A Low-Rank Approach to Off-The-Grid Sparse Deconvolution

Paul Catala⋆, Vincent Duval†§, and Gabriel Peyré⋆

⋆ CNRS, DMA, Ecole Normale Supérieure, PSL, CNRS, UMR 8553
† Inria Paris, MOKAPLAN
§ CEREMADE, Université Paris-Dauphine, PSL

Abstract

We propose a new solver for the sparse spikes deconvolution problem over the space of Radon measures. A common approach to off-the-grid deconvolution considers semidefinite (SDP) relaxations of the total variation (the total mass of the absolute value of the measure) minimization problem. The direct resolution of this SDP is however intractable for large scale settings, since the problem size grows as \( f^2 c \) where \( f_c \) is the cutoff frequency of the filter and \( d \) the ambient dimension. Our first contribution introduces a penalized formulation of this semidefinite lifting, which has low-rank solutions. Our second contribution is a conditional gradient optimization scheme with non-convex updates. This algorithm leverages both the low-rank and the convolutive structure of the problem, resulting in an \( O(f^d c \log f_c) \) complexity per iteration. Numerical simulations are promising and show that the algorithm converges in exactly \( r \) steps, \( r \) being the number of Diracs composing the solution.

1 Introduction

1.1 Sparse spikes deconvolution

Sparse super-resolution problems consist in recovering pointwise sources from low-resolution and possibly noisy measurements. Such issues arise naturally in fields like astronomical imaging [28], fluorescence microscopy [19, 30] or seismic imaging [21], where it may be crucial to correct the physical blur introduced by sensing devices, due to diffraction or photonic noise for instance.

Formally, let \( \delta_x \) be the Dirac measure at position \( x \in \mathbb{T}^d \), where \( \mathbb{T} = \mathbb{R}/\mathbb{Z} \) is the torus. We aim at retrieving a \( d \)-dimensional discrete Radon measure \( \mu_0 = \sum_{k=1}^r a_{0,k} \delta_{x_k} \) (with \( a_0 \in \mathbb{C}^r \), \( x_k \in \mathbb{T}^d \)), given the measurements

\[
y = \Phi \mu_0 + w = y_0 + w
\]

where \( \Phi \) is some known convolution operator, and \( w \) some unknown noise.
1.2 Beurling LASSO

Although this problem is severely ill-posed, sparse estimates can be found by solving the following optimization program, known as the BLASSO [6]

\[
\arg\min_{\mu \in \mathcal{M}(T^d)} \frac{1}{2\lambda} \| \Phi \mu - y \|^2 + |\mu|(T^d)
\]

(Pλ)

where the total variation norm (defined below) naturally extends the ℓ1-norm of finite-dimensional vectors to the continuous setting of measures, favoring in particular the emergence of spikes in the solution. This problem might have non-unique solutions.

The main asset here is that no assumption is made about the support of the measure, contrarily to early superresolution models where the latter was assumed to live on some discrete grid [9, 10]. This grid-free approach has been at the core of several recent works [2, 4, 7], and has offered beneficial mathematical insight on the problem, leading to a better understanding of the impact of minimal separation distance [33] or of the signs of the spikes [4, 6], lacking in first superresolution works, and to sharp criteria for stable spikes recovery [8, 13].

However, the infinite-dimensionality of (Pλ) poses a major numerical challenge, and off-the-grid superresolution, although analytically better, remains far more difficult to perform in practice. This work introduces a new method to numerically solve this optimization problem in a scalable way.

1.3 Related works

Conditional gradient approach In the case of a generic operator Φ, [2] propose to iteratively estimate the support of the initial measure using a conditional gradient (also known as Frank-Wolfe) scheme, which basically consists in greedy stepwise addition of new spikes. To remedy to the low convergence rate of the method, non-convex corrective updates are added as a final step [2, 1]. This method achieves state-of-the-art result in several applications, among which fluorescence microscopy [1].

Semidefinite approaches Another important class of methods consist in introducing convex semidefinite relaxations of (Pλ), which enables the use of tools from semidefinite programming. SDP approaches for the total variation minimization problem were originally introduced in [4, 34], for unidimensional measures (i.e., d = 1) only however. The multivariate case on the other hand raises a far more challenging theoretical problem, which is solved using the so-called Lasserre’s hierarchy [24], consisting in a sequence of increasingly better semidefinite approximations of (Pλ) or its dual [7, 11].

The numerical interest of semidefinite programs holds in that they are well-known, and benefit from efficient solvers; however, these become quickly intractable for large matrices. This is a major obstacle for the method, since the size of the matrices involved in the semidefinite relaxation of (Pλ) are typically of size \( f_c^d \times f_c^d \), where \( f_c \) is the cutoff frequency of the low-pass filter Φ. Up to now, semidefinite relaxations have essentially been successful for combinatorial problems like the MaxCut problem [16], in which the number of variables is high and the degree of the polynomials involved in the constraints is low (it is equal to 2 in the MaxCut case). Imaging problems on the contrary typically involve polynomials with low numbers of variables (essentially the same as the dimension d), but high degrees (\( f_c \)), and to our knowledge, no efficient way of solving the hierarchy in those settings has been proposed yet.
Non variational approaches Although we focus here on $\ell^1$-regularization techniques, there is also a vast literature on non-convex or non-variational superresolution schemes. Many of these methods derive from the same idea proposed by Prony to encode the positions of the spikes as zeros of some polynomial, see [22] for a review. This is the case for MUSIC [32], ESPRIT [29], or finite rate of innovation [37], among others. They remain difficult to extend to multivariate settings, see for instance [23].

1.4 Contributions

We propose a novel algorithm for solving the BLASSO, able to cope with large problem size – i.e. with large cutoff frequency $f_c$. We target cases where $\Phi$ is a convolution operator.

Our first contribution introduces a semidefinite lifting of $(P_\lambda)$, which admits solutions whose rank is bounded by the number of spikes in the solution of $(P_\lambda)$. We prove this result for $d = 1$, and give numerical evidences that it also holds when $d > 1$.

Our second and main contribution is a conditional gradient algorithm for this lifting, which requires only $O(f_c \log f_c)$ elementary computations per iterations, making it scalable for superresolution problems in dimension greater than one. Numerical simulations suggest a $r$-steps convergence, $r$ being the number of spikes in the solution of the BLASSO.

Roadmap We introduce our semidefinite approximation in Section 3 and 4. The low-rank property is proved in Section 3. Sensitivity analysis is exposed in Section 4. Our algorithm is detailed in Section 5. Finally, Section 6 presents our numerical results.

1.5 Notations

Spaces We consider measures defined over the torus $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$. Note that in practice, in order to avoid periodization artifacts, one can use zero-padding or symmetrization at the boundary of the signal and still consider a periodic convolution setting. We denote by $\mathcal{M}(\mathbb{T}^d)$ the space of bounded Radon measures on $\mathbb{T}^d$, endowed with its weak-* topology. It is the topological dual of the space $C(\mathbb{T}^d)$ of continuous functions on $\mathbb{T}^d$.

