On the non-convex sparse spike estimation problem: explicit basins of attractions of global minimizers

Yann Traonmilin
CNRS
Univ. Bordeaux, Bordeaux INP, IMB, UMR 5251
F-33400 Talence, France.
Email: yann.traonmilin@math.u-bordeaux.fr

Jean-François Aujol
Univ. Bordeaux
Bordeaux INP, CNRS, IMB, UMR 5251
F-33400 Talence, France.

Abstract—The sparse spike estimation problem consists in recovering off-the-grid impulsive sources from under-determined linear measurements. Information theoretic results ensure that the minimization of a non-convex functional is able to recover the spikes for adequately chosen measurements (deterministic or random). We study the shape of the global minimum of this non-convex functional: we give an explicit basin of attraction of the global minimum that shows that the problem becomes easier as the number of measurements grows. This has consequences for methods involving descent algorithms and gives insights for potential improvements of such methods, as a potential alternative to convex methods.

I. INTRODUCTION

Sums of sparse off-the-grid spikes can be used to model impulsive sources (e.g. in astronomy, microscopy,...). Estimating such signals from a finite number of measurements is known as the super-resolution (SR) problem [6]. In the space $\mathcal{M}$ of finite signed measure over $\mathbb{R}^d$, we aim at recovering $x = \sum_{i=1}^k a_i \delta_{t_i}$, from

$$y = Ax_0 + e,$$

where $\delta_{t_i}$ is the Dirac measure at position $t_i$, the operator $A$ is a linear observation operator, $y \in \mathbb{C}^m$ are the $m$ noisy measurements and $e$ is a finite energy observation noise. It is possible to estimate spikes from a finite number of regular Fourier measurements as long as their locations are sufficiently separated, using convex minimization variational methods in the space of measures [5], [8], [9]. Other general studies on inverse problems have shown that an ideal non-convex method (unfortunately computionally inefficient) can be used to recover these signals if $A$ has a restricted isometry property (RIP) [3]. In the case of SR, adequately chosen random compressive measurements have been shown to meet the sufficient RIP conditions for separated spikes, thus guaranteeing the success of the ideal non-convex decoder [11]. These RIP results rely heavily on the choice of an adequate kernel metric on $\mathcal{M}$. Greedy heuristics have also been proposed to approach the non-convex minimization problem and they have shown good practical utility [12], [13], [16].

II. PARAMETERIZED FORMULATION AND OBJECTIVES

In this work, we focus on the ideal non-convex minimization:

$$x^* \in \arg\min_{x \in \Sigma_{k,\epsilon}} \|Ax - y\|_2$$

where $\Sigma_{k,\epsilon} := \{\sum_{r=1}^k a_r \delta_{t_r} : (a_1, t_1, \ldots, t_k) \in \mathbb{R}^{k(d+1)}, a \in \mathbb{R}^k, t_r \in \mathbb{R}^d, \forall r \neq i, \|t_i - t_r\|_2 > \epsilon, \|t_r\|_2 \leq R\}$ is a low-dimensional set modeling the separation constraints on the $k$ Diracs. While simple in its formulation, properties of this minimization procedure have not yet been thoroughly studied. We consider the following parameterization of $x \in \Sigma_{k,\epsilon}$: $x = \sum_{i=1}^k a_i \delta_{t_i} = \phi(\theta)$ with $\theta = (a_1, \ldots, a_k, t_1, \ldots, t_k)$ and define $\Theta_{k,\epsilon} := \phi^{-1}(\Sigma_{k,\epsilon})$.

We consider the problem

$$\theta^* \in \arg\min_{\theta \in E} g(\theta) = \arg\min_{\theta \in E} \|A\phi(\theta) - y\|_2.$$  \hspace{1cm} (3)

where $E = \mathbb{R}^{k(d+1)}$ or $E = \Theta_{k,\epsilon}$ and $g(\theta) = \|A\phi(\theta) - y\|_2$. We remark that (3) is a smooth non-convex minimization problem.

The objective of our work is to study this optimization problem. We focus on the basin of attraction of its global minimizers (i.e. sets were gradient descent intialized in these sets converges to $\theta^*$).

Related work. While original for the sparse spikes estimation problem, the theoretical study of non-convex optimization schemes for linear inverse problems has gained attraction recently for different kinds of low-dimensional models. For low-rank matrix estimation [18], [1], phase recovery [17], blind deconvolution and bi-convex programming, recent works have exploited similar ideas [14], [4]. In the case of SR, the idea of gradient descent has been studied in an asymptotic regime ($k \to \infty$) in [7].

III. CONTRIBUTIONS AND PERSPECTIVES

In this work\footnote{This abstract is a summary of the complete article [15] by the same authors.}, we link a restricted isometry property (guaranteed by a finite number of measurements) of measurement operators $A$ with the conditioning of the Hessian of $g$ at the global minimum. This allows us to give an explicit basin of attractions of the global minimum in the noisless and noisy cases, i.e. an explicit set whose shape depends on the parameters of the problem where the gradient descent converges. The dependency on RIP constant shows that when the number of measurement increases (i.e. the RIP constants improves), the size of the basin of attraction increases. This study has direct consequences for the theoretical study of greedy approaches. Indeed a basin of attraction permits to give recovery guarantees for descent methods (the initialization must fall within the basin), since the gradient descent is a step in the iterations of the greedy approach.

Future work. The missing element to obtain complete theoretical guarantees of descent methods is an adequate initialization procedure, that must be found, as is done for other low dimensional models (see Related works). Also, our theoretical analysis shows that enforcing a separation constraint ($E = \Theta_{k,\epsilon}$) in the descent algorithms might improve them. The idea of projected gradient descent then emerges naturally. This idea has been explored for general low-dimensional models [2]. It has been shown that the RIP guarantees the convergence of such algorithms with an ideal (often non practical) projection. Approached projected gradient descents have also been studied and shown to be successful for some particular applications [10]. These works along with our results suggest to explore projected gradient descent for the spikes SR problem possibly both numerically and theoretically.
REFERENCES


