!]$ denote any index such that $B(x_s^\star, R) \cap \omega \ne \emptyset$. Point (i) in Proposition 23 implies the existence of a unique point $x_{k,s}$ in $X_k \cap B(x_s^\star, R)$. Proposition 23, point (iv) implies that $|A^*q_k(x_{k,s})| \le 1 + c_2\,\mathrm{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k)^2$. Moreover, point (ii) in Proposition 23 states that $|A^*q_k|$ is strongly concave in the balls $B(x_s^\star, R)$. Therefore:

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

Using the above inequality and point (iii) in Proposition 23 gives:

$$\begin{aligned} &\sup_{x\in\omega\cap\mathcal{B}_R}|A^*q_k|(x) \le 1 + c_2\,\mathrm{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k)^2 - c_1\,\mathrm{dist}(\omega, X_k)^2, \\ &\sup_{x\in\omega\cap\mathcal{B}_R^c}|A^*q_k|(x) \le 1 - \frac{\gamma R^2}{4}. \end{aligned} \tag{31}$$

We now need to show that (31) actually implies

$$|A^*q_k|(x) \le 1 + c_2\,\mathrm{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k)^2 - c_1'\,\mathrm{dist}(\omega, X_k)^2.$$

for $x \in \omega \cap B_R^c$. To this end, first notice that point i) in Proposition 23 shows that $\mathrm{dist}_{\mathcal{H}}(X_k|X^\star) \le R$. Therefore, if $\mathrm{dist}(\omega, X^\star) \le R$, we get

$$\mathrm{dist}(\omega, X_k) \overset{(6)}{\le} \mathrm{dist}(\omega, X^\star) + \mathrm{dist}_{\mathcal{H}}(X_k|X^\star) \le 2R.$$

Hence, we get $\mathrm{dist}(\omega, X_k)^2 \lesssim R^2$, and by (31) we get

$$\sup_{x\in\omega\cap\mathcal{B}_R^c}|A^*q_k|(x) \le 1 - \frac{\gamma R^2}{4} \le 1 - c_1'\,\mathrm{dist}(\omega, X_k)^2$$

for some other constant c_1'. In particular, we get

$$\sup_{x\in\omega}|A^*q_k|(x) \le 1 + c_2\,\mathrm{dist}_{\mathcal{H}}(\mathcal{V}_k|X_k)^2 - c_1'\,\mathrm{dist}(\omega, X_k)^2.$$ ◀

Now we can prove a crucial proposition about the structure of the partitions Ω_k.

▶ **Proposition 25** (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. $\operatorname{dist}(\omega, X^\star) \geq R \Rightarrow |\omega| \geq r$.

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

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

Proof of Proposition 25. Let k_2 be the number of iterations referenced in Proposition 24. 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 $\operatorname{dist}(\omega, X^\star) \geq R$. 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 24 states that

$$\|A^* q_k\|_{L^\infty(\omega)} \leq 1 - c_3 R^2.$$

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 $|\omega| \overset{\text{def.}}{=} \ell$ 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 $\Omega^\star_{k_\ell}$, it means that every cell in $\Omega^\star_{k_\ell}$ has an edge-length smaller or equal than 2ℓ. By construction, every cell that contains an element of X_k is in $\Omega^\star_{k_\ell}$, hence

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

4. By Proposition 22, equation (23), we have $\operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k | X_k) \asymp \operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k | X^\star)$. We get $\operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k | X^\star) \lesssim \ell$ for $k = k_\ell$. By monotonicity of the sequence $(\operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k | X^\star))_{k \in \mathbb{N}}$, we also get the inequality for all $k \geq k_\ell$.
5. Noticing that $\operatorname{dist}(\omega_p, X^*) \leq \operatorname{dist}(\omega, X^*) \leq R$, we apply Proposition 24 to ω_p to get

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

   Since ω_p is refined at iteration k_ℓ, it belongs to $\Omega^\star_{k_\ell}$ 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 $\operatorname{dist}(\omega_p, X_k) \lesssim \ell$.
6. To conclude, remark that

   $$\begin{aligned} \ell &\gtrsim \operatorname{dist}(\omega_p, X_k) \\ &\overset{(6)}{\geq} \operatorname{dist}(\omega, X_k) - \operatorname{dist}_{\mathcal{H}}(\omega | \omega_p) \geq \operatorname{dist}(\omega, X_k) - \sqrt{D}\ell \\ &\overset{(6)}{\geq} \operatorname{dist}(\omega, X^\star) - \operatorname{dist}_{\mathcal{H}}(X_k | X^\star) - \sqrt{D}\ell \geq \operatorname{dist}(\omega, X^\star) - c_5 \ell, \end{aligned}$$

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

$$\min_{\omega\in\Omega_k} |\omega| = 2^{-K_{\text{end}}}$$

for some $K_{\text{end}} \in \mathbb{N}$ with $K_{\text{end}} \geq k_2$ in Proposition 25. Then the number of cells in Ω_k satisfies:

$$|\Omega_k| \leq c_0 + c_1 S K_{\text{end}}$$

for some constants $c_0, c_1 > 0$ independent of K_{end} and S.