Linear operator We focus on the case where the measurements $y$ consist in a low-pass filtering of the initial measure. In our periodic framework, measures are uniquely defined by their Fourier moments, and convolution with some kernel boils down to the product of these moments with the Fourier coefficients of the kernel. For a filter of cutoff frequency $f_c$, the measurements are thus simply a finite collection of weighted trigonometric moments of the initial measure. We define the linear operator $\Phi : \mathcal{M}(\mathbb{T}^d) \rightarrow \mathbb{C}^{(2f_c+1)^d}$ as

$$\Phi \mu = \int_{\mathbb{T}^d} \varphi(x) d\mu(x), \quad \text{with} \quad \varphi(x) = \left[c_k e^{-2i\pi \langle k, x \rangle}\right]_{k \in [-f_c, f_c]^d},$$

where $f_c$ is the cutoff frequency of the filter, and $(c_k)$ its transfer function. Its adjoint $\Phi^* : \mathbb{C}^{(2f_c+1)^d} \rightarrow C(\mathbb{T}^d)$ is given by

$$\forall p \in \mathbb{C}^{(2f_c+1)^d}, \quad \forall x \in \mathbb{T}^d, \quad (\Phi^* p)(x) = (\varphi(x), p) = \sum_{k \in [-f_c, f_c]^d} p_k \overline{c_k} e^{2i\pi \langle k, x \rangle}.$$
In the following, we call \( n = 2f_c + 1 \) the number of observations in each dimension.

This model encompasses filters which have compact support in the Fourier domain. Note that without loss of generality, since one can still renormalize the observations by posing \( \tilde{y}_k = y_k/c_k \) for all \( k \), in the rest of the paper, we restrict our attention to the ideal low-pass filter case, i.e., \( c_k = 1 \) for all \( k \). \( \Phi \) is then simply the Fourier operator.

Note that while using a periodic boundary condition might seem an issue, provided that the filtering kernel has a bounded support in the Fourier domain, one can always assume this holds, at the price of possibly extending the domain size to avoid periodization issues (e.g. with padding).

**Trigonometric polynomials** We call multivariate Laurent polynomial of degree \( l \) a function of the form

\[
P(z) = \sum_{k \in [-f_c, f_c]^d} p_k z^k = \sum_{-l \leq k_1, \ldots, k_d \leq l} p_{k_1, \ldots, k_d} z_1^{k_1} \cdots z_d^{k_d},
\]

defined for \( z \in \mathbb{C}^d \). Note that the exponents may be negative. When \( p_k = \overline{p_{-k}} \), \( P \) is said to be Hermitian; in particular, it is real-valued on the unit circle, and satisfies \( P(z) = \overline{P(1/\overline{z})} \).

When \( \Phi \) is a low-pass filter, \( \Phi^* p \) is a trigonometric polynomial defined on \( \mathbb{T}^d \) such that

\[
\forall x \in \mathbb{T}^d, \quad (\Phi^* p)(x) = \sum_{k \in [-f_c, f_c]^d} p_k e^{2i\pi \langle k, x \rangle}.
\]

In the following, we call \( P_p \) the Laurent polynomial associated to \( \Phi^* p \), simply defined as \( P_p(z) = \sum_{k \in [-f_c, f_c]^d} p_k z^k \), for \( z \in \mathbb{C}^d \), so that \( P_p(e^{2i\pi x}) = \Phi^* p(x) \) for all \( x \in \mathbb{T}^d \).

**Remark 1.** In all the paper, we write \( \overline{z} \) to designate the conjugate value of \( z \), \( V^h \) the conjugate transpose of a vector or a matrix \( V \), and \( \Phi^* \) the hermitian adjoint of an operator \( \Phi \).

## 2 Preliminaries

### 2.1 Total variation norm

The total variation norm of a measure is defined as

\[
|\mu|(\mathbb{T}^d) \overset{\text{def}}{=} \sup \left\{ \Re \left( \int_{\mathbb{T}^d} \eta d\mu \right) \mid \eta \in \mathcal{C}(\mathbb{T}^d), \ |\eta|_\infty \leq 1 \right\}.
\]

It is the natural extension of the discrete \( \ell^1 \)-norm to the space of Radon measures: indeed, if for instance \( \mu \) is some discrete measure \( \mu = \mu_{a,x} \overset{\text{def}}{=} \sum a_k \delta_{x_k} \), then \( |\mu|(\mathbb{T}^d) = \|a\|_1 \). Similarly, if \( \mu \) admits a density \( f \) with respect to the Lebesgue measure, then \( |\mu|(\mathbb{T}^d) = \|f\|_1 \).

It can be proved that the subdifferential of the total variation norm reads

\[
\partial |\mu|(\mathbb{T}^d) = \left\{ \eta \in \mathcal{C}(\mathbb{T}^d) \mid |\eta|_\infty \leq 1, \ \Re \left( \int_{\mathbb{T}^d} \eta d\mu \right) = |\mu|(\mathbb{T}^d) \right\}.
\]

In particular, one has

\[
\partial |\mu_{a,x}|(\mathbb{T}^d) = \left\{ \eta \in \mathcal{C}(\mathbb{T}^d) \mid |\eta|_\infty \leq 1, \ \eta(x_k) = \frac{a_k}{|a_k|} \ k = 1, \ldots, r \right\}.
\]
2.2 Primal and Dual problems

**Dual problem** The dual of the BLASSO reads \([\mathcal{D}_\lambda]\)

\[
\arg\max_{p \in \mathbb{C}^{(2\lfloor n \rfloor + 1)d}} \Re \langle y, p \rangle - \frac{\lambda}{2} |p|^2 \quad \text{s.t.} \quad |\Phi^* p|_\infty \leq 1. \tag{D_\lambda}
\]

Here the computational challenge comes from the continuum of constraints, which asks for a trigonometric polynomial to be bounded by one in magnitude.

**Dual certificate** Strong duality holds between \([\mathcal{P}_\lambda]\) and \([\mathcal{D}_\lambda]\) \([2, \text{ Proposition } 2]\), and optimality relations between any pair of solutions \((\mu_\lambda, p_\lambda)\) read

\[
\begin{cases}
\Phi^* p_\lambda \in \partial |\mu_\lambda| (\mathbb{T}^d) \\
p_\lambda = \frac{1}{\lambda} (y - \Phi \mu_\lambda)
\end{cases} \quad \text{(Opt)}
\]

The trigonometric polynomial \(\eta_\lambda = \Phi^* p_\lambda\) is called the **dual certificate** \([3]\). The associated complex polynomial \(P_\lambda\) is Hermitian.

We now recall some important basic results about the primal and dual solutions, with brief proof so that the paper be self-contained.

**Proposition 1.** Let \(\Sigma [p] = \{x \in \mathbb{T}^d : |\Phi^* p(x)| = 1\}\). Then \(\text{Supp} \mu_\lambda \subset \Sigma [p_\lambda]\).

