A Low-rank Approach to Off-the-Grid Sparse Deconvolution

Paul Catala
DMA, ENS
pcatala@dma.ens.fr

Vincent Duval
Mokaplan, INRIA Paris
vincent.duval@inria.fr

Gabriel Peyré
CNRS and DMA, ENS
gabriel.peyre@ens.fr

***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 deconvoluton considers semidefinite (SDP) relaxations of the total variation (i.e. the total mass 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_c^2 where f_c is the cutoff frequency of the filter. Our first contribution introduces an unconstrained dualization of this semidefinite lifting, which has low-rank solutions. Our second contribution is a conditional gradient (a.k.a. Frank-Wolfe) optimization scheme with non-convex updates. This algorithm leverages both the low-rank and the convolutive structure of the involved variable, resulting in an $O(f_c \log f_c)$ complexity per iterations. Our numerical simulations are promising and show that this algorithm converges in exactly k steps, where k is the number of Diracs composing the solution.**

I. Introduction

Our goal is to recover accurately the amplitudes a_i and positions x_i of a discrete measure $\mu_0 = \sum_i a_i \delta_{x_i}$ given low-resolution and noisy measurements $y = \Phi\mu_0 + w$. We model the degradation operator Φ as an ideal low-pass filter with cutoff frequency f_c. Thus, for any measure on the torus $\mu \in \mathcal{M}(\mathbb{T})$, we have $\Phi\mu = \int_{\mathbb{T}} \varphi(x)\,\mathrm{d}\mu(x)$, where $\varphi(x) = (e^{2i\pi kx})_{-f_c \leq k \leq f_c}$. This ill-posed inverse problem can be tackled using total variation regularization, which generalizes ℓ^1-regularization to measures. We consider the following problem:

$$\mu_\lambda \in \operatorname*{argmin}_{\mu \in \mathcal{M}(\mathbb{T})} \frac{1}{2\lambda}\|y - \Phi\mu\|^2 + |\mu|\,(\mathbb{T}) \qquad (\mathcal{P}_\lambda)$$

where the parameter $\lambda > 0$ should be adapted to the noise level $\|w\|$.

Its dual formulation $(\mathcal{D}_\lambda)$ offers both practical and theoretical insights on the solutions μ_λ. It aims at recovering a dual certificate [2], [3], [6] of the form $\Phi^* p_\lambda$ – which in our case is a trigonometric polynomial – where p_λ is the solution to

$$p_\lambda = \operatorname*{argmax}_{p \in \mathbb{C}^{2f_c+1}} \left\{ \langle y,\, p\rangle - \frac{\lambda}{2}\|p\|^2 \;;\; \|\Phi^* p\|_\infty \leq 1 \right\}. \qquad (\mathcal{D}_\lambda)$$

Primal-dual optimality conditions give a simple characterization of the support and amplitudes of μ_λ, depending only on the roots of a trigonometric polynomial: the support S must satisfy $S \subset \{x \;;\; |\Phi^* p_\lambda(x)| = 1\}$ (Fig. 1), and the amplitudes can be deduced using $a_\lambda = (\Phi_S^* \Phi_S)^{-1} \Phi_S^*(y - \lambda p_\lambda)$, where Φ_S restricts Φ to S.

II. Low-rank semidefinite formulations

Both problems $(\mathcal{P}_\lambda)$ and $(\mathcal{D}_\lambda)$ are numerically challenging: $(\mathcal{P}_\lambda)$ is infinite-dimensional while $(\mathcal{D}_\lambda)$ has an infinite number of constraints. In this one-dimensional case however, $(\mathcal{P}_\lambda)$ can be cast as a semidefinite program using the Caratheodory-Toeplitz theorem [10]:

$$\begin{aligned} \mathcal{R}_\lambda \in \operatorname*{argmin}_{R,r,\alpha} \quad & \alpha + \frac{\operatorname{tr} R}{n} + \|\frac{y}{\lambda} + 2r\|^2 \\ \text{s.t.} \quad & \mathcal{R} = \begin{pmatrix} R & r \\ r^H & \alpha \end{pmatrix} \in \mathcal{H}^+_{n+1}, \quad R \in \mathcal{T} \end{aligned} \qquad (\widetilde{\mathcal{P}}_\lambda)$$

where $n \overset{\text{def.}}{=} 2f_c + 1$, and $\mathcal{T}$ is the set of Toeplitz matrices. The coefficients p_λ of the dual polynomial may then be retrieved from a solution r_λ of $(\widetilde{\mathcal{P}}_\lambda)$ through the optimality relation $p_\lambda = y/\lambda + 2r_\lambda$. An inspection of the proof in [10] reveals that the rank of the solutions of $(\widetilde{\mathcal{P}}_\lambda)$ is bounded by the number of input spikes (Fig. 2). This low-rank structure is crucial in the algorithm detailed in section IV.