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

$$\Omega_{k,0} \overset{\text{def.}}{=} \{\omega \in \Omega_k, \operatorname{dist}(\omega, X^\star) > R\},$$

$$\Omega_{k,s} \overset{\text{def.}}{=} \{\omega \in \Omega_k, \operatorname{dist}(\omega, x_s^\star) = \operatorname{dist}(\omega, X^\star), \operatorname{dist}(\omega, X^\star) \leq R\} \quad s \in [\![1, S]\!].$$

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

$$|\Omega_k| \leq \sum_{s=0}^{S} |\Omega_{k,s}|. \tag{32}$$

We first use point (i) in Proposition 25 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

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

Now let us derive a bound for $|\Omega_{k,s}|$. Let $\omega(x)$ denote the cell in Ω_k containing x and $\ell \overset{\text{def.}}{=} 2^{-K_{\text{end}}}$. 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 25, point ii) $|\omega(x)| \geq c \operatorname{dist}(\omega(x), x_s^\star)$. Therefore

$$|\omega(x)| \gtrsim \operatorname{dist}(\omega(x), x_s^\star) \overset{(6)}{\geq} \operatorname{dist}(x, x_s^\star) - \operatorname{dist}_{\mathcal{H}}(x|\omega(x)) \geq \|x - x_s^\star\|_2 - \sqrt{D}|\omega(x)|.$$

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

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

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

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

Summing up everything, we obtain $|\Omega_k| \lesssim c_0 + c_1 S K_{\text{end}}$ for some constants $c_0, c_1 \geq 0$. ◀

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

Proof of Theorem 16. Take $K_{\text{end}} \geq k_2$. The Algorithm terminates whenever a cell of size $2^{-(K_{\text{end}}+1)}$ has to be refined. When it stops, all the cells therefore have a size larger than $2^{-(K_{\text{end}}+1)}$ by construction and $\min_{\omega\in\Omega_k} |\omega| = 2^{-(K_{\text{end}}+1)}$.

Proposition 26 therefore indicates that $|\Omega_k| \leq c_0 + c_1 S K_{\text{end}}$. Since at least one cell is refined per iteration, we reached the termination criterion for a number of iterations $k \leq c_0 + c_1 S K_{\text{end}}$. Point (iii) in Proposition 25 allows us to conclude that $\operatorname{dist}_{\mathcal{H}}(\mathcal{V}_k|X^\star) \lesssim 2^{-K_{\text{end}}}$. The list of inequalities in Proposition 22 yield the conclusion. ◀

A.3 Further proofs

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

A.3.1 Proof of Proposition 3

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

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

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

A.3.2 Proof of Proposition 4

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 $\operatorname{int}(\operatorname{dom}(f))$. In particular, f is continuous at $A\mu$. Using Proposition 1.3.9 in [22], we conclude that

$$g^*(q) \overset{\text{def.}}{=} f^*(q) - \langle A\mu, q \rangle \tag{33}$$

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. ◀

A.3.3 Proof of Proposition 7

Proof. Any choice of $\kappa_1(q_k, \omega)$ in (8) satisfying the Lipschitz inequality

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

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

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

where we consider that $|A^* q_k|'(x) = 0$ on the points of non differentiability of $|A^* q_k|$. To obtain the expression (9), we use a Hölder inequality:

$$\sup_{x \in \omega} |A^* q_k|'(x) = \sup_{x \in \omega} \left| \sum_{m=1}^{M} q_k[m] a_m'(x) \right| \leq \sum_{m=1}^{M} |q_k[m]| \sup_{x \in \omega} |a_m'|(x). \tag{34}$$

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. ◀

A.3.4 Proofs of Proposition 9 and Proposition 12

Our proofs rely on the following well-known (see e.g. [26, Lemmas 1.2.3, 1.2.4]) statements about Taylor-expansions. If $f : C \to \mathbb{R}$ 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

$$|f(x) - f(y) - \langle f'(y), x - y\rangle| \leq \kappa_2 \frac{\|x-y\|_2^2}{2} \tag{35}$$

$$\|f'(x) - f'(y)\|_2 \leq \kappa_2 \|x - y\|_2. \tag{36}$$

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

$$\sup_{x\in\omega} \|(A^*q_k)''(x)\|_{2\to 2} = \sup_{x\in\omega} \left\| \sum_{m=1}^{M} q_k[m] a_m''(x) \right\|_{2\to 2} \leq \sum_{m=1}^{M} |q_k[m]| \sup_{x\in\omega} \|a_m''\|_{2\to 2}(x). \tag{37}$$

