A Low Rank Approach to Off-The-Grid Sparse Super-Resolution

Paul Catala[1], Vincent Duval[2,3] and Gabriel Peyré[1].

[1]DMA, Ecole Normale Supérieure, PSL, UMR 3553. [2]Mokaplan, INRIA Paris. [3]CEREMADE, Université Paris-Dauphine, PSL, UMR 7534.

***Abstract*— We consider the sparse spikes super-resolution problem over the space of Radon measures. When the observations consist in a low-pass filtering of the input, a common approach to off-the-grid super-resolution considers semidefinite (SDP) relaxations of the total variation (the total mass of the measure) minimization problem. Solving this SDP is often intractable for large scale settings, since the problem size grows as f_c^{2d} where f_c is the cutoff frequency of the filter, and d the dimension of the signal. We propose a solver applicable to a class of observations larger than simply convolution, and scalable with the dimension. Our first contribution is a Fourier approximation scheme of the forward operator, making the TV-minimization problem expressible as a SDP. Our second contribution introduces a penalized formulation of this semidefinite lifting, which has low-rank solutions. Our last contribution is a conditional gradient approach with non-convex updates. This algorithm leverages both the low-rank and the Fourier structure of the problem, resulting in an $O(f_c^d \log f_c)$ complexity per iteration.**

1 Introduction

Super-resolution problems aim at accurately recovering sparse signals from low-resolution and possibly noisy measurements. This is an important challenge in medical imaging, microscopy or astronomical imaging, where it may be crucial to overcome the physical limitations of sensing devices. It is also related to several statistical problems, for instance compressive statistical learning [10], or Gaussian mixture estimation [9].

Formally, our goal is to retrieve a d-dimensional discrete measure $\mu_0 = \sum_{k=1}^r a_i \delta_{x_i}$ ($a_i \in \mathbb{R}$, $x_i \in \mathbb{T}^d$, where $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ is the torus) given the observations

$$y = \Phi\mu_0 + w \in \mathbb{C}^m,$$

where w is some noise, and $\Phi : \mathcal{M}(\mathbb{T}^d) \to \mathbb{C}^m$ is a known linear operator defined as

$$\Phi\mu \overset{\text{def.}}{=} \int_{\mathbb{T}^d} \varphi(x)\mathrm{d}\mu(x), \quad \text{where } \varphi(x) = (\varphi_1(x), \dots, \varphi_m(x)),$$

with operator norm $\|\Phi\| \overset{\text{def.}}{=} \sum_{i=1}^p \|\varphi_i\|^2_{\mathrm{L}^2(\mathbb{T}^d)}$. An important example is when Φ is the Fourier operator $\mathcal{F}$, defined as

$$\mathcal{F} : \mu \mapsto \Big(\int_{\mathbb{T}^d} e^{-2i\pi\langle \omega, x\rangle} \mathrm{d}\mu(x) \Big)_{\omega \in \Omega_c}.$$

where $\Omega_c \overset{\text{def.}}{=} [\![-f_c, f_c]\!]^d$, for some cutoff frequency $f_c \in \mathbb{N}^*$. In that case, $\Phi\mu_0$ is a vector of trigonometric moments of μ_0, at freqencies $\omega \in \Omega_c$. This is the framework considered in [6].

2 Fourier approximation of operators

The method we introduce in this paper is based on an approximate factorization of Φ under the form $\mathcal{A} \circ \mathcal{F}$, where $\mathcal{A} : \mathbb{C}^{|\Omega_c|} \to \mathbb{C}^m$ is some linear operator. Given an arbitrary forward operator Φ, we define $\mathcal{A}$ as the operator minimizing $\|\Phi - \mathcal{A} \circ \mathcal{F}\|^2$. The approximation error can be made arbitrarily small as f_c goes to infinity.

A typical example is the convolution with some kernel ψ. This corresponds to

$$\mathcal{A} = \mathrm{Diag}(\hat{\psi}(\omega))_{\omega \in \Omega_c}, \quad \text{and} \quad m = |\Omega_c|.$$

For instance, if ψ is the Dirichlet kernel, $\mathcal{A} = \mathrm{Id}$. In the case where $\mathrm{Supp}\,\hat{\psi} \not\subset \Omega_c$, $\hat{\psi}$ must be sampled over Ω_c.