**Proof.** The polar decomposition lemma for complex measures ensures that there exists a measurable function \(\xi : \mathbb{T}^d \to \mathbb{C}\) such that \(|\xi| = 1\) and \(\mu = \xi |\mu|\). Let \(\eta_\lambda = \Phi^* p_\lambda\). Using (Opt) and (1), we get

\[
0 = |\mu_\lambda|(\mathbb{T}^d) - \Re \left( \int_{\mathbb{T}^d} \overline{\eta_\lambda} d\mu_\lambda \right) = \int_{\mathbb{T}^d} (1 - \Re (\overline{\eta_\lambda} \xi)) d|\mu_\lambda|.
\]

Since \(|\eta_\lambda|_\infty \leq 1\) and \(|\xi(x)| = 1\) for all \(x \in \mathbb{T}^d\), it follows that \(1 \geq \Re (\overline{\eta_\lambda} \xi)\). The equality above thus implies \(\Re (\overline{\eta_\lambda} \xi) = 1\) \(\mu_\lambda\)-a.e., which leads to \(|\eta_\lambda| = 1\) \(\mu_\lambda\)-a.e. since \(|\eta_\lambda|^2 = |\overline{\eta_\lambda} \xi|^2 \geq \Re (\overline{\eta_\lambda} \xi)^2\). \(\square\)

The dual polynomial, bounded by one in magnitude, thus interpolates the signs of the spikes on the support of \(\mu_\lambda\), see Fig. [4]. One can then recover the measure \(\mu_\lambda\) using the following result:

**Proposition 2.** Assume that \(\Sigma_\lambda = \Sigma [p_\lambda]\) is finite. Let \(\Phi_{\Sigma_\lambda}\) be the operator defined as \(\Phi_{\Sigma_\lambda} : a \in \mathbb{R}^{[\Sigma_\lambda]} \mapsto \sum_{s \in \Sigma_\lambda} a_s \Phi (\delta_s)\). If \(\Phi_{\Sigma_\lambda}\) is injective, \([\mathcal{P}_\lambda]\) has a unique solution \(\mu_\lambda\) given by

\[
\mu_\lambda = \sum_{s \in \Sigma_\lambda} a_s \delta_s, \quad \text{where} \quad a = (\Phi_{\Sigma_\lambda}^* \Phi_{\Sigma_\lambda})^{-1} \Phi_{\Sigma_\lambda}^* (y - \lambda p_\lambda).
\]

**Proof.** Let \(\mu_\lambda\) and \(\mu'_\lambda\) be solutions of \([\mathcal{P}_\lambda]\). Using Proposition 1, we may write \(\mu_\lambda = \sum_{s \in \Sigma_\lambda} a_s \delta_s\) and \(\mu'_\lambda = \sum_{s \in \Sigma_\lambda} a'_s \delta_s\). Thus

\[
\Phi_{\Sigma_\lambda} a = \Phi \mu_\lambda = y - \lambda p_\lambda = \Phi \mu'_\lambda = \Phi_{\Sigma_\lambda} a'
\]

which implies \(a = a'\), and therefore \(\mu_\lambda = \mu'_\lambda\). Moreover, the previous relation immediately yields

\[
a = (\Phi_{\Sigma_\lambda}^* \Phi_{\Sigma_\lambda})^{-1} \Phi_{\Sigma_\lambda}^* (y - \lambda p_\lambda).
\]

\(\square\)
Figure 1: Left: $d = 1$, $f_c = 13$, $r = 5$, $|w| = 0$, $\lambda = 5 \times 10^{-3}$. Right: $d = 2$, $f_c = 7$, $r = 5$, $|w| = 0$, $\lambda = 10^{-2}$. Dual polynomials; true positions of Diracs are shown in red (positive spikes) or cyan (negative spikes).

Support recovery In order to determine the support of a solution $\mu_\lambda$ of $(P_\lambda)$ from a dual certificate, one can therefore solve the polynomial equation $1 - |\eta_\lambda(x)|^2 = 0$.

**Proposition 3.** Suppose that the hypothesis of Proposition 2 hold. Let $\tilde{P}_\lambda = 1 - |P_{\mu_\lambda}|^2$. Then:

$$\text{Supp } \mu_\lambda \subset \frac{1}{2\pi} \arg \left\{ z \in \mathbb{C}^d ; \tilde{P}_\lambda(z) = 0 \text{ and } |z_i| = 1, \ i = 1, \ldots, d \right\}$$

**Proof.** From Proposition 2 we know that $\text{Supp } \mu_\lambda \subset \Sigma_\lambda$. Then we have

$$x \in \Sigma_\lambda \iff \tilde{P}_\lambda(e^{2i\pi x}) = 0 \iff x \in \frac{1}{2\pi} \arg \left\{ z \in \mathbb{C}^d ; \tilde{P}_\lambda(z) = 0 \text{ and } |z_i| = 1, \ i = 1, \ldots, d \right\}$$

which yields the result.

**Remark 2.** Should the inclusion above ever be strict, solving the linear system (2) which gives the amplitudes leads to null amplitudes for points located in $\Sigma_\lambda \setminus \text{Supp } \mu_\lambda$, provided $\Phi_{\Sigma_\lambda}$ injective.

**Remark 3.** Finding the support of $\mu_\lambda$ thus amounts to finding the roots of $\tilde{P}_\lambda$ that are located on the unit circle. For multivariate polynomials, this root-finding problem is very difficult. We discuss our numerical approach further on, see Section 6.

3 Semidefinite hierarchies

An important feature of $(P_\lambda)$ is that although it is defined over the space of Radon measures, only moments are actually involved. It would thus seem only natural to reduce it to a problem over sequences of moments. But this raises the difficult question of finding conditions to ensure that a sequence coincides with the moments of some measure (such sequence is said to admit a representing measure). This problem, known as the moment problem, has been extensively studied, and it is
well-known that such conditions exist for measures defined on semi-algebraic sets, i.e. sets described by polynomials inequalities.

The numerical assets of this theory have only recently been investigated by Lasserre, with Parrilo \cite{Parrilo2007} making similar contributions independently. His pioneering work \cite{Lasserre2001} introduces a generic method based on semidefinite programming, now known as the Lasserre’s hierarchy, for solving a large class of problems, referred to as generalized moment problems \cite{Lasserre2001} – of which the BLASSO is an instance. The core idea of this method is that for measures defined on a semi-algebraic set $K$, one can derive a hierarchy of SDP-representable necessary conditions for any sequence to admit a representing measure on $K$, such that at each order of the hierarchy, the size of the semidefinite program increases as the corresponding condition sharpens. In other words, one can approximate to arbitrary precision the cone of sequences admitting a representing measure by either larger semidefinite relaxations (thus semidefinite strenghtenings) or thinner ones (semidefinite relaxations). We refer to \cite{Lasserre2001} for a detailed presentation of Lasserre’s hierarchy.

Semidefinite hierarchies as studied by Lasserre are concerned with measures defined on real algebraic sets. The case of trigonometric algebraic sets (i.e. described by trigonometric polynomials inequalities), of greater interest to us, is studied by Dumitrescu \cite{Dumitrescu2013}. It can actually be cast into a real framework, since any trigonometric polynomial $R_n(\omega) = \sum_{k=1}^{n} c_{k} \prod_{i=1}^{d} (\cos \omega_i)^{k_i} (\sin \omega_i)^{l_i}$ can be transformed into a real polynomial $P_n$ of $2d$ variables by posing $t_i = \cos \omega_i$, $t_{i+d} = \sin \omega_i$, $k_{i+d} = l_i$, and $P_n(t) = \sum_{k} c_k t^{k}$. We thus obtain an equivalent real semialgebraic set by considering the set $\{ t \in \mathbb{R}^{2d} ; P_n(t) \geq 0, n = 1 \ldots N \} \cap \{ t \in \mathbb{R}^{2d} ; t_i^2 + t_{i+d}^2 = 1, i = 1 \ldots d \}$, which is indeed a real semialgebraic set (see \cite{Dumitrescu2013} for further details).