III. Toeplitz relaxation

The interaction between the SDP and the Toeplitz constraints is numerically challenging. We introduce a penalized approximation:

$$\begin{aligned} \mathcal{R}_{\lambda,\rho} \in \operatorname*{argmin}_{R,r,\alpha} \quad & f(\mathcal{R}) \quad \text{s.t.} \quad \begin{pmatrix} R & r \\ r^H & \alpha \end{pmatrix} \in \mathcal{H}^+_{n+1}, \qquad (\widetilde{\mathcal{P}}_{\lambda,\rho}) \\ \text{where } f(\mathcal{R}) \overset{\text{def.}}{=} & \alpha + \frac{\operatorname{tr} R}{n} + \|\frac{y}{\lambda} + 2r\|^2 + \frac{1}{2\rho}\|R - P_{\mathcal{T}}(R)\|^2 \end{aligned}$$

for some (small) relaxatation parameter $\rho > 0$.

Following an approach similar to [11], under some mild non-degeneracy hypothesis, one can show that for ρ small enough, the solutions of $(\widetilde{\mathcal{P}}_{\lambda,\rho})$ have the same rank as those of $(\widetilde{\mathcal{P}}_\lambda)$. Numerical observations confirm that this regime exists (Fig. 3), and that one can then recover the solution of $(\widetilde{\mathcal{P}}_\lambda)$ with sufficient accuracy.

IV. Algorithm

We propose a Frank-Wolfe scheme [7] which exploits the low-rank structure of the solutions, storing our iterates as $\mathcal{R} = zz^H$. We add a non-convex step similar to [1], which consists in a gradient descent on $F : z \mapsto f(zz^H)$ – we use a BFGS. The performance of the method is briefly described in Fig. 5.

set: $z_0 = [0 \dots 0]^\top$, D_0 s.t. $\operatorname{tr}(z^\star) \leq D_0$ (bound on the domain)
for $i = 1 : N$ (where $N \geq 2f_c + 1$ is fixed) **do**
1. Compute: $v_i = D_0 \arg\min_{\|v\| \leq 1} v^\top \cdot \nabla f(z_i z_i^H) \cdot v$
2. Update: $\hat{z}_{i+1} = [\alpha_i z_i, \beta_i v_i]$,
   where $\alpha, \beta = \arg\min_{\alpha+\beta \leq 1} f(\alpha z_i z_i^H + \beta v_i v_i^H)$
3. Non-convex corrective step:
   $z_{i+1} = \textbf{descent}((z, F(z)) : z \in \mathbb{C}^{(n+1)\times(i+1)})$

end for
return $p = \frac{y}{\lambda} + 2r$, where $\begin{bmatrix} r \\ \alpha \end{bmatrix}$ is the last column of $z_{N+1} z_{N+1}^H$.

The first step amounts to finding the lowest eigenvalue of $\nabla f(z_i z_i^H)$. We use power iterations, which only require matrix-vector multiplications. Owing to both the Toeplitz structure and the low-rank factorization, these operations can be done in $O(n \log n)$ using Fast Fourier Transforms.

When μ_0 is composed of k spikes, and w/λ and λ are small, we observe that k iterations suffice to reach the solution (Fig. 4).

V. Conclusion and perspectives

The proposed algorithm can be generalized to 2D, opening new perspectives for the superresolution of images. In that case, the SDP hierarchy does not collapse immediately, and one has to consider higher relaxation orders [4]. However, due to its low complexity, our method scales well with the dimension of the problem, and may yield a good alternative to MUSIC and Prony's method [8], [9] in the bidimensional case, currently being implemented.

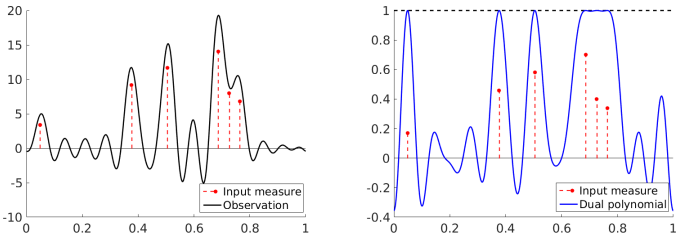

Fig. 1: *Left*: Noisy and low-resolution observations (in black). *Right*: corresponding dual polynomial (in blue). The support of the measure is retrieved by extracting the points where the polynomial reaches 1.

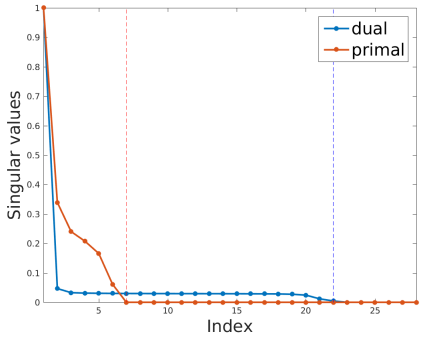

Fig. 2: (6 spikes, $f_c = 13$, $\lambda = 0.05$, $\|w\| = 0$). Display of the singular values of both the primal matrix ($\mathcal{R}_{\lambda,\rho}$, in red) and its dual counterpart (blue). The rank of the primal variable appears to be much lower than the one of the dual variable.