This framework also encompasses some non-convolution problems, an important example being subsampled convolution. In that case, with $\mathcal{G}$ the sampling grid, $\mathcal{A}$ is given by

$$\mathcal{A} = \Big(\hat{\psi}(\omega) e^{2i\pi\langle \omega, t\rangle} \Big)_{(t,\omega) \in \mathcal{G} \times \Omega_c}, \quad \text{and} \quad m = |\mathcal{G}|.$$

When ψ is a Gaussian filter, this last case corresponds to the setting of the Single Molecule Localization Microscopy (SMLM) data, on which we perform our experiments (see section 6).

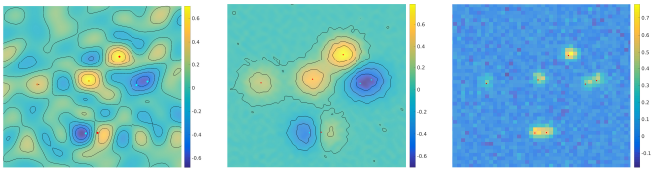

Figure 1: Examples of ideal low-pass filter (left), Gaussian (middle) and subsampled Gaussian (right) observations.

3 Beurling LASSO

Although the inverse problem we consider is severely ill-posed, sparse estimates can be found by solving the following minimization problem, known as the BLASSO [4]:

$$\min_{\mu \in \mathcal{M}(\mathbb{T}^d)} \frac{1}{2\lambda} \|y - \mathcal{A} \circ \mathcal{F}\mu\|^2 + |\mu|(\mathbb{T}^d) \qquad (\mathcal{P}_\lambda)$$

where λ should be adapted to the noise level. The total variation norm is defined as

$$|\mu|(\mathbb{T}^d) \overset{\text{def.}}{=} \sup \left\{ \Re\Big(\int_{\mathbb{T}^d} \overline{\eta} \mathrm{d}\mu \Big) \;;\; \eta \in \mathcal{C}(\mathbb{T}^d), \quad \|\eta\|_\infty \leqslant 1 \right\}.$$

It is the natural extension of the ℓ^1-norm to the space of Radon measures. For instance, if $\mu = \sum a_k \delta_{x_k}$, then $|\mu|(\mathbb{T}^d) = \|a\|_1$.

This paper introduces a novel numerical solver for $(\mathcal{P}_\lambda)$. Indeed, although the BLASSO grid-free approach offers beneficial mathematical insight, its numerical resolution remains challenging. In the seminal paper [6], the authors propose to lift the problem (when $d = 1$, and $\Phi = \mathcal{F}$) to a semidefinite program, solvable via interior points methods. Using Lasserre hierarchy, this semidefinite lifting can also be applied in higher dimensions [8, 11]. However, usual interior points solvers scale poorly with d. We propose a novel method, based on these semidefinite hierarchies, scalable with the dimension.

4 Semidefinite formulations

Since $(\mathcal{P}_\lambda)$ only involves a few trigonometric moments of measures, it may be cast as a quadratic program over the cone of moment sequences. Semidefinite characterizations of this *moment cone* exists when the underlying measures are defined on semi-algebraic sets, *i.e.* sets described by polynomials inequalities. These results have important numerical assets: in [2], a generic method is proposed to approximate to arbitrary precision the moment cone by semidefinite liftings of increasing size – the so-called Lasserre hierarchy.

The semidefinite relaxation of $(\mathcal{P}_\lambda)$ at order $l \in \mathbb{N}$ reads:

$$\min_{\tau,z,u} f(\mathcal{R}) = \tau + u_0 + \frac{1}{2}\|\frac{y}{\lambda} + 2\mathcal{A}z\|^2$$
$$\text{s.t.} \begin{cases} (a) & \mathcal{R} = \begin{bmatrix} R & \tilde{z} \\ \tilde{z}^* & \tau \end{bmatrix} \succeq 0, \\ & \tilde{z}_\omega = z_\omega, \quad \forall\, \omega \in \Omega_c \\ (b) & R = \sum_{k \in]\!]-l-n_c, l+n_c]\!]^d} u_k \Theta_k \end{cases} \qquad (\mathcal{P}_\lambda^{(l)})$$

where $n_c = 2f_c + 1$ and $\Theta_k = \theta_{k_d} \otimes \ldots \otimes \theta_{k_1}$, θ_{k_i} being the Tœplitz matrix with ones on its k_i-th diagonal, zeros elsewhere, and $\otimes$ the Kronecker product. When $d = 1$ for instance, contraint (b) is simply asking that R be a Toeplitz matrix.