A.3.4.1 Proof of Proposition 9

Proof. Now, let us prove that

$$\overline{|A^*q_k|}(\omega) - \kappa_2(q_k,\omega)\operatorname{diam}(\omega)^2 \leq \|A^*q_k\|_{L^\infty(\omega)} \leq \overline{|A^*q_k|}(\omega). \tag{38}$$

By equation (35), 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 (38). We also have

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

Again, taking the supremum in x and then the infimum over the vertices v we obtain the left-hand side of (38). The right hand-side of (38) proves that the second order selection process satisfies Assumption 4, whereas the left hand-side proves that it obeys Assumption 5. ◀

A.3.4.2 Proof of Proposition 12

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

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

which implies

$$\begin{aligned} &\|(A^*q_k)'(x)\| \geq \|(A^*q_k)'(v)\|_2 - \kappa_2(q_k,\omega)\|x-v\|_2 \\ &\qquad \Longrightarrow \inf_{x\in\omega} \|(A^*q_k)'(x)\| \geq \|(A^*q_k)'(v)\|_2 - \kappa_2(q_k,\omega)\operatorname{diam}(\omega). \end{aligned}$$

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

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

In other words, $\underline{\|\nabla A^*q_k\|_2}$ is a lower bound of $\|(A^*q_k)'\|$. Since κ_2 is an upper bound of the Lipschitz constant of $\|(A^*)'q_k\|$, $\underline{\|\nabla A^*q_k\|_2}(\omega)$ 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 $\underline{\|\nabla A^*q_k\|_2}(\omega) \leq 0$. Hence $\Omega_k^\star$ verifies Assumption 4. It is clear that the $\Omega_k^\star$ of Definition 11 is included in the $\Omega_k^\star$ of Definition 8. Because the latter verifies Assumption 5, so does the former. ◀

Table 3 Assumptions related to the general convergence Theorem 14.

Type of assumption	Here	[17]
Regularity of f	f convex lower semi-continuous with $\text{int}(\text{dom}(f)) \neq \emptyset$	f convex with either $\text{dom}(f) = \mathbb{R}^M$ or f polyhedral
Regularity of $J/\mathcal{V}_0$	J coercive and there exists a $\mu \in \mathcal{M}(\mathcal{V}_0)$ with $A\mu \in \text{int}(\text{dom}(f))$.	f lower bounded and *either* f differentiable with Lipschitz gradient *or* A restricted to $\mathcal{M}(\mathcal{V}_0)$ is surjective.
Regularity of A	a_m continuous	a_m continuous
Refinement procedure	All cells containing local maximizers exceeding 1 are candidates.	Global maximizer $x_k^\star$ of $\lvert A^* q_k \rvert$ added to $\mathcal{V}_k$.

Table 4 Assumptions related to the linear convergence Theorem 16.

Type of assumption	Here	[17]
Regularity of A	The a_m are $\mathcal{C}^2$	The a_m are $\mathcal{C}^2$.
Source condition	Assumption 8	Assumption 8
Regularity of f	f differentiable with Lipschitz gradient	f differentiable with Lipschitz gradient
Refinement procedure	Candidate cells ω fulfill $\sup_{x\in\omega} \lvert A^* q_k(x) \rvert > 1 - \kappa \lvert \omega \rvert^2$	All local maximizers exceeding 1 added to $\mathcal{V}_k$.

A.3.5 Proof of Proposition 20

Proof. We have

$$a_m'(x) = -a_m(x)\left[\frac{x - z_m}{\sigma^2}\right]$$
$$a_m''(x) = a_m(x)\left[-\frac{\sigma^2}{\sigma^4}\,\text{Id} + \frac{(x-z_m)(x-z_m)^T}{\sigma^4}\right].$$

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

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

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

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

◀

B Collection of the assumptions needed to prove Theorems 14 and 16

Out of convenience for the reader, we collect all the assumptions made throughout the paper to prove our main results, and in particular compare them to the corresponding ones in [17]. The assumptions related to Theorem 14 are collected in Table 3, and the ones related to Theorem 16 are collected in Table 4.