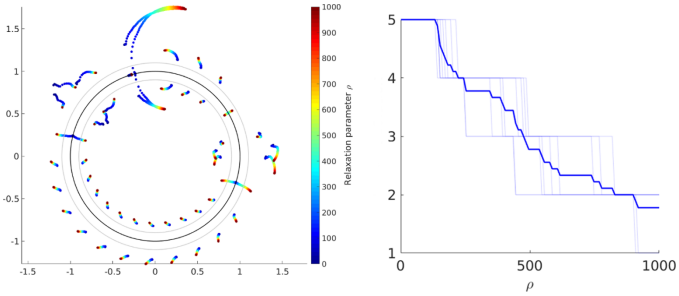

Fig. 3: (5 spikes, $f_c = 13$, $\lambda = 0.05$, $\|w\| = 0.005\|y_0\|$). Behaviour of solutions with respect to ρ. *Left*: trajectory of the roots of the dual polynomial. When $\rho = 0$, the support may be reconstructed from double roots, located on the unit circle. When $\rho > 0$, these may split, but remain identifiable when ρ is small. *Right*: rank of $\mathcal{R}_{\lambda,\rho}$, averaged over small randomized variations around fixed positions.

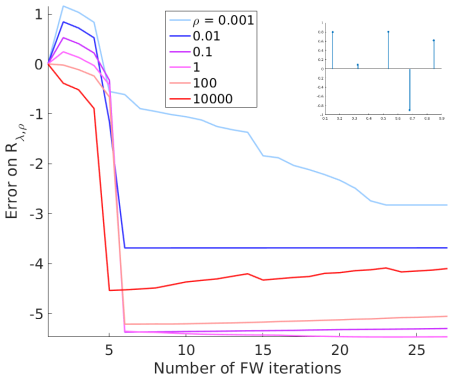

Fig. 4: (5 spikes, $f_c = 13$, $\|w\| = 0$). Relative error (in logarithmic scale) on $\mathcal{R}_{\lambda,\rho}$ at each Frank-Wolfe step, with respect to the solution found with CVX. The input measure is displayed. For a wide range of values of parameter ρ, we observe finite convergence, with exactly as many steps as the number of spikes in the input measure.

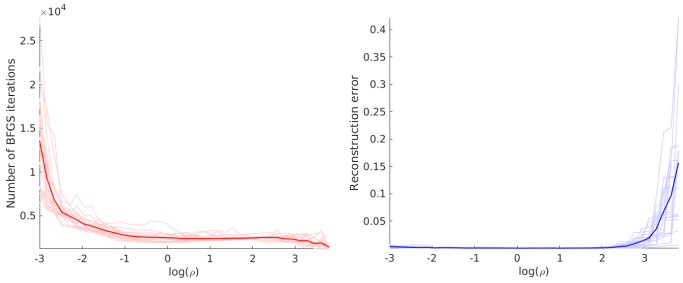

Fig. 5: ($f_c = 13$, $\lambda = 0.05$, $\|w\| = 0$). Randomized tests with 5 spikes of positive amplitudes, separated at least from $\Delta_{min} = 1/f_c$. *Left*: total number of BFGS iterations (tolerance: 10^{-9} in terms of function/parameters changes; maximum number of iterations: 1000), summed over all Frank-Wolfe iterations until convergence, w.r.t. ρ. Convergence is decided when the rank of the iterates stops increasing. *Right*: we measure $\|\mu_{\lambda,\rho} - \mu_\lambda\|^*_{\mathrm{BL}}$ w.r.t. ρ, where $\mu_{\lambda,\rho}$ and μ_λ are the measures retrieved by solving ($\widetilde{\mathcal{P}}_{\lambda,\rho}$) (with our algorithm) or ($\widetilde{\mathcal{P}}_\lambda$) (with CVX) respectively, and $\|\cdot\|^*_{\mathrm{BL}}$ is the dual bounded Lipschitz norm [5]. Our method will be interesting in regimes where the computational time is low and the quality of reconstruction is high.

ACKNOWLEDGMENTS

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

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] 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.

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

[4] 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.

[5] R. Dudley. Convergence of Baire measures. *Studia Mathematica*, 27(3):251–268, 1966.

[6] V. Duval and G. Peyré. Sparse regularization on thin grids I: the Lasso. *Inverse Problems*, 33(5), 2017.

[7] M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2):95–110, 1956.

[8] W. Liao and A. Fannjiang. MUSIC for single-snapshot spectral estimation: Stability and super-resolution. *CoRR*, 2014.

[9] G. Ongie and M. Jacob. Off-the-grid recovery of piecewise constant images from few fourier samples. *SIAM Journal on Imaging Sciences*, 9(3):1004–1041, 2016.

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

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