Proposition 1. *Let $R \in \mathcal{M}_{(l+n_c)^d}(\mathbb{C})$ be a solution of $(\mathcal{P}_\lambda^{(l)})$, and let R_{l-1} be the top left block of size $(l+n_c-1)^d \times (l+n_c-1)^d$ in R. If $\operatorname{rank} R = \operatorname{rank} R_{l-1}$, then $\min(\mathcal{P}_\lambda^{(l)}) = \min(\mathcal{P}_\lambda)$.*

The rank condition in Proposition 1 ensures that R is the moment matrix of some measure supported on $\operatorname{rank}(R)$ points [1]. In the rest of the paper, we assume that we are given l such that the two minima are equal (the hierarchy is said to have *collapsed*). Note that in practice, $l = 0$ seems to suffice.

The following crucial result motivates the use of a conditional gradient (aka Frank-Wolfe) scheme to solve $(\mathcal{P}_\lambda^{(l)})$.

Proposition 2. *When $d = 1$, $(\mathcal{P}_\lambda^{(l)})$ always admits a solution $\mathcal{R}_\star$ such that $\operatorname{rank} \mathcal{R}_\star \leqslant r$, r being the number of spikes in a solution of $(\mathcal{P}_\lambda)$.*

Numerical results seem to show that this holds when $d = 2$.

Remark 1. To extract the support of a measure from its moment matrix, one may use the method described in [3]. We do not discuss it here, but it is the one we use in our implementation.

5 FFT-based Frank-Wolfe solver

Conditional gradient algorithms are well suited to handle low-rank iterates, in contrast for instance with interior points schemes. Following Proposition 2, this is a property we want to exploit. However, their efficiency relies on the simplicity of the geometry of the domain. In our case, the geometry induced by constraint (b) (with the semidefinite constraint) remains too complex. We introduce the following penalized problem:

$$\min_{\tau,z,R} \tau + \frac{1}{l}\operatorname{Tr} R + \frac{1}{2}\|\frac{y}{\lambda} + 2\mathcal{A}z\|^2 + \frac{1}{2\rho}\|P_{V_\Theta^\perp}(R)\|^2$$
$$\text{s.t.} \begin{bmatrix} R & \tilde{z} \\ \tilde{z}^* & \tau \end{bmatrix} \succeq 0, \quad \tilde{z}_\omega = z_\omega, \quad \forall\, \omega \in \Omega_c \qquad (\mathcal{P}_{\lambda,\rho}^{(l)})$$

where $P_{V_\Theta^\perp}$ is simply the orthogonal projection on $(\operatorname{Vect} \Theta_k)^\perp$.

We propose to solve $(\mathcal{P}_{\lambda,\rho}^{(l)})$ using Frank-Wolfe algorithm [5]. We store our iterates as $\mathcal{R} = \mathcal{U}\mathcal{U}^*$. Frank-Wolfe's oracle over the semidefinite cone is given by a leading eigenvector of the matrix M defined as $\nabla f \cdot H = \langle M, H \rangle$ (for simplicity, we also write ∇f instead of M), which we compute using power iterations. This is done efficiently in $O(f_c^d \log f_c)$, exploiting the connection between Tœplitz matrices and the Fast Fourier Transform. We further add a non convex BFGS update on $\mathcal{U}$ similar to [7] after each Frank-Wolfe step. Our algorithm appears to converge in exactly r steps, r being the number of spikes composing the solution of the BLASSO.

Algorithm 1 Building the moment matrix

set: $\mathcal{U}_0 = [0 \ldots 0]^\top$, D_0 s.t. $\operatorname{tr}(\mathcal{U}_\star\mathcal{U}_\star^*) \leqslant D_0$
while $\langle \mathcal{U}_t\mathcal{U}_t^h - v_t v_t^*, \nabla f(v_t v_t^*)\rangle \geqslant \varepsilon f(\mathcal{U}_0\mathcal{U}_0^*)$ **do**
1. *FW oracle*: $v_t = D_0 \arg\min_{\|v\| \leqslant 1} v^\top \cdot \nabla f(\mathcal{U}_t\mathcal{U}_t^*) \cdot v$
2. *Update*: $\widehat{\mathcal{U}}_{t+1} = \left[\sqrt{\alpha_r}\mathcal{U}_t, \sqrt{\beta_r} v_t\right]$, where
$\alpha_t, \beta_t = \arg\min_{0 \leqslant \alpha+\beta \leqslant 1} f(\alpha \mathcal{U}_t\mathcal{U}_t^* + \beta v_t v_t^*)$
3. *Correction*: $\mathcal{U}_{t+1} = \textbf{bfgs}(\mathcal{U} \mapsto f(\mathcal{U}\mathcal{U}^*),$ start at $\widehat{\mathcal{U}}_{t+1})$
end while
return $\mathcal{U}_{t+1}$

Remark 2. In Algorithm 1 in practice, $D_0 = 2f(0)$ works. Furthermore, the linesearch in Step 2. has a closed-form solution.

