The Sliding Frank-Wolfe Algorithm for the BLASSO

Quentin Denoyelle
EPFL, Biomedical Imaging Group
Email: quentin.denoyelle@epfl.ch

Vincent Duval
INRIA Paris, MOKAPLAN
Email: vincent.duval@inria.fr

Gabriel Peyré
CNRS, ENS
Email: gabriel.peyre@ens.fr

Emmanuel Soubies
CNRS, IRIT
Email: emmanuel.soubies@irit.fr

***Abstract*—This paper showcases the Sliding Frank-Wolfe (SFW), which is a novel optimization algorithm to solve the BLASSO sparse spikes super-resolution problem. The BLASSO is the continuous (i.e. off-the-grid or grid-less) counterpart of the well-known ℓ^1 sparse regularisation method (also known as LASSO or Basis Pursuit). Our algorithm is a variation on the classical Frank-Wolfe (also known as conditional gradient) which follows a recent trend of interleaving convex optimization updates (corresponding to adding new spikes) with non-convex optimization steps (corresponding to moving the spikes). We prove theoretically that this algorithm terminates in a finite number of steps under a mild non-degeneracy hypothesis.**

I. Introduction

A. Sparse Spikes Super-Resolution

The sparse spikes super-resolution problem aims at recovering an approximation of an unknown input discrete measure $m_{a_0,x_0} \stackrel{\text{def.}}{=} \sum_{i=1}^N a_{0,i}\delta_{x_{0,i}}$ from noisy measurements $y \stackrel{\text{def.}}{=} y_0 + w \in \mathbb{R}^M$ where $y_0 \stackrel{\text{def.}}{=} \Phi(m_{a_0,x_0})$ and w models the acquisition noise. The linear operator Φ is defined over the space of Radon measures $\mathcal{M}(\mathbb{R}^d)$ by $\Phi(m) \stackrel{\text{def.}}{=} \int_{\mathbb{R}^d} \varphi(x)\mathrm{d}m(x)$ (where φ belongs to a general class of kernels detailed in Definition 1) and models the acquisition process. Here $a_{0,i} \in \mathbb{R}$ are the amplitudes of the Dirac masses at positions $x_{0,i} \in \mathbb{R}^d$. This is an ill-posed inverse problem and the BLASSO

$$\min_{m\in\mathcal{M}(\mathbb{R}^d)} T_\lambda(m) \stackrel{\text{def.}}{=} \frac{1}{2}\left\|\Phi(m)-y\right\|^2_{\mathbb{R}^M} + \lambda|m|(\mathbb{R}^d), \qquad (\mathcal{P}_\lambda(y))$$

is a way to solve it in a stable way by introducing a sparsity-enforcing convex regularization. The problem $\mathcal{P}_\lambda(y)$ is a convex optimization problem that generalizes the LASSO over the non-reflexive Banach space $\mathcal{M}(\mathbb{R}^d)$ where the ℓ^1 norm becomes its continuous counterpart represented by the total variation norm $|\cdot|(\mathbb{R}^d)$.

B. Previous Works

The performance of the BLASSO has been theoretically studied in many papers, see for example [1], [2], [3], [4], [5], [6]. On the numerical standpoint, it is possible to solve the BLASSO by considering its dual and recasting it, for Fourier measurements and in a one dimensional setting, as a finite dimensional SDP [1]. In dimension greater than 2, one need to use the Lasserre's hierarchy [7], [8], [9]. In order to solve directly $\mathcal{P}_\lambda(y)$, algorithms that do not use any Hilbertian structure and can instead deal with measures are required. The authors of [2] proposed a modified version of the Frank-Wolfe (FW) where the amplitudes and positions are updated separately to further decrease the objective function. This idea was numerically studied in [10]. We follow a similar approach.

II. The Algorithm

From now on, we suppose for simplicity that $d = 1$. Our algorithm is presented in Algorithm 1 (see [11] for more details). It consists in recursively adding a new Dirac mass to the estimated measure (Step 3), computing the new amplitudes by solving the LASSO (Step 7), and moving continuously *both* the amplitudes and positions by finding a critical point of a non-convex problem (Step 8). See Figures 1 and 2 for illustrations of the algorithm.