### 3.1 Moment encoding

The semidefinite relaxation at order $m$ ($m \geq n$, with $n = 2f_c + 1$) of (P) reads:

$$
\begin{align*}
\mathcal{R}^{(m)}_\lambda \in \arg \min_{z \in \mathbb{C}^{2m-1}, \tau \in \mathbb{R}, u \in \mathbb{C}^{(2m-1)d}} & \quad \tau + u_0 + \frac{1}{2} \| y \|_2^2 + 2z^2 \\
\text{s.t.} & \quad \mathcal{R} = \begin{bmatrix} R & \tilde{z} \\ \tilde{z}^h & \tau \end{bmatrix} \succeq 0, \\
& \quad \tilde{z}_k = z_k, \\
& \quad k \in [-f, f]^d, \\
& \quad (b) \quad R = \sum_{k=-m,m} u_k \Theta_k
\end{align*}
$$

where $\Theta_k = \theta_{k_1} \otimes \cdots \otimes \theta_{k_d} \in \mathcal{M}_{m \times m}(\mathbb{C})$. Here $\theta_{k_i}$ denotes the $m \times m$ Toeplitz matrix with ones on its $k_i$-th diagonal and zeros everywhere else, and $\otimes$ stands for the Kronecker product. For instance, with $2 \times 2$ matrices, for $d = 2$ and $k = (-1,0)$, one has

$$
\Theta_k = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.
$$

The constraints (a) and (b) above encode a necessary condition for the sequence $z$ to be the moments of some measure on $\mathbb{T}^d$. We often refer to $R$ as a moment matrix. This particular formulation, introducing this block-structured matrix $\mathcal{R}$, can be deduced from the bounded real lemma \cite{Dumitrescu2013}.
Remark 4. The dual point of view of this moment formulation consists in replacing the positivity constraint in \((D_\lambda)\) by a sufficient sum-of-squares condition, which is also SDP representable. Thus, to the above primal relaxation corresponds a dual strengthening, which reads\[ Q^{(m)}_\lambda \in \arg\max_{p \in \mathbb{C}^d} R(y, p) - \frac{\lambda}{2} \|p\|^2 \]

\[ \begin{align*}
Q & = \begin{bmatrix} Q & \hat{p} \\ \hat{p}^b & 1 \end{bmatrix} \succeq 0, \\
\hat{p}_k & = \begin{cases} p_k & \text{if } k \in [-f_c, f_c]^d \\ 0 & \text{otherwise} \end{cases}
\end{align*} \quad \text{for } (D^{(m)}_\lambda)\]

We refer to [11] for further details on this problem.

Collapsing In [25], it is shown that the optimal value of the semidefinite program at level \(m\) of the hierarchy converges to the one of the initial problem as \(m\) goes to infinity. However, in order to recover the correct support in our case, we obviously need the hierarchy to have finite convergence – in which case it is said to collapse. In dimension one, the semidefinite approximations of \((P_\lambda)\) and \((D_\lambda)\) are known to be tight for \(m \geq 2f_c + 1\). This results from Carathéodory-Toeplitz theorem [34] or Fejér-Riesz theorem [11]. In dimension \(d > 1\) however, there are no results on whether the relaxations converge finitely or not in general, except in the case \(d = 2\), for which the hierarchy is known to collapse, but with no a priori information on the order \(m\) at which it occurs, as mentioned in [7, Section 4]: more specifically, one may find instances where the order at which the hierarchy collapses is arbitrarily large. Still, although the theory remains rather pessimistic, numerical simulations tend to show that finite convergence almost always happens at low orders of relaxation.

Remark 5. To detect the collapse of the hierarchy, one may recover a measure \(\mu^{(m)}_\lambda\) from the polynomial \(\Phi^* p^{(m)}_\lambda\) (by root-finding), and check whether the conditions \((\text{Opt})\) are satisfied. If so, \(\mu^{(m)}_\lambda\) is a solution of \((P_\lambda)\) (and \(\Phi^* p^{(m)}_\lambda\) a dual certificate).

The collapse of the hierarchy can also be detected using a rank argument: if there exists an order \(m_0\) such that \(\text{rank}(R^{(m_0)}_\lambda) = \text{rank}(R^{(m_0+1)}_\lambda) = r\), then \((P^{(m_0)}_\lambda)\) and \((P_\lambda)\) have the same optimal value, and \((P_\lambda)\) admits a \(r\)-atomic solution that can be retrieved from the moment matrix \(R^{(m_0)}_\lambda\); we refer to [7, 6] for further details about this property.

From now on, we omit the subscript \((m)\) when referring to solutions of \((P^{(m)}_\lambda)\) or \((D^{(m)}_\lambda)\), assuming that we have reached finite convergence.

Dual certificate As before, we define the dual certificate as \(\Phi^* p_\lambda\), where \(p_\lambda\) is the solution of \((D^{(m)}_\lambda)\). One actually do not need to solve the dual problem to recover this vector, thanks to the following result:

Lemma 1. The vector \(p_\lambda\) of coefficients of the dual polynomial is given by

\[ p_\lambda = \frac{y}{\lambda} + 2z_\lambda, \]

where \(z_\lambda\) is a solution of \((P^{(m)}_\lambda)\).
Proof. This is a direct consequence of primal-dual relations at optimality.

The corresponding support can then be recovered via the aforementioned root-finding techniques, described in Section 6.

As it appears, the size of the above semidefinite program is $m^{2d}$, with $m \geq 2f_c + 1$. Therefore, usual interior points methods are limited when $d > 1$, or even when $d = 1$ for large values of $f_c$. In the rest of this paper, we introduce a method which scales well with the dimension $d$.

### 3.2 Low-rank structure

#### The unidimensional case

We now state a fundamental low-rank property of the moment matrix involved in $(P_{\lambda}^{(m)})$, which we then fully exploit in our algorithm, both in terms of spatial memory and computational efficiency.

**Proposition 4.** When $d = 1$, $(P_{\lambda}^{(m)})$ admits a solution $R_{\lambda}$ whose rank is bounded by the number of spikes $r$ in a solution $\mu_{\lambda}$ of $(P_{\lambda})$.

**Proof.** This proof is entirely based on the proof of Proposition 2.1 in [34]. Suppose a solution of $(P_{\lambda})$ is given by $\mu_{\lambda} = \sum_{k=1}^{r} a_k \delta_{x_k}$, with $a_k \in \mathbb{R}$. Let $e(x, \varphi) = \frac{e^{2i\pi j x + i\varphi}}{f_c \leq j \leq f_c}$. Then

$$\Phi \mu_0 = \left[ \sum_k a_k e^{2i\pi j x_k} \right]_{-f_c \leq j \leq f_c} = \sum_k |a_k| e^{j \varphi_k}.$$

Let $z = \sum_k |a_k| e(x_k, \varphi_k)$, $u = \sum_k |a_k| e(x_k, 0)$, $\tau = \sum_k |a_k|$ and consider the matrix

$$R = \begin{bmatrix} \text{Trace}(u) & z \\ z^* & \tau \end{bmatrix} = \sum_{k=1}^{r} |a_k| \begin{bmatrix} e(x_k, \varphi_k) \\ 1 \end{bmatrix} \begin{bmatrix} e(x_k, \varphi_k) \\ 1 \end{bmatrix}^*.$$

Then

$$f(R) = \frac{1}{2} \sum |a_k| + \frac{1}{2} \sum |a_k| + \frac{1}{2\lambda} \|w\|^2 = |\mu_{\lambda}^\top T_d + \frac{1}{2\lambda} \|\nu\|^2.$$

Therefore $R$ is indeed a solution of $(P_{\lambda}^{(m)})$ since $\mu_{\lambda}$ is a solution of $(P_{\lambda})$, and one has rank $R \leq r$. □

#### Higher dimensions

For $d > 1$, although we do not have theoretical proof for this result, numerical experiments seem to indicate that the same rank property holds. Figure 2 shows the singular values of both a solution of $(P_{\lambda}^{(m)})$ and a solution of its dual [4], computed using a standard interior points solver, interfaced with the CVX software [17].