6 Numerics

Our tests are performed on the Contest Dataset 2, from the 2013 SMLM challenge, *bigwww.epfl.ch/smlm/*. We measure the performance with the Jaccard index J (optimal performance is achieved when $J = 1$). Figure 2 (left) shows the Jaccard index on high density images (*i.e.* $r \geqslant 9$) with respect to several values of relaxation parameters λ, ρ, up to a normalization factor ($\|\Phi^* y\|_\infty$ for λ, $\|\Phi^* y\|_\infty^{-1}$ for ρ). Figure 2 (right) shows the perfomance (in blue), as well as the computational time (in red) as the maximum number of BFGS iterations increases. Figure 3 shows an example of spikes reconstruction.

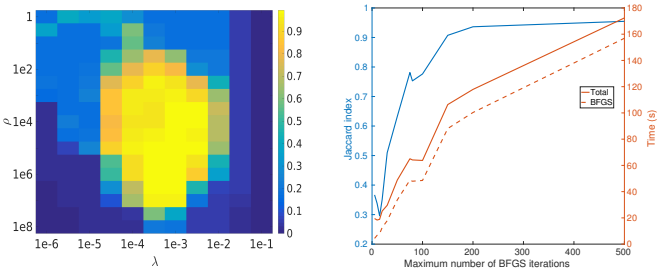

Figure 2: Performance evaluation. Results are averaged over 20 images.

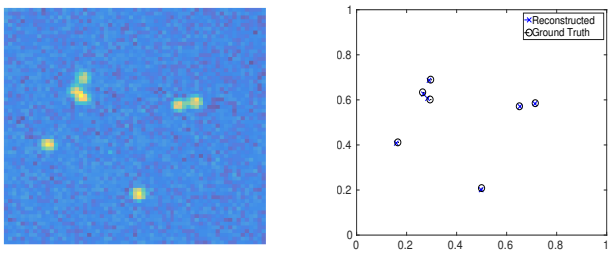

Figure 3: Support localization example. Error is $\frac{\|x_0 - x_{\text{recov}}\|}{\|x_0\|} = 1.57 * 10^{-2}$.

Acknowledgment This work was supported by grants from Région Ile-de-France.

References

[1] R.E. Curto and L.A. Fialkow, "Solution of the Truncated Complex Moment Problem for Flat Data", Memoirs of the AMS, **568**, 1996

[2] "Global optimization with polynomials and the problem of moments" IAM Journal on Optimization, **11**(3):103–122, 2001

[3] M. Laurent, "Sums of Squares, Moment Matrices and Optimization Over Polynomials", Emerging Applications of Algebraic Geometry, **149**, 2010

[4] Y. de Castro and F. Gamboa, "Exact reconstruction using Beurling minimal extrapolation", J. Math Anal. Appl., **395**(1):336–354, 2012

[5] M. Jaggi, "Revisiting Frank-Wolfe: Projection-free sparse convex optimization", in ICML, **28**, 2013

[6] E.J. Candès and C. Fernandez-Granda, "Towards a Mathematical theory of super-resolution", Comm. Pure Appl. Math, **67**(6):906–956, 2014

[7] N. Boyd, G. Schiebinger and B. Recht, "The Alternating Descent Conditional Gradient method for sparse inverse problems", in CAMSAP, 57–60, 2015

[8] 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. Inf. Theory, **63**(1):621–630, 2017

[9] C. Poon and G. Peyé, "Multi-dimensional Sparse Super-resolution", arXiv:1709.03157, 2017

[10] R. Gribonval, G. Blanchard, N. Keriven and Y. Traonmilin, "Compressive Statistical Learning with Random Feature Moments", arXiv:1706.07180, 2017

[11] B.A. Dumitrescu, "Positive trigonometric polynomials and signal processing applications", Signals and Communication Technology, Springer, 2017