Algorithm 1 Sliding Frank-Wolfe (SFW) Algorithm

1: Initialize with $m^{[0]} = 0$ and $n = 0$.
2: **for** $k = 0, \ldots, n$ **do**
3: $\quad m^{[k]} = \sum_{i=1}^{N^{[k]}} a_i^{[k]}\delta_{x_i^{[k]}}$, $a_i^{[k]} \in \mathbb{R}$, $x_i^{[k]}$ pairwise distincts, find $x_*^{[k]} \in \mathbb{R}$ s.t.:
$$x_*^{[k]} \in \arg\max_{x\in\mathbb{R}} |\eta^{[k]}(x)| \quad \text{where} \quad \eta^{[k]} \stackrel{\text{def.}}{=} \frac{1}{\lambda}\Phi^*(y-\Phi(m^{[k]})),$$
4: $\quad$ **if** $|\eta^{[k]}(x_*^{[k]})| \leqslant 1$ **then**
5: $\quad\quad m^{[k]}$ is a solution of $\mathcal{P}_\lambda(y)$. Stop.
6: $\quad$ **else**
7: $\quad\quad$ Get $m^{[k+1/2]} = \sum_{i=1}^{N^{[k]}} a_i^{[k+1/2]}\delta_{x_i^{[k]}} + a_{N^{[k]}+1}^{[k+1/2]}\delta_{x_*^{[k]}}$, s.t.
$$a^{[k+1/2]} \in \arg\min_{a\in\mathbb{R}^{N^{[k]}+1}} \frac{1}{2}\left\|\Phi(m_{a,x^{[k+1/2]}})-y\right\|^2_{\mathbb{R}^M} + \lambda\left\|a\right\|_1$$
$$\text{where} \quad x^{[k+1/2]} = (x_1^{[k]}, \ldots, x_{N^{[k]}}^{[k]}, x_*^{[k]})$$
8: $\quad\quad$ Find $m^{[k+1]} = \sum_{i=1}^{N^{[k]}+1} a_i^{[k+1]}\delta_{x_i^{[k+1]}}$ by initializing with $(a^{[k+1/2]}, x^{[k+1/2]})$ and obtaining a critical point of
$$(a,x) \in \mathbb{R}^{N^{[k]}+1}\times\mathbb{R}^{N^{[k]}+1} \mapsto \frac{1}{2}\left\|\Phi(m_{a,x})-y\right\|^2_{\mathbb{R}^M} + \lambda\left\|a\right\|_1.$$
9: $\quad$ **end if**
10: **end for**

III. Convergence Result

Definition 1 (Admissible kernels φ)**.** *We denote by* $\mathrm{KER}^{(k)}$*, the set of admissible kernels of order* k*. A function* $\varphi : X \to \mathbb{R}^M$ *belongs to* $\mathrm{KER}^{(k)}$ *if:*

- $\varphi \in \mathscr{C}^k(\mathbb{R}, \mathbb{R}^M)$,
- *For all* $p \in \mathbb{R}^M$, $x \in X \mapsto \langle\varphi(x), p\rangle_{\mathbb{R}^M}$ *vanishes at infinity,*
- *for all* $0 \leqslant i \leqslant k$, $\sup_{x\in\mathbb{R}} \left\|\varphi^{(i)}(x)\right\|_{\mathbb{R}^M} < +\infty$.

The next theorem gives a finite time convergence guarantee under mild assumptions that significantly improves the previously known convergence for this kind of algorithm (weak-* convergence).

Theorem 1. *Suppose that* $\varphi \in \mathrm{KER}^{(2)}$*, that* $m_{a,x} = \sum_{i=1}^N a_i\delta_{x_i}$ *is the unique solution of* $\mathcal{P}_\lambda(y)$*, and that* $\eta_\lambda = \frac{1}{\lambda}\Phi^*(y-\Phi(m_{a,x}))$ *is nondegenerate, i.e. for all* $x \in \mathbb{R}\setminus\bigcup_{i=1}^N\{x_i\}$

$$|\eta_\lambda(x)| < 1 \quad \text{and} \quad \forall i \in \{1,\ldots,N\}, \quad \eta_\lambda''(x_i) \neq 0. \qquad (1)$$

Then Algorithm 1 recovers $m_{a,x}$ *after a finite number of steps (i.e. there exists* $k \in \mathbb{N}$ *such that* $m^{[k]} = m_{a,x}$*).*

Acknowledgement

The work of Gabriel Peyré has been supported by the European Research Council (ERC project NORIA). The work of Emmanuel Soubies has been supported by the European Research Council (ERC project GlobalBioIm).