The search space of $(P_{\lambda}^{(m)})$ may thus be restricted to rank-deficient matrices. Such geometry is well exploited by conditional gradient algorithms. However, these methods are not able to handle the SDP contraint $(a)$ together with the affine constraint $(b)$. Furthermore, the intersection between manifolds of fixed rank matrices and the affine space defined by $(b)$ quickly becomes unanalyzable for matrices larger than $2 \times 2$, and non-convex optimization schemes on this search space would likely be difficult to implement. Instead, we propose to smooth the geometry of the problem.
Figure 2: \( f_c = 7, r = 3, d = 2, m = 2f_c + 1 \). Singular values (relative to the maximal singular value) of primal (blue) and dual (black) matrices. The rank of the primal matrices appears to be bounded by \( r \), and that of the dual matrices to be bounded below by \( n - r \). Not all dual matrices satisfy this property: when \( d = 1 \) for instance, there always exists a dual solution of rank at most 3; however, such low-rank matrices are unlikely to be found by interior points methods, since these tend to maximize the rank of their solutions.

4 Toeplitz penalization

To overcome the difficulty induced by constraint (b) in \( (P^{(m)}_{\lambda}) \), we introduce a penalized version of \( (P^{(m)}_{\lambda}) \) – which can also be seen as a perturbation of the atomic norm regularizer used in \([34, 35]\). We call \( T \) the Toeplitz projector, defined as

\[
T : R \mapsto \sum_{|k| \leq m} \frac{(R, \Theta_k)}{||\Theta_k||^2} \Theta_k
\]

where the matrix \( \Theta_k \) are introduced in Section 3.1. In dimension one, the operator \( T \) replaces each entry of \( R \) by the mean of the corresponding diagonal. We consider the following program:

\[
\mathcal{R}_{\lambda, \rho} \in \arg\min_{\tau, z, R} \left\{ \tau + \frac{1}{m} \text{tr}(R) + \frac{1}{2} \frac{y}{\lambda} + 2z \right\} + \frac{1}{2\rho} \| R - T(R) \|^2
\]

\[
\text{s.t.} \quad \begin{cases} \begin{bmatrix} R & \hat{z} \\ \hat{z}^T & \tau \end{bmatrix} & \succeq 0 \\ \hat{z}_k = z_k, & k \in [-f_c, f_c]^d \end{cases}
\]

(\( P^{(m)}_{\lambda, \rho} \))

where the parameter \( \rho \) controls the penalization of the Toeplitz constraint. We study numerically the validity of this approach. For simplicity, the numerical experiments of this section are all in dimension one.

4.1 Sensitivity analysis

Obviously, the dual polynomial \( \eta_{\lambda, \rho} \) associated with \( (P^{(m)}_{\lambda, \rho}) \) (defined as \( \Phi^*(\frac{\tau}{\lambda} + 2z_{\lambda, \rho}) \)) may differ from \( \eta_{\lambda} \), in particular with regard to their saturation points, thus compromising the support recovery. However, following an approach similar to \([38]\), under some mild non-degeneracy hypothesis on \( \eta_{\lambda, \rho} \), it is possible to show that, for small enough values of \( \rho \), \( \mathcal{R}_{\lambda, \rho} \) is sufficiently close to \( \mathcal{R}_{\lambda} \) to allow accurate support reconstruction – in particular, they have the same rank. Numerical observations confirm that this regime exists: Figure 3 shows the stability of the rank for low values of \( \rho \).
3.4
3.6
3.8
4
4.2
4.4
4.6
4.8
5
rank $R$, min = $1/(10^f c)$

Figure 3: $r = 5$, $f_c = 13$, $d = 1$, $\lambda = 0.05$, $\|w\| = 0$. Here, we have rank($R_\lambda$) = 5 (resulting from $r = 5$). We plot the evolution of the rank of $R_{\lambda, \rho}$ w.r.t. $\rho$, averaged over 50 randomized tries (with positive amplitudes), and minimal separation distance $\Delta_{\text{min}}$. We see that for low values of $\rho$, we have rank($R_{\lambda, \rho}$) = rank($R_\lambda$). We lose some stability as the minimal distance between spikes decreases.

### 4.2 Support recovery

The rank stability result however do not provide explicit information on the behavior of the roots of the Laurent polynomial $\tilde{P}_{\lambda, \rho}$ (see Section 2.2), in which we are interested to recover the support. Although such a link may exist, we do not have any theoretical results. Nonetheless, we observe experimentally that for reasonable values of $\rho$, we are still able to recover the roots of $\tilde{P}_{\lambda, \rho}$ which derive from unit-modulus roots of $\tilde{P}_\lambda$, see Fig. 4. Moreover, in such regimes, the argument of those new roots appear to be sufficiently close to the new ones to still allow for accurate support recovery.

Figure 4: $r = 5$, $f_c = 13$, $d = 1$, $\lambda = 0.05$, $\|w\| = 0$. Roots of $\tilde{P}_{\lambda, \rho}$ w.r.t. $\rho$. When $\rho = 0$, support is recovered from double roots of $\tilde{P}_\lambda$, which are of unit modulus. When $\rho > 0$, these may split but remain identifiable (in the sense that they are close to the unit circle) when $\rho$ is small. We use a manually tuned tolerance around the unit circle to detect them.

### 5 Fast Fourier Transform-based Conditional Gradient Solver

In this section, we take advantage of the low-rank property of the solutions as well as of the convolutive structure of the Toeplitz constraint to build a robust and efficient numerical scheme for
solving \( \mathcal{P}_\lambda^{(m)} \). In the following, we note
\[
f(R, z, \tau) = \tau + \frac{1}{m} \text{tr}(R) + \frac{1}{2} \|y\|_\lambda^2 + 2z \|\| + \frac{1}{2\rho} \|R - \sum_{k \in [-m, m]} \frac{\langle R, \Theta_k \rangle}{||\Theta_k||^2} \Theta_k \|^2.
\]

### 5.1 Conditional gradient

Conditional gradient algorithm, aka Frank-Wolfe algorithm [15], aims at minimizing a convex and continuously differentiable function \( f \) over a compact convex subset \( K \) of a Hilbert space. The essence of the method is as follows: linearize \( f \) at the current position \( x_r \), solve the auxiliary linear problem of minimizing \( s \mapsto \langle \nabla f(x_r), s \rangle \) on \( K \), and move toward the minimizer to obtain the next position. This scheme ensures the sparsity of its iterates, since the solution after \( k \) iterations is a convex combination of at most \( k \) atoms. We refer to [20] for a detailed overview of the method.