References

1 Akshay Agrawal, Robin Verschueren, Steven Diamond, and Stephen Boyd. A rewriting system for convex optimization problems. *J. Control Decis.*, 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*. Society for Industrial and Applied Mathematics, 2014.

3 Johathan M. Borwein and Adrian S. Lewis. Partially finite convex programming, Part I: Quasi relative interiors and duality theory. *Math. Program.*, 57(1):15–48, 1992.

4 Claire Boyer, Antonin Chambolle, Yohann De Castro, Vincent Duval, Frédéric De Gournay, and Pierre Weiss. On Representer Theorems and Convex Regularization. *SIAM J. Optim.*, 29(2):1260–1281, 2019.

5 Kristian Bredies and Hanna Katriina Pikkarainen. Inverse problems in spaces of measures. *ESAIM, Control Optim. Calc. Var.*, 19(1):190–218, 2013.

6 Emmanuel J Candès and Carlos Fernandez-Granda. Towards a Mathematical Theory of Super-resolution. *Commun. Pure Appl. Math.*, 67(6):906–956, 2014.

7 Lenaic Chizat. Sparse optimization on measures with over-parameterized gradient descent. *Math. Program.*, 194(1):487–532, 2022.

8 Lenaic Chizat and Francis Bach. On the global convergence of gradient descent for over-parameterized models using optimal transport. *Adv. Neural Inf. Process. Syst.*, 31:3036–3046, 2018.

9 Christian Clason and Karl Kunisch. A measure space approach to optimal source placement. *Comput. Optim. Appl.*, 53(1):155–171, 2012.

10 Yohann De Castro and Fabrice Gamboa. Exact reconstruction using Beurling minimal extrapolation. *J. Math. Anal. Appl.*, 395(1):336–354, 2012.

11 Yohann De Castro, Fabrice Gamboa, Didier Henrion, and Jean-Bernard Lasserre. Exact solutions to Super Resolution on semi-algebraic domains in higher dimensions. *IEEE Trans. Inf. 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. *J. Comput. Appl. Math.*, 422: article no. 114937 (13 pages), 2023.

13 Quentin Denoyelle, Vincent Duval, Gabriel Peyré, and Emmanuel Soubies. The sliding Frank–Wolfe algorithm and its application to super-resolution microscopy. *Inverse Probl.*, 36(1): article no. 014001 (42 pages), 2019.

14 Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *J. Mach. Learn. Res.*, 17(83):1–5, 2016.

15 Vincent Duval and Gabriel Peyré. Exact support recovery for sparse spikes deconvolution. *Found. Comput. Math.*, 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 J. Optim.*, 29(2):1329–1349, 2019.

17 Axel Flinth, Frédéric De Gournay, and Pierre Weiss. On the linear convergence rates of exchange and continuous methods for total variation minimization. *Math. Program.*, 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, 2018.

19 Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Nav. Res. Logist. Q.*, 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 Rev.*, 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, 2013.

23 Karl Kunisch, Philip Trautmann, and Boris Vexler. Optimal control of the undamped linear wave equation with measure valued controls. *SIAM J. Control Optim.*, 54(3):1212–1244, 2016.

24 Jean Bernard Lasserre. *An introduction to polynomial and semi-algebraic optimization*, volume 52 of *Cambridge Texts in Applied Mathematics*. Cambridge University Press, 2015.

25 Eugene L Lawler and David E Wood. Branch-and-bound methods: A survey. *Oper. Res.*, 14(4):699–719, 1966.

26 Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87 of *Applied Optimization*. Kluwer Academic Publishers, 2004.

27 Konstantin Pieper and Daniel Walter. Linear convergence of accelerated conditional gradient algorithms in spaces of measures. *ESAIM, Control Optim. Calc. Var.*, 27: article no. 38 (37 pages), 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, 2019.

29 Rembert Reemtsen and Stephan Görner. Numerical methods for semi-infinite programming: A survey. In *Semi-Infinite Programming*, pages 195–262. Springer, 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 *2013 Asilomar Conference on Signals, Systems and Computers*, pages 1043–1047. IEEE, 2013.

32 Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht. Compressed sensing off the grid. *IEEE Trans. Inf. 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 Probl.*, 36(4): article no. 045003 (27 pages), 2020.

34 Joel A Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inf. 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 Rev.*, 59(4):769–793, 2017.

36 Yinyu Ye and Edison Tse. An extension of Karmarkar's projective algorithm for convex quadratic programming. *Math. Program.*, 44:157–179, 1989.

37 Alp Yurtsever, Joel A Tropp, Olivier Fercoq, Madeleine Udell, and Volkan Cevher. Scalable semidefinite programming. *SIAM J. Math. Data Sci.*, 3(1):171–200, 2021.