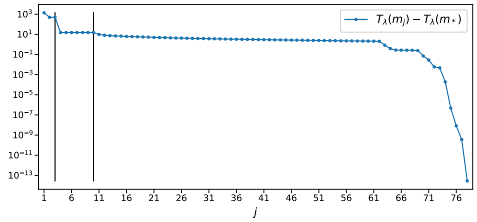
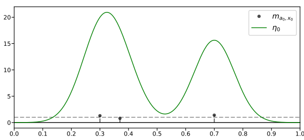
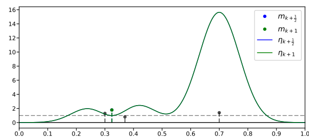


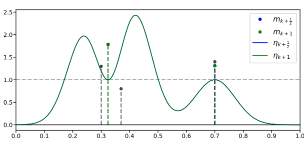
Fig. 1: Values of the objective function throughout the SFW algorithm (cumulative iterations of the BFGS used to find a critical point of the non-convex problem of Step 8). The vertical black lines separate the main outer iterations of the algorithm. The measure m_* represents the unique solution of the BLASSO. It is the same example as in Figure 2.



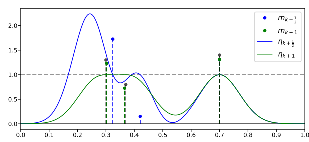
$k = 0$. Start of the loop.



$k = 0$. End of the loop.



$k = 1$. End of the loop.



$k = 2$. End of the loop.

Fig. 2: Main steps of the SFW algorithm.

References

[1] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Comm. Pure Appl. Math.*, vol. 67, no. 6, pp. 906–956, 2014. [Online]. Available: https://doi.org/10.1002/cpa.21455

[2] K. Bredies and H. K. Pikkarainen, "Inverse problems in spaces of measures," *ESAIM Control Optim. Calc. Var.*, vol. 19, no. 1, pp. 190–218, 2013. [Online]. Available: https://doi.org/10.1051/cocv/2011205

[3] E. J. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," *J. Fourier Anal. Appl.*, vol. 19, no. 6, pp. 1229–1254, 2013. [Online]. Available: https://doi.org/10.1007/s00041-013-9292-3

[4] C. Fernandez-Granda, "Support detection in super-resolution," *Proc. Proceedings of the 10th International Conference on Sampling Theory and Applications*, pp. 145–148, 2013. [Online]. Available: http://arxiv.org/abs/1302.3921

[5] V. Duval and G. Peyré, "Exact support recovery for sparse spikes deconvolution," *Found. Comput. Math.*, vol. 15, no. 5, pp. 1315–1355, 2015. [Online]. Available: https://doi.org/10.1007/s10208-014-9228-6

[6] Q. Denoyelle, V. Duval, and G. Peyré, "Support recovery for sparse super-resolution of positive measures," *Journal of Fourier Analysis and Applications*, vol. 23, no. 5, pp. 1153–1194, 2017.

[7] J. B. Lasserre, "Moments, positive polynomials and their applications," *Imperial College Press Optimization Series*, vol. 1, pp. xxii+361, 2010.

[8] ——, "Global optimization with polynomials and the problem of moments," *SIAM J. Optim.*, vol. 11, no. 3, pp. 796–817, 2000/01. [Online]. Available: https://doi.org/10.1137/S1052623400366802

[9] 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. Inform. Theory*, vol. 63, no. 1, pp. 621–630, 2017. [Online]. Available: https://doi.org/10.1109/TIT.2016.2619368

[10] N. Boyd, G. Schiebinger, and B. Recht, "The alternating descent conditional gradient method for sparse inverse problems," *SIAM J. Optim.*, vol. 27, no. 2, pp. 616–639, 2017. [Online]. Available: https://doi.org/10.1137/15M1035793

[11] Q. Denoyelle, V. Duval, G. Peyré, and E. Soubies, "The sliding frank-wolfe algorithm and its application to super-resolution microscopy," *arXiv preprint arXiv:1811.06416*, 2018.