A remarkable asset of the method is the simplicity of the linear minimization step, which amounts to extracting an extremal point of the set \( K \). When the latter consists in the convex hull of an atomic domain, this subproblem may thus be solved efficiently. In particular, the positive semidefinite cone being linearly spanned by unit-rank matrices, the linear minimization in this case results in computing a leading eigenvector of \( \nabla f \), which can be done using power iterations.

### 5.2 Super-resolution algorithm

We propose a Frank-Wolfe scheme to minimize \( f \) over the cone of positive semidefinite matrices. In order to boil the problem down to optimization over a compact set, we impose an upper bound \( D_0 \) on the trace of the iterates. This does not change the solution of the problem provided that \( \text{tr} \mathcal{R}_{\lambda, \rho} \leq D_0 \). In practice, one may choose \( D_0 = f(0) \).

It is known [20] that if \( s \) is a solution of the linear minimization oracle, then it satisfies the inequality \( \langle x - s, \nabla f(x) \rangle \geq f(x) - f(x_*) \) for any \( x \). We use this property as a stopping criterion, ceasing the iterations if \( \langle x - s, \nabla f(x) \rangle/f(x_0) \) goes below some tolerance \( \varepsilon \).

To take advantage of the low-rank structure of the solutions, we store our iterates as \( \mathcal{R} = UU^h \). Consequently, at each step of the algorithm, the update consists in adding a column (namely a leading eigenvector of \( \nabla f \), as mentioned above) at the end of the matrix \( U \), thus increasing by one the rank of \( \mathcal{R} \) each time. We further add a non-convex step similar to [1], consisting in a gradient descent on \( F : s \mapsto f(ss^h) \), which can only improve the speed of convergence of the algorithm. We use a BFGS descent in our implementation, see Algorithm 1.

**Algorithm 1** Recovering dual polynomial

```plaintext
set: \( U_0 = [0 \ldots 0]^\top \), \( D_0 \) s.t. \( \text{tr}(U_0U_0^h) \leq D_0 \) (bound on the domain)
while \( \langle U_rU_r^h - uv_r^h, \nabla f(s_r, s_r^h) \rangle \geq \varepsilon f(x_0) \) do
1. Compute next atom: \( v_r = D_0 \arg\min_{\|v\| \leq 1} v^\top \cdot \nabla f(U_rU_r^h) \cdot v \)
2. Update: \( \hat{U}_{r+1} = [\sqrt{\alpha_r}U_r, \sqrt{\beta_r}v_r] \), where \( \alpha_r, \beta_r = \arg\min_{0 \leq \alpha, \beta \leq 1} f(\alpha U_rU_r^h + \beta v_rv_r^h) \)
3. Corrective step: \( U_{r+1} = \text{bfgs}(F(U) : U \in \mathbb{C}^{(n+1) \times (r+1)}, \text{initialize at } \hat{U}_{r+1}) \)
end while
return \( p_\lambda = \frac{y}{\lambda} + 2z \), where \( \begin{bmatrix} 2 \\ \tau \end{bmatrix} \) is the last column of \( U_{N+1}U_{N+1}^h \).
```
Both the positions and amplitudes of the solution may then be recovered from the coefficients \( p_\lambda \). Algorithm 2 recalls the main steps of the reconstruction process. As explained in Section 4.2, in order to recover the correct roots of \( \tilde{P}_{\lambda,\rho} \), we fix a tolerance \( \delta \) around the set of roots whose components are all of unit-modulus, in which they should be located for small enough values of \( \rho \).

**Algorithm 2**

**Reconstruction via root-finding**

**input:** \( p_\lambda \) coefficients of the dual polynomial  
**set** \( \delta > 0 \), \( Z = \text{roots}(p_\lambda) \subset \mathbb{C}^d \)  
- \( R_1 \leftarrow \{ z \in Z : |1 - |z_i|| < \delta, \ i = 1, \ldots, d \} \) + remove symmetric roots (paired as \( z \) and \( 1/\bar{z} \))  
- **positions:** \( \Sigma_\lambda \leftarrow \frac{1}{\pi} \arg(R_1) \)  
- **amplitudes:** \( a \leftarrow (\Phi_{\Sigma_\lambda}^\ast \Phi_{\Sigma_\lambda})^{-1} \Phi_{\Sigma_\lambda}^\ast (y - \lambda p_\lambda) \), where \( \Phi_{\Sigma_\lambda} = \Phi|_{\Sigma_\lambda} \)

**return** \( \Sigma_\lambda, a \)

### 5.3 Fast-Fourier-Transform-based computations

Step 1. in Algorithm 1 simply amounts to computing a leading eigenvector of \( \nabla f(U_rU_h^\ast_U) \), using power iterations, which is based on matrix-vector products only. Given our objective \( f \) defined in (5), the gradient reads  
\[
\nabla f(x) = \begin{bmatrix} \frac{1}{n} I_n & p \\ p^\ast & 1 \end{bmatrix} + \frac{1}{\rho} x - \frac{1}{\rho} T(x).
\]

The sole costly operations of the power iterations steps thus consist in

(i) computing \( T(U_rU_h^\ast_U) \) the projection of \( U_rU_h^\ast_U \) on Toeplitz tensors,

(ii) multiplying the latter with a vector.

Both steps can be done efficiently using only Fast Fourier Transforms: indeed, one can show that the projector in (i) has a simple factorization involving only padding, un-padding, FFT and inverse FFT operators; on the other hand, the Toeplitz-vector product [ii] may also be computed using FFT and padding operators, since multiplication with a Toeplitz matrix is simply a circular convolution combined with zero-padding.

**Proposition 5.** Let \( U \in \mathbb{C}^{m \times d} \times r \), and denote \( U = [U_1, \ldots, U_r] \). Then \( T(UU^\ast_U) = \text{Toep}_t \) with

\[
t_i = C_t(\sum_k F^{-1}([F \circ \text{Pad}(U_k) e_t]^2), e_t).
\]

**Proof.** Let \( \Omega = \| - m, m \|_d \) and \( \omega = \{0, m \|_d \} \). Then \( \| \text{Toep}_t \|_2^2 = \sum_{i,j} |t_i - j|^2 \). Thus we have:

\[
|\text{Toep}_t - UU^\ast_U|^2_2 = \sum_{i,j} |t_i - j|^2 - \sum_k |U_{i,k}U_{j,k}|^2 = \sum_{l \in \Omega} \sum_{i,j} |t_i - \sum_k U_{i,k}U_{i-1,k}|^2.
\]

Therefore, with \( t' \overset{\text{def}}{=} \arg\min_{t'} \|\text{Toep}_{t'} - UU^\ast_U\|_2^2 \), one has

\[
\forall l \in \Omega, \ t_i = \frac{1}{\text{card}\{ (i,j) \in \omega : i - j = l \}} \sum_{i \in \omega} \sum_k U_{i,k}U_{i-1,k},
\]

13
with $U_{i-l,k} = 0$ if $i - l \in \Omega \setminus \omega$. In particular,

$$\sum_{i \in \omega} U_{i,k}U_{i-l,k} = \langle \text{Pad}(U_{.k}) * \text{Pad}(U_{-.k}), e_l \rangle$$

$$= \langle \mathcal{F}^{-1}(\mathcal{F}(\text{Pad}(U_k)) \cdot \mathcal{F}(\text{Pad}(U_k))), e_l \rangle.$$

Thus $t_l = C_l \sum_k \langle \mathcal{F}^{-1}((\mathcal{F} \circ \text{Pad}(U_k))^2), e_l \rangle$. \hfill \qed

**Remark 6.** The value $1/C_l = \text{card} \{ (i, j) \in \omega : i - j = l \}$ can be reckoned easily by computing $\mathbb{1} * \mathbb{1}$.

**Proposition 6.** Let $T = \text{Toep}_t$, where $t \in \mathbb{C}^{(2m+1)^d}$, and let $x \in \mathbb{C}^{(m+1)^d}$. Then

$$T \cdot x = \text{Pad}^{-1} \circ \mathcal{F}^{-1}((\mathcal{F}(t), \mathcal{F} \circ \text{Pad}(x))).$$

**Proof.** The periodic convolution $c * x$ can be written as the product between the circulant matrix $C$ whose diagonal values are given by $c$, and the vector $x$. In that case, one has $C \cdot x = \mathcal{F}^{-1}((\mathcal{F}(c), \mathcal{F}(x)))$. The product $\text{Toep}_t \cdot x$ on the other hand corresponds to an aperiodic convolution, and as such can be equivalently reformulated as a periodic convolution between $t$ and a zero-padded version of $x$, hence the result. \hfill \qed

**Remark 7.** Let $U$ be a $(n^d + 1) \times r$-matrix and $v$ a $(n^d + 1)$-vector. Regarding the computation of $\alpha_r, \beta_r$ for the linesearch step 2. in Algorithm 1 we note that $f(\alpha U U^h + \beta v v^h) = c_{uu}\alpha^2 + c_{uv}\beta^2 + c_{uv}\alpha\beta + c_u\alpha + c_v\beta + 0.5|y/\lambda|^2$, with

$$\begin{cases}
    c_{uu} = 2\|U_1^h u\|^2 + \frac{1}{2\rho} \|U_1 U_1^h - T(U_1 U_1^h)\|^2 \\
    c_{uv} = 2t^2 \|v_1\|^2 + \frac{1}{2\rho} \|v_1 v_1^h - T(v_1 v_1^h)\|^2 \\
    c_{uv} = 4t \Re(U_1^h u, v_1) + \frac{1}{\rho} \Re(U_1 U_1^h - T(U_1 U_1^h), v_1 v_1^h - T(v_1 v_1^h)) \\
    c_u = \|U_1\|^2 + \frac{1}{n} \|U_1\|^2 + \frac{2}{\lambda} \Re(y, U_1^h u) \\
    c_v = \|v_1\|^2 + \frac{1}{n} \|v_1\|^2 + \frac{2t}{\lambda} \Re(y, v_1)
\end{cases}$$

where

$$U = \begin{bmatrix} U_1 & u \end{bmatrix}^h, \quad U_1 \in \mathbb{C}^{r \times n^d}, \quad u \in \mathbb{C}^r$$

$$v = \begin{bmatrix} v_1 & t \end{bmatrix}^h, \quad v_1 \in \mathbb{C}^{n^d}, \quad t \in \mathbb{C}.$$

We thus need to solve

$$(\alpha_r, \beta_r) = \arg\min_{\alpha, \beta \in \Delta} c_{uu}\alpha^2 + c_{uv}\beta^2 + c_{uv}\alpha\beta + c_u\alpha + c_v\beta + \frac{1}{2} \frac{|y/\lambda|^2}{\lambda}$$

with $\Delta = \left\{ x, y \in [0, 1]^2 : x + y \leq 1 \right\}$. This problem admits a closed form solution.
5.4 Complexity

At step $r$ in algorithm 1, the iterate $s_r$ has size $(n + 1)^d \times r$, with $n = 2f_c + 1$. Step 1 is done with power iteration, using the FFT-scheme described in section 5.3 and costs $O(n^d(r + \log n))$ per iteration, hence $O(C_{\text{PI}}(n) \cdot n^d(r + \log n))$ overall, where $C_{\text{PI}}(n)$ depends on the number of iterations. The BFGS iterations in Step 3 costs $O(C_{\text{BFGS}}(n) \cdot n^d(r + \log n))$. Hence, for $r$ iterations, the overall costs is $O((C_{\text{PI}}(n) + C_{\text{BFGS}}(n)) \cdot r n^d(r + \log n))$. The dependency of the factors $C_{\text{PI}}$ and $C_{\text{BFGS}}$ with respect to $n$ is still unclear; $C_{\text{BFGS}}$ should at least decrease as $n$ increases, since the problem gets better conditioned for larger values of $f_c$.

6 Numerics

**Power Iteration step** The tolerance for the power iteration step is set to $10^{-4}$, with a maximum of 500 iterations. In order to determine the lowest singular value of $\nabla f$ (simultaneously with a corresponding singular vector), one may need to run the algorithm two times consecutively: if after the first run, the algorithm returns a value $\lambda_1 < 0$, this is indeed the lowest one (since power iterations retrieve the singular value whose magnitude is the greatest); if not, we re-run it on $\nabla f - \lambda_1 I$, and add the resulting value to $\lambda_1$ to obtain the lowest singular value of $\nabla f$.

**BFGS step** We use Mark Schmidt’s code for the BFGS solver [31]. The tolerance for this step is set to $10^{-8}$ (in terms of functions or parameters changes), with 500 as the maximum number of iterations (except in Figure 5). This step is crucial to ensure the finite convergence of the algorithm, see Fig. 6. When $\rho$ tends to zero, plain Frank-Wolfe steps become significantly insufficient, and the number of BFGS iterations necessary to converge at each step dramatically increases, Fig. 5.

**Root-finding** Because of the Toeplitz relaxation described in Section 4, the quality of our reconstruction depends on our ability to reliably detect the roots of $P_{\lambda,\rho}$ that are truly associated with a Dirac. In our simulations, we manually fix a tolerance $\delta$ around the unit circle in which we look for those roots, as already mentioned. In practice, we take $\delta$ of the order of $10^{-2}$. In future works, it is important that we find a sharper criteria to separate the $r$-th root from the $(r+1)$-th one.

In dimension 1, the root-finding step may be done using a standard polynomial root-finding algorithm, whose complexity is of the order of $O(f^3)$. However, for multivariate polynomials, the problem of root-finding becomes numerically challenging, and could be a bottleneck for our approach. To our knowledge, two methods are possible to determine the support, only one we implemented:

- we locate the extremas of the dual polynomial using a Newton algorithm. As an initialization, we select points on a chosen grid such that the modulus of $\eta_{\lambda,\rho}$ at those points is not farther than $\delta$ from 1; in our experiments, we use a grid of size $200 \times 200$. Finally, we average the different clusters of points returned by the Newton steps to get the locations. An example of reconstruction is given in Fig. 7.

- it is also possible to determine the support of the initial measure at a lower cost directly from the moment matrix $R_{\lambda}$, using tools from algebra. Schematically, this is done by solving a generalized eigenvalue problem, whose associated matrices are morally of size $r \times r$, $r$ being the sparsity [25, 26]. Such methods are already implemented in the software Gloptipoly [18] for
Figure 5: $r = 5$, $f_c = 13$, $d = 1$, $\lambda = 0.05$, $\|w\| = 0$. Left: Error between $\mu_{\lambda, \rho}$ and $\mu_\lambda$, measured with the dual bounded Lipschitz norm (defined as $\sup\{|\int f d\mu| : \text{lip} f + \|f\|_\infty \leq 1\}$, and also known as flat-norm [14]), with respect to $\rho$. $\mu_\lambda$ is computed using CVX. Right: Total number of BFGS iterations with respect to $\rho$. Here we fix the maximum number of BFGS iterations by Frank-Wolfe iteration to $10^5$. The tests are done over randomized positive measures, with minimal separation distance $1/f_c$. There is obviously a sweet spot to find between reasonable computational cost and sufficient quality reconstruction.

real measures, and can be readily extended to measures on the torus. We did not investigate these methods further yet, but they are likely to provide a viable alternative to root-finding techniques. However, it is still unclear what would be the effect of the Toeplitz relaxation on the recovery algorithm.

References

[1] N. Boyd, G. Schiebinger, and B. Recht. The alternating descent conditional gradient method for sparse inverse problems. In CAMSAP, pages 57–60, 2015.

[2] K. Bredies and H. K. Pikkarainen. Inverse problems in spaces of measures. ESAIM: Control, Optimization and Calculus of Variations, 19(1):190–218, 2013.

[3] E. J. Candès and Y. Plan. A probabilistic and ripless theory of compressed sensing. IEEE Trans. Information Theory, 57(11):7235–7524, 2010.

[4] E.J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. Communications on pure and applied Mathematics, 67(6):906–956, 2014.

[5] R.E. Curto and L.A. Fialkow. Solution of the truncated complex moment problem for flat data. Memoirs of the AMS, (568), 1996.

[6] Y. de Castro and F. Gamboa. Exact reconstruction using Beurling minimal extrapolation. Journal of Mathematical analysis and applications, 395(1):336–354, 2012.

[7] Y. de Castro, F. Gamboa, D. Henrion, and J.B. Lasserre. Exact solutions to super resolution on semi-algebraic domains in higher dimensions. IEEE Trans. Information Theory, 63(1):621–630, 2017.
Figure 6: $r = 5$, $f_c = 8$, $d = 2$, $m = 2f_c + 1$, $\lambda = 5.10^{-1}$, $|w| = 5.10^{-4}||y_0||$, $\rho = 50$. We plot the relative error between the energy associated with $\mathcal{P}_{\lambda, \rho}$ and the energy at the optimal point for $\mathcal{P}_\lambda$ for ten randomized trials with same signal-to-noise ratio and same initial measure, and with positive amplitudes. In almost all cases, finite convergence appears to occur after $r$ iterations of Frank-Wolfe.

Figure 7: $r = 7$, $f_c = 10$, $m = 2f_c + 1$, $\rho = 1$. Synthetic examples of reconstruction in a noiseless (left) and noisy (right) case. In the noisy case, positions are accurately recovered, but not amplitudes, and a spike is missed (its amplitude is too low).
[8] Q. Denoyelle, V. Duval, and G. Peyré. Support recovery for sparse deconvolution of positive measures. CoRR, 2015.
[9] D. L. Donoho. Superresolution via sparsity constraints. SIAM Journal on Mathematical Analysis, 23(5):1309–1331, 1992.
[10] C. Dossal and S. Mallat. Sparse spike deconvolution with minimum scale. In Proceedings of SPARS, pages 123–126, November 2005.
[11] B. A. Dumitrescu. Positive trigonometric Polynomials and Signal Processing Applications. Signals and Communication Technology. Springer International Publishing, 2017.
[12] V. Duval and G. Peyré. Exact support recovery for sparse spikes deconvolution. Foundations of Computational Mathematics, 15(5):1315–1355, 2015.
[13] V. Duval and G. Peyré. Sparse regularization on thin grids I: the Lasso. Inverse Problems, 33(5), 2017.
[14] M. Frank and P. Wolfe. An algorithm for quadratic programming. Naval Research Logistics Quarterly, 3(1-2):95–110, 1956.
[15] M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. J. ACM, 42(6):1115–1145, November 1995.
[16] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx, 2014.
[18] D. Henrion, J.B. Lasserre, and J. Löfberg. Gloptipoly 3: moments, optimization and semidefinite programming. Optimization Methods and Software, 24(4–5):761–779, 2009.
[19] S.T. Hess, T.P.K Girirajan, and M.D. Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. Biophysical Journal, 91(11):4258–4272, 2006.
[20] M. Jaggi. Revisiting frank-wolfe: Projection-free sparse convex optimization. In ICML, volume 28, 2013.
[22] H. Krim and M. Viberg. Two decades of array signal processing research: the parametric approach. IEEE Signal Processing Magazine, 13(4):67–94, 1996.
[23] S. Kunis, T. Peter, T. Römer, and U. Von der Ohe. A multivariate generalization of prony’s method. Linear Algebra and its Applications, 490:31–47, 2016.
[24] J. B. Lasserre. Global optimization with polynomials and the problem of moments. SIAM Journal on Optimization, 11(3):796–817, 2001.
[25] J. B. Lasserre. Moments, positive polynomials and their applications. Imperial College Press Optimization Series. Imperial College Press, London, 2010.
[26] M. Laurent. Sums of squares, moment matrices and optimization over polynomials. In Emerging Applications of Algebraic Geometry, volume 149. Springer new York, 2009.
[27] P. A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. Mathematical Programming, 96(2):293–320, 2003.
[28] K.G. Puschmann and F. Kneer. On super-resolution in astronomical imaging. Astronomy and Astrophysics, 436(1):373–378, 2005.
[30] M.J. Rust, M. Bates, and Xiaowei Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, 3:793–796, 2006.

[31] M. Schmidt. minFunc: Unconstrained differentiable multivariate optimization in Matlab. [http://www.cs.ubc.ca/~schmidt/Software/minFunc.html](http://www.cs.ubc.ca/~schmidt/Software/minFunc.html), 2005.

[32] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, 1986.

[33] G. Tang. Resolution limits for atomic decompositions via Markov-Bernstein type inequalities. *2015 International Conference on Sampling Theory and Applications (SampTA)*, pages 548–552, 2015.

[34] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *IEEE Trans. Information Theory*, 59(11):7465–7490, 2013.

[35] G. Tang, B.N. Bhaskar, and B. Recht. Near minimax line spectral estimation. *IEEE Trans. Information Theory*, 61(1):499–512, 2015.

[36] S. Vaiter, G. Peyré, and J. Fadili. *Low Complexity Regularization of Linear Inverse Problems*, pages 103–153. Springer International Publishing, 2015.

[37] X. Wei and P.L. Dragotti. FRESH – FRI-based single-image super-resolution algorithm. *IEEE Trans. Image Processing*, 25(8), 2016.

[38] W. Xu, J.-F. Cai, K.V. Mishra, M. Cho, and A. Kruger. Precise semidefinite programming formulation of atomic norm minimization for recovering d-dimensional (d ≥ 2) off-the-grid frequencies. In *Information Theory and Applications Workshop*, 2014.